NON COMPACT ESTIMATION OF THE CONDITIONAL DENSITY FROM DIRECT OR NOISY DATA

F. COMTE AND C. LACOUR

Abstract. In this paper, we propose a nonparametric estimation strategy for the conditional density function of Y given X, from independent and identically distributed observations $(X_i, Y_i)_{1\le i\le n}$. We consider a regression strategy related to projection subspaces of $\mathbb{L}^2$ generated by non compactly supported bases. This first study is then extended to the case where Y is not directly observed, but only $Z = Y + \varepsilon$, where ε is a noise with known density. In these two settings, we build and study collections of estimators, compute their rates of convergence on anisotropic space on non-compact supports, and prove related lower bounds. Then, we consider adaptive estimators for which we also prove risk bounds.

Key words and phrases. Anisotropic Sobolev spaces. Conditional density. Deconvolution. Hermite basis. Laguerre basis. Noisy data. Nonparametric estimation.
AMS 2020 Classification. 62G05-62G07-62G08.

1. Introduction

The purpose of this paper is to estimate the conditional density of a response Y given a variable X, with or without directly observing Y. We may assume that a noise ε spoils the response so that only $Z = Y+\varepsilon$ is available. From independent and identically distributed couples of variables $(X_i, Y_i)_{1\le i\le n}$ first, and $(X_i, Z_i)_{1\le i\le n}$ in a second step, we estimate the conditional density $\pi(x, y)$ of Y given X defined by

$$\pi(x, y)dy = \mathbb{P}(Y \in dy|X = x).$$

In this framework, the regression function $\mathbb{E}[Y|X = x]$ is often studied, but this information is more restrictive than the entire distribution of Y given X, in particular when the distribution is asymmetric or multimodal. Thus the problem of conditional density estimation is found in various application fields: meteorology, insurance, medical studies, geology, astronomy (see Nguyen (2018) and Izbicki and Lee (2017) and references therein).

1.1. Bibliographical elements on conditional density estimation.

The estimation of the conditional density has often been studied with kernel strategies, initiated by Rosenblatt (1969). The idea is to define the estimator as a quotient of two kernel density estimators: we can cite among others Youndjé (1996), Fan et al. (1996), Hyndman and Yao (2002), De Gooijer and Zerom (2003), Fan and Yim (2004). Also with kernel tools, Ferraty et al. (2006) or Laksaci (2007) are interested in the conditional density estimation when X is a functional random variable. Using histograms on partitions, Györfi and Kohler (2007) estimate the conditional distribution of Y given X consistently in total variation, see also Sart (2017). Then several papers proposed strategies to estimate the conditional density π as an anisotropic function under the Mean Integrated Squared error criterion. They give oracle inequalities and adaptive minimax results. For instance Efromovich (2007) uses a Fourier decomposition to construct a blockwise-shrinkage Efromovich-Pinsker estimator, whereas Brunel et al. (2007) and Akakpo and Lacour (2011) use projection estimators and model selection. Next, Efromovich (2010) developed a strategy relying on conditional characteristic function estimation, and Chagny (2013) studied a warped basis estimator while Bertin et al. (2016) used a Lepski-type method. Specific methods for higher dimensional covariates were recently developed by Fan et al. (2009), Holmes et al. (2010), Cohen and Le Pennec (2013), Izbicki and Lee (2016), Otneim and Tjostheim (2018), Nguyen et al. (2021).

The problem of estimating the conditional density when the response is observed with noise has been much less studied. Ioannides (1999) considers the estimation of the conditional density of Y given X for strongly mixing processes when both X and Y are noisy, in order to estimate the conditional mode. Using a quotient of deconvoluting kernel estimators, he establishes a convergence rate for an ordinary smooth noise (see Assumption **A5** below for the definition of ordinary smooth and supersmooth noise) when x belongs to a compact set.

1.2. **About non compact support specificity.** Our specific aim in this paper is to deal with variables lying in a non-compact domain. Many authors assume that X and Y belong to a bounded and known interval. In practice, this interval is estimated from the data and so it is not deterministic. As explained in Reynaud-Bouret et al. (2011), "this problem is not purely theoretical since the simulations show that the support-dependent methods are really affected in practice by the size of the density support, or by the weight of the density tail". They show in their paper that the minimax rate of convergence for density estimation may deteriorate when the support becomes infinite and they name it the "curse of support". This phenomenon had been previously highlighted by Juditsky and Lambert-Lacroix (2004), and has been extended in the mutivariate case by Goldenshluger and Lepski (2014). When using a $\mathbb{R}$-supported basis for density estimation, Belomestny et al. (2019) obtain a nonstandard variance order; however it is associated to a nonstandard bias, which leads to classical rates; the same kind of result holds for $\mathbb{R}^+$-Laguerre basis, see Comte and Genon-Catalot (2018). For regression function estimation, Comte and Genon-Catalot (2020) introduce a specific method adapted to the non-compact case, which allows them to obtain new minimax results; our study is inspired by their work.

1.3. **Conditional density as a mixed regression-density framework.** Here we study the estimation of a conditional density: we can think of it as a regression issue in the first direction and a density issue in the second. We show that the rate of convergence is again modified in the case of a non-compact support. To do this, we define an estimator $\hat{\pi}_{\mathbf{m}}^{(D)}$, $\mathbf{m} = (m_1, m_2)$, by minimization of a least squares contrast on a subspace $S_{\mathbf{m}}$ with finite dimension. This estimator is a classical projection estimator expanded on an orthogonal basis $(\varphi_j \otimes \varphi_k)_{0\leq j\leq m_1-1, 0\leq k\leq m_2-1}$. The coefficients are written with the same kind of formula as in standard linear regression, with the use of matrix

$$\widehat{\Psi}_m = \widehat{\Psi}_m(\mathbf{X}) = \frac{1}{n}\, {}^t\widehat{\Phi}_m \widehat{\Phi}_m, \quad \text{where } \widehat{\Phi}_m = (\varphi_j(X_i))_{1\leq i\leq n, 0\leq j\leq m-1}.$$

The point is to use specific bases adapted to the non-compact problem. Two cases are of special interest: the case where the support is $\mathbb{R}$, for which we use the Hermite basis, and the case where the support is $\mathbb{R}_+$, for which we use the Laguerre basis. This last case is very useful in various applications as reliability, economics, survival analysis. Note that we also consider the trigonometric basis to include the compactly-supported case in our study. We detail the properties of the Hermite and Laguerre bases in Section 2. In particular, these bases are associated to Sobolev-type functional spaces, and this allows us to define the smoothness of the target function. Moreover a second motivation to study the non-compactly supported case is to allow an extension to the noisy case, when Y is not directly observed. Indeed the classical use of Fourier transform for nonparametric deconvolution requires to work on the whole real line. And actually these two bases can be used in the deconvolution setting when considering noisy observations, see Mabon (2017) for Laguerre deconvolution and Sacko (2020) for the Hermite case. Note that a conditional density is an intrinsically anisotropic object, with possibly anisotropic smoothness. That is why we use bases with different cardinalities m_1 in the x-direction and m_2 in the y-direction, where $\mathbf{m} = (m_1, m_2)$.

1.4. **Anisotropic (conditional) model selection.** In this paper we compute the integrated squared risk for our estimator, in particular the variance is of order $m_1\sqrt{m_2}/n$ instead of m_1m_2/n in the compact case. We derive the anisotropic rate of convergence for the conditional density estimation with non-compact support. We recover classical rates in the compacted supported case, and obtain different ones in the Hermite and Laguerre cases, for which we provide lower bounds, under some condition. Moreover, we tackle the problem of model selection: what is the better choice for m_1 and m_2, and how to select it only from the data? Here we use the Goldenshluger-Lepski method (Goldenshluger and Lepski, 2011), which consists in minimizing some penalized differences criterion over a collection of models $\mathcal{M}_n$. In our framework this collection has to be random because of the very importance of the normal matrix $\widehat{\Psi}_{m_1}$ if we do not assume that the distribution of X has a lower-bounded density, contrary to what is almost always supposed in regression or conditional distribution issues. Instead, similarly to the non-compact regression case, our results depend on a condition on $\widehat{\Psi}_{m_1}$ called stability condition, which bounds the operator norm $\|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}$ in term of n and m_1. Here we improve the condition required by Comte and Genon-Catalot (2020) for the adaptive procedure in the regression context. Despite this inherent difficulty of the role of $\widehat{\Psi}_{m_1}$, we provide an adaptive method with no unknown quantity, and easy to implement. This is worthy since adaptive penalized methods in complex models often involve unknown quantities in the penalty. For example Brunel et al. (2007) have a penalty which depends on an upperbound on π, or on a lowerbound on the design density. Here we avoid it by a judicious use of conditioning.

1.5. **Extensions to noisy case.** Last but not least, we extend all the previous results to the noisy case, where Y is not observed, and only $Y+\varepsilon$ is available. As usual, we assume that the distribution of ε is known for identifiability reasons. This brings us to a deconvolution issue in the y-direction: see Meister (2009) for an overview on nonparametric deconvolution. We divide our study of this noisy case in two parts. In the first part (see Section 4), we consider the case where all the variable are positive, including the noise. In another part (see Section 5), we consider variables in $\mathbb{R}$, with the classical hypothesis that the characteristic function of the noise does not vanish. We study both cases of ordinary smooth noise and supersmooth noise. For these two noisy cases (variables in $\mathbb{R}_+$ or in $\mathbb{R}$), we provide new estimators $\hat{\pi}_{\mathbf{m}}^{(L)}$ and $\hat{\pi}_{\mathbf{m}}^{(H)}$ and study their integrated risk. The rates of convergence are more involved than in the direct (non-noisy) case since they depend on the smoothness of the noise density. Indeed the smoother the noise distribution, the smoother the distribution of Z, so that the true signal is difficult to recover. We also propose an adaptive model selection and we obtain again an oracle inequality, using an entirely known penalty term. Thus (unlike Ioannides (1999)) our method reaches an automatic squared bias-variance compromise, without requiring the knowledge of the regularity order of the function to estimate.

1.6. **Content of the paper.** The paper is organized as follows. After describing in Section 2 the study framework (notation, bases functions and their useful properties, regularity spaces, model of the observations), Section 3 is devoted to the definition and study of the estimation procedure in the direct case (the Y_i's are observed). A risk bound is given in this setting, and the rates of convergence of the estimators both in the usual and in new bases are given, together with Laguerre and Hermite lower bounds as these cases correspond to nonstandard rates. Section 4 defines and studies the estimator corresponding to the noisy case when all random variables are nonnegative and the Laguerre basis is used, while the more general $\mathbb{R}$-supported case is considered in Section 5, relying on an estimator defined in the Hermite basis. Lastly, Section 6, states a general adaptive result, based on a Goldenshluger-Lepski method, see Goldenshluger and Lepski (2011). A few concluding remarks are stated in Section 7. All proofs are postponed in Section 8, while some useful results are given in Appendix.

2. Model and assumptions

2.1. **Notation.** We denote by f the density of the covariate X, so that the joint density of (X, Y) is $f(x)\pi(x, y)$. We consider the weighted $\mathbb{L}^2$ norm of a bivariate measurable function T, defined by:

$$\|T\|_f^2 := \iint T^2(x,y) f(x) dx dy \tag{1}$$

and the associated dot product $\langle T_1, T_2\rangle_f = \iint T_1(x,y)T_2(x,y)f(x)dxdy$. The usual (non-weighted) $\mathbb{L}^2$ norm is denoted by $\|.\|_2$. We also introduce the empirical norm of T:

$$\|T\|_n^2 := \frac{1}{n}\sum_{i=1}^n \int T^2(X_i, y)dy. \tag{2}$$

Note that for any deterministic function T, $\mathbb{E}\|T\|_n^2 = \|T\|_f^2$. For two functions $x \mapsto t(x)$ and $y \mapsto s(y)$, defined on $\mathbb{R}$ or $\mathbb{R}^+$, we set $(t \otimes s)(x, y) = t(x)s(y)$.

Let $\mathcal{M}_n$ be a subset of $\{1, \dots, n\} \times \{1, \dots, n\}$ and let $\mathbf{m} = (m_1, m_2)$ denote an element of $\mathcal{M}_n$. We construct a sequence $(\hat\pi_{\mathbf{m}})_{\mathbf{m}\in\mathcal{M}_n}$ of estimators of π, each $\hat\pi_{\mathbf{m}}$ belonging to a subspace $S_{\mathbf{m}} = S_{m_1} \otimes S_{m_2}$ where each linear space S_{m_i}, $i = 1, 2$ is generated by m_i functions,

$$S_{m_i} = \operatorname{span}\{\varphi_j, j = 0, \dots, m_i - 1\},\ i = 1, 2,$$

and the φ_j are known orthonormal functions with respect to the standard $\mathbb{L}^2$-scalar product:

$$\langle \varphi_j, \varphi_k\rangle = \int \varphi_j(u)\varphi_k(u)du = \delta_{j,k}.$$

Here $\delta_{j,k}$ is the Kronecker symbol, equal to 0 if $j \neq k$ and to 1 if $j = k$. Thus $S_{\mathbf{m}}$ is spanned by $\{\varphi_j \otimes \varphi_k, j = 0, \dots, m_1 - 1, k = 0, \dots, m_2 - 1\}$. A key quantity associated to the basis $(\varphi_j)_j$ is

$$L(m) = \sup_{t\in S_m} \left(\|t\|_\infty^2/\|t\|_2^2\right) = \sup_{x\in\mathbb{R}} \sum_{j=0}^{m-1} \varphi_j^2(x). \tag{3}$$

Clearly, for the tensorized basis, $L(\mathbf{m}) = L(m_1)L(m_2)$.

Lastly, for a non necessarily square matrix M with real coefficients, we define its operator norm $\|M\|_{\mathrm{op}}$ as $\sqrt{\lambda_{\max}(M\,{}^tM)}$ where ${}^tM$ is the transpose of M and $\lambda_{\max}$ denotes the largest eigenvalue. Its Frobenius norm is defined by $\|M\|_F^2 = \operatorname{Tr}(M\,{}^tM)$ where $\operatorname{Tr}(A)$ denotes the trace of the square matrix A.

2.2. **Bases.** We give now the examples of basis functions we consider in the sequel: the trigonometric basis as an example of compactly supported basis for comparison with previous results, and the Laguerre and Hermite bases which are respectively $\mathbb{R}_+$ and $\mathbb{R}$-supported.

- Trigonometric basis functions are supported by $[0, 1]$, with $t_0(x) = \mathbf{1}_{[0,1]}(x)$, and for $j \geq 1$, $t_{2j-1}(x) = \sqrt{2}\cos(2\pi jx)\mathbf{1}_{[0,1]}(x), t_{2j}(x) = \sqrt{2}\sin(2\pi jx)\mathbf{1}_{[0,1]}(x)$. For the basis $(t_j)_{0\leq j\leq m-1}$, if m is odd, then $L(m) = m$ with $L(m)$ defined by (3).

- The Laguerre functions are defined as follows:

$$\ell_j(x) = \sqrt{2}L_j(2x)e^{-x}\mathbf{1}_{x\geq 0} \quad \text{with} \quad L_j(x) = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{x^k}{k!}.$$

The functions ℓ_j are orthonormal, and are bounded by $\sqrt{2}$ (see 22.14.12 in Abramowitz and Stegun (1964)). So $\sum_{j=0}^{m-1} \ell_j^2(x) \leq 2m$ and as $\ell_j(0) = \sqrt{2}$, it holds that the supremum value $2m$ is reached in 0 and $L(m) = 2m$. The convolution product of two Laguerre functions has the following useful property (see 22.13.14 in Abramowitz and Stegun (1964)):

$$\ell_j \star \ell_k(x) = \int_0^x \ell_j(u)\ell_k(x-u)du = \frac{1}{\sqrt{2}}(\ell_{j+k}(x) - \ell_{j+k+1}(x)), \quad \forall x \geq 0. \tag{4}$$

Moreover, by Comte and Genon-Catalot (2018), Lemma 8.2, if

$$\exists C > 0, \forall x \geq 0, \quad \mathbb{E}\left(\frac{1}{\sqrt{Y}}|X = x\right) < C \tag{5}$$

then for $j \geq 1$

$$\forall x \geq 0, \quad \mathbb{E}(\ell_j^2(Y)|X = x) = \int_0^\infty \ell_j^2(y)\pi(x,y)dy \leq \frac{c}{\sqrt{j}}. \tag{6}$$

For instance, condition (5) holds if $Y = g(X) + U$ with $g \geq 0$, X and U independent, and $\mathbb{E}U^{-1/2} < \infty$. Under (5), for $m \geq 1$, for $x \geq 0$, $\mathbb{E}\left(\sum_{j=0}^{m-1} \varphi_j^2(Y)|X = x\right) \leq c'\sqrt{m}$ for $c' > 0$ a constant.

- The Hermite functions are defined as follows:

$$h_j(x) = \frac{1}{\sqrt{2^j j! \sqrt{\pi}}} H_j(x) e^{-x^2/2}, \quad \text{with} \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}).$$

The functions h_j are orthonormal, and are bounded by $1/\pi^{1/4}$. The Hermite functions have the following Fourier transform:

$$\forall x \in \mathbb{R}, \quad h_j^*(x) := \int e^{ixu} h_j(u)du = \sqrt{2\pi}(i)^j h_j(x), \text{ where } i^2 = -1. \tag{7}$$

Moreover, from Askey and Wainger (1965) or Markett (1984), it holds

$$|h_j(x)| \leq Ce^{-\xi x^2}, \quad \text{for } |x| \geq \sqrt{2j+1}, \tag{8}$$

where C and ξ are positive constants independent of x and j, $0 < \xi < \frac{1}{2}$. Note that with (7), h_j^* satisfies the same inequality, with constant multiplied by $\sqrt{2\pi}$.
Relying on these results, we can prove the following Lemma, (see Section 8.1):

Lemma 1. *There exists a constant $K > 0$ such that* $\sup_{x\in\mathbb{R}} \sum_{j=0}^{m-1} h_j^2(x) \leq K\sqrt{m}$, *for any $m \geq 1$.*

As a consequence, for this basis $L(m) \leq K\sqrt{m}$.
In the sequel, $\varphi_j = t_j$ or $\varphi_j = \ell_j$ or $\varphi_j = h_j$. Note that, for simplicity, we tensorize twice the same basis but we could mix two different bases.

2.3. **Anisotropic Laguerre and Hermite Sobolev spaces.** To study the bias term, we assume that π belongs to a Sobolev-Laguerre or a Sobolev-Hermite space. In dimension $d = 1$, these functional spaces have been introduced by Bongioanni and Torrea (2009) to study the Laguerre operator. The connection with Laguerre or Hermite coefficients was established later and are summarized in Comte and Genon-Catalot (2018). They were extended to multidimensional case in Dussap (2021). Following the same idea, we define Sobolev-Laguerre balls on $\mathbb{R}_+^d$ and Sobolev-Hermite balls on $\mathbb{R}^d$.

Definition 1. *(Sobolev-Laguerre or Hermite ball). Let $L > 0$ and $\mathbf{s} \in (0, +\infty)^d$, we define the Sobolev-Laguerre with $A = \mathbb{R}_+^d$ or Sobolev-Hermite with $A = \mathbb{R}^d$ ball of order $\mathbf{s} = (s_1, \ldots, s_d)$ and* *radius L by:*

$$W_{\mathbf{s}}(A,L) := \left\{ g \in \mathbb{L}^2(A), \sum_{\mathbf{k}\in\mathbb{N}^d} a_{\mathbf{k}}^2(g)\mathbf{k}^{\mathbf{s}} \leq L \right\}, \quad k^{\mathbf{s}} = k_1^{s_1} \dots k_d^{s_d},$$

with $a_{\mathbf{k}}(g) := \langle g, \varphi_{\mathbf{k}} \rangle = \langle g, \varphi_{k_1} \otimes \cdots \otimes \varphi_{k_d} \rangle$*, the Laguerre coefficients of* g *if* $\varphi_{\mathbf{k}} = \ell_{\mathbf{k}} = \ell_{k_1} \otimes \cdots \otimes \ell_{k_d}$ *or the Hermite coefficients of* g *if* $\varphi_{\mathbf{k}} = h_{\mathbf{k}} = h_{k_1} \otimes \cdots \otimes h_{k_d}$.

We refer to Belomestny et al. (2019) for details about this space and the link with usual Sobolev space. Note in particular that when $d = 1$ and $\mathbf{s}$ is an integer, g belongs to the Sobolev-Hermite space if and only if g admits derivatives up to order $\mathbf{s}$ and the functions $g, g', \dots, g^{(\mathbf{s})}, x^{\mathbf{s}-k}g^{(k)}, k = 0, \dots, \mathbf{s}-1$ belongs to $\mathbb{L}^2(A)$

Assuming that g belongs to $W_{\mathbf{s}}(A,L)$, the approximation term decreases to 0 with polynomial rate. Indeed, for $\mathbf{m} = (m_1, \dots, m_d) \in (\mathbb{N}^*)^d$ and $g_{\mathbf{m}}$ the orthogonal projection of g on $S_{\mathbf{m}}$, we have:

$$\|g - g_{\mathbf{m}}\|_2^2 = \sum_{\mathbf{k}\in\mathbb{N}^d, \exists q, k_q \geq m_q} a_k^2(g) \leq \sum_{q=1}^{d} \sum_{\mathbf{k}\in\mathbb{N}^d, k_q \geq m_q} a_k^2(g) k_q^{s_q} k_q^{-s_q} \leq L \sum_{q=1}^{d} m_q^{-s_q}.$$

Remark. In the present bivariate context, mixed cases involving basis $(\ell_j)_{j\geq 0}$ in one direction and basis $(h_j)_{j\geq 0}$ in the other, with coefficients of a function g defined by $a_{\mathbf{k}}(g) := \langle g, \ell_{k_1} \otimes h_{k_2} \rangle$ would be possible. The link between regularity spaces defined by the rate of decay of the coefficients and derivability properties is then undocumented, contrary to the "homogeneous" case described in Definition 1.

Supersmooth sub-classes. We mention here that in the context of Laguerre one-dimensional developements, functions ψ defined as mixtures of Gamma densities constitute a class of supersmooth functions in the sense that $\|\psi - \psi_m\|^2$ has exponential rate of decrease, see Lemma 3.9 in Mabon (2017). Continuous mixtures are also studied in section 3.2 of Comte and Genon-Catalot (2018).

We choose to be more explicit in the context of Hermite expansions. Let us define

$$\psi_{p,\sigma}(x) = \frac{x^{2p}}{\sigma^{2p+1}\sqrt{2\pi}c_{2p}} \exp(-\frac{x^2}{2\sigma^2})$$

where $c_{2p} = \mathbb{E}[N^{2p}]$ for $N \sim \mathcal{N}(0,1)$, $\sigma^2 \neq 1$ (cases with $\sigma^2 = 1$ have finite developments in the basis, and null bias for m_i larger that p). It is proved in Belomestny et al. (2019) (Proposition 12) that, for $i = 1, 2$,

$$\|\psi_{p,\sigma} - (\psi_{p,\sigma})_{m_i}\|^2 \leq C(p,\sigma^2) m_i^{p-1/2} \exp(-\lambda_0\, m_i), \quad \lambda_0 = \log\left[\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2\right] > 0,$$

where $(\psi_{p,\sigma})_{m_i}$ are the orthogonal projections of $\psi_{p,\sigma}$ on S_{m_i}.

By tensorization, we can thus consider the class $WSS_{\mathbf{s},\lambda}(L)$ for $\mathbf{s} = (s_1, s_2)$ and $\lambda = (\lambda_1, \lambda_2)$ for real numbers s_1, s_2 and positive λ_1, λ_2, of functions g such that,

$$\|g - g_{\mathbf{m}}\|^2 \leq L(m_1^{-s_1} \exp(-\lambda_1\, m_1) + m_2^{-s_2} \exp(-\lambda_2\, m_2)) \tag{9}$$

where $\mathbf{m} = (m_1, m_2)$. Mixed cases with ordinary smooth decay in one direction and super smooth in the other may also be possible.

2.4. **Direct and noisy cases.** In the sequel, we consider two settings.

• In the direct case, we observe independent and identically distributed couples of variables (X_k, Y_k), $k = 1, \dots n$ with the same law as (X, Y). It is studied in section 3, under the Assumption

Assumption A1. *The random variables $(X_i, Y_i)_{1\le i\le n}$ are i.i.d. and the X_i, $i = 1 \dots, n$ are almost surely distinct.*

• In the noisy case, the observations are (X_k, Z_k), $k = 1, \dots n$ with the same distribution as (X, Z), where Z can be written as

$$Z = Y + \varepsilon.$$

This case is studied in Section 4 (nonnegative random variables and Laguerre basis) and Section 5 (general case and Hermite basis), under the additional assumption:

Assumption A2. *The distribution of ε is known, ε is independent of X and independent of Y conditionally to X.*

Notice that this implies the independence of Y and ε.

In both direct and noisy settings, we estimate the function π on $\mathbb{R}\times\mathbb{R}$ or $\mathbb{R}_+\times\mathbb{R}_+$. In the direct case, we also consider the case of π estimated on $[0,1]\times[0,1]$ already studied in the literature for comparison.

We state a general result of adaptive model selection gathering all cases in Section 6.

3. Minimum contrast estimation procedure without noise

3.1. **Definition of the contrast and estimators in the direct case.** We consider the contrast function

$$\gamma_n^{(D)}(T) := \|T\|_n^2 - \frac{2}{n}\sum_{i=1}^{n} T(X_i, Y_i),$$

where $\|T\|_n^2$ is defined by (2) and the estimator

$$\widehat{\pi}_{\mathbf{m}}^{(D)} := \arg\min_{T\in S_{\mathbf{m}}} \gamma_n^{(D)}(T) \tag{10}$$

for $\mathbf{m} = (m_1, m_2)$. This contrast function has already been considered in Brunel et al. (2007). It can be understood by computing its expectation, for any deterministic function T:

$$\mathbb{E}\gamma_n^{(D)}(T) = \|T\|_f^2 - 2\int T(x,y)\pi(x,y)f(x)dx = \|T-\pi\|_f^2 - \|\pi\|_f^2,$$

where $\|T\|_f^2$ is defined by (1), and by observing that it is minimum for $T = \pi$.

To give an explicit formula for $\widehat{\pi}_{\mathbf{m}}^{(D)}$, we define

$$\widehat{\Phi}_m = (\varphi_j(X_i))_{1\le i\le n, 0\le j\le m-1}, \quad \widehat{\Psi}_m = \frac{1}{n}\,{}^t\widehat{\Phi}_m\,\widehat{\Phi}_m. \tag{11}$$

Note that $\Psi_m := \mathbb{E}(\widehat{\Psi}_m) = (\langle\varphi_j, \varphi_k\rangle_f)_{0\le j,k\le m-1}$. We find, assuming that $\widehat{\Psi}_{m_1}$ is invertible,

$$\widehat{\pi}_{\mathbf{m}}^{(D)}(x,y) = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} \widehat{a}_{j,k}^{(D)}\varphi_j(x)\varphi_k(y), \quad \widehat{A}_{\mathbf{m}}^{(D)} = (\widehat{a}_{j,k}^{(D)})_{0\le j\le m_1-1, 0\le k\le m_2-1}$$

with

$$\widehat{A}_{\mathbf{m}}^{(D)} = \frac{1}{n}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\widehat{\Theta}_{m_2}(\mathbf{Y}), \quad \text{with} \quad \widehat{\Theta}_m(\mathbf{Y}) = (\varphi_j(Y_i))_{1\le i\le n, 0\le j\le m-1}. \tag{12}$$

Remark. In the Laguerre and Hermite case, conditions ensuring *a.s.* inversibility of $\widehat{\Psi}_{m_1}$ are weak: $m_1 \le n$ and *a.s.* distinct observations, see Comte and Genon-Catalot (2020). The same conditions work in the trigonometric case. These conditions are ensured by Assumption **A1** and $n \geq m_1$, taken for granted in the sequel.

3.2. **Bound on the empirical MISE of $\widehat{\pi}_{\mathbf{m}}$.** First we study the quadratic (empirical) risk of the estimator $\widehat{\pi}_{\mathbf{m}}$, on a given space $S_{\mathbf{m}} = S_{m_1} \otimes S_{m_2}$ as described in Sections 2.1 and 2.2.

We denote $\pi_{\mathbf{m},n}$ the orthogonal projection of π on $S_{\mathbf{m}}$ for the empirical norm, and $\pi_{\mathbf{m}}$ the orthogonal projection for the $\mathbb{L}^2$-norm. Then we can write

$$\|\widehat{\pi}_{\mathbf{m}}^{(D)} - \pi\|_n^2 = \|\widehat{\pi}_{\mathbf{m}}^{(D)} - \pi_{\mathbf{m},n}\|_n^2 + \|\pi_{\mathbf{m},n} - \pi\|_n^2, \tag{13}$$

and then note that

$$\|\pi_{\mathbf{m},n} - \pi\|_n^2 = \inf_{T \in S_{\mathbf{m}}} \|T - \pi\|_n^2 \leq \|\pi_{\mathbf{m}} - \pi\|_n^2.$$

Thus

$$\mathbb{E}(\|\pi_{\mathbf{m},n} - \pi\|_n^2) \leq \|\pi_{\mathbf{m}} - \pi\|_f^2.$$

Note that the following Lemma (proved in Section 8.2) is useful, here and further:

Lemma 2. *Assume that Assumption* **A1** *holds and $n \geq m_1$. Then it holds that $\mathbb{E}[\widehat{\pi}_{\mathbf{m}}^{(D)}(X_i, y)|\mathbf{X}] = \pi_{\mathbf{m},n}(X_i, y) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} [\mathbf{D}_{\mathbf{m}}]_{j,k} \varphi_j(X_i) \varphi_k(y)$ with*

$$\mathbf{D}_{\mathbf{m}} = \frac{1}{n} \widehat{\Psi}_{m_1}^{-1} \, {}^t\widehat{\Phi}_{m_1} \left(\int \varphi_k(y) \pi(X_i, y) dy \right)_{1 \leq i \leq n, 0 \leq k \leq m_2 - 1}. \tag{14}$$

Using this result, we obtain the following risk bound (proved in Section 8.3).

Proposition 1. *Let $\widehat{\pi}_{\mathbf{m}}$ be defined by (10)-(12), and assume that Assumption* **A1** *is fulfilled. Then, for any $\mathbf{m} = (m_1, m_2)$ such that $m_1 \leq n$,*

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}}^{(D)} - \pi\|_n^2 \leq \|\pi - \pi_{\mathbf{m}}\|_f^2 + \frac{m_1 L(m_2)}{n}. \tag{15}$$

If moreover (6) holds for Laguerre basis, then

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}}^{(D)} - \pi\|_n^2 \leq \|\pi - \pi_{\mathbf{m}}\|_f^2 + c\frac{m_1 \sqrt{m_2}}{n}, \tag{16}$$

where c is a positive constant.

The bound in (15) is obtained under weak conditions with explicit and optimal constants. It involves a bias term $\|\pi - \pi_{\mathbf{m}}\|_f^2$ and a variance term $m_1 L(m_2)/n$. Recall that $m_1 L(m_2) = \Phi_0^2 m_1 m_2$ for trigonometric basis ($\Phi_0^2 = 1$ for odd m_2) or Laguerre basis ($\Phi_0^2 = 2$), and $m_1 L(m_2) \leq c m_1 \sqrt{m_2}$ for Hermite basis. Consequently, the order of the variance is not the same for all bases.

Note that for any estimator $\widehat{\pi}_{\mathbf{m}}$, we can set $\widehat{\pi}_{\mathbf{m}}^+(x, y) = \sup(\widehat{\pi}_{\mathbf{m}}(x, y), 0) = \widehat{\pi}_{\mathbf{m}}(x, y) \mathbf{1}_{\widehat{\pi}_{\mathbf{m}}(x,y) \geq 0}$, and we have $\|\widehat{\pi}_{\mathbf{m}}^+ - \pi\|_n^2 \leq \|\widehat{\pi}_{\mathbf{m}}^+ - \pi\|_n^2$ so the $\widehat{\pi}_{\mathbf{m}}^+$ is well defined, nonnegative, and inherits of the risk bound proved for $\widehat{\pi}_{\mathbf{m}}$.

Remark. The variance order $m_1 \sqrt{m_2}/n$ in the Hermite case is coherent with the following facts:

- when estimating a regression function $b(\cdot)$ in a model $V_i = b(U_i) + \eta_i$, for i.i.d. centered η_i independent of U_i, from observations $(U_i, V_i)_{1 \leq i \leq n}$ with a least square projection estimator on the space S_{m_1} generated by $h_0, \ldots, h_{m_1-1}$, the resulting integrated variance is of order m_1/n, see Comte and Genon-Catalot (2020).
- when estimating a density f_U from n i.i.d. observations $U_1, \ldots, U_n$ with a projection estimator on S_{m_2} generated by $h_0, \ldots, h_{m_2-1}$, the integrated variance of the estimator is of order $\sqrt{m_2}/n$, see Comte and Genon-Catalot (2018);

The same kind of property can hold in the Laguerre basis under the additional condition $\mathbb{E}(1/\sqrt{Y}|X) < +\infty$ for the density estimator.

3.3. **Anisotropic rates.** In this setting, we obtain the following bound:

Proposition 2. *Assume that the density f of X is bounded, and that π belongs to $W_{\mathbf{s}}(A, L)$ with $\mathbf{s} = (s_1, s_2)$, see Section 2.3, and consider the estimator $\widehat{\pi}_{\mathbf{m}}^{(D)}$ defined by (10)-(12) under Assumption* **A1** *in Laguerre or Hermite basis. Then choosing, in the Laguerre basis $\mathbf{m}^o = (m_1^o, m_2^o)$ with*

$$m_1^o \propto n^{\frac{s_2}{s_1 s_2 + s_1 + s_2}} \text{ and } m_2^o \propto n^{\frac{s_1}{s_1 s_2 + s_1 + s_2}},$$

provides

$$\mathbb{E}(\|\widehat{\pi}_{\mathbf{m}^o}^{(D)} - \pi\|_n^2) = O(n^{-\frac{\bar{\mathbf{s}}}{\bar{\mathbf{s}}+2}}), \quad \frac{1}{\bar{\mathbf{s}}} = \frac{1}{2}\left(\frac{1}{s_1} + \frac{1}{s_2}\right).$$

If $s_1 = s_2 = s$, the rate becomes $n^{-\frac{s}{s+2}}$. If moreover (6) holds for Laguerre basis, or if the basis is the Hermite basis, then choosing

$$m_1^\star \propto n^{\frac{s_2}{s_1 s_2 + s_1/2 + s_2}} \text{ and } m_2^\star \propto n^{\frac{s_1}{s_1 s_2 + s_1/2 + s_2}}, \tag{17}$$

we obtain

$$\mathbb{E}(\|\widehat{\pi}_{\mathbf{m}^\star}^{(D)} - \pi\|_n^2) = O\left(n^{-\frac{1}{1+\frac{1}{s_1}+\frac{1}{2s_2}}}\right)$$

Note that these rates are different from the rates on periodic Sobolev spaces associated to the trigonometric basis (or on Besov spaces associated to piecewise polynomials basis), $n^{-2\bar{\alpha}/(2\bar{\alpha}+2)}$ for regularity $\alpha = (\alpha_1, \alpha_2)$, that we may also recover, see Brunel et al. (2007); a lower bound corresponding to this rate is proved in Lacour (2007).

We remark that, under the assumptions of Proposition 2, if $\pi \in WSS_{\mathbf{s},\lambda}(L)$, see (9), then choosing $m_i^\star \propto [\log(n) - (s_i + 3/2)\log\log(n)]/\lambda_i$ gives

$$\mathbb{E}(\|\widehat{\pi}_{\mathbf{m}^\star}^{(D)} - \pi\|_n^2) = O\left(\frac{\log^{3/2}(n)}{n}\right).$$

This is an almost parametric rate, which is classical over analytic classes for instance.

Proof of Proposition 2. We start from Inequality (15). For the bias term, we have, as f is upper bounded by $\|f\|_\infty < +\infty$, that

$$\|\pi - \pi_{\mathbf{m}}\|_f^2 \leq \|f\|_\infty \|\pi - \pi_{\mathbf{m}}\|_2^2.$$

We can then use regularity assumptions on π on Laguerre or Hermite Sobolev spaces to get the order of the bias term, with the result recalled in Section 2.3. This gives $\|\pi - \pi_{\mathbf{m}}\|_2^2 \leq L[m_1^{-s_1} + m_2^{-s_2}]$. Therefore

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}}^{(D)} - \pi\|_n^2 \leq C(m_1^{-s_1} + m_2^{-s_2} + \frac{m_1 m_2}{n}).$$

Let $\tau(m_1, m_2) = m_1^{-s_1} + m_2^{-s_2} + \frac{m_1 m_2}{n}$. Then solving in m_1, m_2 the sytem

$$\frac{\partial \tau(m_1, m_2)}{\partial m_1} = \frac{\partial \tau(m_1, m_2)}{\partial m_2} = 0$$

gives the first result. The second result is obtained analogously, by using the new variance order $m_1\sqrt{m_2}/n$. □

We mention that we may prove a risk bound measured in $\mathbb{L}^2(A, f(x)dxdy)$-norm, in the spirit of Comte and Genon-Catalot (2020), but this would require the so-called "stability condition" (see also Cohen et al. (2013)), namely

$$L(m_1)\|\Psi_{m_1}^{-1}\|_{\rm op} \le \mathfrak{d}_0 \frac{\log(n)}{n}, \tag{18}$$

for a well defined constant $\mathfrak{d}_0$; we omit this result for sake of conciseness. Condition (18) also appears in the model selection setting, see Section 6.

3.4. **Lower bound.** As the rate obtained in Proposition 2 is not standard, we need to check if it is optimal in some sense. The answer can be positive, but on Sobolev-Laguerre or Hermite regularity spaces taking the weighted norm into account.
More precisely, we consider as weighted Laguerre or Hermite Sobolev spaces regularity spaces the ones defined by:

$$W_{\mathbf{s}}^f(A,R) = \{g \in \mathbb{L}^2(A, f(x)dxdy), \forall \mathbf{m}, m_1, m_2 \ge 1, \|g - g_{\mathbf{m}}^{(f)}\|_f^2 \le R(m_1^{-s_1} + m_2^{-s_2})\} \tag{19}$$

where $g_{\mathbf{m}}^{(f)}$ is the orthogonal projection in $\mathbb{L}^2(A, f(x)dxdy)$ of g on $S_{\mathbf{m}} = S_{m_1} \otimes S_{m_2}$.

Note that the rates in Proposition 2 are unchanged by considering that π belongs to $W_{\mathbf{s}}^f(A,R)$, without requiring f bounded. On the other hand, for f bounded, $W_{\mathbf{s}}(A,L) \subset W_{\mathbf{s}}^f(A,R)$ with $R = L\|f\|_\infty$.

We assume that the regularity orders (s_1, s_2) are integer. We denote by $W_{s_1}(A_1, R)$ the univariate Laguerre-Sobolev or Hermite-Sobolev ball, where $A_1 = \mathbb{R}_+$ in the Laguerre case, and $A_1 = \mathbb{R}$ in the Hermite case.

Theorem 1. *Let R be a positive real and L be a large enough positive real. Then, for any density $f \in W_{s_1}(A_1, R) \cap \mathbb{L}^\infty(A_1)$, there exists a constant c such that for any estimator $\hat\pi_n$, $A = \mathbb{R}^2$ or $A = \mathbb{R}_+^2$ and for n large enough,*

$$\sup_{\pi \in W_{\mathbf{s}}^f(A,L)} \mathbb{E}_\pi\left[\|\hat\pi_n - \pi\|_f^2\right] \ge c\psi_n^2$$

where

$$\psi_n^2 = n^{-\frac{1}{1+\frac{1}{s_1}+\frac{1}{2s_2}}}$$

if, for $m_1^\star = \lfloor \psi_n^{-2/s_1} \rfloor$,

$$L(m_1^\star)\|\Psi_{m_1^\star}^{-1}\|_{\rm op} \le \psi_n^{-2}. \tag{20}$$

This result proves the optimality of the rate obtained in Proposition 2 for Hermite basis or Laguerre basis under (6).

Let us comment condition (20). This condition is stronger than the stability condition, which restricts the collection of models for the adaptive method and would appear for controlling the integrated risk instead of the empirical one: see Comte and Genon-Catalot (2020) where it is also proved that $\|\Psi_{m_1}^{-1}\|_{\rm op} \le cm_1^\beta$ if f has some polynomial decay. Recall that in the Laguerre case, $L(m_1) = 2m_1$ and in the Hermite case, $L(m_1) = K\sqrt{m_1}$. Therefore, if in addition $\|\Psi_{m_1}^{-1}\|_{\rm op} = cm_1^\beta$, then (20) is fulfilled if $\beta + 1 \le s_1$ in the Laguerre case and if $\beta + 1/2 \le s_1$ in the Hermite case.

4. Indirect Laguerre case

Now, the observations are (X_k, Z_k) with $Z_k = Y_k + \varepsilon_k$, $k = 1, \dots, n$, under Assumptions **A1** and **A2**. In this Section, we assume that $X_k \ge 0$, $Y_k \ge 0$, $\varepsilon_k \ge 0$ a.s., thus it is legit to use the Laguerre basis, defined on $\mathbb{R}_+$ only. This framework of non-negative variables can be found in many applications, in particular in survival analysis. Note in particular that ε is not centered. More precisely we assume

Assumption A3. *The distribution of the noise ε admits a density with respect to the Lebesgue measure, denoted by f_ε. Moreover $X \geq 0$, $Y \geq 0$, $\varepsilon \geq 0$ a.s.*

4.1. **Definition of the estimators in the noisy-Laguerre case.** In this context, computations rely on property (4), specifically fulfilled by the Laguerre functions, see also Comte et al. (2017) and Mabon (2017) in regression and density context respectively. First we denote by $\pi_{Z|X}(x,z)$ the conditional density of Z given X. We have

$$\pi_{Z|X}(x,z) = \int \pi(x, z-u) f_\varepsilon(u) du.$$

This means that we can estimate the conditional density of Z given X and then invert the convolution link to obtain the coefficients of π.

Let us define the matrix the $m_2 \times m_2$ lower triangular matrix $\mathbf{G}_{m_2} = (g_{j,k})_{0\leq j,k\leq m_2-1}$ with coefficients

$$g_{j,k} = \frac{1}{\sqrt{2}} \left(\langle f_\varepsilon, \ell_{j-k} \rangle \mathbf{1}_{j-k\geq 0} - \langle f_\varepsilon, \ell_{j-k-1} \rangle \mathbf{1}_{j-k-1\geq 0} \right).$$

The diagonal elements of $\mathbf{G}_{m_2}$ are $\langle f_\varepsilon, \ell_0 \rangle/\sqrt{2} = \int_0^{+\infty} f_\varepsilon(u) e^{-u} du > 0$. As a consequence $\mathbf{G}_{m_2}$ is invertible. Relying on equation (4), in this noisy model we find

$$\begin{aligned}
\pi_{Z|X}(x,z) &= \sum_{j\geq 0} \sum_{k\geq 0} \langle \pi_{Z|X}, \ell_j \otimes \ell_k \rangle \ell_j(x) \ell_k(z) \\
&= \sum_{j\geq 0} \sum_{k\geq 0} \sum_{p\geq 0} \langle \pi, \ell_j \otimes \ell_k \rangle \langle f_\varepsilon, \ell_p \rangle \ell_j(x) \int \ell_k(z-u) \ell_p(u) du \\
&= \sum_{j,k,p\geq 0} \langle \pi, \ell_j \otimes \ell_k \rangle \langle f_\varepsilon, \ell_p \rangle \ell_j(x) 2^{-1/2} \left(\ell_{p+k}(z) - \ell_{p+k+1}(z) \right) \\
&= \sum_{j,k\geq 0} \left(\sum_{p=0}^{k} 2^{-1/2} (\langle f_\varepsilon, \ell_{k-p} \rangle - \langle f_\varepsilon, \ell_{k-p-1} \rangle) \langle \pi, \ell_j \otimes \ell_p \rangle \right) \ell_j(x) \ell_k(z) \\
&= \sum_{j,k\geq 0} \left[(\langle \pi, \ell_j \otimes \ell_p \rangle)_{p\geq 0} \, {}^t\mathbf{G}_\infty \right]_k \ell_j(x) \ell_k(z) \qquad (21)
\end{aligned}$$

In other words, we have

$$\pi_{Z|X}(x,z) = \sum_{j,k\geq 0} \left(\sum_{p=0}^{k} \langle \pi, \ell_j \otimes \ell_p \rangle g_{k,p} \right) \ell_j(x) \ell_k(z).$$

The partial $\mathbb{L}^2$-projection on $S_{(\infty, m_2)}$ of $\pi_{Z|X}$ can thus be written

$$(\pi_{Z|X})_{(\infty,m_2)}(x,z) = \sum_{j\geq 0} \sum_{k=0}^{m_2-1} \left[(\langle \pi, \ell_j \otimes \ell_p \rangle)_{0\leq p\leq m_2-1} \, {}^t\mathbf{G}_{m_2} \right]_k \ell_j(x) \ell_k(z),$$

thanks to the triangular structure of $\mathbf{G}_{m_2}$. This explains why a two-step strategy gives in this basis:

$$\widehat{\pi}_{\mathbf{m}}^{(L)}(x,y) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \widehat{a}_{j,k}^{(L)} \ell_j(x) \ell_k(y), \quad \widehat{A}_{\mathbf{m}}^{(L)} = (\widehat{a}_{j,k}^{(L)})_{0\leq j\leq m_1-1, 0\leq k\leq m_2-1}$$

with

$$\widehat{A}_{\mathbf{m}}^{(L)} = \frac{1}{n} \widehat{\Psi}_{m_1}^{-1} \, {}^t\widehat{\Phi}_{m_1} \widehat{\Theta}_{m_2}(\mathbf{Z}) \, {}^t\mathbf{G}_{m_2}^{-1}, \quad \text{with} \quad \widehat{\Theta}_m(\mathbf{Z}) = (\ell_j(Z_i))_{1\leq i\leq n, 0\leq j\leq m-1} \qquad (22)$$

where $\widehat{\Psi}_{m_1}$ and $\widehat{\Phi}_{m_1}$ are defined by (11).

4.2. **Bound on the empirical MISE of $\widehat{\pi}_{\mathbf{m}}^{(L)}$.** Let us note that

$$\mathbb{E}[\widehat{A}_{\mathbf{m}}^{(L)}|\mathbf{X}] = \frac{1}{n}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\mathbb{E}[\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}]\,{}^t\mathbf{G}_{m_2}^{-1}.$$

A first useful property is given by the lemma:

Lemma 3. *We have* $\mathbb{E}[\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}]\,{}^t\mathbf{G}_{m_2}^{-1} = \mathbb{E}[\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X}]$ *and thus* $\mathbb{E}[\widehat{\pi}_{\mathbf{m}}^{(L)}|\mathbf{X}] = \pi_{\mathbf{m},n}$.

Thanks to this result, we can prove the following risk bound (see Section 8.6).

Proposition 3. *Assume that Assumptions* **A1**, **A2** *and* **A3** *hold. Then the estimator* $\widehat{\pi}_{\mathbf{m}}^{(L)}$ *defined by (22) satisfies:*

$$\mathbb{E}(\|\widehat{\pi}_{\mathbf{m}}^{(L)} - \pi\|_n^2) \le \|\pi - \pi_{\mathbf{m}}\|_f^2 + \frac{\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2\, m_1 L(m_2)}{n}$$

where $L(m_2) = 2m_2$ *here. If in addition the condition*

$$\exists C > 0, \forall x, \quad \mathbb{E}\left(\frac{1}{\sqrt{Z}}|X = x\right) < C \tag{23}$$

holds, then

$$\mathbb{E}(\|\widehat{\pi}_{\mathbf{m}}^{(L)} - \pi\|_n^2) \le \|\pi - \pi_{\mathbf{m}}\|_f^2 + C\frac{\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2\, m_1\sqrt{m_2}}{n}. \tag{24}$$

As $\mathbf{G}_{m_2}$ is lower triangular, its eigenvalues are given by the diagonal terms, which are all equal to $2^{-1/2}\langle f_\varepsilon, \ell_0\rangle = \int_{\mathbb{R}^+} e^{-u} f_\varepsilon(u)du \le 1$. Therefore

$$\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2 = \lambda_{\max}(\mathbf{G}_{m_2}^{-1}\,{}^t\mathbf{G}_{m_2}^{-1}) \ge [\lambda_{\max}(\mathbf{G}_{m_2}^{-1})]^2 \ge 1.$$

Therefore, as expected, the variance order in the inverse problem increases compared to the variance order in the direct case. Moreover, it is proved in Lemma 3.4 of Mabon (2017) that $m_2 \mapsto \|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2$ is increasing.

Note that condition (23) holds if condition (5) holds or if $\mathbb{E}(1/\sqrt{\varepsilon})$ is finite.

4.3. **Rates in the noisy-Laguerre case.** Now let us assess the order of the variance term and more specifically of $\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2$. Comte et al. (2017) show that we can recover the order of this spectral norm, under the conditions on the density f_ε. First we define an integer $\alpha \ge 1$ such that

$$\frac{d^j}{dx^j} f_\varepsilon(x)|_{x=0} = \begin{cases} 0 & \text{if} \quad j = 0, 1, \dots, \alpha - 2 \\ B_\alpha \neq 0 & \text{if} \quad j = \alpha - 1. \end{cases}$$

Consider the two following assumptions:

Assumption A4.

(1) $f_\varepsilon \in \mathbb{L}^1(\mathbb{R}_+)$ *is* α *times differentiable and* $f_\varepsilon^{(\alpha)} \in \mathbb{L}^1(\mathbb{R}_+)$.
(2) *The Laplace transform of* f_ε, $z \mapsto \mathbb{E}\left[e^{-z\varepsilon}\right]$ *has no zero with non-negative real part except for the zeros of the form* $\infty + ib$, $b \in \mathbb{R}$.

It follows from Comte et al. (2017) that, under Assumptions **A4**, there exists positive constants C and C' such that:

$$C' m_2^{2\alpha} \le \|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2 \le C m_2^{2\alpha}.$$

For example a Gamma distribution with $\Gamma(p,\theta)$ satisfies Assumptions **A4** for $\alpha = p$ ($\alpha = 1$ for an exponential distribution). If f_ε follows a $\beta(a,b)$ and $b > a$, then $\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2 = O(m_2^{2a})$ (see Mabon (2017)). On the contrary an Inverse Gamma distribution does not satisfy Assumptions **A4** because there exists no value of α such that the derivative is nonzero at 0.

These assumptions allow to deduce from Proposition 3 the following rates of convergence of the estimator.

Proposition 4. *Assume that f is bounded. Under Assumptions* **A1**–**A4**, *for $\pi \in W_{\mathbf{s}}(\mathbb{R}_+^2, L)$, and $\mathbf{m}^\star = (m_1^\star, m_2^\star)$ such that*

$$m_1^\star \propto n^{s_2/[(2\alpha+1)s_1+s_2+s_1s_2]} \quad \text{and} \quad m_2^\star \propto n^{s_1/[(2\alpha+1)s_1+s_2+s_1s_2]}$$

then

$$\mathbb{E}[\|\widehat{\pi}_{\mathbf{m}^\star}^{(L)} - \pi\|^2] \leq C(\mathbf{s}, L, \|f\|_\infty) n^{-1/[\frac{2\alpha+1}{s_2}+\frac{1}{s_1}+1]}.$$

5. Indirect Hermite case

Now we consider the general case where we observe $(X_i, Z_i)_{1\leq i\leq n}$ from $Z_i = Y_i + \varepsilon_i$ and all variables take values in $\mathbb{R}$. Then, we define the estimator in the Hermite basis, and use standard deconvolution methods in the y-direction, while still the regression strategy in the x-direction.

5.1. Assumption related to the noise.

We denote by f_ε^* the characteristic function of the noise ε:

$$\forall u \in \mathbb{R} \qquad f_\varepsilon^*(u) = \mathbb{E}[e^{-iu\varepsilon}].$$

The following assumptions are required for f_ε^*:

Assumption A5.

(1) *Function f_ε^* never vanishes, i.e. $\forall u \in \mathbb{R}, f_\varepsilon^*(u) \neq 0$.*
(2) *There exist $\alpha \in \mathbb{R}$, $\beta > 0, 0 \leq \gamma \leq 2$, ($\alpha > 0$ if $\gamma = 0$), $\beta < \xi$ if $\gamma = 2$ for ξ defined in (8), and $k_0, k_1 > 0$ such that $\forall u \in \mathbb{R}$,*

$$k_0(u^2+1)^{-\alpha/2} \exp(-\beta|u|^\gamma) \leq |f_\varepsilon^*(u)| \leq k_1(u^2+1)^{-\alpha/2} \exp(-\beta|u|^\gamma)$$

If $\gamma = 0$, the noise is called ordinary smooth, and super smooth for $\gamma > 0, \beta > 0$. For instance, Laplace or Gamma distributions are ordinary smooth. On the other hand, Gaussian or Cauchy noises are supersmooth.

Remark. If f_ε is a density, it is known that $\gamma \leq 2$ (at least for $\alpha = 0$).[1]

5.2. Definition of the contrasts in the noisy-Hermite case.

To begin with, we recall that the Fourier transform t^* of $t \in S_m$ is defined by

$$t^*(u) = \int e^{ixu} t(x) dx.$$

For a bivariate function $T \in S_{\mathbf{m}}$, we denote by $T^{(*,2)}$ the Fourier transform with respect to the second variable :

$$T^{(*,2)}(x,u) = \int e^{iyu} T(x,y) dy.$$

Definition 2. *For any function $t \in S_m$, we denote by v_t the inverse Fourier transform of $t^*/f_\varepsilon^*(-.)$, i.e.*

$$v_t(x) = \frac{1}{2\pi} \int e^{-ixu} \frac{t^*(u)}{f_\varepsilon^*(-u)} du.$$

For any bivariate function $T \in S_{\mathbf{m}}$, we denote by Φ_T the following bivariate function

$$\Phi_T(x,z) = \frac{1}{2\pi} \int e^{-iuz} \frac{T^{(*,2)}(x,u)}{f_\varepsilon^*(-u)} du$$

We can also write $\Phi_T^{(*,2)}(x,u) = T^{(*,2)}(x,u)/f_\varepsilon^*(-u)$.

---

[1] According to Lukacs (1970), Theorem 4.1.1, the only characteristic function ϕ with $\phi(u) = 1 + o(u^2)$, as $u \to 0$, is the function $\phi(u) = 1$ for all u. This rules out characteristic functions of the form $e^{-\beta|u|^\gamma}$ with $\gamma > 2$. This implies that if $|f_\varepsilon^*(u)|^2 = c\exp(-2\beta|u|^\gamma)$ then necessarily $\gamma \leq 2$. Indeed, $|f_\varepsilon^*(u)|^2$ is the characteristic function of a probability density function (it is a characteristic function of $\varepsilon_1 - \varepsilon_1'$ where ε_1' is an independent copy of ε_1).

Note that v_{h_k} is well defined for all ordinary smooth noise distributions and for a wide range of super-smooth distributions also, thanks to property (8) of the Hermite basis and Assumption **A5**-(2). Moreover, the operators v and Φ are linked via the formula

$$\Phi_{t\otimes s}(x,y) = t(x)v_s(y), \quad \Phi_{h_j\otimes h_k}(x,y) = h_j(x)v_{h_k}(y)$$

and are helpful because of the following properties.

$$\forall t \in S_m, \qquad \mathbb{E}[v_t(Z_1)|Y_1] = t(Y_1) \text{ and } \mathbb{E}[v_t(Z_1)|X_1] = \int t(z)\pi(X_1,z)dz,$$

$$\forall T \in S_{\mathbf{m}} \qquad \mathbb{E}[\Phi_T(X_1,Z_1)|X_1] = \mathbb{E}[T(X_1,Y_1)|X_1] = \int T(X_1,z)\pi(X_1,z)dz.$$

Now we can define our estimators by:

$$\widehat{\pi}_{\mathbf{m}}^{(H)} = \arg\min_{T\in S_{\mathbf{m}}} \gamma_n^{(H)}(T), \tag{25}$$

with the following contrast $\gamma_n^{(H)}$:

$$\gamma_n^{(H)}(T) = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{\mathbb{R}} T^2(X_i,y)dy - 2\Phi_T(X_i,Z_i)\right]. \tag{26}$$

The interest of this contrast can be easily understood by the computation of $\mathbb{E}[\gamma_n^{(H)}(T)]$. Indeed, using the previous properties, we can write

$$\begin{aligned}\mathbb{E}[\gamma_n^{(H)}(T)] &= \mathbb{E}[\int T^2(X,y)dy - 2\Phi_T(X,Z)] = \iint T^2(x,y)f(x)dxdy - 2\mathbb{E}[T(X,Y)] \\ &= \iint [(T(x,y)-\pi(x,y))^2 - \pi^2(x,y)]f(x)dxdy = \|T-\pi\|_f^2 - \|\pi\|_f^2.\end{aligned}$$

We obtain the following new estimator of π:

$$\widehat{\pi}_{\mathbf{m}}^{(H)}(x,y) = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} \widehat{a}_{j,k}^{(H)} h_j(x)h_j(y) \quad \widehat{A}_{\mathbf{m}}^{(H)} = (\widehat{a}_{j,k}^{(H)})_{0\le j\le m_1-1, 0\le k\le m_2-1}$$

with

$$\widehat{A}_{\mathbf{m}}^{(H)} = \frac{1}{n}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\widehat{\Upsilon}_{m_2}(\mathbf{Z}), \text{ with } \widehat{\Upsilon}_{m_2}(\mathbf{Z}) = \big(v_{h_j}(Z_i)\big)_{1\le i\le n, 0\le j\le m_2-1}, \tag{27}$$

with $\widehat{\Psi}_{m_1}$, $\widehat{\Phi}_{m_1}$ still defined by (11).

Note that if $X_i \in \mathbb{R}_+$ and $Y_i \in \mathbb{R}$ we may use a product basis $(\ell_j \otimes h_k)_{j,k}$ for estimation purpose. The formulae above would still hold.

5.3. **Bound on the empirical MISE of $\widehat{\pi}_{\mathbf{m}}^{(H)}$ and rates.** Here we can prove the following bound:

Proposition 5. *Under Assumptions* **A1**, **A2** *and* **A5**, *we have* $\mathbb{E}[\widehat{\pi}_{\mathbf{m}}^{(H)}|\mathbf{X}] = \pi_{\mathbf{m},n}$ *and*

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}}^{(H)} - \pi\|_n^2 \le \|\pi - \pi_{\mathbf{m}}\|_f^2 + \frac{m_1\Delta(m_2)}{n}, \quad \text{where } \Delta(m_2) = \frac{1}{\pi}\left(4\int_{|u|\le\sqrt{2m_2}} \frac{du}{|f_\varepsilon^*(u)|^2} + \mathfrak{c}\right)$$

and $\mathfrak{c}$ *is a constant only depending on* ξ *(see* (8)*) and on* f_ε^*.

Note that, under Assumption **A5**, we can compute that

$$\Delta(m_2) \le k m_2^{\alpha+\frac{1-\gamma}{2}} \exp[2\beta(2m_2)^{\gamma/2}].$$

By computations similar to the ones to prove Proposition 2, we obtain the following rates.

Proposition 6. *Assume that f is bounded and Assumptions* **A1**, **A2** *and* **A5** *hold. Let $\pi \in W_{\mathbf{s}}(\mathbb{R}^2, L)$.*

(1) *If $\gamma = 0$, then for $m_i^\star \propto n^{s_i/[(\alpha+1/2)s_1+s_2(s_1+1)]}$, $i = 1, 2$, we have*

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}^\star}^{(H)} - \pi\|_n^2 \le Cn^{-\frac{1}{1+\frac{\alpha+1/2}{s_2}+\frac{1}{s_1}}}.$$

(2) *If $\gamma, \beta > 0$, then for $m_1^\star = (\log n)^{2s_2/(\gamma s_1)}$ and $m_2^\star = (1/2)\,(\log n/4\beta)^{2/\gamma}$, we have*

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}^\star}^{(H)} - \pi\|_n^2 \le C\,(\log n)^{-2s_2/\gamma}.$$

Let $\pi \in WSS_{\mathbf{s},\lambda}(L)$, see (9), and $\gamma = 0$, then for $m_i^\star = [\log(n) - (\alpha + s_i)\log\log(n)]/\lambda_i$, $i = 1, 2$

$$\mathbb{E}\|\widehat{\pi}_{\mathbf{m}^\star}^{(H)} - \pi\|_n^2 \le C\frac{(\log n)^{1+\alpha}}{n}.$$

Here we find a usual phenomenon in deconvolution: if the noise is supersmooth and the target function is only Sobolev, the rate of convergence is logarithmic. For more details about the rates see Comte and Lacour (2013).

6. Adaptive estimators with Goldenschluger-Lepski method

In the previous sections, we have described collections of estimators $\hat{\pi}_{\mathbf{m}}$ and computed their rates of convergence for optimal $\mathbf{m} = \mathbf{m}^\star$. Nevertheless these values of $\mathbf{m}^\star$ depend of the smoothness $\mathbf{s}$ of the unknown conditional density π. Now we aim at selecting $\mathbf{m}$ in a purely data-driven way. In this section, we define adaptive estimators of the conditional density for the three settings described previously, and prove a risk bound for them, showing that they realize the adequate compromise between bias and variance.

More precisely, we define the collections of models and estimators, and give a general result with superscript $^{(\mathfrak{S}up)}$ where $(\mathfrak{S}up) = (D)$ (direct case), or $(\mathfrak{S}up) = (L)$ (Laguerre-noisy case) or $(\mathfrak{S}up) = (H)$ (Hermite-noisy case).

6.1. **Collection of models.** First we define

$$V^{(D)}(\mathbf{m}) = K_0\frac{m_1 L(m_2)}{n}, V^{(L)}(\mathbf{m}) = K_0\frac{m_1 L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\text{op}}^2}{n}, V^{(H)}(\mathbf{m}) = K_0\frac{m_1\Delta(m_2)}{n}, \tag{28}$$

where K_0 is a numerical constant ($K_0 = 12(1+\epsilon)$ for $\epsilon > 0$ suits, from the proof here).

These terms are of order of the variance of the corresponding estimators in the trigonometric ($L(m_2) = m_2$ for odd m_2) or in the Hermite case ($L(m_2) = K\sqrt{m_2}$). For the Laguerre case, we have $L(m_2) = 2m_2$ while we know that, under condition (5) the optimal order is $\sqrt{m_2}$. These quantities will be used as penalty in a criterion to be minimized.

Then we consider the following collection of models

$$\mathcal{M}_n^{(\mathfrak{S}up)} = \left\{\mathbf{m} \in \{1, \dots, n\}^2, V^{(\mathfrak{S}up)}(\mathbf{m}) \le 1,\ \ L(m_1)\|\Psi_{m_1}^{-1}\|_{\text{op}} \le \frac{\mathfrak{d}^\star}{2}\frac{n}{\log^2(n)}\right\}$$

where $\mathfrak{d}^\star$ a well-chosen numerical constant such that $\mathfrak{d}^\star/\log(n) \le \mathfrak{d}$ with $\mathfrak{d} = (3\log(3/2) - 1)/10$ and $\mathfrak{d}^\star \le \epsilon C(\epsilon^2)/42$, $C(\epsilon^2) = \min(\sqrt{1+\epsilon^2} - 1, 1)$. Recall that $\Psi_{m_1} = \mathbb{E}(\widehat{\Psi}_{m_1}) = (\langle \varphi_j, \varphi_k\rangle_f)_{0 \le j,k \le m_1 - 1}$. Moreover, note that for a non-zero vector $\mathbf{x} = {}^t(x_0, \dots, x_{m_1-1}) \in \mathbf{R}^{m_1}$, then ${}^t\mathbf{x}\Psi_{m_1}\mathbf{x} = \int(\sum_{j=0}^{m_1-1} x_j\varphi_j(x))^2 f(x)dx > 0$ under our assumptions, so that Ψ_{m_1} is invertible.

We also introduce its empirical random counterpart

$$\widehat{\mathcal{M}}_n^{(\mathfrak{S}up)} = \{\mathbf{m} \in \{1, \dots, n\}^2, V^{(\mathfrak{S}up)}(\mathbf{m}) \le 1,\ L(m_1)\|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}} \le \mathfrak{d}^\star\frac{n}{\log^2(n)}\}. \tag{29}$$

Note that $m_1 \mapsto L(m_1)\|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}$ is increasing, and $\mathbf{m} \mapsto V^{(\mathfrak{S}\mathrm{up})}(\mathbf{m})$ also, with respect to each variable. Thus both collection are such that, if they contain $\mathbf{m}$ and $\mathbf{m}'$, then they also contain

$\mathbf{m} \wedge \mathbf{m}'$ defined as component-wise minimum.

Comments.
1. The definition of the collection of models involves two constraints. The first one is standard and means that the variance remains bounded. As this term is known, it is the same for the two sets, $\mathcal{M}_n^{(\mathfrak{S}up)}$ and $\widehat{\mathcal{M}}_n^{(\mathfrak{S}up)}$. The second constraint must be compared to the so-called "stability condition" introduced by Cohen et al. (2013), Cohen et al. (2019) and also used in Comte and Genon-Catalot (2020): $L(m_1)\|\Psi_{m_1}^{-1}\|_{\text{op}} \leq \frac{\mathfrak{d}}{2}\frac{n}{\log(n)}$. Obviously, it is here slightly reinforced. However, when dealing with adaptive estimation, Comte and Genon-Catalot (2020) had a stronger condition: $L(m_1)\|\Psi_{m_1}^{-1}\|_{\text{op}}^2 \leq \mathfrak{d}^\star \frac{n}{\log(n)}$. The improvement here is substantial, specifically for non campactly supported bases where $\|\Psi_{m_1}^{-1}\|_{\text{op}}$ can be large. This is due to conditional preliminary result. As the matrix Ψ_m is unknown, it has to be replaced by its empirical version and leads to a random model collection $\widehat{\mathcal{M}}_n^{(\mathfrak{S}up)}$.
2. Let us now mention specific properties in the direct case. If the support of the basis used for estimation along x is compact, say $[0,1]$, then we can assume that $f(x) \geq f_0, \forall x \in [0,1]$. In that case $\|\Psi_{m_1}^{-1}\|_{\text{op}} \leq 1/f_0$, see Comte and Genon-Catalot (2020). The collection of models no longer need to involve neither $\|\Psi_{m_1}^{-1}\|_{\text{op}}$ nor $\|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}$ and reduces to the standard one, and is similar to Brunel et al. (2007). However, the penalty we obtain here, $V^{(D)}(m)$, does not depend on f_0 nor $\|\pi\|_\infty$, and this is an important improvement compared to this work.

Now we present constraints on the model collection.

Case (D). Assume that, for any $c_1 > 0$, there exists $\Sigma > 0$ such that

$$\sum_{\mathbf{m} \in \{1,\ldots,n\}^2} e^{-c_1 m_1 L(m_2)} \leq \Sigma < +\infty. \tag{30}$$

Case (L). Assume that, for any $c_1 > 0$, there exists $\Sigma > 0$ such that

$$\sum_{\mathbf{m}} \|\mathbf{G}_{m_2}^{-1}\|_{\text{op}}^2 e^{-c_1 m_1 L(m_2)} \leq \Sigma < +\infty \tag{31}$$

Case (H). Let $\delta(m_2) := \sup_{|z| \leq \sqrt{2m_2}} \dfrac{1}{|f_\varepsilon^*(z)|^2} + \mathfrak{c}$, and assume that, for any $c_1 > 0$, there exists $\Sigma > 0$ such that

$$\sum_{\mathbf{m}} \delta(m_2) \exp\left(-c_1 m_1 \frac{\Delta(m_2)}{\delta(m_2)}\right) \leq \Sigma < +\infty. \tag{32}$$

Let us comment these conditions. First, condition (30) if fulfilled for all our bases.
Under Assumption **A4**, $\|\mathbf{G}_{m_2}^{-1}\|_{\text{op}}^2 = O(m_2^{2\alpha})$ and condition (31) is fulfilled.
Now we discuss condition (32). In the ordinary smooth case where $\delta = \gamma = 0$,

$$\delta(m_2)\exp(-c_1 m_1 \frac{\Delta(m_2)}{\delta(m_2)}) \sim m_2^\alpha \exp(-c_1 m_1 \sqrt{m_2})$$

is indeed summable and condition (32) is fulfilled. In the super-smooth case, with Lemma 1 in Comte and Lacour (2013), $\Delta(m_2) \sim C m_2^{\alpha+(1-\gamma)/2} \exp(\beta m_2^{\gamma/2})$; then condition (32) is fulfilled if $\gamma < 1/2$; otherwise, the penalty must be slightly changed, see Comte and Lacour (2013).

6.2. **General adaptive estimator and result.**

Assumption A6. *The conditional density π of Y given X is bounded on $\mathbb{R}^2$.*

For $\mathbf{m} = (m_1, m_2)$ and $\mathbf{m}' = (m_1', m_2')$, we define $S_{\mathbf{m}\wedge\mathbf{m}'} := (S_{m_1} \cap S_{m_1'}) \otimes (S_{m_2} \cap S_{m_2'})$ where $S_{m_i \wedge m_i'} := S_{m_i} \cap S_{m_i'}$ is well defined with trigonometric, Laguerre and Hermite bases, which are our leading examples. These collections are regular and nested in each direction, with at most one model for each m_i. Thus $S_{\mathbf{m}\wedge\mathbf{m}'}$ is well defined, and we denote by $\widehat{\pi}^{(\mathfrak{S}up)}_{\mathbf{m},\mathbf{m}'}$ the minimum contrast estimator on $S_{\mathbf{m}\wedge\mathbf{m}'}$.

We propose a model selection relying on the strategy initiated by Goldenshluger and Lepski (2011) adapted to model selection in the spirit of Chagny (2013). Let then

$$A^{(\mathfrak{S}up)}(\mathbf{m}) = \sup_{\mathbf{m}' \in \widehat{\mathcal{M}}_n^{(\mathfrak{S}up)}} \left(\|\widehat{\pi}^{(\mathfrak{S}up)}_{\mathbf{m}',\mathbf{m}} - \widehat{\pi}^{(\mathfrak{S}up)}_{\mathbf{m}'}\|_n^2 - V^{(\mathfrak{S}up)}(\mathbf{m}') \right)_+$$

with $V^{(\mathfrak{S}up)}(\mathbf{m}')$ defined by (28) and $a_+ = \max(a, 0)$ denotes the positive part of a. We select the model $\mathbf{m}$ with the following rule

$$\hat{\mathbf{m}}^{(\mathfrak{S}up)} = \arg \min_{\mathbf{m} \in \widehat{\mathcal{M}}_n^{(\mathfrak{S}up)}} \left\{ A^{(\mathfrak{S}up)}(\mathbf{m}) + V^{(\mathfrak{S}up)}(\mathbf{m}) \right\}.$$

Our final estimator is

$$\tilde{\pi}^{(\mathfrak{S}up)} = \widehat{\pi}^{(\mathfrak{S}up)}_{\hat{\mathbf{m}}^{(\mathfrak{S}up)}}.$$

The first result is obtained conditionally on $X_1, \dots, X_n$.

Theorem 2. *Assume that Assumption* **A1** *and* **A6** *hold. Assume that condition (30) for* $(\mathfrak{S}up) = (D)$*, Assumptions* **A2**, **A3** *and condition (31) for* $(\mathfrak{S}up) = (L)$ *and Assumptions* **A2**, **A5** *and condition (32) for* $(\mathfrak{S}up) = (H)$ *hold. Then, for any* $\mathbf{m} \in \widehat{\mathcal{M}}_n^{(\mathfrak{S}up)}$*, we have a.s.*

$$\mathbb{E}\left[\|\pi - \tilde{\pi}^{(\mathfrak{S}up)}\|_n^2 \,|\mathbf{X} \right] \le C \inf_{\mathbf{m} \in \widehat{\mathcal{M}}_n^{(\mathfrak{S}up)}} \{ \|\pi - \pi_{\mathbf{m},n}\|_n^2 + V^{(\mathfrak{S}up)}(\mathbf{m}) \} + \frac{C'}{n}, \tag{33}$$

where C is a numerical constant and C' is a constant which depends on $\|\pi\|_\infty$, Σ, but not on $(X_1, \dots, X_n)$ nor on n.

The same assumptions and the method of proof used in the direct case lead to the following non conditional result.

Corollary 1. *Under the Assumptions of Theorem 2, for any $\mathbf{m} \in \mathcal{M}_n^{(\mathfrak{S}up)}$, we have*

$$\mathbb{E}\|\pi - \tilde{\pi}^{(\mathfrak{S}up)}\|_n^2 \le C \inf_{\mathbf{m} \in \mathcal{M}_n^{(\mathfrak{S}up)}} \{ \|\pi - \pi_{\mathbf{m}}\|_f^2 + V^{(\mathfrak{S}up)}(\mathbf{m}) \} + \frac{C''}{n}, \tag{34}$$

where C is a numerical constant and C'' is a constant which depends on $\|\pi\|_\infty$, Σ.

Inequality (34) states that the estimator is adaptive in the sense it performs a squared-bias/variance compromise over the collection $\mathcal{M}_n^{(\mathfrak{S}up)}$, up to the multiplicative constant C and the additive negligible term C"$/n$. In the direct case (D), and for compactly supported basis along x, optimal rates are then automatically reached under $f(x) \ge f_0 > 0$ for x in the support, see section 3.3. Compared to previous results, we mention that the penalty term does not depend on f_0 nor $\|f\|_\infty$. Moreover, the additional novelty is that more general non compact supports are admitted, with size of the model collection depending on $\|\Psi_{m_1}^{-1}\|_{\text{op}}$. The optimal rate may not be reached, depending on the order of this term. We emphasize that the results obtained in the noisy cases are entirely new.

Note that a compactly supported basis case be used in x and the Hermite basis for deconvolving in y, even if this would make the bias term of particular feature.

7. Concluding remarks

We have proposed in this paper adaptive estimation method for the conditional density of Y given $X = x$, when the observations are $(X_i, Y_i)_{1\le i\le n}$ so-called direct observations, or $(X_i, Z_i)_{1\le i\le n}$ with $Z_i = Y_i + \varepsilon_i$ so-called noisy observations. The difficulty in the noisy case, is to use the same basis in the two directions, the regression direction in x and the density direction with deconvolving in y. Indeed, until recently, efficient regression methods with projection spaces rely on compactly supported bases, while deconvolution requires Fourier transforms and inversions which are more convenient with non compact support. This is why we first studied conditional density estimation in the direct case with possibly non compactly supported bases, which, thanks to the ideas in Comte and Genon-Catalot (2020) and Goldenshluger and Lepski (2011) conducted to new risk bounds for simple (fixed projection space) and adaptive estimator. Then, our two extensions to noisy cases, either with $\mathbb{R}^+$-supported or with real valued variables, lead to new estimators and risk bounds.

Now, the most standard basis for deconvolution is the sinus cardinal basis, $\varphi_{m,j} = \sqrt{m}\varphi(mx-j)$ for $j \in \mathbb{Z}$, and $\varphi(x) = \sin(\pi x)/(\pi x)$, and the question of using this basis in regression setting remains unsolved. Another extension would be to take into account multidimensional covariates; this has been studied in deconvolution setting with Laguerre basis in Dussap (2021), but the regression context is to be considered.

8. Proofs

8.1. **Proof of Lemma 1.** We first use (7) to write

$$\sum_{j=0}^{m-1} h_j^2(x) = \frac{1}{2\pi}\sum_{j=0}^{m-1} |h_j^*(x)|^2$$

Now by splitting $h_j^*(x) = \int_{|u|\le\sqrt{2m+1}} e^{iux} h_j(u)du + \int_{|u|>\sqrt{2m+1}} e^{iux} h_j(u)du$ and using (8), we get, for $j \le m-1$,

$$|h_j^*(x)|^2 \le 2\langle h_j, e^{i\cdot x}\mathbf{1}_{|.|\le\sqrt{2m+1}}\rangle^2 + 2C\int_{|u|>\sqrt{2m+1}} e^{-\xi u^2}du.$$

Thus

$$\sum_{j=0}^{m-1} h_j^2(x) \le 2\|\mathbf{1}_{|.|\le\sqrt{2m+1}}\|_2^2 + 2Cme^{-\xi(2m+1)/2}\int e^{-\xi u^2/2}du = 2\sqrt{2m+1} + \frac{2\sqrt{2\pi}C}{\sqrt{\xi}}Cme^{-\xi(2m+1)/2}.$$

This implies the result of Lemma 1 with $K = K(C,\xi)$.□

8.2. **Proof of Lemma 2.** We compute $\pi_{\mathbf{m},n}$, the orthogonal projection of π w.r.t. the empirical scalar product. We have

$$\pi_{\mathbf{m},n}(X_i, y) = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} [\mathbf{D_m}]_{j,k}\varphi_j(X_i)\varphi_k(y)$$

where $\mathbf{D_m}$ is such that $\langle \pi_{\mathbf{m},n} - \pi, \varphi_j \otimes \varphi_k\rangle_n = 0$ for $0 \le j \le m_1-1$ and $0 \le k \le m_2-1$. Therefore writing that the terms

$$\begin{aligned}\langle \pi, \varphi_j \otimes \varphi_k\rangle_n &= \frac{1}{n}\sum_{i=1}^n \int \pi(X_i, y)\varphi_j(X_i)\varphi_k(y)dy = \frac{1}{n}\sum_{i=1}^n \varphi_j(X_i)\int \pi(X_i, y)\varphi_k(y)dy \\ &= \frac{1}{n}\left[\,{}^t\widehat{\Phi}_{m_1}\left(\int \pi(X_i, y)\varphi_k(y)dy\right)_{1\le i\le n, 0\le k\le m_2-1}\right]_{j,k},\end{aligned}$$

and

$$\begin{aligned}\langle \pi_{\mathbf{m},n}, \varphi_j \otimes \varphi_k \rangle_n &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j'=0}^{m_1-1}\sum_{k'=0}^{m_2-1}[\mathbf{D_m}]_{j',k'}\varphi_j(X_i)\varphi_{j'}(X_i)\int \varphi_k(y)\varphi_{k'}(y)dy \\ &= \sum_{j'=0}^{m_1-1}[\mathbf{D_m}]_{j',k}\frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i)\varphi_{j'}(X_i) = \sum_{j'=0}^{m_1-1}[\mathbf{D_m}]_{j',k}[\widehat{\Psi}_{m_1}]_{j,j'} = \left[\widehat{\Psi}_{m_1}\mathbf{D_m}\right]_{j,k}\end{aligned}$$

are equal, implies formula (14). The last part of the result follows from

$$\left(\int \varphi_k(y)\pi(X_i,y)dy\right)_{1\le i\le n, 0\le k\le m_2-1} = \mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X}\right), \quad \mathbf{X} = (X_1,\dots,X_n),$$

where $\Theta_{m_2}(\mathbf{Y})$ is defined by (12). □

8.3. **Proof of Proposition 1.** We start from equation (13). By elementary algebraic computation, we find

$$\begin{aligned}\|\widehat{\pi}_{\mathbf{m}} - \pi_{\mathbf{m},n}\|_n^2 &= \frac{1}{n}\sum_{i=1}^{n}\int \left(\widehat{\pi}_{\mathbf{m}}(X_i,y) - \pi_{\mathbf{m},n}(X_i,y)\right)^2 dy \\ &= \frac{1}{n}\sum_{i=1}^{n}\int \left(\sum_{j,k}([\widehat{A}_{\mathbf{m}}]_{j,k} - [\mathbf{D_m}]_{j,k})\varphi_j(X_i)\varphi_k(y)\right)^2 dy \\ &= \frac{1}{n}\sum_{k=0}^{m_2-1}\sum_{i=1}^{n}\left(\sum_{j=0}^{m_1-1}([\widehat{A}_{\mathbf{m}}]_{j,k} - [\mathbf{D_m}]_{j,k})\varphi_j(X_i)\right)^2 \\ &= \mathrm{Tr}\left[\, {}^t(\widehat{A}_{\mathbf{m}} - \mathbf{D_m})\widehat{\Psi}_{m_1}(\widehat{A}_{\mathbf{m}} - \mathbf{D_m})\right].\end{aligned}$$

Replacing the matrix coefficients by their formula, we get

$$\|\widehat{\pi}_{\mathbf{m}} - \pi_{\mathbf{m},n}\|_n^2 = \frac{1}{n^2}\mathrm{Tr}\left[\, {}^t\!\left(\widehat{\Theta}_{m_2}(\mathbf{Y}) - \mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X}\right)\right)\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\, {}^t\widehat{\Phi}_{m_1}\left(\widehat{\Theta}_{m_2}(\mathbf{Y}) - \mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X}\right)\right)\right].$$

Then

$$\mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{m_2-1}\mathbb{E}\left[\left(\varphi_j(Y_i) - \mathbb{E}(\varphi_j(Y_i)|X_i)\right)^2|X_i\right][\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\, {}^t\widehat{\Phi}_{m_1}]_{i,i}. \tag{35}$$

Now, note that $[\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\, {}^t\widehat{\Phi}_{m_1}]_{i,i} \ge 0$ as it is of the form ${}^te_iMe_i = \|M^{1/2}e_i\|_2^2$ for M positive definite. Under (6) for Laguerre basis or by Lemma 1 for Hermite basis,

$$\sum_{j=0}^{m_2-1}\mathbb{E}\left[\left(\varphi_j(Y_i) - \mathbb{E}(\varphi_j(Y_i)|X_i)\right)^2|X_i\right] \le \sum_{j=0}^{m_2-1}\mathbb{E}\left(\varphi_j^2(Y_i)|X_i\right) \le c\sqrt{m_2}$$

and

$$\mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] \le c\frac{\sqrt{m_2}}{n^2}\mathrm{Tr}\left(\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\, {}^t\widehat{\Phi}_{m_1}\right) = c\frac{m_1\sqrt{m_2}}{n},$$

as $\mathrm{Tr}\left(\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\, {}^t\widehat{\Phi}_{m_1}\right) = n\mathrm{Tr}\left(\widehat{\Phi}_{m_1}({}^t\widehat{\Phi}_{m_1}\widehat{\Phi}_{m_1})^{-1}\, {}^t\widehat{\Phi}_{m_1}\right) = n\mathrm{Tr}\left(({}^t\widehat{\Phi}_{m_1}\widehat{\Phi}_{m_1})^{-1}\, {}^t\widehat{\Phi}_{m_1}\widehat{\Phi}_{m_1}\right) = n\, m_1$.

In the general case, we have

$$\sum_{j=0}^{m_2-1}\mathbb{E}\left[\left(\varphi_j(Y_i) - \mathbb{E}(\varphi_j(Y_i)|X_i)\right)^2|X_i\right] \le \sum_{j=0}^{m_2-1}\mathbb{E}\left(\varphi_j^2(Y_i)|X_i\right) \le L(m_2),$$

and the variance bound becomes:

$$\mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] \leq \frac{m_1 L(m_2)}{n}. \quad \square$$

8.4. **Proof of Theorem 1.**

8.4.1. *Core of the proof.* We give the proof both in the Laguerre and Hermite settings. For the Laguerre case, we note that the assumption $\mathbb{E}(1/\sqrt{Y}|X = x) \leq C$ for all x is fulfilled for π_0 and π_θ below. This is why the lower bound concerns the rate associated with the variance $m_1\sqrt{m_2}/n$, in both cases.

As usual in the proofs of lower bounds, we build a set of conditional densities (π_θ) quite distant from each other in terms of the weighted $\mathbb{L}_2$-norm, but whose distance between the resulting models is small. Let us define in the Laguerre case

$$\pi_0(x, y) = \pi_0(y) = \frac{1}{2}\mathbf{1}_{[0,1]}(y) + P_L(y)\mathbf{1}_{]1,2]}(y)$$

where P_L is a polynomial, $P_L(y) \geq 0$ on $[1, 2]$, $\int_1^2 P_L(y)dy = 1/2$, $P_L(1) = 1/2$, $P_L(2) = 0$, $P_L^{(k)}(1) = P_L^{(k)}(2) = 0$ for $k = 1, \ldots, s_2 + 1$. In the Hermite setting, we define

$$\pi_0(x, y) = \pi_0(y) = P_H(y)\mathbf{1}_{[-1,0[}(y) + \frac{1}{2}\mathbf{1}_{[0,1]}(y) + Q_H(y)\mathbf{1}_{]1,2]}(y),$$

where $P_H(-1) = Q_H(2) = 0$, $P_H(0) = Q_H(1) = 1/2$, $P_H, Q_H \geq 0$ on $[-1, 0]$ and $[1, 2]$ respectively, $\int_{-1}^0 P_H = \int_1^2 Q_H = 1/4$, and $P_H^{(k)}(-1) = P_H^{(k)}(0) = Q_H^{(k)}(1) = Q_H^{(k)}(2) = 0$, for $k = 1, \ldots, s_2 + 1$.

Next, we assume without loss of generality that $\sqrt{m_2}$ is an integer and we define in both Laguerre and Hermite cases

$$\pi_\theta(x, y) = \pi_0(x, y) + \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{k=0}^{\sqrt{m_2}-1} \mathbf{A}_{j,k}\varphi_j(x)(m_2^{1/4}\psi(\sqrt{m_2}y - k)),$$

with

$$\mathbf{A} = \Psi_{m_1}^{-1/2}\Theta, \quad \Theta = (\theta_{j,k})_{1\leq j\leq m_1, 1\leq k\leq\sqrt{m_2}} \in \{0, 1\}^{m_1\sqrt{m_2}},$$

for $\delta > 0$ small enough, where ψ is a bounded function with support $[0, 1]$ such that $\int_0^1 \psi(u)du = 0$. Moreover, we assume that ψ admits continuous bounded derivatives up to order s_2. We use the notation Θ for the matrix with m_1 lines and $\sqrt{m_2}$ columns, and $\theta = vec(\Theta)$ the associated vector with $m_1\sqrt{m_2}$ components. Lastly, $\varphi_j = \ell_j$ in the Laguerre case and $\varphi_j = h_j$ in the Hermite case.

Now we shall use the following lemmas:

Lemma 4. (a) *Assume that $f \in W_{s_1}(A_1, R) \cap \mathbb{L}^\infty(A_1)$. Then there exists $L > 0$ such that π_0 is a conditional density belonging to $W_{\mathbf{s}}^f(A, L)$.*

(b) *If $\delta \leq 1/(4\|\psi\|_\infty)$ and*

$$L(m_1)\|\Psi_{m_1}^{-1}\|_{\text{op}} \leq n/(m_1\sqrt{m_2}) \tag{36}$$

then for all $\theta \in \{0, 1\}^{m_1\sqrt{m_2}}$, π_θ is a conditional density.

(c) *If δ is small enough, for all $\theta \in \{0, 1\}^{m_1\sqrt{m_2}}$, $\pi_\theta - \pi_0$ belongs to $W_{\mathbf{s}}^f(A, L)$ as soon as*

$$\frac{m_1^{s_1+1}\sqrt{m_2}}{n} = O(1) \text{ and } \frac{m_1 m_2^{s_2+1/2}}{n} = O(1).$$

Then under the conditions of this lemma, the π_θ's are conditional densities belonging to $W_{\mathbf{s}}^f(A, 4L)$.

Lemma 5. *We denote $\rho(\theta, \theta')$ the Hamming distance between θ and θ'.*

- *For all* $\theta \in \{0,1\}^{m_1\sqrt{m_2}}$*, the Kullback divergence between the distribution of* $(X_i, Y_i)_{1\leq i\leq n}$ *under* π_θ *and under* π_0 *verifies* $K(P_\theta^{\otimes n}, P_0^{\otimes n}) \leq 2\delta^2\|\psi\|^2 m_1\sqrt{m_2}$.
- *For all* $\theta, \theta' \in \{0,1\}^{m_1\sqrt{m_2}}$, $\|\pi_\theta - \pi_{\theta'}\|_f^2 = \delta^2\|\psi\|^2 n^{-1}\rho(\theta,\theta')$

We also recall the Varshamov-Gilbert bound (see Lemma 2.9 p.104 in Tsybakov (2009)), that we use with $K = m_1\sqrt{m_2}$.

Lemma 6. *Fix some even integer* $K > 0$*. There exists a subset* $\{\theta^{(0)}, \ldots, \theta^{(M)}\}$ *of* $\{0,1\}^K$ *and a constant* $\mathfrak{a}_1 > 0$*, such that* $\theta^{(0)} = (0,\ldots,0)$*,* $\rho(\theta^{(j)}, \theta^{(l)}) \geq \mathfrak{a}_1 K$*, for all* $0 \leq j < l \leq M$*. Moreover it holds that, for some constant* $\mathfrak{a}_2 > 0$*,* $M \geq 2^{\mathfrak{a}_2 K}$*.*

Thus we have built M conditional densities $\pi_{\theta^{(0)}}, \ldots, \pi_{\theta^{(M)}}$ belonging to $W_{\mathbf{s}}^f(A,R)$ such that $\|\pi_{\theta^{(j)}} - \pi_{\theta^{(k)}}\|_f^2 \geq (\delta^2\|\psi\|^2\mathfrak{a}_1)m_1\sqrt{m_2}/n$ and $K(P_{\theta^{(j)}}^{\otimes n}, P_{\theta^{(0)}}^{\otimes n}) \leq [2\delta^2\|\psi\|^2/\mathfrak{a}_2\log(2))]\log(M)$. To conclude it is sufficient to use Theorem 2.5 of Tsybakov (2009) with $m_1 = m_1^\star$ and $m_2 = m_2^\star$ given by (17). Note that $m_1^\star\sqrt{m_2^\star}/n = \psi_n^2$ is the targeted rate, and that condition (36) comes from (20). □

8.4.2. *Proof of Lemma 5.* We start by proving Lemma 5 since its provides a computation useful for the proof of Lemma 4.

• Note that the Kullback divergence between the distribution of $(X_i, Y_i)_{1\leq i\leq n}$ under π_θ and under π_0 verifies

$$K(P_\theta^{\otimes n}, P_0^{\otimes n}) \leq nK(P_\theta, P_0) \leq n\chi^2(\pi_\theta, \pi_0)$$

where $\chi^2(\pi_\theta,\pi_0) = \iint \frac{(\pi_\theta(x,y)-\pi_0(x,y))^2}{\pi_0(x,y)} f(x)dxdy$. Now, using that the $\psi(\sqrt{m_2}y - k)$ have disjoint supports

$$\begin{aligned}
\chi^2(\pi_\theta,\pi_0) &= \int\int_0^1 \frac{(\pi_\theta(x,y)-\pi_0(x,y))^2}{\pi_0(x,y)} f(x)dxdy \\
&= 2\frac{\delta^2}{n}\int\int_0^1 \left(\sum_{j=0}^{m_1-1}\sum_{k=0}^{\sqrt{m_2}-1} \mathbf{A}_{j,k}\varphi_j(x) m_2^{1/4}\psi(\sqrt{m_2}y-k)\right)^2 f(x)dxdy \\
&= 2\frac{\delta^2}{n}\int\sum_{k=0}^{\sqrt{m_2}-1}\left(\sum_{j=0}^{m_1-1}\mathbf{A}_{j,k}\varphi_j(x)\right)^2 f(x)dx\int_0^1\sqrt{m_2}\psi^2(\sqrt{m_2}y-k)dy \\
&= 2\frac{\delta^2}{n}\|\psi\|^2\sum_{k=0}^{\sqrt{m_2}-1}\sum_{j,\ell=0}^{m_1-1}\mathbf{A}_{j,k}\mathbf{A}_{\ell,k}[\Psi_{m_1}]_{j,\ell} \\
&= 2\frac{\delta^2}{n}\|\psi\|^2\mathrm{Tr}[\,{}^t\mathbf{A}\Psi_{m_1}\mathbf{A}] = 2\frac{\delta^2}{n}\|\psi\|^2\mathrm{Tr}[\,{}^t\Theta\Theta] \\
&\leq 2\delta^2\|\psi\|^2\frac{m_1\sqrt{m_2}}{n}
\end{aligned}$$

• Let θ and θ' in $\{0,1\}^{m_1\sqrt{m_2}}$. Denoting $\mathbf{A}' = \Psi_{m_1}^{-1/2}\Theta'$,

$$\begin{aligned}
\|\pi_\theta - \pi_{\theta'}\|_f^2 &= \frac{\delta^2}{n}\iint \left[\sum_{j=0}^{m_1-1}\sum_{k=0}^{\sqrt{m_2}-1} (\mathbf{A}_{j,k} - \mathbf{A'}_{j,k})\varphi_j(x) m_2^{1/4}\psi(\sqrt{m_2}y - k)\right]^2 f(x)dxdy \\
&= \frac{\delta^2}{n}\sum_{j,\ell,k}(\mathbf{A}_{j,k} - \mathbf{A'}_{j,k})(\mathbf{A}_{\ell,k} - \mathbf{A'}_{\ell,k})[\Psi_{m_1}]_{j,\ell}\int \sqrt{m_2}\psi^2(\sqrt{m_2}y - k)dy \\
&= \frac{\delta^2\|\psi\|^2}{n}\sum_{j,\ell,k}(\mathbf{A}_{j,k} - \mathbf{A'}_{j,k})[\Psi_{m_1}]_{j,\ell}(\mathbf{A}_{\ell,k} - \mathbf{A'}_{\ell,k}) \\
&= \frac{\delta^2\|\psi\|^2}{n}\mathrm{Tr}\left[\,{}^t(\mathbf{A} - \mathbf{A}')\Psi_{m_1}(\mathbf{A} - \mathbf{A}')\right] \\
&= \frac{\delta^2\|\psi\|^2}{n}\mathrm{Tr}\left[\,{}^t(\Theta - \Theta')(\Theta - \Theta')\right] = \frac{\delta^2\|\psi\|^2}{n}\rho(\theta,\theta'). \qquad (37)
\end{aligned}$$

8.4.3. *Proof of Lemma 4.*
In this proof, for univariate functions g, h, the dot product $\langle g, h\rangle_f$ means naturally $\int g(x)h(x)f(x)dx$ and $\langle g, h\rangle = \int g(y)h(y)dy$.

(a) First, $\int \pi_0(x,y)dy = 1$ and $\pi_0(x,y) \geq 0$, $\forall x, y$ and thus π_0 is a conditional density. Now we have to prove that the functions π_0 are in $W_{\mathbf{s}}^f(A, L)$ for some $L > 0$.

In the Laguerre case, it is proved in Belomestny et al. (2017), proof of Lemma 4.1, that $y \mapsto \pi_0(y)$ is in the univariate Sobolev-Laguerre space $W_{s_2}(\mathbb{R}^+, L_2)$ for some $L_2 > 0$. In the Hermite case, we use the property proved in Bongioanni and Torrea (2006) stating that the functions in the usual Sobolev space

$$W^{s_2} = \{f \in \mathbb{L}^2(\mathbb{R}), f \text{ admits derivatives up to order } s, \text{ such that } \sum_{j=0}^{s_2}\|f^{(j)}\|^2 < +\infty\}$$

which have compact support also belong to Sobolev-Hermite space with same regularity index. Thus, $y \mapsto \pi_0(y)$ is in $W_{s_2}(\mathbb{R}, L_2)$ for some $L_2 > 0$.

Now we want to prove that $(x,y) \mapsto \pi_0(x,y) = \pi_0(y)$ belongs to the weighted bivariate space $W_{\mathbf{s}}^f(A, L)$ for some $L > 0$, for $\mathbf{s} = (s_1, s_2)$. We have

$$(\pi_0)^{(f)}_{(\ell_1,\ell_2)}(x,y) = \sum_{j=0}^{\ell_1-1}\sum_{k=0}^{\ell_2-1} a_{j,k}^{(f)}\varphi_j(x)\varphi_k(y)$$

avec ceux de la definition des π_θ with

$$a_{j,k}^{(f)} = \langle \pi_0, \varphi_j \otimes \varphi_k\rangle_f = \left(\int \varphi_j(x)f(x)dx\right)\left(\int \pi_0(y)\varphi_k(y)dy\right).$$

Thus, it holds

$$\begin{aligned}
\|\pi_0-(\pi_0)^{(f)}_{(\ell_1,\ell_2)}\|_f^2 &= \iint\left(\sum_{j\geq\ell_1 \text{ or } k\geq\ell_2} a^{(f)}_{j,k}\varphi_j(x)\varphi_k(y)\right)^2 f(x)dxdy\\
&\leq 2\int\left(\sum_{j\geq 0}\langle\varphi_j,1\rangle_f\varphi_j(x)\right)^2 f(x)dx\int\left(\sum_{k\geq\ell_2}\langle\pi_0,\varphi_k\rangle\varphi_k(y)\right)^2 dy\\
&\quad +2\int\left(\sum_{j\geq \ell_1}\langle\varphi_j,1\rangle_f\varphi_j(x)\right)^2 f(x)dx\int\left(\sum_{k\geq 0}\langle\pi_0,\varphi_k\rangle\varphi_k(y)\right)^2 dy
\end{aligned}$$

Now we have $\int\left(\sum_{k\geq 0}\langle\pi_0,\varphi_k\rangle\varphi_k(y)\right)^2 dy=\int\pi_0^2(y)dy$ which is a finite constant, and the regularity of π_0 implies $\int\left(\sum_{k\geq\ell_2}\langle\pi_0,\varphi_k\rangle\varphi_k(y)\right)^2 dy\leq L_2\ell_2^{-s_2}$. On the other hand,

$$\int\left(\sum_{j\geq\ell_1}\langle\varphi_j,1\rangle_f\varphi_j(x)\right)^2 f(x)dx\leq\|f\|_\infty\int\left(\sum_{j\geq\ell_1}\langle\varphi_j,1\rangle_f\varphi_j(x)\right)^2 dx=\|f\|_\infty\sum_{j\geq\ell_1}\langle\varphi_j,f\rangle^2.$$

The assumption that $f\in W_{s_1}(A_1,R)$ implies $\sum_{j\geq\ell_1}\langle\varphi_j,f\rangle^2\leq R\ell_1^{-s_1}$. In the same way

$$\int\left(\sum_{j\geq 0}\langle\varphi_j,1\rangle_f\varphi_j(x)\right)^2 f(x)dx\leq\|f\|_\infty\|f\|_2^2\leq\|f\|_\infty^2.$$

Gathering all terms yields $\|\pi_0-(\pi_0)^{(f)}_{(\ell_1,\ell_2)}\|_f^2\leq L(\ell_1^{-s_1}+\ell_2^{-s_2})$ and thus $\pi_0\in W^f_{\mathbf{s}}(A,L)$ for some $L>0$ depending on $\|f\|_\infty,R,s_2$.

(b) Since $\int\psi=0$, we have $\int\pi_\theta(x,y)dy=1$ and we prove hereafter that $\pi_\theta(x,y)\geq 0$.

In the Laguerre case, for $y\in]1,2]$, $\pi_\theta(x,y)=P_L(y)\geq 0$. Analogously, $\pi_\theta(x,y)\geq 0$ for $y\in\mathbb{R}\setminus[0,1]$ in the Hermite case. Now, take $y\in[0,1]$ in both Laguerre and Hermite case. If for $k_0=0,\dots,\sqrt{m_2}-1$, $y\in[k_0/\sqrt{m_2},(k_0+1)/\sqrt{m_2}[$, then

$$\pi_\theta(x,y)=\frac{1}{2}+\frac{\delta}{\sqrt{n}}\sum_{j=0}^{m_1-1}\mathbf{A}_{j,k_0}\varphi_j(x)(m_2^{1/4}\psi(\sqrt{m_2}y-k_0)).$$

Denoting $\vec{\varphi}(x)={}^t(\varphi_0(x),\dots,\varphi_{m_1-1}(x))$ and $\|.\|$ the Euclidean norm on $\mathbb{R}^{m_1}$, we have

$$\begin{aligned}
\left|\pi_\theta(x,y)-\frac{1}{2}\right| &= \frac{\delta m_2^{1/4}}{\sqrt{n}}\left|\sum_{j=0}^{m_1-1}\mathbf{A}_{j,k_0}\varphi_j(x)\psi(\sqrt{m_2}y-k_0)\right|\\
&\leq \frac{\delta m_2^{1/4}}{\sqrt{n}}\|\psi\|_\infty|[{}^t\mathbf{A}\vec{\varphi}(x)]_{k_0}|=\frac{\delta m_2^{1/4}}{\sqrt{n}}\|\psi\|_\infty|[{}^t\Theta\Psi_{m_1}^{-1/2}\vec{\varphi}(x)]_{k_0}|\\
&\leq \frac{\delta m_2^{1/4}}{\sqrt{n}}\|\psi\|_\infty|{}^te_{k_0}\,{}^t\Theta\Psi_{m_1}^{-1/2}\vec{\varphi}(x)|\leq\frac{\delta m_2^{1/4}}{\sqrt{n}}\|\psi\|_\infty\|\Theta e_{k_0}\|\|\Psi_{m_1}^{-1/2}\|_{\mathrm{op}}\sqrt{L(m_1)}\\
&\leq \frac{\delta m_2^{1/4}}{\sqrt{n}}\|\psi\|_\infty\sqrt{L(m_1)\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}\sum_{j=1}^{m_1}\theta_{j,k_0}^2}\leq\delta\|\psi\|_\infty\sqrt{\frac{m_1\sqrt{m_2}L(m_1)\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}}{n}}
\end{aligned}$$

Then, if $L(m_1)\|\Psi_{m_1}^{-1}\|_{\rm op} \le n/(m_1\sqrt{m_2})$

$$\left|\pi_\theta(x,y) - \frac{1}{2}\right| \le \delta\|\psi\|_\infty$$

which is less than 1/4 for $\delta \le 1/(4\|\psi\|_\infty)$. For this choice of δ, we deduce $\pi_\theta(x,y)\ge 0$.

(c) Next, it remains to prove that $h := \pi_\theta - \pi$ belongs to $W_{\mathbf{s}}^f(A,L)$; this will give $\pi_\theta \in W_{\mathbf{s}}^f(A,4L)$. We note that for any function h,

$$\begin{aligned}\|h-h^f_{(\ell_1,\ell_2)}\|_f &= \|h-h^f_{(\ell_1,\infty)}+h^f_{(\ell_1,\infty)}-h^f_{(\ell_1,\ell_2)}\|_f = \|h-h^f_{(\ell_1,\infty)}+\Pi^{\perp f}_{S_{\ell_1}\otimes S_\infty}(h-h^f_{(\infty,\ell_2)})\|_f\\ &\le \|h-h^f_{(\ell_1,\infty)}\|_f+\|h-h^f_{(\infty,\ell_2)}\|_f.\end{aligned}$$

So, to check that $\pi_\theta-\pi_0$ belongs to $W_{\mathbf{s}}^f(A,L)$, we prove

(i) $\ell_1^{s_1}\|(\pi_\theta-\pi_0)-(\pi_\theta-\pi_0)^{(f)}_{(\ell_1,\infty)}\|_f^2 \le L/2$ and (ii) $\ell_2^{s_2}\|(\pi_\theta-\pi_0)-(\pi_\theta-\pi_0)^{(f)}_{(\infty,\ell_2)}\|_f^2 \le L/2$.

Let us first check condition (i). For the case $\ell_1 \le m_1$, we write, using the same computation as (37),

$$\|(\pi_\theta-\pi_0)-(\pi_\theta-\pi_0)^{(f)}_{(\ell_1,\infty)}\|_f^2 \le \|\pi_\theta-\pi_0\|_f^2 = \frac{\delta^2\|\psi\|^2}{n}\mathrm{Tr}[\,{}^t\Theta\,\Theta] \le \delta^2\|\psi\|^2\frac{m_1\sqrt{m_2}}{n}$$

so that

$$\ell_1^{s_1}\|(\pi_\theta-\pi_0)-(\pi_\theta-\pi_0)^{(f)}_{\ell_1,\infty}\|_f^2 \le \delta^2\|\psi\|^2\frac{\ell_1^{s_1}m_1\sqrt{m_2}}{n} \le \delta^2\|\psi\|^2\frac{m_1^{s_1+1}\sqrt{m_2}}{n} = O(1)$$

if $\frac{m_1^{s_1+1}\sqrt{m_2}}{n} = O(1)$. On the other hand, for $m_1 < \ell_1$, then $(\pi_\theta-\pi_0)^{(f)}_{\ell_1,\infty} = \pi_\theta-\pi_0$ and

$$\|(\pi_\theta-\pi_0)-(\pi_\theta-\pi_0)^{(f)}_{\ell_1,\infty}\|_f^2 = 0.$$

Therefore, (i) is proved.

Now, we turn to condition (ii). Let us be more precise on the computation of the projection of $\pi_\theta-\pi_0$. We have

$$(\pi_\theta-\pi_0)^{(f)}_{(\ell_1,\ell_2)}(x,y) = \sum_{j=0}^{\ell_1-1}\sum_{k=0}^{\ell_2-1} B_{j,k}\varphi_j(x)\varphi_k(y)$$

with for $0\le p\le \ell_1-1$, $0\le q\le \ell_2-1$,

$$\langle \pi_\theta-\pi_0,\varphi_p\otimes\varphi_q\rangle_f = \langle(\pi_\theta-\pi_0)^{(f)}_{\ell_1,\ell_2},\varphi_p\otimes\varphi_q\rangle_f.$$

Denote $\psi_{m_2,k}(y) = m_2^{1/4}\psi(\sqrt{m_2}y-k)$. The left hand side is equal to

$$\begin{aligned}\langle \pi_\theta-\pi_0,\varphi_p\otimes\varphi_q\rangle_f &= \frac{\delta}{\sqrt{n}}\sum_{j=0}^{m_1-1}\sum_{k=0}^{\sqrt{m_2}-1}\mathbf{A}_{j,k}\langle\varphi_j,\varphi_p\rangle_f\langle\psi_{m_2,k},\varphi_q\rangle\\ &= \frac{\delta}{\sqrt{n}}\sum_{j=0}^{m_1-1}\left(\sum_{k=0}^{\sqrt{m_2}-1}\mathbf{A}_{j,k}\langle\psi_{m_2,k},\varphi_q\rangle\right)\langle\varphi_j,\varphi_p\rangle_f.\end{aligned}$$

On the other hand

$$\langle(\pi_\theta-\pi_0)^{(f)}_{(\ell_1,\ell_2)},\varphi_p\otimes\varphi_q\rangle_f = \sum_{j=0}^{\ell_1-1}\sum_{k=0}^{\ell_2-1}B_{j,k}\langle\varphi_j,\varphi_p\rangle_f\delta_{k,q} = \sum_{j=0}^{\ell_1-1}B_{j,q}\langle\varphi_j,\varphi_p\rangle_f.$$

So, for $\ell_1 \geq m_1$, a solution is

$$B_{j,q} = \frac{\delta}{\sqrt{n}} \sum_{k=0}^{\sqrt{m_2}-1} A_{j,k} \langle \psi_{m,k}, \varphi_q \rangle \text{ for } j = 0, \ldots, m_1 - 1, \text{ and } B_{j,q} = 0 \text{ for } j = m_1, \ldots, \ell_1 - 1.$$

We obtain for $\ell_1 \geq m_1$,

$$\begin{aligned}
(\pi_\theta - \pi_0)^{(f)}_{(\ell_1,\ell_2)}(x,y) &= \sum_{j=0}^{m_1-1} \sum_{k=0}^{\ell_2-1} B_{j,k} \varphi_j(x) \varphi_k(y) \\
&= \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{k=0}^{\ell_2-1} \left(\sum_{p=0}^{\sqrt{m_2}-1} \mathbf{A}_{j,p} \langle \psi_{m_2,p}, \varphi_k \rangle \right) \varphi_j(x) \varphi_k(y) \\
&= \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{p=0}^{\sqrt{m_2}-1} \mathbf{A}_{j,p} \left(\sum_{k=0}^{\ell_2-1} \langle \psi_{m_2,p}, \varphi_k \rangle \varphi_k(y) \right) \varphi_j(x) \\
&= \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{p=0}^{\sqrt{m_2}-1} \mathbf{A}_{j,p} \varphi_j(x) \psi^{(S_{\ell_2})}_{m_2,p}(y),
\end{aligned}$$

where

$$\psi^{(S_{\ell_2})}_{m_2,p}(y) = \sum_{k=0}^{\ell_2-1} \langle \psi_{m_2,p}, \varphi_k \rangle \varphi_k(y)$$

is the $\mathbb{L}^2(dy)$-orthogonal projection on S_{ℓ_2} of $y \mapsto \psi_{m_2,p}(y)$. Thus, with $\ell_1 = +\infty$, we obtain

$$\begin{aligned}
\|(\pi_\theta - \pi_0) - (\pi_\theta - \pi_0)^{(f)}_{(\infty,\ell_2)}\|_f^2 &= \|(\pi_\theta - \pi_0) - (\pi_\theta - \pi_0)^{(f)}_{(m_1,\ell_2)}\|_f^2 \\
&= \frac{\delta^2}{n} \| \sum_{j=0}^{m_1-1} \sum_{p=0}^{\sqrt{m_2}-1} \mathbf{A}_{j,p} \varphi_j(x) (\psi_{m_2,p} - \psi^{(S_{\ell_2})}_{m_2,p}) \|_f^2 \\
&= \frac{\delta^2}{n} \sum_{k,k'=0}^{\sqrt{m_2}-1} [\,{}^t\Theta\Theta]_{k,k'} \langle \psi_{m_2,k} - \psi^{(S_{\ell_2})}_{m_2,k}, \psi_{m_2,k'} - \psi^{(S_{\ell_2})}_{m_2,k'} \rangle \\
&= \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} \left\| \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} \psi_{m_2,k} - \left(\sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} \psi_{m_2,k} \right)^{(S_{\ell_2})} \right\|^2 \\
&= \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} \|\xi_j - \xi_j^{(S_{\ell_2})}\|^2
\end{aligned}$$

where $\xi_j = \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} \psi_{m_2,k}$ and $\xi_j^{(S_{\ell_2})}$ is the $\mathbb{L}^2$-orthogonal projection of ξ_j on S_{ℓ_2}. We denote, for a function $h \in \mathbb{L}^2(\mathbb{R}^+)$ (Laguerre) or $h \in \mathbb{L}^2(\mathbb{R})$ (Hermite) by

$$|h|_s^2 := \sum_{k \geq 0} k^s a_k^2(h), \quad a_k(h) := \langle h, \varphi_k \rangle.$$

Then

$$\|\xi_j - \xi_j^{(S_{\ell_2})}\|^2 = \sum_{p \geq \ell_2} a_p^2(\xi_j) \leq \ell_2^{-s_2} \sum_{p \geq \ell_2} a_p^2(\xi_j) p^{s_2} \leq \ell_2^{-s_2} |\xi_j|_{s_2}^2.$$

For the Laguerre case, we use the result proved in (Belomestny et al., 2016, Appendix), stating that the norm $|\xi_j|_{s_2}$ is equivalent to $|||\xi_j|||_{s_2}$ where $|||\xi_j|||_{s_2}^2 := \sum_{r=0}^{s_2} \|\xi_j\|_r^2$ and

$$\|\xi_j\|_r^2 = \left\| x^{r/2} \sum_{q=0}^{r} \binom{r}{q} \xi_j^{(q)} \right\|^2$$

and here $\xi_j^{(q)}$ is the derivative of order q of ξ_j. For $r \le s_2$, we have

$$\begin{aligned}
\|\xi_j\|_r^2 &= \int \left(x^{r/2} \sum_{q=0}^{r} \binom{r}{q} \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} m_2^{1/4} m_2^{q/2} \psi^{(q)}(\sqrt{m_2}x - k) \right)^2 dx \\
&\le 2^r \sum_{q=0}^{r} \binom{r}{q} \int_0^{+\infty} \left(x^{r/2} \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} m_2^{1/4+q/2} \psi^{(q)}(\sqrt{m_2}x - k) \right)^2 dx \\
&= 2^r \sum_{q=0}^{r} \binom{r}{q} \sum_{k=0}^{\sqrt{m_2}-1} \int_0^{+\infty} x^r \theta_{j,k}^2 m_2^{q+1/2} (\psi^{(q)}(\sqrt{m_2}x - k))^2 dx
\end{aligned}$$

as the $\psi^{(q)}(\sqrt{m_2}x - k), \psi^{(q)}(\sqrt{m_2}x - k')$ have disjoint supports for $k \neq k'$. As they are bounded, we get, for $r \le s_2$,

$$\begin{aligned}
\|\xi_j\|_r^2 &\le 2^r m_2^{r+1/2} \sum_{q=0}^{r} \binom{r}{q} \sum_{k=0}^{\sqrt{m_2}-1} \int_{k/\sqrt{m_2}}^{(k+1)/\sqrt{m_2}} x^r (\psi^{(q)}(\sqrt{m_2}x - k))^2 dx \\
&\le 2^r c m_2^{r+1/2} \sum_{q=0}^{r} \binom{r}{q} \sum_{k=0}^{\sqrt{m_2}-1} \int_{k/\sqrt{m_2}}^{(k+1)/\sqrt{m_2}} x^r dx = 2^r c m_2^{r+1/2} \sum_{q=0}^{r} \binom{r}{q} \frac{1}{r+1} = \frac{c 2^{2r}}{r+1} m_2^{r+1/2} \\
&\le C m_2^{s_2+1/2}.
\end{aligned}$$

For the Hermite case, we use the result in (Belomestny et al., 2019, Sec. 4.1) (see Proposition 4 and its proof, Sec. 7.4), which states that, for a compactly supported function h, the squared norm $|h|_s^2$ is equivalent to the squared norm $N_s^2(h) := \|h\|^2 + \|h'\|^2 + \cdots + \|h^{(s)}\|^2$. Here ξ_j is compactly supported and it is easy to see that the same computation as above yields for $r \le s_2$, $\|\xi_j^{(r)}\|^2 \le \|\psi^{(r)}\|_\infty^2 m_2^{r+1/2} \le \|\psi^{(r)}\|_\infty^2 m_2^{s_2+1/2}$. Consequently, in both Laguerre and Hermite cases, we obtain

$$\ell_2^{s_2} \|(\pi_\theta - \pi_0) - (\pi_\theta - \pi_0)_{\infty,\ell_2}^{(f)}\|_f^2 \le \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} |\xi_j|_{s_2}^2 \le C(\psi, s_2) \frac{\delta^2}{n} m_1 m_2^{s_2+1/2},$$

and this quantity is bounded using our assumption. □

8.5. **Proof of Lemma 3.** We check that the coefficients of the $n \times m_2$ matrix, $\mathbb{E}(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}) = (\mathbb{E}(\ell_j(Z_i)|X_i))_{1\le i\le n, 0\le j\le m_2-1}$ are the same as those of $\mathbb{E}(\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X})\ {}^t\mathbf{G}_{m_2}$. On the one hand, by using Formula (21), we have for $i = 1, \ldots, n$ and $j = 0, \ldots, m_2 - 1$,

$$\begin{aligned}
\mathbb{E}(\ell_j(Z_i)|X_i) &= \int \pi_{Z|X}(X_i,z)\ell_j(z)dz \\
&= \sum_{j',k\geq 0}\left(\sum_{p=0}^{k}\langle \pi, \ell_{j'}\otimes \ell_p\rangle g_{k,p}\right)\ell_{j'}(X_i)\underbrace{\int \ell_k(z)\ell_j(z)dz}_{=\delta_{j,k}} \\
&= \sum_{j'\geq 0}\left(\sum_{p=0}^{j}\langle \pi, \ell_{j'}\otimes \ell_p\rangle g_{j,p}\right)\ell_{j'}(X_i) \\
&= \sum_{j'\geq 0}\left[(\langle \pi, \ell_{j'}\otimes \ell_p\rangle)_{0\leq p\leq m_2-1}\,{}^t\mathbf{G}_{m_2}\right]_j \ell_{j'}(X_i) \qquad (38)
\end{aligned}$$

as $\left[(\langle \pi, \ell_{j'}\otimes \ell_p\rangle)_p\,{}^t\mathbf{G}_\infty\right]_j = \left[(\langle \pi, \ell_{j'}\otimes \ell_p\rangle)_{0\leq p\leq m_2-1}\,{}^t\mathbf{G}_{m_2}\right]_j$ for $j=0,\dots,m_2-1$.
On the other hand,

$$\begin{aligned}
\left[\mathbb{E}(\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X})\right]_{i,j} &= \mathbb{E}(\ell_j(Y_i)|X_i) = \int \pi(X_i,z)\ell_j(z)dz \\
&= \int \sum_{j',k\geq 0}\langle \pi, \ell_{j'}\otimes \ell_k\rangle \ell_{j'}(X_i)\ell_k(z)\ell_j(z)dz = \sum_{j'\geq 0}\langle \pi, \ell_{j'}\otimes \ell_j\rangle \ell_{j'}(X_i).
\end{aligned}$$

Therefore

$$\begin{aligned}
\left[\mathbb{E}(\widehat{\Theta}_{m_2}(\mathbf{Y})|\mathbf{X})\,{}^t\mathbf{G}_{m_2}\right]_{i,j} &= \sum_{p=0}^{m_2-1}\mathbb{E}(\ell_p(Y_i)|X_i)[\mathbf{G}_{m_2}]_{j,p} = \sum_{p=0}^{m_2-1}\sum_{j'\geq 0}\langle \pi, \ell_{j'}\otimes \ell_p\rangle \ell_{j'}(X_i)[\mathbf{G}_{m_2}]_{j,p} \\
&= \sum_{j'\geq 0}\left(\sum_{p=0}^{m_2-1}\langle \pi, \ell_{j'}\otimes \ell_p\rangle[\mathbf{G}_{m_2}]_{j,p}\right)\ell_{j'}(X_i) \\
&= \sum_{j'\geq 0}\left[(\langle \pi, \ell_{j'}\otimes \ell_p\rangle)_{0\leq p\leq m_2-1}\,{}^t\mathbf{G}_{m_2}\right]_j \ell_{j'}(X_i). \qquad (39)
\end{aligned}$$

The equality of (38) and (39) gives the result. □

8.6. **Proof of Proposition 3.** It follows from Lemma 3 that $\mathbb{E}(\widehat{\pi}_{\mathbf{m}}^{(L)}|\mathbf{X}) = \pi_{m,n}$ and

$$\|\widehat{\pi}_{\mathbf{m}}^{(L)}-\pi\|_n^2 = \|\widehat{\pi}_{\mathbf{m}}^{(L)}-\pi_{\mathbf{m},n}\|_n^2 + \|\pi_{\mathbf{m},n}-\pi\|_n^2.$$

The last term is the announced bias term, and we consider the variance term:

$$\begin{aligned}
&\|\widehat{\pi}_{\mathbf{m}}^{(L)}-\pi_{\mathbf{m},n}\|_n^2 = \|\widehat{\pi}_{\mathbf{m}}^{(L)}-\mathbb{E}(\widehat{\pi}_{\mathbf{m}}^{(L)}|\mathbf{X})\|_n^2 \\
= \;&\frac{1}{n^2}\mathrm{T}r\left[\mathbf{G}_{m_2}^{-1}\,{}^t\!\left(\widehat{\Theta}_{m_2}(\mathbf{Z})-\mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})-\mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\,{}^t\mathbf{G}_{m_2}^{-1}\right].
\end{aligned}$$

Recall that for the matrix-norms: $\|A\|_F^2 = \mathrm{Tr}(\,{}^tA\,A)$ (Frobenius norm) and $\|A\|_{\mathrm{op}}^2 = \lambda_{\max}({}^tA\,A)$ (operator norm), we have $\|AB\|_F^2 \leq \|A\|_{\mathrm{op}}^2\|B\|_F^2$. Thus

$$\begin{aligned}
&\mathrm{Tr}\left[\mathbf{G}_{m_2}^{-1}\,{}^t\!\left(\widehat{\Theta}_{m_2}(\mathbf{Z})-\mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})-\mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\,{}^t\mathbf{G}_{m_2}^{-1}\right] \\
\leq\; &\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2\mathrm{Tr}\left[\,{}^t\!\left(\widehat{\Theta}_{m_2}(\mathbf{Z})-\mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})-\mathbb{E}\left(\widehat{\Theta}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\right].
\end{aligned}$$

Therefore, we obtain, analogously to (35),

$$\mathbb{E}[\|\widehat{\pi}_{\mathbf{m}}^{(L)} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}] \leq \frac{\|\mathbf{G}_{m_2}^{-1}\|_{\text{op}}^2 L(m_2)m_1}{n}.$$

This gives the first result.

Next it is easy to see that similarly to (35), we have

$$\mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}}^{(L)} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] \leq \frac{\|\mathbf{G}_{m_2}^{-1}\|_{\text{op}}^2}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{m_2-1}\mathbb{E}\left[(\varphi_j(Z_i) - \mathbb{E}(\varphi_j(Z_i)|X_i))^2|X_i\right][\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}]_{i,i}.$$

and condition (23) implies that $\sum_{j=0}^{m_2-1}\mathbb{E}\left[(\varphi_j(Z_i) - \mathbb{E}(\varphi_j(Z_i)|X_i))^2|X_i\right] \leq C\sqrt{m_2}$, with the same argument as for (5). As $[\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}]_{i,i} \geq 0$ a.s., this yields (24). As $Y \geq 0$ and $\varepsilon \geq 0$,

$$\mathbb{E}\left(\frac{1}{\sqrt{Z}}|X = x\right) \leq \min\left(\mathbb{E}\left(\frac{1}{\sqrt{Y}}|X = x\right), \mathbb{E}(1/\sqrt{\varepsilon})\right),$$

which explains the comment. □

8.7. **Proof of Proposition 5.** We write again

$$\|\widehat{\pi}_{\mathbf{m}}^{(H)} - \pi\|_n^2 = \|\pi - \pi_{\mathbf{m},n}\|_n^2 + \|\widehat{\pi}_{\mathbf{m}}^{(H)} - \pi_{\mathbf{m},n}\|_n^2$$

and note that $\mathbb{E}\left(\widehat{\Upsilon}_{m_2}(\mathbf{Z})|\mathbf{X}\right) = \left(\int \pi(X_k, y)h_j(y)dy\right)_{1\leq k\leq n, 0\leq j\leq m_2-1}$ so that $\mathbb{E}(\widehat{\pi}_{\mathbf{m}}^{(H)}|\mathbf{X}) = \pi_{\mathbf{m},n}$. Next,

$$\begin{aligned}&\|\widehat{\pi}_{\mathbf{m}}^{(H)} - \pi_{\mathbf{m},n}\|_n^2\\ =\ & \frac{1}{n^2}\text{Tr}\left[\,{}^t\left(\widehat{\Upsilon}_{m_2}(\mathbf{Z}) - \mathbb{E}\left(\widehat{\Upsilon}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}\left(\widehat{\Upsilon}_{m_2}(\mathbf{Z}) - \mathbb{E}\left(\widehat{\Upsilon}_{m_2}(\mathbf{Z})|\mathbf{X}\right)\right)\right].\end{aligned}$$

We have

$$\mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}}^{(H)} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=0}^{m_2-1}\mathbb{E}\left[\left(v_{h_j}(Z_k) - \mathbb{E}(v_{h_j}(Z_k)|X_k)\right)^2|X_k\right][\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}]_{k,k}.$$

Now, let us study $\sum_{j=0}^{m_2-1}\mathbb{E}\left[\left(v_{h_j}(Z_k) - \mathbb{E}(v_{h_j}(Z_k)|X_k)\right)^2|X_k\right]$

$$\begin{aligned}\sum_{j=0}^{m_2-1}\left(v_{h_j}(Z_k) - \mathbb{E}(v_{h_j}(Z_k)|X_k)\right)^2 &= \sum_{j=0}^{m_2-1}\left(\frac{1}{2\pi}\int[e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)\frac{h_j^*(u)}{f_\varepsilon^*(-u)}du\right)^2\\ &= \sum_{j=0}^{m_2-1}\left(\frac{i^j}{\sqrt{2\pi}}\int h_j(u)\frac{[e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)]}{f_\varepsilon^*(-u)}du\right)^2 \quad \text{with (7)}\\ &\leq \frac{1}{2\pi}\sum_{j=0}^{m_2-1}\left(\int h_j(u)\frac{[e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)]}{f_\varepsilon^*(-u)}(\mathbf{1}_{|u|\leq\sqrt{2m_2}} + \mathbf{1}_{|u|>\sqrt{2m_2}})du\right)^2\end{aligned}$$

Now, since $(h_j)_{0\leq j\leq m_2-1}$ is an orthonormal basis,

$$\begin{aligned}\sum_{j=0}^{m_2-1}\left(\int h_j(u)\frac{[e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)]}{f_\varepsilon^*(-u)}\mathbf{1}_{|u|\leq\sqrt{2m_2}}du\right)^2 &\leq \int\left|\frac{e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)}{f_\varepsilon^*(-u)}\mathbf{1}_{|u|\leq\sqrt{2m_2}}\right|^2du\\ &\leq 4\int_{|u|\leq\sqrt{2m_2}}\frac{du}{|f_\varepsilon^*(u)|^2}.\end{aligned}$$

On the other hand, using (8), we have, for $|u| > \sqrt{2m_2} = \sqrt{(2m_2-1)+1} > \sqrt{2j+1}$ for any $j \le m_2 - 1$, $|h_j(u)| \le Ce^{-\xi u^2}$ and thus, as, under Assumption **A5**, $\eta = \xi - \beta > 0$,

$$\begin{aligned}\sum_{j=0}^{m_2-1}\left(\int h_j(u)\frac{[e^{-iZ_k u}-\mathbb{E}(e^{-iZ_k u}|X_k)]}{f_\varepsilon^*(-u)}\mathbf{1}_{|u|>\sqrt{2m_2}}du\right)^2 &\le \sum_{j=0}^{m_2-1}4\left(\int_{|u|>\sqrt{2m_2}}\frac{Ce^{-(\beta+\eta)u^2}}{|f_\varepsilon^*(u)|}du\right)^2\\ &\le C'\sum_{j=0}^{m_2-1}e^{-4\eta m_2}\left(\int\frac{Ce^{-(\beta+\eta/2)u^2}}{|f_\varepsilon^*(u)|}du\right)^2 \le \mathfrak{c},\end{aligned}$$

for a constant $\mathfrak{c}$ depending on f_ε but not on m_2.

Gathering the two parts, we obtain

$$\sum_{j=0}^{m_2-1}\mathbb{E}\left[\left(v_{h_j}(Z_k)-\mathbb{E}(v_{h_j}(Z_k)|X_k)\right)^2|X_k\right] \le \Delta(m_2)$$

and thus

$$\mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}}^{(H)}-\pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] \le \frac{1}{n^2}\sum_{k=1}^{n}\Delta(m_2)[\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\widehat{\Phi}_{m_1}]_{k,k} = \frac{m_1\Delta(m_2)}{n}. \quad \square$$

8.8. **Proof of Theorem 2.** In this proof, we shall denote by $\mathbb{E}_{\mathbf{X}}[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\mathbf{X}]$.
Let $\mathbf{m}$ be an arbitrary element of $\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}$. First, we write

$$\begin{aligned}\|\tilde{\pi}^{(\mathfrak{Sup})}-\pi\|_n^2 &\le 3\left(\|\widehat{\pi}_{\hat{\mathbf{m}}}^{(\mathfrak{Sup})}-\widehat{\pi}_{\hat{\mathbf{m}},\mathbf{m}}^{(\mathfrak{Sup})}\|_n^2+\|\widehat{\pi}_{\hat{\mathbf{m}},\mathbf{m}}^{(\mathfrak{Sup})}-\widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})}\|_n^2+\|\widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})}-\pi\|_n^2\right)\\ &\le 3((A^{(\mathfrak{Sup})}(\mathbf{m})+V^{(\mathfrak{Sup})}(\hat{\mathbf{m}}))+(A^{(\mathfrak{Sup})}(\hat{\mathbf{m}})+V^{(\mathfrak{Sup})}(\mathbf{m}))+\|\widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})}-\pi\|_n^2)\\ &\le 6A^{(\mathfrak{Sup})}(\mathbf{m})+6V^{(\mathfrak{Sup})}(\mathbf{m})+3\|\widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})}-\pi\|_n^2.\end{aligned}$$

The bound on $\mathbb{E}(\|\widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})}-\pi\|_n^2)$ follows from Proposition 1 for $(\mathfrak{Sup}) = (D)$ and Proposition 3 for $(\mathfrak{Sup}) = (L)$ or Proposition 5 for $(\mathfrak{Sup}) = (H)$. The term $V^{(\mathfrak{Sup})}(\mathbf{m})$ has in each case the order of the variance.

We have to study $A(\mathbf{m})$. Thus the result of Theorem 2 follows if we can prove the result:

Proposition 7. *Under the assumptions of Theorem 2, conditionnally to $\mathbf{X} = (X_1, \dots, X_n)$, we have*

$$\mathbb{E}_{\mathbf{X}}(A^{(\mathfrak{Sup})}(\mathbf{m})) \le 12\|\pi_{\mathbf{m},n}-\pi\|_n^2+\frac{C}{n}.$$

Proof of Proposition 7. We decompose $A^{(\mathfrak{Sup})}(\mathbf{m})$ as follows

$$\begin{aligned}A^{(\mathfrak{Sup})}(\mathbf{m}) \le\ & 3\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}_{\mathbf{m}'}^{(\mathfrak{Sup})}-\pi_{\mathbf{m}',n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}')/6)_+ + 3\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}\|\pi_{\mathbf{m}',n}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n^2\\ &+3\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}_{\mathbf{m},\mathbf{m}'}^{(\mathfrak{Sup})}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}')/6)_+\end{aligned}$$

and Proposition 7 holds if we have

$$\begin{aligned}&(a)\quad \mathbb{E}_{\mathbf{X}}\left[\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}_{\mathbf{m}'}^{(\mathfrak{Sup})}-\pi_{\mathbf{m}',n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}')/6)_+\right] \le \frac{C}{n}\\ &(b)\quad \mathbb{E}_{\mathbf{X}}\left[\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}_{\mathbf{m},\mathbf{m}'}^{(\mathfrak{Sup})}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}')/6)_+\right] \le \frac{C}{n}\\ &(c)\quad \sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}\|\pi_{\mathbf{m}',n}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n^2 \le \|\pi-\pi_{\mathbf{m},n}\|_n^2.\end{aligned}$$

We state here a Lemma proved in Section 8.9.

Lemma 7. $\|\widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})} - \pi_{\mathbf{m},n}\|_n = \sup_{T\in B_{\mathbf{m}}} < \widehat{\pi}_{\mathbf{m}}^{(\mathfrak{Sup})} - \pi_{\mathbf{m},n}, T >_n = \sup_{T\in B_{\mathbf{m}}} \nu_n^{(\mathfrak{Sup})}(T)$, *where* $B_{\mathbf{m}} = \{T \in S_{\mathbf{m}},\ \|T\|_n = 1\}$ *and*

Case (D) $\quad \nu_n^{(D)}(T) = \frac{1}{n}\sum_{i=1}^n [T(X_i, Y_i) - \mathbb{E}_{\mathbf{X}}(T(X_i, Y_i))]$,

Case (L) $\quad \nu_n^{(L)}(T) = \frac{1}{n}\sum_{i=1}^n [\Psi_T(X_i, Z_i) - \mathbb{E}_{\mathbf{X}}(\Psi_T(X_i, Z_i))]$,

where for $T(x,y) = \sum_{j,k} b_{j,k}\varphi_j(x)\varphi_k(y)$ *and* $B = (b_{j,k})_{0\le j\le m_1-1, 0\le k\le m_2-1}$,

$$\Psi_T(x,z) = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} [B\mathbf{G}_{m_2}^{-1}]_{j,k}\varphi_j(x)\varphi_k(z). \tag{40}$$

Case (H) $\quad \nu_n^{(H)}(T) = \frac{1}{n}\sum_{i=1}^n [\Phi_T(X_i, Z_i) - \mathbb{E}_{\mathbf{X}}(\Phi_T(X_i, Z_i))]$,

where $\Phi_T(x,z)$ *is defined by Definition 2.*

Moreover, note that the following result holds.

Lemma 8. *If* $T \in S_m$ *then* $\|T\|_\infty^2 \le L(m_1)L(m_2)\|T\|_2^2$, *and* $\|T\|_2^2 \le \|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}\|T\|_n^2$.

Proof of Lemma 8. If $T(x,y) = \sum_{j,k} b_{jk}\varphi_j(x)\varphi_k(y)$, then

$$|T(x,y)|^2 \le \sum_{j,k} b_{j,k}^2 \sum_j \varphi_j^2(x)\sum_k \varphi_k^2(y) \le \|T\|_2^2 L(m_1)L(m_2),$$

which gives the first inequality. Moreover, we have $\|T\|_2^2 = \text{Tr}({}^tBB)$, where B is the matrix (b_{jk}), and $\|T\|_n^2 = \text{Tr}({}^tB\widehat{\Psi}_{m_1}B)$. Then

$$\|T\|_2^2 \le \|\widehat{\Psi}_{m_1}^{-1/2}\|_{\text{op}}^2\|T\|_n^2 \le \|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}\|T\|_n^2,$$

which is the second inequality.□

Proof of (a).
First, using Lemma 7,

$$\begin{aligned}
\mathbb{T}_1^{(\mathfrak{Sup})} &:= \mathbb{E}_{\mathbf{X}}\left[\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}\left(\|\widehat{\pi}_{\mathbf{m}'}^{(\mathfrak{Sup})} - \pi_{\mathbf{m}',n}\|_n^2 - \frac{V^{(\mathfrak{Sup})}(\mathbf{m}')}{6}\right)_+\right] \\
&\le \sum_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}} \mathbb{E}_{\mathbf{X}}\left[\left(\|\widehat{\pi}_{\mathbf{m}'}^{(\mathfrak{Sup})} - \pi_{\mathbf{m}',n}\|_n^2 - \frac{V^{(\mathfrak{Sup})}(\mathbf{m}')}{6}\right)_+\right] \\
&\le \sum_{\mathbf{m}\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}} \left[\left(\sup_{T\in S_{\mathbf{m}}, \|T\|_n=1}[\nu_n(T)^{(\mathfrak{Sup})}]^2 - \frac{V^{(\mathfrak{Sup})}(\mathbf{m})}{6}\right)_+\right].
\end{aligned}$$

Now we consider separately the three different cases.
Direct case (D). We use Talagrand inequality recalled in Lemma 9, conditionally to $\mathbf{X}$. Remember that we have already proved (see the proof of Proposition 1), that

$$\mathbb{E}\left[\left(\sup_{T\in B_{\mathbf{m}}}\frac{1}{n}\sum_{i=1}^n [T(X_i, Y_i) - \mathbb{E}_X(T(X_i, Y_i))]\right)^2 |\mathbf{X}\right] = \mathbb{E}\left[\|\widehat{\pi}_{\mathbf{m}} - \pi_{\mathbf{m},n}\|_n^2|\mathbf{X}\right] \le \frac{m_1L(m_2)}{n} := H^2.$$

Moreover

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(T^2(X_i,Y_i)|\mathbf{X}) \leq \|\pi\|_\infty \|T\|_n^2$$

so that $v = \|\pi\|_\infty$. To compute b, we use Lemma 8:

$$\|T\|_\infty^2 \leq L(m_1)L(m_2)\|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}} \leq \mathfrak{d}^\star \frac{n\,L(m_2)}{\log^2(n)} =: b^2$$

Thus we apply Lemma 9 (Talagrand) so that for $K_0 \geq 12(1+2\epsilon^2)$ we get

$$\mathbb{T}_1^{(D)} \leq \frac{c_0(\|\pi\|_\infty)}{n}\sum_{\mathbf{m}\in\{1,\dots,n\}^2}\left(e^{-c_1 m_1 L(m_2)} + \frac{L(m_2)}{\log^2(n)}e^{-c_2\log(n)\sqrt{m_1}}\right) \lesssim \frac{1}{n},$$

where $c_2 = 2\epsilon C(\epsilon^2)K_1/(7\sqrt{\mathfrak{d}^\star})$. Thus, use that $\sum_{m_1\geq 1} e^{-\kappa\sqrt{m_1}} \leq Se^{-\kappa}$ to get $\sum_{m_1\geq 1} e^{-c_2\log(n)\sqrt{m_1}} \leq S/n^{c_2}$ and choose $\mathfrak{d}^\star$ such that $c_2 \geq 2$ i.e. $\sqrt{\mathfrak{d}^\star} \leq \epsilon C(\epsilon^2)K_1/7$. So, using that for $\mathbf{m} \in \widehat{\mathcal{M}}_n^{(D)}$, $L(m_2) \leq n/K_0$, the result holds for a well chosen constant $\mathfrak{d}^\star$ under condition (30).

Noisy-Laguerre case (L).
Now we can apply Talagrand Inequality (Theorem 9) to

$$\nu_n^{(L)}(T) = \frac{1}{n}\sum_{i=1}^{n}\left[\Psi_T(X_i,Z_i) - \mathbb{E}_{\mathbf{X}}(\Psi_T(X_i,Z_i))\right].$$

First, we get from the proof of Proposition 3,

$$\mathbb{E}\left(\sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\nu_n^2(T)|\mathbf{X}\right) = \mathbf{E}(\|\hat{\pi}_{\mathbf{m}}^{(L)} - \pi_{m,n}\|_n^2|\mathbf{X}) \leq \frac{m_1 L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2}{n} := H^2.$$

Next we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Psi_T^2(X_i,Z_i)|\mathbf{X}] \;=\; \frac{1}{n}\sum_{i=1}^{n}\int \Psi_T^2(X_i,z)\pi_{Z|X}(X_i,z)dz \leq \frac{\|\pi\|_\infty}{n}\sum_{i=1}^{n}\int \Psi_T^2(X_i,z)dz,$$

as $\pi_{Z|X}(x,z) = \int \pi(x,z-u)f_\varepsilon(u)du \leq \|\pi\|_\infty$. Thus,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Psi_T^2(X_i,Z_i)|\mathbf{X}] \;&\leq\; \frac{\|\pi\|_\infty}{n}\sum_{i=1}^{n}\int\left(\sum_{j,k}[B\mathbf{G}_{m_2}^{-1}]_{j,k}\varphi_j(X_i)\varphi_k(z)\right)^2 dz \\
&=\; \|\pi\|_\infty \sum_{j,j',k}[B\mathbf{G}_{m_2}^{-1}]_{j,k}[B\mathbf{G}_{m_2}^{-1}]_{j',k}[\widehat{\Psi}_{m_1}]_{j,j'} = \|\pi\|_\infty \mathrm{Tr}[\,{}^t\mathbf{G}_{m_2}^{-1}\,{}^tB\widehat{\Psi}_{m_1}B\mathbf{G}_{m_2}^{-1}] \\
&\leq\; \|\pi\|_\infty\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2\mathrm{Tr}[\,{}^tB\widehat{\Psi}_{m_1}B] = \|\pi\|_\infty\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2\|T\|_n^2
\end{aligned}$$

which implies that $v = \|\pi\|_\infty\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2$. Lastly we write

$$\begin{aligned}
\|\Psi_T\|_\infty \;&=\; \sup_{x,z}|\sum_{j,k}[B\mathbf{G}_{m_2}^{-1}]_{j,k}\varphi_j(x)\varphi_k(z)| \leq \sqrt{L(m_1)L(m_2)\|B\mathbf{G}_{m_2}^{-1}\|_F^2} \\
&\leq\; \sqrt{L(m_1)L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2\|B\|_F^2}
\end{aligned}$$

and by Lemma 8,

$$\|B\|_F^2 = \mathrm{Tr}[\,{}^tB\,B] = \|T\|_2^2 \leq \|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}}\|T\|_n^2.$$

Therefore, we get

$$\|\Psi_T\|_\infty^2 \le L(m_1)L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\rm op}^2\|\widehat{\Psi}_{m_1}^{-1}\|_{\rm op}\|T\|_n^2 \le \frac{\mathfrak{d}^\star\, n}{\log^2(n)} L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\rm op}^2 := b^2$$

by using that $\mathbf{m} \in \widehat{\mathcal{M}}_n^{(L)}$ (second constraint).

Therefore, by applying Talagrand Inequality (Theorem 9) that for $K_0 \ge 12(1+2\epsilon^2)$, we get

$$\mathbb{T}_1^{(L)} \le \frac{c_0'(\|\pi\|_\infty)}{n} \sum_{\mathbf{m}\in\widehat{\mathcal{M}}_n} \left(\|\mathbf{G}_{m_2}^{-1}\|_{\rm op}^2 e^{-c_1' m_1 L(m_2)} + \frac{L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\rm op}^2}{\log^2(n)} e^{-c_2\log(n)\sqrt{m_1}} \right),$$

where c_2 is the same as in case (D). So, using that $\mathbf{m} \in \widehat{\mathcal{M}}_n^{(m)}$, $L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\rm op}^2 \le n/K_0$ (first constraint), the result holds if (31) holds.

Case Noisy-Hermite (H).

We now proceed to the application of Talagrand inequality to $\nu_n^{(H)}(T)$ conditionally to $\mathbf{X} = (X_1, \dots, X_n)$, where we already saw that

$$\mathbb{E}_{\mathbf{X}} \sup_{T\in S_{\mathbf{m}}, \|T\|_n=1} [\nu_n^{(H)}(T)]^2 = \mathbb{E}_{\mathbf{X}}\|\hat\pi_{\mathbf{m}}^{(H)} - \mathbb{E}_{\mathbf{X}}\hat\pi_{\mathbf{m}}^{(H)}\|_n^2 \le \frac{m_1\Delta(m_2)}{n} := H^2.$$

Next we determine v. Let $T = \sum_{j,k} b_{j,k}\varphi_j\otimes\varphi_k \in S_{\mathbf{m}}$, $B = (b_{j,k})_{j,k}$, such that $\|T\|_n = 1$.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathbf{X}}[\Phi_T^2(X_i, Z_i)] &= \frac{1}{n}\sum_{i=1}^n \int \Phi_T^2(X_i, z)\pi_{Z|X}(X_i, z)dz \le \|\pi_{Z|X}\|_\infty \frac{1}{n}\sum_{i=1}^n \int \Phi_T^2(X_i, z)dz \\
&\le \|\pi\|_\infty \frac{1}{n}\sum_{i=1}^n \sum_{j,k,j',k'} b_{j,k}b_{j',k'}(T)\varphi_j(X_i)\varphi_{j'}(X_i)\int v_{\varphi_k}(z)v_{\varphi_{k'}}(z)dz \\
&= \|\pi\|_\infty {\rm Tr}\left[\,{}^tB\widehat{\Psi}_{m_1}B(\langle v_{\varphi_k}, v_{\varphi_{k'}}\rangle)_{0\le k,k'\le m_2-1}\right].
\end{aligned}$$

As $\Sigma_0 := {}^tB\widehat{\Psi}_{m_1}B$ is square symmetric positive definite and $S_0 := (\langle v_{\varphi_k}, v_{\varphi_{k'}}\rangle)_{0\le k,k'\le m_2-1}$ is symmetric, we can prove that ${\rm Tr}(\Sigma_0 S_0) \le \|S_0\|_{\rm op}{\rm Tr}(\Sigma_0)$. Indeed, $S_0 = {}^tPD_{S_0}P$ with $D_{S_0} = diag(d_i(S_0))_i$ diagonal and P orthogonal, and

$$\begin{aligned}
{\rm Tr}[\Sigma_0 S_0] &= {\rm Tr}[\Sigma_0\, {}^tPD_{S_0}P] = {\rm Tr}[P\Sigma_0\, {}^tPD_{S_0}] \\
&= \sum_{i=1}^{m_2} d_i(S_0)[P\Sigma_0\, {}^tP]_{i,i} \quad \text{with } [P\Sigma_0\, {}^tP]_{i,i} = \|\Sigma_0^{1/2}Pe_i\|^2 \ge 0 \\
&\le \max_i(|d_i(S_0)|)\sum_{i=1}^{m_2}[P\Sigma_0\, {}^tP]_{i,i} = \max_i(|d_i(S_0)|){\rm Tr}(P\Sigma_0\, {}^tP) = \max_i(|d_i(S_0)|){\rm Tr}(\Sigma_0).
\end{aligned}$$

Therefore

$${\rm Tr}\left[\,{}^tB\widehat{\Psi}_{m_1}B(\langle v_{\varphi_k}, v_{\varphi_{k'}}\rangle)_{0\le k,k'\le m_2-1}\right] \le \|(\langle v_{\varphi_k}, v_{\varphi_{k'}}\rangle)_{0\le k,k'\le m_2-1}\|_{\rm op}{\rm Tr}\left[\,{}^tB\widehat{\Psi}_{m_1}B\right].$$

Then ${\rm Tr}\left[\,{}^tB\widehat{\Psi}_{m_1}B\right] = \|T\|_n^2 = 1$ and we have to bound the operator norm. First

$$\|(\langle v_{\varphi_k}, v_{\varphi_{k'}}\rangle)_{0\le k,k'\le m_2-1}\|_{\rm op} = \sup_{x\in\mathbb{R}^{m_2}, \|x\|=1} {}^tx(\langle v_{\varphi_k}, v_{\varphi_{k'}}\rangle)_{0\le k,k'\le m_2-1}x = \sup_{t\in S_{m_2}, \|t\|=1}\|v_t\|^2.$$

Next, as $v_t = (1/2\pi)(t^*/f_\varepsilon^*)^*(-.)$, we have

$$\begin{aligned}\|v_t\|^2 &= \frac{1}{2\pi}\int \left|\frac{t^*(z)}{f_\varepsilon^*(z)}\right|^2 dz \\ &\leq \frac{1}{2\pi}\sup_{|z|\leq\sqrt{2m_2}} \frac{1}{|f_\varepsilon^*(z)|^2}\int |t^*(z)|^2\, dz + \frac{1}{2\pi}\int_{|z|>\sqrt{2m_2}} \left|\frac{t^*(z)}{f_\varepsilon^*(z)}\right|^2 dz \\ &\leq \sup_{|z|\leq\sqrt{2m_2}} \frac{1}{|f_\varepsilon^*(z)|^2} + \sum_{j=0}^{m_2-1}\frac{1}{2\pi}\int_{|z|>\sqrt{2m_2}} \left|\frac{\varphi_j^*(z)}{f_\varepsilon^*(z)}\right|^2 dz \\ &\leq \sup_{|z|\leq\sqrt{2m_2}} \frac{1}{|f_\varepsilon^*(z)|^2} + \mathfrak{c} = \delta(m_2),\end{aligned}$$

by proceeding as in the proof of Proposition 5. As a consequence,

$$\sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathbf{X}}[\Phi_T^2(X_i,Z_i)]\right) \leq \sup_{|z|\leq\sqrt{2m_2}} \frac{1}{|f_\varepsilon^*(z)|^2} + \mathfrak{c} := v.$$

Lastly, for $T(x,y) = \sum_{j,k} b_{j,k}\varphi_j(x)\varphi_k(y)$

$$\begin{aligned}&\sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\ \sup_{x,z}|\Phi_T(x,z)|^2 = \sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\ \sup_{x,z}\left|\sum_{j,k} b_{j,k}\varphi_j(x)v_{\varphi_k}(z)\right|^2 \\ &\leq \sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\ \sup_{x,z}\mathrm{Tr}[\,{}^tB\,B]\sum_j \varphi_j^2(x)\sum_k |v_{\varphi_k}(z)|^2 \\ &\leq \sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\ \sup_{x,z}\|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}}\mathrm{Tr}[\,{}^tB\widehat{\Psi}_{m_1}B]L(m_1)\Delta(m_2) \\ &\quad= \|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}}L(m_1)\Delta(m_2) \leq \frac{\mathfrak{d}^\star n}{\log(n)}\Delta(m_2) := b^2\end{aligned}$$

as $\mathrm{Tr}[\,{}^tB\widehat{\Psi}_{m_1}B] = \|T\|_n^2 = 1$ and using that $\mathbf{m}\in\widehat{\mathcal{M}}_n^{(H)}$.
As a consequence, by Talagrand inequality,

$$\begin{aligned}&\sum_{\mathbf{m}}\mathbb{E}\left[\left(\sup_{T\in S_{\mathbf{m}},\|T\|_n=1}[\nu_n^{(H)}(T)]^2 - V^{(H)}(\mathbf{m})\right)_+\right] \\ \leq\ & \frac{c_0''}{n}\sum_{\mathbf{m}}\left\{\delta(m_2)\exp(-c_1''m_1\frac{\Delta(m_2)}{\delta(m_2)}) + \frac{\Delta(m_2)}{\log^2(n)}\exp(-c_2\log(n)\sqrt{m_1})\right\},\end{aligned}$$

where c_2 is the same as previously. As for $\mathbf{m}\in\widehat{\mathcal{M}}_n^{(H)}$, we have $\Delta(m_2)\leq n$, and the choice of $\mathfrak{d}$ manage with the second sum. The first one is handled with condition (32).

Consequently

$$\mathbb{E}_{\mathbf{X}}\left[\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(H)}}(\|\hat\pi_{\mathbf{m}'}^{(H} - \pi_{\mathbf{m}',n}\|_n^2 - V^{(H)}(\mathbf{m}')/6)_+\right] \leq C/n.$$

This ends the proof of case (H).□
We proved (a) in the three cases.

Proof of (b).
To prove (b) we simply write, using first the fact that $V(.)$ is nondecreasing with respect to both

m_1 and m_2, and secondly that we assumed that $\mathbf{m}\wedge\mathbf{m}'$ was still in the collection,

$$\begin{aligned}
&\left[\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m},\mathbf{m}'}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}')/6)_+\right]\\
\leq\ &\left[\sup_{\mathbf{m}'\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m},\mathbf{m}'}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}\wedge\mathbf{m}')/6)_+\right]\\
\leq\ &\left[\sup_{\mathbf{m}"\in\widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}}(\|\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m}"}-\pi_{\mathbf{m}",n}\|_n^2-V^{(\mathfrak{Sup})}(\mathbf{m}")/6)_+\right].
\end{aligned}$$

Therofore, the bound on the expectation follows from (a).

Proof of (c)
We already noticed that $\mathbb{E}(\widehat{\pi}^{(D)}_{\mathbf{m}}|\mathbf{X})=\mathbb{E}(\widehat{\pi}^{(L)}_{\mathbf{m}}|\mathbf{X})=\mathbb{E}(\widehat{\pi}^{(H)}_{\mathbf{m}}|\mathbf{X})=\pi_{\mathbf{m},n}$, so the bias terms are exactly the same in the three cases.

Let us define $\mathrm{Proj}^{(n)}_{S_{\mathbf{m}}}$ denotes the empirical projection on $S_{\mathbf{m}}$ which associates to $(x,y)\mapsto T(x,y)$ the function $(x,y)\mapsto(\mathrm{Proj}^{(n)}_{S_{\mathbf{m}}}T)(x,y)=$

$$\sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1}\left[({}^t\widehat{\Phi}_{m_1}\widehat{\Phi}_{m_1})^{-1}\,{}^t\widehat{\Phi}_{m_1}\left(\int\varphi_k(z)T(X_i,z)dz\right)_{\substack{1\leq i\leq n\\0\leq k\leq m_2-1}}\right]_{j,k}\varphi_j(x)\varphi_k(y).$$

For any bivariate function T, the following holds:

$$\mathrm{Proj}^{(n)}_{S_{m_1\wedge m_1'}\otimes S_{m_2\wedge m_2'}}T=\mathrm{Proj}^{(n)}_{S_{m_1'}\otimes S_{m_2'}}\left(\mathrm{Proj}^{(n)}_{S_{m_1}\otimes S_{m_2}}T\right). \tag{41}$$

Then

$$\begin{aligned}
\|\pi_{\mathbf{m}',n}-\pi_{(\mathbf{m},\mathbf{m}'),n}\|_n &= \|\mathrm{Proj}^{(n)}_{S_{m_1'}\otimes S_{m_2'}}\pi-\mathrm{Proj}^{(n)}_{S_{m_1\wedge m_1'}\otimes S_{m_2\wedge m_2'}}\pi\|_n\\
&= \|\mathrm{Proj}^{(n)}_{S_{m_1'}\otimes S_{m_2'}}\pi-\mathrm{Proj}^{(n)}_{S_{m_1'}\otimes S_{m_2'}}\mathrm{Proj}^{(n)}_{S_{m_1}\otimes S_{m_2}}\pi\|_n\\
&\leq \|\pi-\mathrm{Proj}^{(n)}_{S_{m_1}\otimes S_{m_2}}\pi\|_n.
\end{aligned}$$

Thus we obtain (c).

8.9. **Proof of Lemma 7.** We prove now that

$$\|\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m}}-\mathbb{E}_{\mathbf{X}}\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m}}\|_n=\sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\langle\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m}}-\mathbb{E}_{\mathbf{X}}\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m}},T\rangle_n=\sup_{T\in S_{\mathbf{m}},\|T\|_n=1}\nu_n^{(\mathfrak{Sup})}(T).$$

The first equality is standard (bound the scalar product by the norm and choose T to see that the upper bound is reached). For the second equality, we denote $T(x,y)=\sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1}B_{j,k}\varphi_j(x)\varphi_k(y)$, so that

$$\begin{aligned}
\langle\widehat{\pi}^{(\mathfrak{Sup})}_{\mathbf{m}}-\pi_{\mathbf{m},n},T\rangle_n &= \sum_{j,j',k,k'}(\widehat{A}^{(\mathfrak{Sup})}_{\mathbf{m}}-\mathbb{E}_{\mathbf{X}}\widehat{A}^{(\mathfrak{Sup})}_{\mathbf{m}})_{jk}B_{j'k'}\langle\varphi_j\otimes\varphi_k,\varphi_{j'}\otimes\varphi_{k'}>_n\\
&= \sum_{j,j',k}(\widehat{A}^{(\mathfrak{Sup})}_{\mathbf{m}}-\mathbb{E}_{\mathbf{X}}\widehat{A}^{(\mathfrak{Sup})}_{\mathbf{m}})_{jk}B_{j'k}\,(\widehat{\Psi}_{m_1})_{j,j'}\\
&= \mathrm{Tr}\left[{}^t(\widehat{A}^{(\mathfrak{Sup})}_{\mathbf{m}}-\mathbb{E}_{\mathbf{X}}\widehat{A}^{(\mathfrak{Sup})}_{\mathbf{m}})\,\widehat{\Psi}_{m_1}\,B\right].
\end{aligned}$$

For the rest of the proof, we study separately the three cases.

Direct case. Recall that $\widehat{A}^{(D)}_{\mathbf{m}} = \frac{1}{n}\widehat{\Psi}^{-1}_{m_1}\,{}^t\widehat{\Phi}_{m_1}\widehat{\Theta}_{m_2}(\mathbf{Y})$. Then

$$\begin{aligned}\langle \widehat{\pi}^{(D)}_{\mathbf{m}} - \pi_{\mathbf{m},n}, T\rangle_n &= \frac{1}{n}\mathrm{Tr}\left[\,{}^t(\widehat{\Theta}_{m_2}(\mathbf{Y}) - \mathbb{E}_{\mathbf{X}}\widehat{\Theta}_{m_2}(\mathbf{Y}))\widehat{\Phi}_{m_1}\,B\right]\\ &= \frac{1}{n}\sum_{i=1}^n\sum_{jk}(\varphi_k(Y_i) - \mathbb{E}_{\mathbf{X}}(\varphi_k(Y_i)))\varphi_j(X_i)B_{jk}\\ &= \frac{1}{n}\sum_{i=1}^n[T(X_i,Y_i) - \mathbb{E}_{\mathbf{X}}T(X_i,Y_i)].\end{aligned}$$

Laguerre case. In this case $\widehat{A}^{(L)}_{\mathbf{m}} = \frac{1}{n}\widehat{\Psi}^{-1}_{m_1}\,{}^t\widehat{\Phi}_{m_1}\widehat{\Theta}_{m_2}(\mathbf{Z})\,{}^t\mathbf{G}^{-1}_{m_2}$, then

$$\begin{aligned}\langle \widehat{\pi}^{(L)}_{\mathbf{m}} - \pi_{\mathbf{m},n}, T\rangle_n &= \frac{1}{n}\mathrm{Tr}\left[\mathbf{G}^{-1}_{m_2}\,{}^t(\widehat{\Theta}_{m_2}(\mathbf{Z}) - \mathbb{E}_{\mathbf{X}}\widehat{\Theta}_{m_2}(\mathbf{Z}))\widehat{\Phi}_{m_1}\,B\right]\\ &= \frac{1}{n}\sum_{i=1}^n\sum_{jk}(\varphi_k(Y_i) - \mathbb{E}_{\mathbf{X}}(\varphi_k(Y_i)))\varphi_j(X_i)(BG^{-1}_{m_2})_{jk}\\ &= \frac{1}{n}\sum_{i=1}^n[\Psi_T(X_i,Z_i) - \mathbb{E}_{\mathbf{X}}\Psi_T(X_i,Z_i)].\end{aligned}$$

Hermite case. Here we use that $\widehat{A}^{(H)}_{\mathbf{m}} = \frac{1}{n}\widehat{\Psi}^{-1}_{m_1}\,{}^t\widehat{\Phi}_{m_1}\widehat{\Upsilon}_{m_2}(Z)$. Thus

$$\begin{aligned}\langle \widehat{\pi}^{(H)}_{\mathbf{m}} - \pi_{\mathbf{m},n}, T\rangle_n &= \frac{1}{n}\mathrm{Tr}\left[\,{}^t(\widehat{\Upsilon}_{m_2}(\mathbf{Z}) - \mathbb{E}_{\mathbf{X}}\widehat{\Upsilon}_{m_2}(\mathbf{Z}))\widehat{\Phi}_{m_1}\,B\right]\\ &= \frac{1}{n}\sum_{i=1}^n\sum_{jk}(v_{\varphi_k}(Z_i) - \mathbb{E}_{\mathbf{X}}(v_{\varphi_k}(Z_i)))\varphi_j(X_i)B_{jk}\\ &= \frac{1}{n}\sum_{i=1}^n[\Phi_T(X_i,Z_i) - \mathbb{E}_{\mathbf{X}}\Phi_T(X_i,Z_i)]. \qquad \square\end{aligned}$$

8.10. **Proof of Corollary 1.** De-conditionning is justified by Lemma 10 stated and proved in Appendix.

Let $\Lambda^{(\mathfrak{Sup})}_n = \{\mathcal{M}^{(\mathfrak{Sup})}_n \subset \widehat{\mathcal{M}}^{(\mathfrak{Sup})}_n\}$ and write

$$\mathbb{E}[\|\widehat{\pi}^{(\mathfrak{Sup})}_{\hat{m}} - \pi\|^2_n] = \mathbb{E}[\mathbb{E}_{\mathbf{X}}[\|\widehat{\pi}^{(\mathfrak{Sup})}_{\hat{m}} - \pi\|^2_n]\mathbf{1}_{\Lambda^{(\mathfrak{Sup})}_n}] + \mathbb{E}[\|\widehat{\pi}^{(\mathfrak{Sup})}_{\hat{m}} - \pi\|^2_n\mathbf{1}_{(\Lambda^{(\mathfrak{Sup})}_n)^c}] := T_1 + T_2.$$

We first study T_1. On $\Lambda^{(\mathfrak{Sup})}_n$, we have

$$\begin{aligned}\inf_{\mathbf{m}\in\widehat{\mathcal{M}}^{(\mathfrak{Sup})}_n}\left\{\|\pi_{\mathbf{m},n} - \pi\|^2_n + V^{(\mathfrak{Sup})}(\mathbf{m})\right\} &\le \inf_{\mathbf{m}\in\mathcal{M}^{(\mathfrak{Sup})}_n}\left\{\|\pi_{\mathbf{m},n} - \pi\|^2_n + V^{(\mathfrak{Sup})}(\mathbf{m})\right\}\\ &\le \inf_{\mathbf{m}\in\mathcal{M}^{(\mathfrak{Sup})}_n}\left\{\|\pi_{\mathbf{m}} - \pi\|^2_n + V^{(\mathfrak{Sup})}(\mathbf{m})\right\}.\end{aligned}$$

So, for the first term, we have, using Theorem 2 and the definition of $\Lambda^{(\mathfrak{Sup})}_n$,

$$\mathbb{E}_{\mathbf{X}}[\|\widehat{\pi}^{(\mathfrak{Sup})}_{\hat{m}} - \pi\|^2_n\mathbf{1}_{\Lambda^{(\mathfrak{Sup})}_n} \le C\inf_{\mathbf{m}\in\mathcal{M}^{(\mathfrak{Sup})}_n}\{\|\pi - \pi_{\mathbf{m},n}\|^2_n + V^{\mathfrak{Sup}}(\mathbf{m})\} + \frac{C'}{n},$$

and taking the expectation yields

$$\mathbb{E}[\mathbb{E}_{\mathbf{X}}[\|\widehat{\pi}^{(\mathfrak{Sup})}_{\hat{m}} - \pi\|^2_n\mathbf{1}_{\Lambda^{(\mathfrak{Sup})}_n}] \le C\inf_{\mathbf{m}\in\mathcal{M}^{(\mathfrak{Sup})}_n}\{\|\pi - \pi_{\mathbf{m}}\|^2_f + V^{\mathfrak{Sup}}(\mathbf{m})\} + \frac{C'}{n}.$$

Now, T_2 is bounded thanks to the two facts:

(1) $\mathbb{P}[(\Lambda_n^{(\mathfrak{Sup})})^c] \leq C/n^2$ for $\mathfrak{d}$ well chosen,
(2) $\|\widehat{\pi}_{\hat{m}}^{(\mathfrak{Sup})} - \pi\|_n^2 \leq C\,n$ for $\hat{m} \in \widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}$.

First let us prove point (2). We prove that, $\forall \mathbf{m} \in \widehat{\mathcal{M}}_n^{(D)}$, $\|\widehat{\pi}_{\mathbf{m}}^{(D)} - \pi\|_n^2 \leq 2(K_0 n + \|\pi\|_\infty)$. Indeed we have

$$\begin{aligned}
\|\widehat{\pi}_{\mathbf{m}}^{(D)}\|_n^2 &= \frac{1}{n^2}\mathrm{Tr}\left(\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\!\widehat{\Phi}_{m_1}\widehat{\Theta}_{m_2}(\mathbf{Y})\,{}^t\!\widehat{\Theta}_{m_2}(\mathbf{Y})\right)\\
&\leq \frac{1}{n^2}\|\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\!\widehat{\Phi}_{m_1}\|_{\mathrm{op}}\mathrm{Tr}[\widehat{\Theta}_{m_2}(\mathbf{Y})\,{}^t\!\widehat{\Theta}_{m_2}(\mathbf{Y})]\\
&\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{m_2-1}\varphi_k^2(Y_i) \leq L(m_2) \leq K_0 n
\end{aligned}$$

Similarly, for $\mathbf{m} \in \widehat{\mathcal{M}}_n^{(L)}$,

$$\begin{aligned}
\|\widehat{\pi}_{\mathbf{m}}^{(L)}\|_n^2 &= \frac{1}{n^2}\mathrm{Tr}\left(\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\!\widehat{\Phi}_{m_1}\mathbf{G}_{m_2}^{-1}\widehat{\Theta}_{m_2}(\mathbf{Z})\,{}^t\!\widehat{\Theta}_{m_2}(\mathbf{Z})\,{}^t\mathbf{G}_{m_2}^{-1}\right)\\
&\leq L(m_2)\|\mathbf{G}_{m_2}^{-1}\|_{\mathrm{op}}^2 \leq K_0\,n
\end{aligned}$$

and for $\mathbf{m} \in \widehat{\mathcal{M}}_n^{(H)}$,

$$\begin{aligned}
\|\widehat{\pi}_{\mathbf{m}}^{(H)}\|_n^2 &= \frac{1}{n^2}\mathrm{Tr}\left(\widehat{\Phi}_{m_1}\widehat{\Psi}_{m_1}^{-1}\,{}^t\!\widehat{\Phi}_{m_1}\widehat{\Upsilon}_{m_2}(\mathbf{Z})\,{}^t\widehat{\Upsilon}_{m_2}(\mathbf{Z})\right)\\
&\leq \Delta(m_2) \leq K_0\,n.
\end{aligned}$$

To bound $\|\pi\|_n^2$, as π is bounded, we have

$$\int \pi^2(X_1,y)dy \leq \|\pi\|_\infty \int \pi(X_1,y)dy = \|\pi\|_\infty < +\infty,$$

and the result of (2) holds.

Now we study point (1). We have

$$\begin{aligned}
\mathbb{P}((\Lambda_n^{(\mathfrak{Sup})})^c) &= \mathbb{P}\left(\left\{\mathcal{M}_n^{(\mathfrak{Sup})} \subset \widehat{\mathcal{M}}_n^{(\mathfrak{Sup})}\right\}^c\right) = \mathbb{P}(\exists \mathbf{m} \in \mathcal{M}_n^{(\mathfrak{Sup})}, \text{such that } \mathbf{m} \notin \widehat{\mathcal{M}}_n^{(\mathfrak{Sup})})\\
&\leq \sum_{\mathbf{m}\in\mathcal{M}_n^{(\mathfrak{Sup})}} \mathbb{P}\left(L(m_1)\|\Psi_{m_1}^{-1}\|_{\mathrm{op}} \leq \frac{\mathfrak{d}^\star}{2}\frac{n}{\log^2(n)} \text{ and } L(m_1)\|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}} > \mathfrak{d}^\star\frac{n}{\log^2(n)}\right)\\
&\leq \sum_{\mathbf{m}\in\mathcal{M}_n^{(\mathfrak{Sup})}} \mathbb{P}\left(L(m_1)(\|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}} - \|\Psi_{m_1}^{-1}\|_{\mathrm{op}}) \geq \frac{\mathfrak{d}^\star}{2}\frac{n}{\log^2(n)}\right)\\
&\leq \sum_{\mathbf{m}\in\mathcal{M}_n^{(\mathfrak{Sup})}} \mathbb{P}\left(L(m_1)(\|\widehat{\Psi}_{m_1}^{-1} - \Psi_{m_1}\|_{\mathrm{op}}) > L(m_1)\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}\right)\\
&= \sum_{\mathbf{m}\in\mathcal{M}_n^{(\mathfrak{Sup})}} \mathbb{P}\left((\|\widehat{\Psi}_{m_1}^{-1} - \Psi_{m_1}\|_{\mathrm{op}}) > \|\Psi_{m_1}^{-1}\|_{\mathrm{op}}\right)\\
&\leq \sum_{\mathbf{m}\in\mathcal{M}_n^{(\mathfrak{Sup})}} \mathbb{P}\left(\|\Psi_{m_1}^{-1/2}\widehat{\Psi}_{m_1}\Psi_{m_1}^{-1/2} - \mathrm{Id}_{m_1}\|_{\mathrm{op}} > \frac{1}{2}\right),
\end{aligned}$$

where the last inequality follows from Proposition 4 (ii) in Comte and Genon-Catalot (2020). Then the matrix Chernov Inequality (see Tropp (2012)) gives, for $0 \leq \delta \leq 1$,

$$\mathbb{P}\left(\|\Psi_{m_1}^{-1/2}\widehat{\Psi}_{m_1}\Psi_{m_1}^{-1/2} - \mathrm{Id}_{m_1}\|_{\mathrm{op}} > \delta\right) \leq 2m_1 \exp\left(-c(\delta)\frac{n}{L(m_1)(\|\Psi_{m_1}^{-1}\|_{\mathrm{op}} \vee 1)}\right), \tag{42}$$

where $c(\delta) = (1+\delta)\log(1+\delta) - \delta$, which for $\delta = 1/2$ yields $c(1/2) = (3/2)\log(3/2) - 1/2 = 5\mathfrak{d}$, $c(1/2) \sim 0.11$. Thus, under the condition $L(m_1)\|\Psi_{m_1}^{-1}\|_{\mathrm{op}} \leq \mathfrak{d}^\star n/\log^2(n) \leq \mathfrak{d}n/\log(n)$, for n large enough, in the definition of $\mathcal{M}_n^{(\mathfrak{Sup})}$, we get

$$\mathbb{P}((\Lambda_n^{(\mathfrak{Sup})})^c) \leq \sum_{\mathbf{m}\in\mathcal{M}_n^{(\mathfrak{Sup})}} \frac{2m_1}{n^5} \leq 2\frac{|\mathcal{M}_n^{(\mathfrak{Sup})}|}{n^4} \leq \frac{2}{n^2}.$$

This ends the proof. □

Appendix

Lemma 9 (Talagrand Inequality)**.** *Let $Y_1,\dots,Y_n$ be independent random variables and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\epsilon^2 > 0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|^2 - 2(1+2\epsilon^2)H^2\Big]_+ \leq \frac{2}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}} + \frac{49b^2}{4K_1n^2C^2(\epsilon^2)}e^{-\frac{2K_1C(\epsilon^2)\epsilon}{7}\frac{nH}{b}}\right),$$

with $C(\epsilon^2) = (\sqrt{1+\epsilon^2}-1)\wedge 1$, $K_1 = 1/6$, and

$$\sup_{f\in\mathcal{F}}\|f\|_\infty \leq b, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|\Big] \leq H, \quad \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{k=1}^n \mathrm{Var}(f(Y_k)) \leq v.$$

This inequality comes from a concentration Inequality in Klein and Rio (2005) and arguments that can be found in Birgé and Massart Birgé and Massart (1998). Usual density arguments show that this result can be applied to the class of functions of type $\mathcal{F} = B_{\mathbf{m}}(0,1)$.

Lemma 10. *Let $(X_i,Y_i)_{1\leq i\leq n}$ be i.i.d. couples of random variables. Then $(Y_i)_{1\leq i\leq n}$ are independent conditionally to $(X_1,\dots,X_n)$.*

This Lemma legitimates the application of Talagrand inequality conditionally to $(X_1,\dots,X_n)$.
Proof of Lemma 10. First $Y_1,\dots,Y_n$ are independent conditionally to $X_1,\dots,X_n$ if, for all measurable (bounded or nonnegative) functions $f_i:\mathbb{R}\to\mathbb{R}$,

$$\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)|X_1,\dots,X_n\right] = \prod_{i=1}^n \mathbb{E}\left[f_i(Y_i)|X_1,\dots,X_n\right]. \tag{43}$$

As collection of test functions of $X_1,\dots,X_n$ for caracterization of the conditional expectation, we consider $g(X_1,\dots,X_n) = \prod_{i=1}^n g_i(X_i)$ for measurable functions $g_i:\mathbb{R}\to\mathbb{R}$, bounded or nonnegative (density argument: measurable function as monotone limit of linear combinations of indicators of measurable partitions and take as a borelian A of the partition in the product σ-algebra the cartesian product $A = A_1\times\cdots\times A_n$ which are generators). Therefore (43) holds if

$$\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\prod_{i=1}^n g_i(X_i)\right] = \mathbb{E}\left[\prod_{i=1}^n \mathbb{E}(f_i(Y_i)|X_1,\dots,X_n)\prod_{i=1}^n g_i(X_i)\right].$$

To check that this equality holds, let us start from the right-hand-side term.

$$\mathbb{E}\left[\prod_{i=1}^{n}\mathbb{E}(f_i(Y_i)|X_1,\dots,X_n)\prod_{i=1}^{n}g_i(X_i)\right] = \mathbb{E}\left[\prod_{i=1}^{n}g_i(X_i)\mathbb{E}(f_i(Y_i)|X_1,\dots,X_n)\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n}\mathbb{E}(g_i(X_i)f_i(Y_i)|X_1,\dots,X_n)\right] = \mathbb{E}\left[\prod_{i=1}^{n}\underbrace{\mathbb{E}(g_i(X_i)f_i(Y_i)|X_i)}_{=\psi_i(X_i)}\right]$$

$$= \prod_{i=1}^{n}\mathbb{E}[\psi_i(X_i)] \text{ as the } X_i \text{ are independent}$$

$$= \prod_{i=1}^{n}\mathbb{E}\left[\mathbb{E}(g_i(X_i)f_i(Y_i)|X_i)\right] = \prod_{i=1}^{n}\mathbb{E}\left[g_i(X_i)f_i(Y_i)\right] = \mathbb{E}\left[\prod_{i=1}^{n}g_i(X_i)f_i(Y_i)\right]$$

where the last line follows by independence of the $(X_1, Y_1), \dots, (X_n, Y_n)$. This ends the proof of Lemma 10. □

References

Abramowitz, M. and Stegun, I. A. (1964). Handbook of mathematical functions with formulas, graphs, and mathematical tables, volume 55 of National Bureau of Standards Applied Mathematics Series. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Akakpo, N. and Lacour, C. (2011). Inhomogeneous and anisotropic conditional density estimation from dependent data. Electron. J. Stat., 5:1618–1653.

Askey, R. and Wainger, S. (1965). Mean convergence of expansions in Laguerre and Hermite series. Amer. J. Math., 87:695–708.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. Electronic Journal of Statistics, 10(2):3114 – 3152.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2017). Correction to: Nonparametric Laguerre estimation in the multiplicative censoring model [MR3571964]. Electron. J. Stat., 11(2):4845–4850.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2019). Sobolev-Hermite versus Sobolev nonparametric density estimation on ℝ. Ann. Inst. Statist. Math., 71(1):29–62.

Bertin, K., Lacour, C., and Rivoirard, V. (2016). Adaptive pointwise estimation of conditional density function. Ann. Inst. Henri Poincaré Probab. Stat., 52(2):939–980.

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

Bongioanni, B. and Torrea, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. Proc. Indian Acad. Sci. Math. Sci., 116(3):337–360.

Brunel, E., Comte, F., and Lacour, C. (2007). Adaptive estimation of the conditional density in presence of censoring. Sankhya, 69(4):734–763.

Chagny, G. (2013). Warped bases for conditional density estimation. Math. Methods Statist., 22(4):253–282.

Cohen, A., Davenport, M. A., and Leviatan, D. (2013). On the stability and accuracy of least squares approximations. Found. Comput. Math., 13(5):819–834.

Cohen, A., Davenport, M. A., and Leviatan, D. (2019). Correction to: On the stability and accuracy of least squares approximations. Found. Comput. Math., 19(1):239.

Cohen, S. X. and Le Pennec, E. (2013). Partition-based conditional density estimation. ESAIM Probab. Stat., 17:672–697.

Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2017). Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging. J. R. Stat. Soc. Ser. B. Stat. Methodol., 79(1):69–94.

Comte, F. and Genon-Catalot, V. (2018). Laguerre and Hermite bases for inverse problems. J. Korean Statist. Soc., 47(3):273–296.

Comte, F. and Genon-Catalot, V. (2020). Regression function estimation as a partly inverse problem. Ann. Inst. Statist. Math., 72(4):1023–1054.

Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. Ann. Inst. Henri Poincaré Probab. Stat., 49(2):569–609.

De Gooijer, J. G. and Zerom, D. (2003). On conditional density estimation. Statist. Neerlandica, 57(2):159–176.

Dussap, F. (2021). Anisotropic multivariate deconvolution using projection on the Laguerre basis. Journal of Statistical Planning and Inference, 215:23–46.

Efromovich, S. (2007). Conditional density estimation in a regression setting. Ann. Statist., 35(6):2504–2535.

Efromovich, S. (2010). Oracle inequality for conditional density estimation and an actuarial example. Ann. Inst. Statist. Math., 62(2):249–275.

Fan, J., Yao, Q., and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika, 83(1):189–206.

Fan, J. and Yim, T. H. (2004). A crossvalidation method for estimating conditional densities. Biometrika, 91(4):819–834.

Fan, J.-q., Peng, L., Yao, Q.-w., and Zhang, W.-y. (2009). Approximating conditional density functions using dimension reduction. Acta Mathematicae Applicatae Sinica, English Series, 25(3):445–456.

Ferraty, F., Laksaci, A., and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. Stat. Inference Stoch. Process., 9(1):47–76.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. Ann. Statist., 39(3):1608–1632.

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

Györfi, L. and Kohler, M. (2007). Nonparametric estimation of conditional distributions. IEEE Trans. Inform. Theory, 53(5):1872–1879.

Holmes, M. P., Gray, A. G., and Isbell, Jr., C. L. (2010). Fast kernel conditional density estimation: a dual-tree Monte Carlo approach. Comput. Statist. Data Anal., 54(7):1707–1718.

Hyndman, R. J. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. J. Nonparametr. Stat., 14(3):259–278.

Ioannides, D. A. (1999). Estimating the conditional mode of a stationary stochastic process from noisy observations. Metrika, 50(1):19–35.

Izbicki, R. and Lee, A. B. (2016). Nonparametric conditional density estimation in a high-dimensional regression setting. Journal of Computational and Graphical Statistics, 25(4):1297–1316.

Izbicki, R. and Lee, A. B. (2017). Converting high-dimensional regression to high-dimensional conditional density estimation. Electron. J. Stat., 11(2):2800–2831.

Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. Bernoulli, 10(2):187–220.

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. Ann. Probab., 33(3):1060–1077.

Lacour, C. (2007). Adaptive estimation of the transition density of a markov chain. Ann. Inst. H. Poincaré Probab. Statist., 43(5):571–597.

Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. Ann. I.S.U.P., 51(3):69–80 (2008).

Lukacs, E. (1970). Characteristic functions. Hafner Publishing Co., New York. Second edition, revised and enlarged.

Mabon, G. (2017). Adaptive deconvolution on the non-negative real line. Scand. J. Stat., 44(3):707–740.

Markett, C. (1984). Norm estimates for (C, δ) means of Hermite expansions and bounds for δ_{eff}. Acta Math. Hungar., 43(3-4):187–198.

Meister, A. (2009). Deconvolution problems in nonparametric statistics, volume 193. Springer Science & Business Media.

Nguyen, M.-L. J. (2018). Nonparametric method for sparse conditional density estimation in moderately large dimensions. https://hal.archives-ouvertes.fr/hal-01688664.

Nguyen, M.-L. J., Lacour, C., and Rivoirard, V. (2021). Adaptive greedy algorithm for moderately large dimensions in kernel conditional density estimation. https://hal.archives-ouvertes.fr/hal-02085677.

Otneim, H. k. and Tjostheim, D. (2018). Conditional density estimation using the local Gaussian correlation. Stat. Comput., 28(2):303–321.

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

Rosenblatt, M. (1969). Conditional probability density and regression estimators. In Multivariate Analysis, II (Proc. Second Internat. Sympos., Dayton, Ohio, 1968), pages 25–31. Academic Press, New York.

Sacko, O. (2020). Hermite density deconvolution. ALEA Lat. Am. J. Probab. Math. Stat., 17(1):419–443.

Sart, M. (2017). Estimating the conditional density by histogram type estimators and model selection. ESAIM Probab. Stat., 21:34–55.

Tsybakov, A. B. (2009). Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York.

Youndjé, É. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. Rev. Roumaine Math. Pures Appl., 41(7-8):535–566.

*Laboratoire MAP5, Université de Paris, France
E-mail address: fabienne.comte@parisdescartes.fr

LAboratoire LAMA, Université Gustave Eiffel, France
E-mail address: claire.lacour@u-pem.fr