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

F. COMTE AND C. LA COUR

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 \((X_i, Y_i)_{1 \leq i \leq n}\). We consider a regression strategy related to projection subspaces of \( L^2 \) 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 \( \varepsilon \) 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.


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 \( \varepsilon \) spoils the response so that only \( Z = Y + \varepsilon \) is available. From independent and identically distributed couples of variables \((X_i, Y_i)_{1 \leq i \leq n}\) first, and \((X_i, Z_i)_{1 \leq i \leq n}\) 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 \( E[Y | X = x] \) 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 Youndje (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 \( \pi \) as an anisotropic function under the Mean Integrated Squared error criterion. They give oracle inequalities and adaptive minimax results. For instance Efroymovich (2007) uses a Fourier decomposition to construct a blockwise-shrinkage Efroymovich-Pinsker estimator, whereas Brunel et al. (2007) and Asakpo and Lacour (2011) use projection estimators and model selection. Next, Efroymovich (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

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 A3 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 multivariate case by Golkenshluger and Lepskii (2014). When using a \( R \)-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 \( R^+ \)-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 \( \hat{\pi}_m \), \( m = (m_1, m_2) \), by minimization of a least squares contrast on a subspace \( S_m \) 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

\[
\hat{\Psi}_m = \hat{\Psi}_m(X) = \frac{1}{n} \hat{\Phi}_m \hat{\Phi}_m, \quad \text{where} \quad \hat{\Phi}_m = (\varphi_j(X_i))_{1 \leq i \leq n, 0 \leq j \leq m-1}.
\]

The point is to use specific bases suited to the non-compact problem. Two cases are of special interest: the case where the support is \( R \), for which we use the Hermite basis, and the case where the support is \( R_+ \), for which we use the Laguerre basis. This last case is 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 \( m = (m_1, m_2) \).
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 $m_1 m_2/n$ 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 compactly 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 $M_n$. In our framework this collection has to be random because of the very importance of the matrix $\hat{\Psi}_{m_1}$ 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 $\hat{\Psi}_{m_1}$ called stability condition, which bounds the operator norm $\|\hat{\Psi}_{m_1}^{-1}\|_{\text{op}}$ 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 $\hat{\Psi}_{m_1}$, 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 $\pi$, 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 $\varepsilon$ 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 variables in $\mathbb{R}$, with the classical hypothesis that the characteristic function of the noise does not vanish. In another part (see Section 5), we consider the case where all the variables are positive, including the noise. For these two noisy cases (variables in $\mathbb{R}$ or in $\mathbb{R}_+$), we provide new estimators $\hat{\pi}_{m_1}^{(H)}$ and $\hat{\pi}_{m_1}^{(L)}$ and study their empirical and 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 $Z$, which makes the true signal harder to recover. We show that the rate obtained in the Hermite-noisy case is optimal in the minimax sense under some conditions. We also propose an adaptive model selection and we obtain 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 an estimator defined in the Hermite basis in the noisy case, for the general $\mathbb{R}$-supported functions, and provides upper and lower bounds on the integrated risk; the specific case when all random variables are nonnegative and the Laguerre basis is used, is considered in Section 5. Lastly, Section 6 states a general adaptive result, based on a Goldenshuger-Lepski method, see Goldenshuger 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 $L^2$ norm of a bivariate measurable function $T$, defined by:

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

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

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

Note that for any deterministic function $T$, $\mathbb{E} \|T\|^2_n = \|T\|^2_f$. 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 $\{1, \ldots, n\} \times \{1, \ldots, n\}$ and let $\mathbf{m} = (m_1, m_2)$ denote an element of $\mathcal{M}_n$. We construct a sequence $\{\hat{\pi}_m\}_{m \in \mathcal{M}_n}$ of estimators of $\pi$, each $\hat{\pi}_m$ 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,

$$S_{m_i} = \text{span}\{\varphi_j, j = 0, \ldots, m_i - 1\}, \quad i = 1, 2,$$

and the $\varphi_j$ are known orthonormal functions with respect to the standard $L^2$-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 $\{\varphi_j \otimes \varphi_k, j = 0, \ldots, m_1 - 1, k = 0, \ldots, m_2 - 1\}$. A key quantity associated to the basis $(\varphi_j)_j$ is

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

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 $\|M\|_{\text{op}}$ as $\sqrt{\lambda_{\text{max}}(M^t M)}$ where $M$ is the transpose of $M$ and $\lambda_{\text{max}}$ denotes the largest eigenvalue. Its Frobenius norm is defined by $\|M\|^2_F = \text{Tr}(M^t M)$ where $\text{Tr}(A)$ 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 $t_0(x) = 1_{[0,1]}(x)$, and for $j \geq 1$, $t_{2j-1}(x) = \sqrt{2} \cos(2\pi j x)1_{[0,1]}(x)$, $t_{2j}(x) = \sqrt{2} \sin(2\pi j x)1_{[0,1]}(x)$. 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 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:

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

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

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

where $C$ and $\xi$ are positive constants independent of $x$ and $j$, $0 < \xi < \frac{1}{2}$. Note that with (4), $h_j^*$ 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 $L(m) \leq K \sqrt{m}$.

- The Laguerre functions are defined as follows:

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

The functions $\ell_j$ are orthonormal, and are bounded by $\sqrt{2}$ (see 22.14.12 in [Abramowitz and Stegun (1964)]). So $\sum_{j=0}^{m-1} \ell_j^2(x) \leq 2m$. Moreover, 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)]):

\[ \ell_j \ast \ell_k(x) = \int_{0}^{\infty} \ell_j(u)\ell_k(x-u) du = \frac{1}{\sqrt{2}}(\ell_{j+k}(x) - \ell_{j+k+1}(x)), \quad \forall x \geq 0. \]

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

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

then for $j \geq 1$

\[ \forall x \geq 0, \quad \mathbb{E}(\ell_j^2(Y) | X = x) = \int_{0}^{\infty} \ell_j^2(y)\pi(x, y) dy \leq \frac{c}{\sqrt{j}}, \]

For instance, condition (7) holds if $Y = g(X) + U$ with $g \geq 0$, $X$ and $U$ independent, and $\mathbb{E}U^{-1/2} < \infty$. Under (7), 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.

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 it would be possible to mix two different bases.

2.3. **Anisotropic Laguerre and Hermite Sobolev spaces.** To study the bias term, we assume that $\pi$ 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-Hermite ellipsoids on $\mathbb{R}^d$ and Sobolev-Laguerre ellipsoids on $\mathbb{R}^d_+$. 

Indeed, for \( m \) would be possible. \( \int \) In the present bivariate context, mixed cases involving basis \( d \) \( (\) and basis \( = 1 \) \( s \) \( k \) \( \in \mathbb{N}^d \) \( g \) \( \varphi_k = \sum_{k=1}^{d} a_k \varphi_k \) \( or the Hermite coefficients of \( g \) if \( \varphi_k = h_k = h_{k_1} \otimes \cdots \otimes h_{k_d} \).

We refer to Belomestny et al. (2019) for details about this space and the link with usual Sobolev space. Note in particular that when \( d = 1 \) \( a \) \( \varphi_k \) \( \in \mathbb{R}^d \) or the Sobolev-Laguerre ellipsoid. Let \( A \rightarrow \mathbb{R}^d \) \( A, \varphi \) \( \psi \) \( \exists \) \( x \) \( \exp(\cdot) \) \( \sqrt{2\pi\sigma} \) \( \sqrt{2\pi\sigma^2} \) \( \mathbb{E}[N^p] \) \( N \sim N(0,1) \) \( \lambda_0 \) \( \lambda_0 = \log \left( \frac{(\sigma^2 + 1)}{(\sigma^2 - 1)^2} \right) > 0 \), \n
\begin{align*}
W_s(A, L) := \left\{ g \in \mathbb{L}^2(A), \sum_{k \in \mathbb{N}^d} a_k^2(g) k^s \leq L \right\}, \quad k^s = k_1^{s_1} \cdots k_d^{s_d},
\end{align*}

\begin{align*}
\| g - g_m \|^2 &= \sum_{k \in \mathbb{N}^d, k \geq m} a_k^2(g) \leq \sum_{q=1}^{d} \sum_{k \in \mathbb{N}^d, k \geq m} a_k^2(g) k^s_q k^{-s_q} \leq L \sum_{q=1}^{d} m^{-s_q}.
\end{align*}

Since we are interested by the weighted \( \mathbb{L}^2 \) norm, we also consider weighted Laguerre or Hermite Sobolev spaces defined by:

\begin{align}
W_s^f(A, R) = \{ g \in \mathbb{L}^2(A, f(x)dx) dy), \forall m, m_q \geq 1, \| g - g_m \|_f^2 \leq R \sum_{q=1}^{d} m^{-s_q} \}
\end{align}

where \( R = \mathbb{L} \mathbf{L} \) \( f \mathbf{L} \mathbf{L} \) \( g \mathbf{L} \mathbf{L} \) \( | \psi - \psi_m |^2 \) \( \text{has exponential rate of decrease, see Lemma 3.9 in Mabon (2017). Continuous mixtures are also studied in section 3.2 of Comte and Genon-Catalot (2018).} \)

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

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

where \( c_{2p} = \mathbb{E}[N^{2p}] \) for \( N \sim N(0,1) \), \( \sigma^2 \neq 1 \) \( \text{cases with } \sigma^2 = 1 \) \( \text{have finite developments in the basis, and null bias for } m_i \text{ larger than } p \). It is proved in Belomestny et al. (2019) (Proposition 12) that, for \( i = 1, 2 \), \n
\begin{align*}
\| \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( \frac{(\sigma^2 + 1)}{(\sigma^2 - 1)^2} \right) > 0,
\end{align*}

where \( (\psi_{p,\sigma})_{m_i} \) \( \text{are the orthogonal projections of } \psi_{p,\sigma} \text{ on } S_{m_i}. \)
By tensorization, we can thus consider the class $WSS_{k,\lambda}(L)$ for $s = (s_1, s_2)$ and $\lambda = (\lambda_1, \lambda_2)$ for real numbers $s_1, s_2$ and positive $\lambda_1, \lambda_2$, of functions $g$ such that,

$$
\|g - g_m\|^2 \leq L(m_1^{-s_1} \exp(-\lambda_1 m_1) + m_2^{-s_2} \exp(-\lambda_2 m_2))
$$

where $m = (m_1, m_2)$. 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 $(X_k, Y_k)$, $k = 1, \ldots, n$ with the same law as $(X, Y)$. It is studied in Section 3 under the Assumption

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

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

$$
Z = Y + \varepsilon.
$$

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

Assumption A2. The distribution of $\varepsilon$ is known, $\varepsilon$ is independent of $X$ and independent of $Y$ conditionally to $X$.

Notice that this implies the independence of $Y$ and $\varepsilon$.

In both direct and noisy settings, we estimate the function $\pi$ on $\mathbb{R} \times \mathbb{R}$ or $\mathbb{R}_+ \times \mathbb{R}_+$. In the direct case, we also consider the case of $\pi$ 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

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

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

$$
\hat{\pi}_m^{(D)} := \arg\min_{T \in S_m} \gamma_n^{(D)}(T)
$$

for $m = (m_1, m_2)$. 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 $\hat{\pi}_m^{(D)}$, we define

$$
\Phi_m = (\varphi_j(X_i))_{1 \leq i \leq n, 0 \leq j \leq m-1}, \quad \Psi_m = \frac{1}{n} \Phi_m \Phi_m^T.
$$

Note that $\Psi := \mathbb{E}(\tilde{\Psi}_m) = ((\varphi_j, \varphi_k)f)_{0 \leq j, k \leq m-1}$. We find, assuming that $\tilde{\Psi}_{m_1}$ is invertible,

$$
\hat{\pi}_m^{(D)}(x, y) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \hat{a}_{j,k}^{(D)} \Phi_j(x) \varphi_k(y), \quad \hat{A}_m^{(D)} = (\hat{a}_{j,k}^{(D)})_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}
$$
with
\[ \widehat{A}_{m}^{(D)} = \frac{1}{n} \widehat{\Psi}_{m_{1}}^{-1} \widehat{\Phi}_{m_{1}} \widehat{\Theta}_{m_{2}}(Y), \quad \text{with} \quad \widehat{\Theta}_{m}(Y) = (\varphi_{j}(Y))_{1 \leq i \leq n, 0 \leq j \leq m-1}. \]

**Remark.** In the Laguerre and Hermite case, conditions ensuring a.s. inverisibility of \( \widehat{\Psi}_{m_{1}} \) are weak: \( m_{1} \leq n \) 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 \( \widehat{\pi}_{m} \)

First we study the quadratic (empirical) risk of the estimator \( \widehat{\pi}_{m} \), on a given space \( S_{m} = S_{m_{1}} \otimes S_{m_{2}} \) as described in Sections 2.1 and 2.2.

We denote by \( \pi_{m,n} \) the orthogonal projection of \( \pi \) on \( S_{m} \) for the empirical norm, by \( \pi_{m,f} \) the orthogonal projection of \( \pi \) on \( S_{m} \) for the \( L^{2}(A, f(x) dx dy) \)-norm and by \( \pi_{m} \) the orthogonal projection for the Lebesgue \( L^{2} \)-norm on \( A \). Then we can write
\[ \| \widehat{\pi}_{m}^{(D)} - \pi \|_{n}^{2} = \| \widehat{\pi}_{m}^{(D)} - \pi_{m,n} \|_{n}^{2} + \| \pi_{m,n} - \pi \|_{n}^{2}, \]

and then note that
\[ \| \pi_{m,n} - \pi \|_{n}^{2} = \inf_{T \in S_{m}} \| T - \pi \|_{n}^{2} \leq \| \pi_{m,f} - \pi \|_{n}^{2}. \]

Thus
\[ \mathbb{E}(\| \pi_{m,n} - \pi \|_{n}^{2}) \leq \| \pi_{m,f} - \pi \|_{n}^{2}. \]

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 \( \mathbb{E}(\widehat{\pi}_{m}^{(D)}(X_{i}, y) | X) = \pi_{m,n}(X_{i}, y) = \sum_{j=0}^{m_{1}-1} \sum_{k=0}^{m_{2}-1} [D_{m}]_{j,k} \varphi_{j}(X_{i}) \varphi_{k}(y) \) with
\[ D_{m} = \frac{1}{n} \widehat{\Psi}_{m_{1}}^{-1} \widehat{\Phi}_{m_{1}} \left( \int \varphi_{k}(y) \pi(X_{i}, y) dy \right)_{1 \leq i \leq n, 0 \leq k \leq m_{2}-1}. \]

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

**Proposition 1.** Let \( \widehat{\pi}_{m} \) be defined by (11)-(13), and assume that Assumption A1 is fulfilled. Then, for any \( m = (m_{1}, m_{2}) \) such that \( m_{1} \leq n \),
\[ \mathbb{E}(\| \widehat{\pi}_{m}^{(D)} - \pi \|_{n}^{2}) \leq \| \pi - \pi_{m,f} \|_{n}^{2} + \frac{m_{1} L(m_{2})}{n}. \]

If moreover \( \mathbb{E}(\| \pi_{m,n} - \pi \|_{n}^{2}) \leq \| \pi_{m,f} - \pi \|_{n}^{2} + \frac{m_{1} \sqrt{m_{2}}}{n} \) holds for Laguerre basis, then
\[ \mathbb{E}(\| \widehat{\pi}_{m}^{(D)} - \pi \|_{n}^{2}) \leq \| \pi - \pi_{m,f} \|_{n}^{2} + \frac{c m_{1} \sqrt{m_{2}}}{n}, \]

where \( c \) is a positive constant.

The bound in (16) is obtained under weak conditions with explicit and optimal constants. It involves a bias term \( \| \pi - \pi_{m,f} \|_{n}^{2} \) 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 \( \widehat{\pi}_{m} \), we can set \( \widehat{\pi}_{m}^{+}(x, y) = \max(\widehat{\pi}_{m}(x, y), 0) = \widehat{\pi}_{m}(x, y) 1_{\widehat{\pi}_{m}(x, y) \geq 0} \) and we have \( \| \widehat{\pi}_{m}^{+} - \pi \|_{n}^{2} \leq \| \widehat{\pi}_{m} - \pi \|_{n}^{2} \). Therefore, \( \widehat{\pi}_{m}^{+} \) is well defined, nonnegative, and inherits of the risk bound proved for \( \widehat{\pi}_{m} \).

**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 \( \eta_i \) 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)].

3.3. Bound on the weighted integrated MISE. The aim here is to prove a risk bound measured in \( L^2(A, f(x)dx) \)-norm, in the spirit of [Comte and Genon-Catalot (2020)]. This requires the so-called "stability condition" introduced by Cohen et al. (2013). Let us define

\[
\Psi_{m_1} = \mathbb{E}(\hat{\Psi}_{m_1}) = \left( \int \varphi_j(x) \varphi_k(x) f(x) dx \right)_{1 \leq i, j \leq m_1}.
\]

Note that for a non-zero vector \( x = (x_0, \ldots, x_{m_1-1}) \in \mathbb{R}^{m_1} \), then

\[
\langle x, \Psi_{m_1} x \rangle = \int \left( \sum_{j=0}^{m_1-1} x_j \varphi_j(x) \right)^2 f(x) dx > 0
\]

for our bases examples and non-degenerate density \( f \) (see Lemma 1 in [Comte and Genon-Catalot (2020)]), so that \( \Psi_{m_1} \) is invertible. So we consider the stability constraint

\[
L(m_1)\|\Psi_{m_1}^{-1}\|_{op} \leq \frac{\varrho_0}{2} \frac{n}{\log n}, \quad \varrho_0 = \varrho_0(3), \quad \varrho_0(r) = \frac{3 \log(3/2) - 1}{2 + 2r}.
\]

In practice, the constraint has to be replaced by an empirical version, thus we define the truncated (or stabilized) version of \( \hat{\pi}_m \) by

\[
\tilde{\pi}_m^{(D)} = \hat{\pi}_m^{(D)} 1_{L(m_1)\|\Psi_{m_1}^{-1}\|_{op} \leq \varrho_0 \log(n)/n}.
\]

**Theorem 1.** For any \( m = (m_1, m_2) \) such that condition (19) is fulfilled and that \( L(m_2) \leq n \),

\[
\mathbb{E}\|\tilde{\pi}_m^{(D)} - \pi\|_f^2 \leq \left( 1 + \frac{4\varrho_0}{\log n} \right) \|\pi - \pi_{m,f}\|_f^2 + \frac{4m_1L(m_2)}{n} + \frac{c}{n},
\]

where \( c \) is a positive constant depending on \( \pi \), \( \varrho_0 \).

It is most interesting to see that this upper bound almost preserves the constant 1 in front of the bias term. Note that by a slight modification of \( \varrho_0 \) (change the choice \( r = 3 \)) the constraint \( L(m_2) \leq n \) may be weakened into \( L(m_2) \leq n^a \) for some constant \( a \).

3.4. Anisotropic rates. If we consider a conditional density \( \pi \) belonging to a Laguerre or Hermite Sobolev space defined in Section 2.3, we deduce from Theorem 1 the following bound:

**Proposition 2.** Assume that \( \pi \) belongs to \( W^s_f(A, R) \) with \( s = (s_1, s_2) \). Consider the estimator \( \tilde{\pi}_m^{(D)} \) defined by (11)-(13)-(20) under Assumption A1 in Laguerre or Hermite basis. Then choosing, in the Laguerre basis \( m = (m_1^0, m_2^0) \) with

\[
m_1^0 \propto n^{s_{12} + \frac{s_1}{s_2}} \quad \text{and} \quad m_2^0 \propto n^{s_{12} + \frac{s_2}{s_1}},
\]

it holds

\[
\mathbb{E}(\|\tilde{\pi}_m^{(D)} - \pi\|_f^2) = O(n^{-\frac{a}{s_2}}), \quad \frac{1}{s} = \frac{1}{2} \left( \frac{1}{s_1} + \frac{1}{s_2} \right),
\]
provided that $m_1^*$ satisfies condition (19). If in addition, $s_1 = s_2 = s$, the rate becomes $n^{-\frac{1}{1+\frac{1}{s}}}$.

If the basis is the Hermite basis, or if the Laguerre basis is used under condition (8), then choosing

$$m_1^* \propto n^{-\frac{1}{1+\frac{1}{s_1} + \frac{1}{s_2}}} \quad \text{and} \quad m_2^* \propto n^{-\frac{1}{1+\frac{1}{s_1} + \frac{1}{s_2}}} ,$$

we obtain

$$E(\|\hat{\pi}^{(D)}_m - \pi\|_f^2) = O\left(\frac{n^{-1+\frac{1}{s_1} + \frac{1}{s_2}}}{n}\right),$$

provided that $m_1^*$ satisfies condition (19).

Note that we also obtain the same convergence rates for the empirical norm $E(\|\hat{\pi}^{(D)}_m - \pi\|_n^2)$ and $\pi$ belonging to $W_s^f(A, R)$, with no need of the stability constraint. The same rates hold on $W_s^f(A, L)$ if $f$ is bounded.

We also emphasize 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), $n^{-2\alpha/(2\alpha+2)}$ for regularity $\alpha = (\alpha_1, \alpha_2)$, that we also recover, see Brunel et al. (2007). Lastly, we remark that, under the assumptions of Proposition 2, if $f$ is bounded and $\pi$ is supersmooth, that is $\pi \in WSS_{s, \lambda}(L)$, see (10), then choosing $m_1^* \propto \log(n) - (s_i + 3/2) \log \log(n)/\lambda_i$ gives

$$E(\|\hat{\pi}^{(D)}_m - \pi\|_f^2) = O\left(\frac{\log^{3/2}(n)}{n}\right).$$

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

**Proof of Proposition 2.** We start from Inequality (21). For the bias term, we have by the regularity assumptions on $\pi$, that $\|\pi - \pi_m\|_f^2 \leq R[m_1^{-s_1} + m_2^{-s_2}]$. Therefore

$$E(\|\hat{\pi}^{(D)}_m - \pi\|_f^2) \leq R(m_1^{-s_1} + m_2^{-s_2} + \frac{m_1 m_2}{n}).$$

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 system

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

gives the first result. The second result is obtained analogously, by using the new variance order $m_1^*/m_2/n$. $\square$

3.5. **Lower bound.** As the rate obtained in Proposition 2 is not standard, we need to check if it is optimal in some sense.

We assume that the regularity orders $(s_1, s_2)$ are integer. We denote by $W_s^f(A_1, R)$ 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 2.** Let $R$ be a positive real and $L$ be a large enough positive real. Then, for any density $f \in W_s^f(A_1, R) \cap L^\infty(A_1)$, 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,

$$\sup_{\pi \in W_s^f(A, L)} E_{\pi} \left[\|\hat{\pi}_n - \pi\|_2^2\right] \geq c\psi_n^2$$

where

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

if, for $m_1^* = [\psi_n^{-2/s_1}]$,

$$L(m_1^*)\|\Psi_{m_1^*}^{-1}\|_{op} \leq \psi_n^{-2}.$$


This result proves the optimality of the rate obtained in Proposition 2 for Hermite basis or Laguerre basis under (8).

Let us comment condition (25). This condition is stronger than the stability condition (19). In Comte and Genon-Catalot (2020), it is proved that \(|\Psi_{m_1}^{-1}\|_{\text{op}} \leq cm_1^\beta\) if \(f\) has some polynomial decay, and 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 \(\|\Psi_{m_1}^{-1}\|_{\text{op}} = cm_1^\beta\), then (25) is fulfilled if \(\beta + 1 \leq s_1\) in the Laguerre case and if \(\beta + 1/2 \leq s_1\) in the Hermite case.

4. 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.

4.1. Assumption related to the noise. We denote by \(f_\varepsilon^*\) the characteristic function of the noise \(\varepsilon:\)

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

The following assumptions are required for \(f_\varepsilon^*\):

**Assumption A3.**

1. Function \(f_\varepsilon^*\) never vanishes, i.e. \(\forall u \in \mathbb{R}, f_\varepsilon^*(u) \neq 0\).
2. There exist \(\alpha \in \mathbb{R}, \beta > 0, 0 \leq \gamma \leq 2, (\alpha > 0 \text{ if } \gamma = 0), \beta < \xi \text{ if } \gamma = 2\) for \(\xi\) defined in (5), 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 \(f_\varepsilon\) is a density, it is known that \(\gamma \leq 2\) (at least for \(\alpha = 0\)).\footnote{According to Lukacs (1970), Theorem 4.1.1, the only characteristic function \(\phi\) 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 \(|f_\varepsilon^*(u)|^2 = \exp(-2\beta|u|^\gamma)\) 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 \(\varepsilon_1 - \varepsilon_1^\prime\) where \(\varepsilon_1^\prime\) is an independent copy of \(\varepsilon_1\)).}

Let us mention that, if the density of the noise is unknown, \(f_\varepsilon^*\) must be replaced by an estimator; this can be done if replicate measurements of \(Z\) are available or if an independent sample of the noise is available, see Kappus and Mabon (2014) for the density estimation case.

4.2. Definition of the contrast 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 \(T \in S_m\), we denote by \(T^{(\ast, 2)}\) the Fourier transform with respect to the second variable:

\[T^{(\ast, 2)}(x, u) = \int e^{iyu}T(x, y)dy.\]

**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 \( T \in S_m \), we denote by \( \Phi_T \) the following bivariate function

\[
\Phi_T(x, z) = \frac{1}{2\pi} \int e^{-iu_z} T^{(*)}(x, u) \frac{T^{(*)}(x, u)}{f^*_z(-u)} du
\]

We can also write \( \Phi^{(*)}_T(x, u) = T^{(*)}(x, u)/f^*_z(-u) \).

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 (3) of the Hermite basis and Assumption A3 (2). Moreover, the operators \( v \) and \( \Phi \) 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.

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

\[
\forall T \in S_m \quad \mathbb{E}[\Phi_T(X_1, Z_1)|X_1] = \mathbb{E}[T(X_1, Y_1)|X_1] = \int T(X_1, z)\pi(X_1, z)dz.
\]

Now we can define our estimators by:

\[
(26) \quad \hat{\pi}^{(H)}_m = \arg \min_{T \in S_m} \gamma^{(H)}_n(T),
\]

with the following contrast \( \gamma^{(H)}_n \):

\[
(27) \quad \gamma^{(H)}_n(T) = \frac{1}{n} \sum_{i=1}^n \left[ \int _\mathbb{R} T^2(X_i, y)dy - 2\Phi_T(X_i, Z_i) \right].
\]

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

\[
\mathbb{E}[\gamma^{(H)}_n(T)] = \mathbb{E}\left[ \int T^2(X, y)dy - 2\Phi_T(X, Z) \right] = \int \int T^2(x, y)f(x)dxdy - 2\mathbb{E}[T(X, Y)]
\]

\[
= \int \int [(T(x, y) - \pi(x, y))^2 - \pi^2(x, y)]f(x)dxdy = \|T - \pi\|^2 - \|\pi\|^2.
\]

We obtain the following new estimator of \( \pi \):

\[
\hat{\pi}^{(H)}_m(x, y) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \hat{a}_{j,k}^{(H)} h_j(x)h_k(y) \quad \hat{A}^{(H)}_m = (\hat{a}_{j,k}^{(H)})_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}
\]

with

\[
(28) \quad \hat{A}^{(H)}_m = \frac{1}{n} \hat{\Psi}^{-1}_m \hat{\Phi}_m \hat{Y}_m(Z), \quad \text{with} \quad \hat{Y}_m(Z) = (v_{h_j}(Z_i))_{1 \leq i \leq n, 0 \leq j \leq m_2-1},
\]

with \( \hat{\Psi}_m, \hat{\Phi}_m \) still defined by (12).

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.
4.3. Bound on the empirical and integrated MISE of $\tilde{\pi}^{(H)}_m$ and rates. Here we can prove the following bound:

**Proposition 3.** Under Assumptions **A1, A2 and A3** we have $\mathbb{E}[\tilde{\pi}^{(H)}_m | X] = \pi_{m,n}$ and

$$\mathbb{E}\|\tilde{\pi}^{(H)}_m - \pi\|^2 \leq \|\pi - \pi_m\|^2 \frac{m_1 \Delta(m_2)}{n},$$

where $\Delta(m_2) = \frac{1}{\pi} \left(4 \int_{|u| \leq \sqrt{2m_2} \left|f^*_\varepsilon(u)\right|^2} du + c\right)$

and $c$ is a constant only depending on $\xi$ (see (5)) and on $f^*_\varepsilon$.

Again, we can also provide a bound in integrated norm, under the stability condition (19), for the truncated estimator:

$$\tilde{\pi}^{(H)}_m = \tilde{\pi}^{(H)}_m 1_{L(m_1)}(\|\hat{\Psi}'_m\|_{\text{op}}} \leq d_0 \log(n)/n.$$  

**Theorem 3.** For any $m = (m_1, m_2)$ such that condition (19) is fulfilled and $\Delta(m_2) \leq n$,

$$\mathbb{E}\|\tilde{\pi}^{(H)}_m - \pi_A\|^2 \leq \left(1 + \frac{4d_0}{\log n}\right) \|\pi - \pi_{m,f}\|^2 \frac{m_1 \Delta(m_2)}{n} + \frac{c}{n},$$

where $c$ is a positive constant depending on $\pi$, $d_0$.

Note that, under Assumption **A3**, we can compute that

$$\Delta(m_2) \leq km_2^\alpha \frac{1 - \frac{1}{2} - \frac{1}{2}}{2} \exp[2\beta(2m_2)^{\gamma}/2].$$

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

**Proposition 4.** Assume that Assumptions **A1, A2 and A3** hold. Let $\pi \in W_s^j(\mathbb{R}^2, L)$.

1. If $\gamma = 0$, then for $m_i^* \propto n^s_{\alpha_i}/[(\alpha + 1/2)s_1 + s_2(s_1 + 1)]$, $i = 1, 2$, and provided that $m_i^*$ satisfies condition (19), we have

$$\mathbb{E}\|\tilde{\pi}^{(H)}_m^* - \pi\|^2 \leq Cn^{-1} \frac{1}{\alpha_i + 1/2 + 1}.$$  

2. If $\gamma, \beta > 0$, then for $m_i^* = (\log n)^{2s_2/(\gamma s_1)}$ and $m_2^* = (1/2) (\log n/4\beta)^{2/\gamma}$, and provided that $m_i^*$ satisfies condition (19), we have

$$\mathbb{E}\|\tilde{\pi}^{(H)}_m - \pi\|^2 \leq C (\log n)^{-2s_2/\gamma}.$$  

Note that we also obtain the same convergence rates for the empirical norm $\mathbb{E}(\|\tilde{\pi}^{(H)}_m - \pi\|^2)$ and $\pi$ belonging to $W^j_s(\mathbb{R}^2, L)$, with no need of the stability constraint.

We can remark that if $\pi$ is supersmooth, that is $\pi \in WSS_{s,\lambda}(L)$, see (10), and $\gamma = 0$, then for $m_i^* = [\log(n) - (\alpha + 1) \log \log(n)]/\lambda_i$, $i = 1, 2$

$$\mathbb{E}\|\tilde{\pi}^{(H)}_m - \pi\|^2 \leq C \frac{1}{\lambda_i}.$$  

We observe in the result of Proposition 4 a usual feature of 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 the analogous computations in the multivariate density setting studied in Comte and Lacour (2013).

The above rates of convergence are actually minimax optimal, as stated in the Theorem 4 below. Let us first introduce an additional assumption.
Assumption A4. For any $p = 1, \ldots, [s_2/2] + 1$, there exist $\alpha_p \in \mathbb{R}$ (with $\alpha_p \geq \alpha$ in the case $\gamma = 0$), and $k_2 > 0$ such that $\forall u \in \mathbb{R}$,

$$|(f^*_\varepsilon)^{(p)}(u)| \leq k_2(u^2 + 1)^{-\alpha_p/2} \exp(-\beta|u|^\gamma)$$

where $(f^*_\varepsilon)^{(p)}$ is the derivative of order $p$ of $f^*_\varepsilon$.

It is usual to require an assumption on the derivative $(f^*_\varepsilon)'$ for proving lower bounds in deconvolution contexts (see for instance Meister (2009)). Here we need bounds on more derivatives because of our specific notion of regularity.

Theorem 4. Assume that Assumptions A1, A2, A3, A4 hold. Let $R$ be a positive real and $L$ be a large enough positive real. Then, for any density $f \in W_{s_1}(\mathbb{R}, \mathbb{R}) \cap L^\infty(\mathbb{R})$, there exists a constant $c$ such that for any estimator $\hat{\pi}_n$, and for $n$ large enough,

$$\sup_{\pi \in W_{s_1}(\mathbb{R}^2, L)} \mathbb{E}_\pi \left[ \|\hat{\pi}_n - \pi\|_f^2 \right] \geq c\psi_n^2$$

as soon as

$$L(m_1^*)\|\Psi^{-1}\|_{op} \leq \psi_n^{-2} \quad \text{for } m_1^* = [\psi_n^{-2/s_1}]$$

where

- if $\gamma = 0$ (ordinary smooth noise), $\psi_n^2 = n^{-1 + 1/2 + 2}^{\alpha+1/2}$,
- If $\gamma, \beta > 0$ (super smooth noise), $\psi_n^2 = (\log n)^{-2s_2/\gamma}$.

5. Indirect Laguerre case

Now, the observations are $(X_k, Z_k)$ with $Z_k = Y_k + \varepsilon_k$, $k = 1, \ldots, n$, under Assumptions A1 and A2. In this Section, we assume that $X_k \geq 0$, $Y_k \geq 0$, $\varepsilon_k \geq 0$ 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 $\varepsilon$ is not centered. More precisely we assume

Assumption A5. The distribution of the noise $\varepsilon$ admits a density with respect to the Lebesgue measure, denoted by $f_\varepsilon$. Moreover $X \geq 0$, $Y \geq 0$, $\varepsilon \geq 0$ a.s.

5.1. Definition of the estimators in the noisy-Laguerre case. In this context, computations rely on property (6), 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 $\pi$.

Let us define the matrix the $m_2 \times m_2$ lower triangular matrix $G_{m_2} = (g_{j,k})_{0 \leq j, k \leq m_2 - 1}$ 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 $G_{m_2}$ are $(f_\varepsilon, \ell_0)/\sqrt{2} = \int_0^{+\infty} f_\varepsilon(u)e^{-u}du > 0$. As a consequence $G_{m_2}$ is invertible. Relying on equation (32), in this noisy model we find

$$
\pi_{Z|X}(x, z) = \sum_{j \geq 0} \sum_{k \geq 0} \langle \pi_{Z|X}, \ell_j \otimes \ell_k \rangle \ell_j(x) \ell_k(z)
$$

In other words, we have

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

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

$$(\pi_{Z|X})_{(\infty, m_2)}(x, z) = \sum_{j \geq 0} \sum_{k=0}^{m_1-1} \left[ \langle \pi, \ell_j \otimes \ell_k \rangle_0 \leq p \leq m_2-1 G_{m_2} \right] k \ell_j(x) \ell_k(z),$$

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

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

with

$$
\hat{A}^{(L)}_{m} = \frac{1}{n} \Psi_{m_1}^{-1} \Theta_{m_1} \Theta_{m_2}(Z) \Gamma_{m_2}^{-1}, \quad \text{with} \quad \Theta_{m}(Z) = (\ell_j(Z_i))_{1 \leq i \leq n, 0 \leq j \leq m_1-1}
$$

where $\hat{\Psi}_{m_1}$ and $\hat{\Theta}_{m_1}$ are defined by (12).

5.2. Bound on the empirical and integrated MISE of $\hat{\pi}^{(L)}_{m}$. Let us note that

$$
E[\hat{A}^{(L)}_{m} | X] = \frac{1}{n} \Psi_{m_1}^{-1} \Theta_{m_1} E[\Theta_{m_2}(Z) | X] \Gamma_{m_2}^{-1}.
$$

A first useful property is given by the lemma:

**Lemma 3.** We have $E[\Theta_{m_2}(Z) | X] \Gamma_{m_2}^{-1} = E[\Theta_{m_2}(Y) | X]$ and thus $E[\hat{\pi}^{(L)}_{m} | X] = \pi_{m,n}$.

Thanks to this result, we can prove the following risk bound (see Section 8.9).

**Proposition 5.** Assume that Assumptions $A_1$, $A_2$, and $A_5$ hold. Then the estimator $\hat{\pi}^{(L)}_{m}$ defined by (32) satisfies:

$$
E(\|\hat{\pi}^{(L)}_{m} - \pi\|_n^2) \leq \|\pi - \pi_{m,n}\|_f^2 + \frac{\|G_{m_2}^{-1} \|^2_{op} m_1 L(m_2)}{n}
$$
where \( L(m_2) = 2m_2 \) here. If in addition the condition
\[
\exists C > 0, \forall x, \quad \mathbb{E} \left( \frac{1}{\sqrt{Z}} | X = x \right) < C
\]
holds, then
\[
\mathbb{E} \left( \| \tilde{\pi}_m^{(L)} - \pi \|^2 \right) \leq \| \pi - \pi_{m,f} \|^2 + C \frac{\| \mathbf{G}_{m_2}^{-1} \|^2_{op} m_1 \sqrt{m_2}}{n}.
\]

Note that condition (34) holds if condition (7) holds or if \( \mathbb{E}(1/\sqrt{Z}) \) is finite.

As \( \mathbf{G}_{m_2} \) is lower triangular, its eigenvalues are given by the diagonal terms, which are all equal to \( 2^{-1/2}(\beta, \ell_0) = \int_{\mathbb{R}^+} e^{-u} f_\varepsilon(u) du \leq 1 \). Therefore
\[
\| \mathbf{G}_{m_2}^{-1} \|^2_{op} = \lambda_{\max}(\mathbf{G}_{m_2}^{-1} (\mathbf{G}_{m_2}^{-1} \geq [\lambda_{\max}(\mathbf{G}_{m_2}^{-1})]^2 \geq 1.
\]

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
\[
\text{for any } a, b > 0, \text{ under Assumption A6, the term } L(m_2) \text{ in the bound can be replaced by } C \sqrt{m_2} \text{ if (34) holds.}
\]

Let us also mention that, if the density of the noise is unknown, then the matrix \( \mathbf{G}_{m_2} \) must be replaced by an estimator; this can be done if an independent sample of observations of the noise is available, see Comte and Mabon (2017) for the density estimation case.

5.3. Rates in the noisy-Laguerre case. Now let us assess the order of the variance term and more specifically of \( \| \mathbf{G}_{m_2}^{-1} \|^2_{op} \). Comte et al. (2017) show that we can recover the order of this spectral norm, under the conditions on the density \( f_\varepsilon \). First we define an integer \( \alpha \geq 1 \) such that
\[
\frac{d^j}{dx^j} f_\varepsilon(x) |_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \ldots, \alpha - 2 \\ B_\alpha \neq 0 & \text{if } j = \alpha - 1. \end{cases}
\]

Consider the two following assumptions:

Assumption A6.

1. \( f_\varepsilon \in L^1(\mathbb{R}_+) \) is \( \alpha \) times differentiable and \( f_\varepsilon^{(\alpha)} \in L^1(\mathbb{R}_+) \).

2. The Laplace transform of \( f_\varepsilon \), \( z \mapsto \mathbb{E} [e^{-z\varepsilon}] \) 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 A6 there exists positive constants \( C \) and \( C' \) such that:
\[
C'M^{2\alpha} \leq \| \mathbf{G}_{m_2}^{-1} \|^2_{op} \leq CM^{2\alpha}.
\]

For example a Gamma distribution with \( \Gamma(p, \theta) \) satisfies Assumptions A6 for \( \alpha = p (\alpha = 1 \) for an exponential distribution). If \( f_\varepsilon \) follows a \( \beta(a, b) \) and \( b > a \), then \( \| \mathbf{G}_{m_2}^{-1} \|^2_{op} = O(m_2^{2\alpha}) \) (see Mabon (2017)). On the contrary an Inverse Gamma distribution does not satisfy Assumptions
because there exists no value of $\alpha$ such that the derivative is nonzero at 0.

These assumptions allow to deduce from Theorem [5] the following rates of convergence of the estimator.

**Proposition 6.** Under Assumptions A6-A6 for $\pi \in W^f_s([\mathbb{R}^2_+, L)$, and $m^* = (m^*_1, m^*_2)$ such that $m^*_1 \propto n^{s_2/[(2\alpha+1)s_1+s_2+s_1s_2]}$ and $m^*_2 \propto n^{s_1/[(2\alpha+1)s_1+s_2+s_1s_2]}$

then

$$\mathbb{E}[||\hat{\pi}^f_{m^*} - \pi||^2_{\mathcal{F}}] \leq C(s, L, ||f||_\infty) n^{-1/[(2\alpha+1)s_2+1]}$$

provided that $m^*_1$ satisfies condition (19).

Note that the rate is not the same with or without condition (3), and establishing the optimal rate in this noisy-Laguerre framework remains an open question. The method is however specific and interesting.

6. **Adaptive estimators with Goldenschluger-Lepski method**

In the previous sections, we have described collections of estimators $\hat{\pi}_m$ and computed their rates of convergence for optimal $m = m^*$. Nevertheless these values of $m^*$ depend on the smoothness $s$ of the unknown conditional density $\pi$. Now we aim at selecting $m$ 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 $(\mathcal{G}^{up})$ where $(\mathcal{G}^{up}) = (D)$ (direct case), or $(\mathcal{G}^{up}) = (H)$ (Hermite-noisy case) or $(\mathcal{G}^{up}) = (L)$ (Laguerre-noisy case).

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

$$V^{(D)}(m) = K_0 \frac{m_1 \Delta(m_2)}{n}, V^{(H)}(m) = K_0 \frac{m_1 \Delta(m_2)}{n}, V^{(L)}(m) = K_0 \frac{m_1 L(m_2)}{n},$$

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

Then we consider the following collection of models

$$\mathcal{M}^{(\mathcal{G}^{up})}_n = \left\{ m \in \{1, \ldots, n\}^2, V^{(\mathcal{G}^{up})}(m) \leq 1, L(m_1) \|\hat{\Psi}^{-1}_{m_1}\|_{op} \leq \frac{d^*}{2} \frac{n}{\log^2(n)} \right\}$$

where $d^*$ a well-chosen numerical constant such that $d^*/\log(n) \leq d$ with $d = (3 \log(3/2) - 1)/10 = \vartheta_0(4)$, $n \geq 2$, and $d^* \leq \epsilon C(\epsilon^2)/42, C(\epsilon^2) = \min(\sqrt{1+\epsilon^2} - 1, 1)$.

We also introduce its empirical random counterpart

$$\hat{\mathcal{M}}^{(\mathcal{G}^{up})}_n = \left\{ m \in \{1, \ldots, n\}^2, V^{(\mathcal{G}^{up})}(m) \leq 1, L(m_1) \|\hat{\Psi}^{-1}_{m_1}\|_{op} \leq d^* \frac{n}{\log^2(n)} \right\}$$

Note that $m_1 \mapsto L(m_1) \|\hat{\Psi}^{-1}_{m_1}\|_{op}$ is increasing, and $m \mapsto V^{(\mathcal{G}^{up})}(m)$ also, with respect to each variable. Thus both collection are such that, if they contain $m$ and $m'$, then they also contain $m \wedge m'$ defined as component-wise minimum.

**Comment.** 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, $\mathcal{M}^{(\mathcal{G}^{up})}_n$ and $\hat{\mathcal{M}}^{(\mathcal{G}^{up})}_n$. The second constraint must be compared to the "stability condition" (19). Obviously, it is here slightly reinforced (by an additional $\log(n)$ factor). However, when dealing with adaptive estimation, Comte and Genon-Catalot (2020) had a stronger
condition: \( L(m_1)\|\Psi_{m_1}^{-1}\|_{op}^2 \leq \delta^*(n/\log(n)) \). The improvement here is substantial, specifically for non compactly supported bases where \( \|\Psi_{m_1}^{-1}\|_{op} \) can be large.

Now we present additional constraints on the model collections, which need to be expressed differently in function of the case which is considered.

**Case (D).** Assume that, for any \( c_1 > 0 \), there exists \( \Sigma > 0 \) such that
\[
\sum_{\mathbf{m} \in \{1, \ldots, n\}^2} e^{-c_1 m_1 L(m_2)} \leq \Sigma < +\infty.
\]

**Case (H).** Let \( \delta(m_2) := \sup_{|z| \leq \sqrt{2m_2}} \frac{1}{|f^*_c(z)|^2} + \epsilon \), and assume that, for any \( c_1 > 0 \), there exists \( \Sigma > 0 \) such that
\[
\sum_{\mathbf{m}} \delta(m_2) \exp \left( -c_1 m_1 \frac{\Delta(m_2)}{\delta(m_2)} \right) \leq \Sigma < +\infty.
\]

**Case (L).** Assume that, for any \( c_1 > 0 \), there exists \( \Sigma > 0 \) such that
\[
\sum_{\mathbf{m}} \|G_{m_2}^{-1}\|_{op}^2 e^{-c_1 m_1 L(m_2)} \leq \Sigma < +\infty
\]

Let us comment these conditions. First, condition \(39\) if fulfilled for all our bases. Under Assumption \( A6 \), \( \|G_{m_2}^{-1}\|_{op}^2 = O(m_2^{2\alpha}) \) and condition \(41\) is fulfilled. Now we discuss condition \(40\). In the ordinary smooth case where \( \delta = \gamma = 0 \),
\[
\delta(m_2) \exp(-c_1 m_1 \frac{\Delta(m_2)}{\delta(m_2)}) \sim m_2^2 \exp(-c_1 m_1 \sqrt{m_2})
\]
is indeed summable and condition \(40\) is fulfilled. In the super-smooth case, with Lemma 1 in [Comte and Lacour (2013)](ComteLacour2013), \( \Delta(m_2) \sim C m_2^{ \alpha+(1-\gamma)/2 } \exp(2\beta(2m_2)^{\gamma/2}) \); then condition \(40\) is fulfilled if \( \gamma < 1/2 \). Otherwise, the penalty must be slightly increased, see the choice of \( C(h) \) in (19) of [Comte and Lacour (2013)](ComteLacour2013) and the associated comment (2)-(b) p.582.

### 6.2. General adaptive estimator and result.

**Assumption A7.** The conditional density \( \pi \) of \( Y \) given \( X \) is bounded on \( A \).

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 \( \hat{\pi}^{(sup)}_{\mathbf{m}, \mathbf{m}'} \) 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)](GoldenshlugerLepski2011) adapted to model selection in the spirit of [Chagny (2013)](Chagny2013). Let then
\[
A^{(sup)}(\mathbf{m}) = \sup_{\mathbf{m}' \in \mathcal{M}^{(sup)}} \left( \|\hat{\pi}^{(sup)}_{\mathbf{m}', \mathbf{m}} - \hat{\pi}^{(sup)}_{\mathbf{m}'}\|_n^2 - V^{(sup)}(\mathbf{m}') \right)^+ +
\]
with \( V^{(sup)}(\mathbf{m}') \) defined by \(37\) and \( a_+ = \max(a, 0) \) denotes the positive part of \( a \). We select the model \( \mathbf{m} \) with the following rule
\[
\hat{\mathbf{m}}^{(sup)} = \arg \min_{\mathbf{m} \in \mathcal{M}^{(sup)}} \left\{ A^{(sup)}(\mathbf{m}) + V^{(sup)}(\mathbf{m}) \right\}.
\]
Our final estimator is
\[ \hat{\pi}(\mathcal{S}_{up}) = \hat{\pi}(\mathcal{S}_{up})_{\text{opt}}. \]

The first result is obtained conditionally on \( X_1, \ldots, X_n \).

**Theorem 6.** Assume that Assumption A1 and A7 hold. Assume that condition (39) for \((\mathcal{S}_{up}) = (D)\), Assumptions A2, A3 and condition (40) for \((\mathcal{S}_{up}) = (H)\) and Assumptions A2, A5 and condition (41) for \((\mathcal{S}_{up}) = (L)\), hold. Then, we have a.s.

\[ \mathbb{E} \left[ \| \pi - \hat{\pi}(\mathcal{S}_{up}) \|_n^2 | X \right] \leq C \inf_{m \in \mathcal{M}_n(\mathcal{S}_{up})} \{ \| \pi - \pi_m \|_n^2 + V(\mathcal{S}_{up})(m) \} + \frac{C'}{n}, \]

where \( C \) is a numerical constant and \( C' \) is a constant which depends on \( \| \pi \|_{\infty}, \Sigma \), but not on \( (X_1, \ldots, X_n) \) 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 6, we have

\[ \mathbb{E} \| \pi - \hat{\pi}(\mathcal{S}_{up}) \|_n^2 \leq C \inf_{m \in \mathcal{M}_n(\mathcal{S}_{up})} \{ \| \pi - \pi_m \|_n^2 + V(\mathcal{S}_{up})(m) \} + \frac{C''}{n}, \]

where \( C \) is a numerical constant and \( C'' \) is a constant which depends on \( \| \pi \|_{\infty}, \Sigma \).

**Interpretation of Inequality (43).** Inequality (43) states that the estimator is adaptive in the sense it performs automatically the compromise between the squared-bias and the term \( V(\mathcal{S}_{up})(m) \) over the collection \( \mathcal{M}_n(\mathcal{S}_{up}) \), up to the multiplicative constant \( C \) and the additive negligible term \( C''/n \).

Therefore, we can discuss the cases where optimal rates are automatically reached.

- First, let us comment about the direct case and trigonometric basis (i.e. a compactly supported case). In that case, \( V(D)(\mathbf{m}) = K_0 m_1 m_2/n \) has the order of the variance term. Moreover, as 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, obviously from (18), we get that \( \| \Psi_{m_1}^{L} \|_{\text{op}} \leq 1/f_0 \). Therefore \( L(m_1) \| \Psi_{m_1}^{L} \|_{\text{op}} \leq L(m_1)/f_0 \) and as \( L(m_1) = m_1 \), the stability constraint appearing in the definition of \( \mathcal{M}_n^{(D)} \) reduces to \( m_1 \leq (f_0 \delta^*/2)(n/\log^2(n)) \) which does not prevent from choosing any optimal \( m_{1, \text{opt}} \). Therefore, in this case, the estimator reaches the optimal rate on the associated regularity spaces. Note that we recover the results obtained in the compact case in [Brunel et al. 2007]. However, contrary to this work, the penalty we obtain here, \( V(D)(\mathbf{m}) \), does not depend on \( f_0 \) nor on \( \| \pi \|_{\infty} \), which is an important improvement.

- Next for the direct case and the Hermite basis, \( V(D)(\mathbf{m}) = K_0 m_1 \sqrt{m_2}/n \). Therefore, if \( \pi \) is in a Sobolev-Hermite ellipsoid, the estimator automatically reaches the rate obtained in (24) under the condition that \( m_1^* \) belongs to \( \mathcal{M}_n^{(D)} \). In addition, this constraint implies that \( m_1^* \) satisfies the condition (19).

- The same conclusion holds for the Hermite basis in the noisy case, where \( V(H)(\mathbf{m}) = K_0 m_1 \Delta(m_2)/n \) is exactly the variance order and \( \mathcal{M}_n^{(H)} \) contains the stability constraint. A restriction occurs because of condition (40) which requires \( \gamma < 1/2 \); but its impact is in fact minor, see the comment after (40).

- For the Laguerre case, the penalty term in the direct case is \( V(D)(\mathbf{m}) = K_0 m_1 m_2/n \), and in the noisy case is \( V(L)(\mathbf{m}) = K_0 m_1 m_2 |G_m^{-1}\|_{\text{op}}/n \), which correspond to the variance orders obtained, without Assumption (7), in (16) and (33). Therefore, this leads to the rate given in (22) for the direct case, and in Proposition 6 for the noisy case, under
the stability condition and for $\pi$ belonging to some Sobolev-Laguerre ellipsoid. These rates are not proved being optimal, but a relevant squared-bias/variance compromise is automatically performed by $\hat{\pi}_m^{(L)}$.

Note that a compactly supported basis can be used in $x$ and the Hermite basis for deconvolving in $y$; this may suppress stability conditions but this would make the bias term of particular feature.

To end this section, we indicate that a control of the risk expressed in term of the $L^2(A, f(x)dx dy)$-norm is possible but at the price of replacing the stability constraint by a stronger one, both for the theoretical and the empirical one, which requires new definitions of theoretical and empirical collection of models, together with additional elements of proof. The interested reader is referred to Dussap (2022), Theorem 4.4 and its proof, where a similar study is conducted in a multivariate regression setting.

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 $(X_i, Y_i)_{1 \leq i \leq n}$ so-called direct observations, or $(X_i, Z_i)_{1 \leq i \leq n}$ 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 used to rely on compactly supported bases. This is why the non-noisy case, which does not involve any deconvolution problem, was studied before in the compact case only. On the opposite, deconvolution methods require Fourier transforms and inversions which are specifically more convenient with non compactly supported bases. Thus, studying conditional density estimation in the direct case with possibly non compactly supported bases is new and of interest; our penalty proposals do not contain any unknown quantity to estimate. This is done thanks to the ideas in Comte and Genon-Catalot (2020) and Goldenshluger and Lepski (2011), and conducts to new risk bounds for both simple (fixed projection space) and adaptive estimators. Then, our two extensions to noisy cases, either with $\mathbb{R}_+ \times \mathbb{R}_+$-supported or with real valued variables, also 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), and in regression context in Dussap (2022): combining both extensions would be interesting.

8. Proofs

8.1. Proof of Lemma 1 We first use (4) 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 $h_j^*(x) = \int_{|u| \leq \sqrt{2m+1}} e^{iux}h_j(u)du + \int_{|u| > \sqrt{2m+1}} e^{iux}h_j(u)du$ and using (5), we get, for $j \leq m - 1$,

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

Thus

$$\sum_{j=0}^{m-1} h_j^2(x) \leq 2\|1_{|\cdot| \leq \sqrt{2m+1}}\|^2 + 2Cm e^{-\xi(2m+1)/2} \int e^{-\xi u^2/2} du = 2\sqrt{2m+1} + \frac{2\sqrt{2\pi}C}{\sqrt{\xi}} e^{-\xi(2m+1)/2}.$$
This implies the result of Lemma 1 with $K = K(C, \xi)$. □

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

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

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

$$
\langle \pi, \varphi_j \otimes \varphi_k \rangle = \frac{1}{n} \sum_{i=1}^{n} \int \pi(X_i, y) \varphi_j(X_i) \varphi_k(y) \, dy = \frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) \int \pi(X_i, y) \varphi_k(y) \, dy
$$

$$
= \frac{1}{n} \left[ \Phi_{m_1} \left( \int \pi(X_i, y) \varphi_k(y) \, dy \right)_{1 \leq i \leq n, 0 \leq k \leq m_2 - 1} \right]_{j,k},
$$

and

$$
\langle \pi_{m,n}, \varphi_j \otimes \varphi_k \rangle = \frac{1}{n} \sum_{i=1}^{n} \sum_{j'=0}^{m_1-1} \sum_{k'=0}^{m_2-1} [D_m][j',k'] \varphi_j(X_i) \varphi_{j'}(X_i) \int \varphi_k(y) \varphi_{k'}(y) \, dy
$$

$$
= \sum_{j'=0}^{m_1-1} [D_m][j',k] \frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) \varphi_{j'}(X_i) = \sum_{j'=0}^{m_1-1} [D_m][j',k] \Phi_{m_1}[j,j'] = \left[ \Phi_{m_1} D_m \right][j,k]
$$

are equal, implies formula (15). The last part of the result follows from

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

where $\Theta_{m_2}(Y)$ is defined by (13). □

8.3. **Proof of Proposition 1** We start from equation (14). By elementary algebraic computation, we find

$$
\| \bar{\pi} - \pi_{m,n} \|_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left( \bar{\pi}(X_i, y) - \pi_{m,n}(X_i, y) \right)^2
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} \int \left( \sum_{j,k} \left[ \hat{A}_m[j,k] - [D_m][j,k] \right] \varphi_j(X_i) \varphi_k(y) \right)^2
$$

$$
= \frac{1}{n} \sum_{k=0}^{m_2-1} \sum_{i=1}^{n} \left( \sum_{j=0}^{m_1-1} \left[ \hat{A}_m[j,k] - [D_m][j,k] \right] \varphi_j(X_i) \right)^2
$$

$$
= \text{Tr} \left[ \hat{\Phi}_{m_1} \hat{D}_m \hat{\Phi}_{m_1} \right] \left[ \hat{A}_m - D_m \right] \hat{\Phi}_{m_1} \left[ \hat{A}_m - D_m \right] \hat{\Phi}_{m_1}
$$

Replacing the matrix coefficients by their formula, we get

$$
\| \bar{\pi} - \pi_{m,n} \|_n^2 = \frac{1}{n^2} \text{Tr} \left[ \hat{\Phi}_{m_1} \hat{D}_m \hat{\Phi}_{m_1} \left( \Theta_{m_2}(Y) - \mathbb{E} \left( \Theta_{m_2}(Y) | X \right) \right) \right].
$$

Then

$$
\mathbb{E} \left[ \| \bar{\pi} - \pi_{m,n} \|_n^2 | X \right] = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=0}^{m_2-1} \mathbb{E} \left[ \left( \varphi_j(Y_i) - \mathbb{E}(\varphi_j(Y_i) | X_i) \right)^2 \left| X_i \right. \right] \hat{\Phi}_{m_1} \hat{\Phi}_{m_1} \left[ \hat{A}_m - D_m \right] \hat{\Phi}_{m_1} [i,j].
$$
Now, note that \( [\hat{\Phi}_{m_j} \hat{\Psi}_{m_j}^{-1} \hat{\Phi}_{m_j}]_{i,i} \geq 0 \) as it is of the form \( \epsilon_i M \epsilon_i = \|M^{1/2} \epsilon_i\|_2^2 \) for \( M \) positive definite. Under \( \mathcal{C} \) for Laguerre basis or by Lemma \( \mathcal{A} \) for Hermite basis,
\[
\sum_{j=0}^{m_2-1} \mathbb{E} \left[ (\varphi_j(Y_i) - \mathbb{E}(\varphi_j(Y_i)|X_i))^2 |X_i \right] \leq \sum_{j=0}^{m_2-1} \mathbb{E} \left( \varphi_j^2(Y_i)|X_i \right) \leq c \sqrt{m_2}
\]
and
\[
\mathbb{E} \left[ \|\hat{\pi}_m - \pi_{m,n}\|^2_n |X \right] \leq c \frac{m_2}{n^2} \text{Tr} \left( \hat{\Phi}_{m_j} \hat{\Psi}_{m_j}^{-1} \hat{\Phi}_{m_j} \right) = c \frac{m_1 m_2}{n},
\]
as \( \text{Tr} \left( \hat{\Phi}_{m_j} \hat{\Psi}_{m_j}^{-1} \hat{\Phi}_{m_j} \right) = n \text{Tr} \left( \hat{\Phi}_{m_j} \left( \hat{\Phi}_{m_j} \hat{\Phi}_{m_j} \right)^{-1} \hat{\Phi}_{m_j} \right) = n \text{Tr} \left( \left( \hat{\Phi}_{m_j} \hat{\Phi}_{m_j} \right)^{-1} \hat{\Phi}_{m_j} \hat{\Phi}_{m_j} \right) = n m_1 \).
In the general case, we have
\[
\sum_{j=0}^{m_2-1} \mathbb{E} \left[ (\varphi_j(Y_i) - \mathbb{E}(\varphi_j(Y_i)|X_i))^2 |X_i \right] \leq \sum_{j=0}^{m_2-1} \mathbb{E} \left( \varphi_j^2(Y_i)|X_i \right) \leq L(m_2),
\]
and the variance bound becomes:
\[
\mathbb{E} \left[ \|\hat{\pi}_m - \pi_{m,n}\|^2_n |X \right] \leq \frac{m_1 L(m_2)}{n}. \quad \square
\]

### 8.4. Proof of Theorems \( \mathcal{B} \), \( \mathcal{E} \) and \( \mathcal{G} \)

Let us define the sets
\[
\Lambda_{m_1} = \left\{ L(m_1) \left( \|\hat{\Psi}_{m_j}^{-1}\|_{\text{op}} \vee 1 \right) \leq \mathfrak{d}_0 \frac{n}{\log(n)} \right\} \quad \text{and} \quad \Omega_{m_1} = \left\{ \|\Psi_{m_j}^{-1/2} \hat{\Psi}_{m_j} \Psi_{m_j}^{-1/2} - \text{Id}_{m_1}\|_{\text{op}} \leq 1 \right\}.
\]
Then we know from [Comte and Genon-Catalot, 2020] that for \( r = 3 \) in \( \mathfrak{d}_0(r) \)
\[
\mathbb{P}(\Lambda_{m_1}^c) \leq \mathbb{P}(\Omega_{m_1}^c) \leq 2n^{-r} = 2n^{-3}.
\]
In our setting, we can prove the following Lemma.

**Lemma 4.** Denote by \( (\varphi_j)_{0 \leq j \leq m_1-1} \) an \( L^2(f(x)dx) \) orthonormal basis of \( S_{m_1} \).

\[
\Omega_{m_1} = \left\{ \sup_{T \in S_{m_1} \|T\|_{f} \neq 0} \left| \frac{\|T\|_n^2}{\|T\|_f^2} - 1 \right| \leq 1 \right\} = \left\{ \|\hat{H}_{m_1} - \text{Id}_{m_1}\|_{\text{op}} \leq 1 \right\}
\]
where \( \hat{H}_{m_1} = ((\varphi_j, \varphi_k)_{1 \leq j, k \leq m_1}) \) is the Gram matrix of \( (\varphi_j)_{1 \leq j \leq m_1} \).

**Proof of Lemma 4.** We prove the first equality. Let \( T \in S_{m_1} \) and denote \( T(x,y) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} a_{j,k} \varphi_j(x) \varphi_k(y) \).
\[
A_m = (a_{j,k})_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}.
\]
Then we have \( \|T\|_f^2 = \text{Tr} \left[ t A_m \Psi_{m_1} A_m \right] \) and \( \|T\|_n^2 = \text{Tr} \left[ t A_m \hat{\Psi}_{m_1} A_m \right] \).
Let \( T \) be such that \( \|T\|_f = 1 \). Then
\[
\frac{\|T\|_n^2}{\|T\|_f^2} - 1 = \|T\|_n^2 - \|T\|_f^2 = \text{Tr} \left[ t A_m \left( \hat{\Psi}_{m_1} - \Psi_{m_1} \right) A_m \right]
\]
\[
= \text{Tr} \left[ (\Psi_{m_1}^{-1/2} \hat{\Psi}_{m_j} \Psi_{m_j}^{-1/2} - \text{Id}_{m_1}) \Psi_{m_1}^{1/2} A_m \Psi_{m_1}^{1/2} \right]
\]
\[
\leq \|\Psi_{m_1}^{-1/2} \hat{\Psi}_{m_j} \Psi_{m_j}^{-1/2} - \text{Id}_{m_1}\|_{\text{op}} \|\Psi_{m_1}^{1/2} A_m \Psi_{m_1}^{1/2}\|
\]
using that the first matrix is symmetric and the second is symmetric positive (thus with positive diagonal elements). Now
\[
\text{Tr} \left[ \Psi_{m_1}^{1/2} A_m \Psi_{m_1}^{1/2} \right] = \text{Tr} \left[ t A_m \Psi_{m_1} A_m \right] = \|T\|_f^2 = 1.
\]
As a consequence

$$\sup_{T \in S_m, \|T\| \neq 0} \left| \frac{\|T\|^2_n}{\|T\|_F^2} - 1 \right| \leq \|\Psi_{m_1}^{-1/2} \hat{\Psi}_{m_1} \Psi_{m_1}^{-1/2} - \text{Id}_{m_1} \|_{op}.$$  

Now we show that the supremum is reached for a well chosen $T$ or equivalently $A_m$. Write $\Psi_{m_1}^{-1/2} \hat{\Psi}_{m_1} \Psi_{m_1}^{-1/2} - \text{Id}_{m_1} = PDP$ where $P^TP = PP = \text{Id}_{m_1}$ is orthogonal and $D = \text{diag}(d_1, \ldots, d_{m_1})$ is diagonal. We choose $d_1$ such that $|d_1| = \max(|d_i|, i = 1, \ldots, m_1)$ and thus $\|\Psi_{m_1}^{-1/2} \hat{\Psi}_{m_1} \Psi_{m_1}^{-1/2} - \text{Id}_{m_1}\|_{op} = |d_1|$. Then $T$ defined by $A_m = \Psi_{m_1}^{-1/2} P E_m$ where $E_m$ is a $m_1 \times m_2$ matrix with all coefficients equal to 0 except the $(1,1)$-one which is equal to 1. Then it is easy to see that $\|T\|^2 = 1$ and $\|\frac{T}{T}\|^2 - 1 = |d_1|$ and the upper bound is reached, hence the equality. This ends the proof of the first part of Lemma [4]

The second equality holds by noting that in the new basis $(\varphi_j \otimes \varphi_k)_{j,k}$ of decomposition of $T$, the analogous of $\hat{\Psi}_{m_1}$ is $\hat{H}_{m_1}$ and the analogous of $\Psi_{m_1}$ is the $m_1 \times m_1$ identity matrix. \(\square\)

Now we start the proof of the theorems with Theorem [1] first in mind, and the two others next, for the points where there are differences. We write the decomposition (recall that $\pi_A = \pi_{1A}$)

$$\|\pi_m^{(D)} - \pi_A\|_f^2 = \|\pi_m^{(D)} - \pi_A\|_f^2 \hat{\Phi}_{m_1} \cap \Omega_{m_1} + \|\pi_m^{(D)} - \pi_A\|_f^2 \hat{\Phi}_{m_1} \cap \Omega_{m_1} + \|\pi_A\|_f^2 \hat{\Phi}_{m_1} \cap \Omega_{m_1} := T_1 + T_2 + T_3.$$

To start with the simplest term, we note that $E(T_3) = \|\pi_A\|_f^2 \mathbb{P}(\Lambda_{m_1}^c) \leq \|\pi_A\|_f^2 \mathbb{P} n^{-3}.$

For $T_2$, we split it in two terms $T_2 \leq 2(\|\pi_m^{(D)}\|_f^2 + \|\pi_A\|_f^2) \hat{\Phi}_{m_1} \cap \Omega_{m_1}$ where

$$E \left[ \|\pi_A\|_f^2 \hat{\Phi}_{m_1} \cap \Omega_{m_1} \right] \leq \|\pi_A\|_f^2 \mathbb{P} \left( \Omega_{m_1} \right) \leq \frac{2\|\pi_A\|_f^2}{n^3}.$$

On the other hand,

$$\|\pi_m^{(D)}\|_f^2 = \text{Tr} \left( \hat{A}_m^{(D)} \Psi_{m_1} \hat{A}_m^{(D)} \right) = \frac{1}{2} \text{Tr} \left( \hat{\Theta}_{m_2}(Y) \hat{\Phi}_{m_1} \Psi_{m_1} \hat{\Phi}_{m_1} \Theta_{m_2}(Y) \right)$$

$$= \frac{1}{n} \|\Psi_{m_1}\|_{op} \frac{1}{n} \|\hat{\Phi}_{m_1} \Psi_{m_1}^{2} \hat{\Phi}_{m_1}\|_{op} \text{Tr} \left( \hat{\Theta}_{m_2}(Y) \hat{\Theta}_{m_2}(Y) \right)$$

$$\leq \frac{L(m_1)}{n} \|\hat{\Psi}_{m_1}\|_{op} n L(m_2).$$

Note that for $\pi_m^{(H)}$, we get at that point

$$\|\pi_m^{(H)}\|_f^2 = \text{Tr} \left( \hat{A}_m^{(H)} \Psi_{m_1} \hat{A}_m^{(H)} \right) = \frac{1}{n^2} \text{Tr} \left( \hat{\Upsilon}_{m_2}(Z) \hat{\Phi}_{m_1} \Psi_{m_1}^{2} \hat{\Phi}_{m_1} \Upsilon_{m_2}(Z) \right)$$

$$= \frac{1}{n} \|\Psi_{m_1}\|_{op} \frac{1}{n} \|\hat{\Phi}_{m_1} \Psi_{m_1}^{2} \hat{\Phi}_{m_1}\|_{op} \text{Tr} \left( \hat{\Upsilon}_{m_2}(Z) \hat{\Upsilon}_{m_2}(Z) \right)$$

$$\leq \frac{L(m_1)}{n} \|\hat{\Psi}_{m_1}\|_{op} n \Delta(m_2).$$

Lastly, for $\pi_m^{(L)}$, we get

$$\|\pi_m^{(L)}\|_f^2 = \text{Tr} \left( \hat{A}_m^{(L)} \Psi_{m_1} \hat{A}_m^{(L)} \right) = \frac{1}{n^2} \text{Tr} \left( \hat{G}_{m_2} \hat{\Theta}_{m_2}(Z) \hat{\Phi}_{m_1} \Psi_{m_1}^{2} \hat{\Phi}_{m_1} \Theta_{m_2}(Z) \hat{G}_{m_2}^{-1} \right)$$

$$\leq \frac{1}{n} \|\Psi_{m_1}\|_{op} \frac{1}{n} \|\hat{\Phi}_{m_1} \Psi_{m_1}^{2} \hat{\Phi}_{m_1}\|_{op} \text{Tr} \left( \hat{G}_{m_2} \hat{\Theta}_{m_2}(Z) \hat{\Theta}_{m_2}(Z) \hat{G}_{m_2}^{-1} \right)$$

$$\leq \frac{L(m_1)}{n} \|\hat{\Psi}_{m_1}\|_{op} n L(m_2) \|\hat{G}_{m_2}\|_f^2.$$
As a consequence, for \( L(m_2) \leq n \) in the direct case,
\[
\mathbb{E}(\|\hat{\pi}^{(D)}_\mathbf{m}\|_f^2 \mathbf{I}_{\Lambda_{m_1} \cap \Omega_{m_1}}) \leq \mathbb{E} \left( L(m_1) \|\tilde{\Psi}^{-1}_\mathbf{m} \|_{\text{op}} L(m_2) \mathbf{I}_{\Lambda_{m_1} \cap \Omega_{m_1}} \right) \\
\leq \frac{\delta_0 n}{\log(n)} L(m_2) \mathbb{P}(\Omega_{m_1}) \leq \frac{2\delta_0}{n^{r-2}} = \frac{2\delta_0}{n}.
\]

The same result holds for \( \mathbb{E}(\|\hat{\pi}^{(L)}_\mathbf{m}\|_f^2 \mathbf{I}_{\Lambda_{m_1} \cap \Omega_{m_1}}) \) under the condition \( L(m_2) \|G^{-1}_{m_2}\|_{\text{op}} \leq n \) and for \( \mathbb{E}(\|\hat{\pi}^{(H)}_\mathbf{m}\|_f^2 \mathbf{I}_{\Lambda_{m_1} \cap \Omega_{m_1}}) \) under \( \Delta(m_2) \leq n \). Consequently, we have \( \mathbb{E}(T_2) \leq c/n \) for \( c \) a positive constant depending on \( \delta_0 \) and \( \|\pi_A\|_f^2 \).

Now we turn to the study of \( T_1 \). Recall that \( \pi_{m,f} \) is the orthogonal projection of \( \pi \) on \( S_m \) for the \( L^2(A, f(x)dx dy) \) scalar product. We write, for \( T_1 \), that is on \( \Omega_{m_1} \cap \Lambda_{m_1} \)
\[
\|\hat{\pi}^{(D)}_\mathbf{m} - \pi_A\|_f^2 = \|\hat{\pi}^{(D)}_\mathbf{m} - \pi_{m,f}\|_f^2 + \|\pi_{m,f} - \pi_A\|_f^2 \\
\leq 2\|\hat{\pi}^{(D)}_\mathbf{m} - \pi_{m,n}\|_f^2 + 2\|\pi_{m,n} - \pi_{m,f}\|_f^2 + \|\pi_{m,f} - \pi_A\|_f^2 \\
\leq \|\pi_{m,f} - \pi_A\|_f^2 + 4\|\hat{\pi}^{(D)}_\mathbf{m} - \pi_{m,n}\|_f^2 + 2\|\pi_{m,n} - \pi_{m,f}\|_f^2
\]

(45)

where we used that \( \|\hat{\pi}^{(D)}_\mathbf{m} - \pi_{m,n}\|_f^2 \leq 2\|\hat{\pi}^{(D)}_\mathbf{m} - \pi_{m,n}\|_n^2 \) on \( \Omega_{m_1} \). Now the first term is a bias term with coefficient 1, and the second term is the variance term in the three contexts and will have the variance order for the three estimates.

The aim is to bound \( \mathbb{E}(\|\pi_{m,n} - \pi_{m,f}\|_f^2 \mathbf{I}_{\Omega_{m_1} \cap \Lambda_{m_1}}) \). Set \( g = \pi - \pi_{m,f} \) and denote by \( \Pi_{\mathbf{m}}^{(n)} \) the orthogonal projection on \( S_m \) with respect to the empirical norm. Clearly \( \Pi_{\mathbf{m}}^{(n)} g = \pi_{m,n} - \pi_{m,f} \). Denote by \( (\hat{\varphi}_j)_{0 \leq j \leq m_1 - 1} \) an \( L^2(f(x)dx) \) orthonormal basis of \( S_m \). Then \( (\hat{\varphi}_j \otimes \varphi_k)_{0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1} \) is an \( L^2(f(x)dx dy) \) orthonormal basis of \( S_m \). If we decompose \( \Pi_{\mathbf{m}}^{(n)} g = \sum_{j,k} c_{j,k} \hat{\varphi}_j \otimes \varphi_k \), writing that for all \( j' = 0, \ldots, m_1 - 1 \) and \( k' = 0, \ldots, m_2 - 1 \), \( \langle g, \hat{\varphi}_{j'} \otimes \varphi_{k'} \rangle_n = \langle \Pi_{\mathbf{m}}^{(n)} g, \hat{\varphi}_{j'} \otimes \varphi_{k'} \rangle_n \) implies that \( C = (c_{j,k})_{j,k} = \hat{H}^{-1}_{m_1} (\langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n)_{j,k} \) where \( [\hat{H}_{m_1}]_{j,k} = \frac{1}{n} \sum_{i=1}^n \langle \hat{\varphi}_j(X_i) \hat{\varphi}_j(X_i) \rangle \). Note that \( \hat{H}_{m_1} \) is a Gram matrix and we have, by Lemma 4, \( \Omega_{m_1} = \{ \|\hat{H}_{m_1} - \text{Id}_{m_1}\|_\text{op} \leq \frac{1}{2} \} \). Therefore we get
\[
\|\Pi_{\mathbf{m}}^{(n)} g\|_f^2 = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} c_{j,k}^2 = \text{Tr}[C' C] = \text{Tr} \left[ (\langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n)_{j,k} \right]^{-1} \left[ (\langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n)_{j,k} \right]^{-1} \\
\leq \|\hat{H}^{-1}_{m_1}\|_\text{op} \sum_{j,k} \langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n^2.
\]

Therefore, noting that \( \mathbb{E}(\langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n) = \langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_f = 0 \) as \( g = \pi - \pi_{m,f} \) is orthogonal to \( S_m \), and using that \( \|\hat{H}^{-1}_{m_1}\|_\text{op} \leq 2 \) on \( \Omega_{m_1} \), we get
\[
\mathbb{E} \left[ \|\Pi_{\mathbf{m}}^{(n)} g\|_f^2 \mathbf{I}_{\Omega_{m_1} \cap \Lambda_{m_1}} \right] \leq 4 \mathbb{E} \left[ \sum_{j,k} \langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n^2 \right] \\
= 4 \sum_{j,k} \text{Var} \left[ \langle g, \hat{\varphi}_j \otimes \varphi_k \rangle_n \right] = \frac{4}{n} \sum_{j,k} \text{Var} \left( \int g(X_1, y) \hat{\varphi}_j(X_1) \varphi_k(y) dy \right) \\
\leq \frac{4}{n} \mathbb{E} \left[ \sum_{j} \hat{\varphi}_j^2(X_1) \sum_k \left( \int g(X_1, y) \varphi_k(y) dy \right)^2 \right].
\]
Now we note that
\[
\sum_k \left( \int g(X_1, y) \varphi_k(y) dy \right)^2 = \| \Pi_{S_{m_2}} g(X_1, \cdot) \|^2 \leq \| g(X_1, \cdot) \|^2 = \int g^2(X_1, y) dy
\]
and
\[
\sup_x \sum_j \varphi_j^2(x) = \sup_{t \in S_{m_1}} \| t \|^2 \leq \sup_{t \in S_{m_1}} \| t \|^2 \sup_{t \in S_{m_1}} \| t \|^2 \leq L(m_1) \| \Psi_{m_1}^{-1} \|_{op}.
\]
Therefore we get
\[
\mathbb{E} \left[ \| \Pi_m^{(n)} g \|_f^2 1_{\Omega_{m_1} \cap \Lambda_{m_1}} \right] \leq \frac{4}{n} L(m_1) \| \Psi_{m_1}^{-1} \|_{op} \mathbb{E}(g^2(X_1, y) dy).
\]
Lastly the equality
\[
\mathbb{E}(g^2(X_1, y) dy) = \int \int g^2(x, y) f(x) dx dy = \| g \|_f^2 = \| \pi - \pi_{m,f} \|_f^2
\]
and the stability condition (19) imply
\[
\mathbb{E} \left[ \| \pi_{m,n} - \pi_{m,f} \|_f^2 1_{\Omega_{m_1} \cap \Lambda_{m_1}} \right] = \mathbb{E} \left[ \| \Pi_m^{(n)} g \|_f^2 1_{\Omega_{m_1} \cap \Lambda_{m_1}} \right] \leq \frac{2 \delta_0}{\log(n)} \| \pi - \pi_{m,f} \|_f^2.
\]
Plugging this in (45) yields
\[
\mathbb{E}(T_1) \leq \left( 1 + \frac{4 \delta_0}{\log(n)} \right) \| \pi - \pi_{m,f} \|_f^2 + \frac{4 m_1 L(m_2)}{n},
\]
which is the last piece of the announced result for the direct estimator, and with the adequate variance bounds replacing $m_1 L(m_2)/n$ in the Laguerre and Hermite indirect cases. \(\square\)

8.5. Proof of Theorem 2\(\). In the Hermite case, it is sufficient to follow the lines of the proof of Theorem 4\(\) below (lower bound in the noisy case), taking for the noise distribution the Dirac measure at 0. The Lemma 7 is identical, and the Lemma 8 remains valid by setting $\pi_0 * f_x = \pi_0$, $f_x = 1$, $(f_x)' = 0$, as well as $\alpha = \beta = \gamma = \alpha' = 0$. The conclusion and the parameter values are the same with $\alpha = 0$.

In the Laguerre case, the approach is the same as the proof of Theorem 4\(\) but with different functions $\pi_0$. Note that the assumption $\mathbb{E}(1/\sqrt{\psi}|X = x) \leq C$ for all $x$ is fulfilled for $\pi_0$ and $\pi_\theta$ below. This is why the lower bound concerns the rate under condition (8). Let us define
\[
\pi_0(x, y) = \pi_0(y) = \frac{1}{2} 1_{[0,1]}(y) + P_L(y) 1_{[1,2]}(y)
\]
where $P_L$ is a polynomial, $P_L(y) \geq 0$ on $[1,2]$, $\int_1^2 P_L(y) dy = 1/2$, $P_L(1) = 1/2$, $P_L(2) = 0$, $P_L'(1) = P_L'(2) = 0$ for $k = 1, \ldots, s_2 + 1$. We set again
\[
\pi_\theta(x, y) = \pi_0(x, y) + \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} A_{j,k} \varphi_j(x) (m_2^{1/4} \psi(\sqrt{m_2 y} - k)),
\]
with the same $A$ as in the proof of Theorem 4\(\) but with $\psi$ is a bounded function with support $[0,1]$ such that $\int_0^1 \psi(u) du = 0$. Moreover, we assume that $\psi$ admits continuous bounded derivatives up to order $s_2$. We can prove

Lemma 5. \(\) \(\text{(a)}\) Assume that $f \in W_{s_1}(\mathbb{R}_+, R) \cap L^\infty(\mathbb{R}_+)$. Then there exists $L > 0$ such that $\pi_0$ is a conditional density belonging to $W_{s_2}(\mathbb{R}_+, L)$.
Lemma 6. We denote $\rho(\theta, \theta')$ the Hamming distance between $\theta$ and $\theta'$.

- 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 $\pi_\theta$ and under $\pi_0$ verifies $K(P_\theta^{\otimes n}, P_0^{\otimes n}) \leq 2\delta^2 \|\psi\|^2 m_1 \sqrt{m_2}$.

- For all $\theta, \theta' \in \{0, 1\}^{m_1 \sqrt{m_2}}$, $\|\pi_\theta - \pi_\theta'\|^2 = \delta^2 \|\psi\|^2 n^{-1} \rho(\theta, \theta')$

To conclude it is sufficient to use Theorem 2.5 of [Tsybakov (2009)] with $m_1 = m_1^*$ and $m_2 = m_2^*$ given by (23). The proofs of Lemma 5 and Lemma 6 can be found in an earlier version of this paper, see [Comte and La Cour (2021)]. They are very similar to the one of Lemma 2 and Lemma 8 but instead of Fourier analysis they use that the functions $y \mapsto \psi(\sqrt{m_2}y - k)$ have disjoint supports for different integers $k$. □

8.6. Proof of Proposition 3

We write again

$$\|\hat{\pi}_m^{(H)} - \pi\|^2_n = \|\pi - \pi_{m,n}\|^2_n + \|\hat{\pi}_m^{(H)} - \pi_{m,n}\|^2_n$$

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

Next,

$$\frac{1}{n^2} \text{Tr} \left[ \left(\hat{\mathcal{M}}_{m_2}(Z) - E\left(\hat{\mathcal{M}}_{m_2}(Z) | X\right)\right) \hat{\Phi}_{m_2}^{-1} \hat{\Phi}_{m_2} \left(\hat{\mathcal{M}}_{m_2}(Z) - E\left(\hat{\mathcal{M}}_{m_2}(Z) | X\right)\right) \right].$$

We have

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

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

$$\sum_{j=0}^{m_2-1} \left(\pi_{h_j}(Z_k) - E(\pi_{h_j}(Z_k) | X_k)\right)^2 = \sum_{j=0}^{m_2-1} \left(\frac{1}{2\pi} \int_{-2\pi}^{2\pi} h_j(u) \frac{e^{-iZ_k u} - E(e^{-iZ_k u} | X_k)}{f_{\hat{\xi}}(-u)} du \right)^2$$

$$= \sum_{j=0}^{m_2-1} \left(\frac{1}{2\pi} \int_{-2\pi}^{2\pi} h_j(u) \frac{e^{-iZ_k u} - E(e^{-iZ_k u} | X_k)}{f_{\hat{\xi}}(-u)} du \right)^2 \text{ with } (4)$$

$$\leq \frac{1}{2\pi} \sum_{j=0}^{m_2-1} \left(\int_{-2\pi}^{2\pi} h_j(u) \frac{e^{-iZ_k u} - E(e^{-iZ_k u} | X_k)}{f_{\hat{\xi}}(-u)} du \right)^2 \text{ with } (4)$$

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

$$\sum_{j=0}^{m_2-1} \left(\int_{-2\pi}^{2\pi} h_j(u) \frac{e^{-iZ_k u} - E(e^{-iZ_k u} | X_k)}{f_{\hat{\xi}}(-u)} du \right)^2 \leq \int \left| e^{-iZ_k u} - E(e^{-iZ_k u} | X_k) \right|_{|u| \leq \sqrt{2m_2}}^2 du$$

$$\leq 4 \int_{|u| \leq \sqrt{2m_2}} \frac{du}{|f_{\hat{\xi}}(u)|^2}. $$
On the other hand, using (5), we have, for \(|u| > \sqrt{2m_2} = \sqrt{(2m_2 - 1) + 1} > \sqrt{2j + 1}\) for any \(j \leq m_2 - 1\), \(|h_j(u)| \leq Ce^{-\xi u^2}\) and thus, as, under Assumption A, \(\eta = \xi - \beta > 0\),

\[
\sum_{j=0}^{m_2-1} \left( \int h_j(u) \frac{|e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)|}{f_\xi(-u)} 1_{|u| > \sqrt{2m_2}du} \right)^2 \leq \sum_{j=0}^{m_2-1} 4 \left( \int |e^{-iZ_ku} - \mathbb{E}(e^{-iZ_ku}|X_k)| \frac{Ce^{-\beta + \eta u^2}}{|f_\xi(u)|} du \right)^2 \\
\leq C' \sum_{j=0}^{m_2-1} e^{-4\eta m_2} \left( \int C e^{-\beta + \eta u^2} |u| \frac{Ce^{-\beta + \eta u^2}}{|f_\xi(u)|} du \right)^2 \leq c,
\]

for a constant \(c\) depending on \(f_\xi\) but not on \(m_2\).

Gathering the two parts, we obtain

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

and thus

\[
\mathbb{E} \left[ \left\| \pi_m^{(H)} - \pi_m, n \right\|_n^2 |X \right] \leq \frac{1}{n^2} \sum_{k=1}^{n} \Delta(m_2) \leq m_1 \Delta(m_2) n.
\]

8.7. Proof of Theorem 4

8.7.1. Core of the proof. As usual in the proofs of lower bounds, we build a set of conditional densities \((\pi_\theta)\) quite distant from each other in terms of the weighted \(L^2\)-norm, but whose distance between the resulting models is small. More precisely, let \(\psi\) be the Meyer wavelet built with with \(C^2\)-conjugate mirror filters (see for instance Section 7.7.2 of Mallat (2009)). We shall use in particular that \(\int f \psi = 0\) and there exist positive constants \(c_\psi, C_{\psi}\) such that

- \(|\psi(x)| \leq c_\psi(1 + |x|)^{-2}\) for any \(x \in \mathbb{R}\),
- \(\psi^*\) has support \(S = [-8\pi/3, -2\pi/3] \cup [2\pi/3, 8\pi/3]\),
- \(\psi^*\) is \(C^{s_2}\) and the functions \(|(\psi^*)^{(q)}|, q = 0, \ldots, s_2\) are upperbounded by \(C_{\psi} > 0\),
- \(|\psi^*|\) is lowerbounded by \(c_\psi > 0\) on the set \([-2\pi, -\pi] \cup [\pi, 2\pi]\),

where \(\psi^*(u) = \int e^{ixu} \psi(x) dx\) is the Fourier transform of \(\psi\). Now, let \(r = [s_2/2] + 1\) and

\[
\pi_0(x, y) = \pi_0(y) = \frac{c_r}{(1 + y^2)^r}
\]

with \(c_r\) such that \(y \mapsto \pi_0(y)\) is a density. Next, we assume without loss of generality that \(\sqrt{m_2}\) is a positive integer and we define

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

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

for \(\delta = \delta(m_2) > 0\) to be defined later. Now we shall use the following lemmas:

**Lemma 7.** (a) Assume that \(f \in W_{s_1}(\mathbb{R}, \mathbb{R}) \cap L^\infty(\mathbb{R})\). Then there exists \(L > 0\) such that \(\pi_0\) is a conditional density belonging to \(W^f_{s_2}(\mathbb{R}, L)\).

(b) If

\[
\delta^2 L(m_1) ||\Psi^{-1/2}\pi_0||_{op} \leq C_1 \frac{n}{m_1 \sqrt{m_2}}
\]

for \(C_1\) some positive constant only depending on \(\psi\), then for all \(\theta \in \{0, 1\}^{m_1 \sqrt{m_2}}\), \(\pi_\theta\) is a conditional density.
For all $\theta \in \{0,1\}^{m_1\sqrt{m_2}}$, $\pi_\theta - \pi_0$ belongs to $W^f_2(\mathbb{R}^2, L)$ as soon as
\[
\delta^2 m_1^{s_1+1} \sqrt{m_2} \leq \frac{L}{6C_\psi} \quad \text{and} \quad \delta^2 m_1^{s_2+1/2} \leq \frac{L}{2C_2}
\]
for some positive constant only depending on $\psi$ and $s_2$.

Then under the conditions of this lemma, the $\pi_{\theta}$'s are conditional densities belonging to $W^f_2(\mathbb{R}^2, 4L)$.

**Lemma 8.** We denote $\rho(\theta, \theta')$ the Hamming distance between $\theta$ and $\theta'$.

- For all $\theta \in \{0,1\}^{m_1\sqrt{m_2}}$, the Kullback divergence between the distribution of $(X_i, Z_i)_{1 \leq i \leq n}$ under $\pi_{\theta}$ and under $\pi_0$ verifies
\[
K(P^\otimes_\theta, P^\otimes_0) \leq C_3 \delta^2 m_1 m_2^{-\alpha' + 1/2} e^{-2\beta' m_2^{\gamma/2}}
\]
where $C_3 > 0$ only depends on $C_\psi$ and $f_\varepsilon$, $\alpha' = \min(\alpha, \alpha_1, \ldots, \alpha_r)$ and $\beta' = (2\pi/3)^\gamma$.

- For all $\theta, \theta' \in \{0,1\}^{m_1\sqrt{m_2}}$, $\|\pi_{\theta} - \pi_{\theta'}\|_2^2 \geq c_\varepsilon^2 \delta^2 n^{-1} \rho(\theta, \theta')$.

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 9.** Fix some even integer $K > 0$. There exists a subset $\{\theta^{(0)}, \ldots, \theta^{(M)}\}$ of $\{0,1\}^K$ and a constant $a_1 > 0$, such that $\theta^{(0)} = (0, \ldots, 0)$, $\rho(\theta^{(j)}, \theta^{(l)}) \geq a_1 K$, for all $0 \leq j < l \leq M$. Moreover it holds that, for some constant $a_2 > 0$, $M \geq 2^{a_2 K}$.

Now we choose, in the ordinary smooth case ($\gamma = \beta = 0$ and $\alpha' = \alpha$):
\[
\delta^2 = \eta^2 m_2^\alpha, \quad m_1 = \lceil n^{s_2}/[(\alpha+1/2)s_1+s_2(s_1+1)] \rceil, \quad \sqrt{m_2} = \lceil n^{s_2}/[(\alpha+1/2)s_1+s_2(s_1+1)] \rceil,
\]
and in the supersmooth case ($\gamma > 0$):
\[
\delta^2 = \eta^2 m_2^\alpha e^{2\beta' m_2^{\gamma/2}}, \quad m_1 = \lceil (\log n)^{(2s_2/(\gamma s_1))} \rceil, \quad \sqrt{m_2} = \left[\left(\frac{1}{2\beta'}(\log n - b \log \log n)\right)^{1/\gamma}\right],
\]
with $b = (2s_2/s_1 + 2\alpha' + 1 + 2s_2)/\gamma$ and $\eta$ a small enough constant only depending on $\psi, L, s_2$. We can then compute that if $\eta$ is small enough, the condition (c) of Lemma 7 is verified. Moreover the condition (b) is ensured by $L(m_1)\|\Psi_m^{-1}\|_{op} \leq \psi_n^{-2}$ if $\eta$ is small enough.

Finally, we have built $M$ conditional densities $\pi_{\theta^{(0)}}, \ldots, \pi_{\theta^{(M)}}$ belonging to $W^f_2(\mathbb{R}^2, R)$ such that
\[
\|\pi_{\theta^{(j)}} - \pi_{\theta^{(l)}}\|_2^2 \geq c_\varepsilon^2 \delta^2 a_1 K/n = c_\varepsilon^2 a_1 \delta^2 m_1 \sqrt{m_2}/n = C\psi_n^2
\]
and
\[
K(P^\otimes_{\theta^{(j)}}, P^\otimes_{\theta^{(l)}}) \leq C'\eta^2 m_1 \sqrt{m_2} = C'\log(M)
\]
with $C'' < 1/8$ if $\eta$ small enough. To conclude it is sufficient to use Theorem 2.5 of [Tsybakov (2009)].

### 8.7.2. Proof of Lemma 8

We start by proving Lemma 8 since it provides a computation useful for the proof of Lemma 7. We use the notation $\lesssim$ and $\gtrsim$ for inequalities up to a constant.

- Note that the density of $(X_i, Z_i)$ is
\[
f(x) \int \pi(x, y - u)f_\varepsilon(u)du = f(x)(\pi(x, \cdot) * f_\varepsilon)(y).
\]
The Kullback divergence between the distribution of $(X_i, Z_i)_{1 \leq i \leq n}$ under $\pi_\theta$ and under $\pi_0$ verifies
\[
K(P^\otimes_\theta, P^\otimes_0) \leq nK(P_\theta, P_0) \leq n\chi^2(P_\theta, P_0)
\]
where
\[ \chi^2(P_0, P_0) = \iint \frac{(\pi_0(x, \cdot) * f_\varepsilon - \pi_0 * f_\varepsilon)^2(y)}{\pi_0 * f_\varepsilon(y)} f(x)dx dy. \]

Denoting \( \psi_{m_2,k} = m_2^{1/4} \psi(\sqrt{m_2}y - k) \),
\[ \pi_0(x, \cdot) * f_\varepsilon - \pi_0 * f_\varepsilon = \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{k=0}^{\sqrt{m_2} - 1} A_{j,k} \varphi_j(x)(\psi_{m_2,k} * f_\varepsilon) \]

Then we compute

\[
\begin{align*}
\chi^2(P_0, P_0) &= \iint \left( \frac{\delta}{\sqrt{n}} \sum_{j=0}^{m_1-1} \sum_{k=0}^{\sqrt{m_2} - 1} A_{j,k} \varphi_j(x)(\psi_{m_2,k} * f_\varepsilon)(y) \right)^2 (\pi_0 * f_\varepsilon)^{-1}(y) f(x)dx dy \\
&= \frac{\delta^2}{n} \sum_{j,j'} \sum_{kk'} A_{j,k} A_{j',k'} \int \varphi_j(x) \varphi_{j'}(x)f(x)dx \int (\psi_{m_2,k} * f_\varepsilon)(\psi_{m_2,k'} * f_\varepsilon)(\pi_0 * f_\varepsilon)^{-1} \\
&= \frac{\delta^2}{n} \sum_{kk'} [A(\Psi_{m_1})A]_{k,k'} \int (\psi_{m_2,k} * f_\varepsilon)(\psi_{m_2,k'} * f_\varepsilon)(\pi_0 * f_\varepsilon)^{-1} \\
&= \frac{\delta^2}{n} \sum_{kk'} \sum_{j} \vartheta_{jk} \vartheta_{jk'} \int (\psi_{m_2,k} * f_\varepsilon)(\psi_{m_2,k'} * f_\varepsilon)(\pi_0 * f_\varepsilon)^{-1} \\
&= \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} \int \left( \sum_{k=0}^{\sqrt{m_2} - 1} \theta_{j,k} \psi_{m_2,k'} * f_\varepsilon \right)^2 (\pi_0 * f_\varepsilon)^{-1}. \end{align*}
\]

Using Lemme 5.1 of Fan (1991) (with \( F = F_\varepsilon \)) there exists \( A > 0 \) such that for \( |y| > A \),
\[ \pi_0 * f_\varepsilon(y) \gtrsim |y|^{-2r}. \]

Moreover for \( |y| \leq A, \pi_0 * f_\varepsilon(y) \gtrsim 1 \) since for any \( B > 0 \)
\[ \pi_0 * f_\varepsilon(y) \geq \int \frac{\pi^{-1} f_\varepsilon(u)}{1 + (y - u)^2} du \geq \int_{|u| < B} \frac{\pi^{-1} f_\varepsilon(u)}{1 + (y - u)^2} du \geq \frac{\pi^{-1}}{1 + (A + B)^2} \int_{|u| < B} f_\varepsilon(u) du \]
which is a positive quantity if \( B \) is well chosen. Now we denote
\[
I_{1j} = \int_{|y| \leq A} \frac{(\sum_k \theta_{j,k} \psi_{m_2,k} * f_\varepsilon)^2}{\pi_0 * f_\varepsilon}, \quad I_{2j} = \int_{|y| > A} \frac{(\sum_k \theta_{j,k} \psi_{m_2,k} * f_\varepsilon)^2}{\pi_0 * f_\varepsilon}
\]
such that \( \chi^2(P_0, P_0) = \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} (I_{1j} + I_{2j}) \). Let us bound these integrals. First, using Fourier analysis
\[
I_{2j} \lesssim \int_{|y| > A} \left( \sum_k \theta_{j,k} \psi_{m_2,k} * f_\varepsilon \right)^2 (y)|y|^{2r} dy \lesssim \int \left| \sum_k \theta_{j,k} (\psi_{m_2,k} * f_\varepsilon)(y) \right|^2 dy \\
\lesssim \int \left| \left( \sum_k \theta_{j,k} (\psi_{m_2,k} * f_\varepsilon) \right)^* \right|^2.
\]
Note that the Fourier transform of $\psi_{m_2,k}$ is $\psi_{m_2,k}^*(u) = e^{iuk/\sqrt{m_2}}m_2^{-1/4}\psi^*(u/\sqrt{m_2})$ then

$$\sum_k \theta_{j,k}(\psi_{m_2,k} * f_\epsilon)^* = m_2^{-1/4}D_j\left(\frac{u}{\sqrt{m_2}}\right)\psi^*\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)$$

denoting

$$D_j(v) = \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k}e^{ivk}.$$  

By differentiating the previous equality, we obtain

$$\left(\sum_k \theta_{j,k}(\psi_{m_2,k} * f_\epsilon)^*\right)^{(r)} = m_2^{-1/4} \sum_{q=0}^{r} \binom{r}{q} D_j\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)^{(q)} \cdot m_2^{(q-r)/2}(\psi^*)^{(r-q)}\left(\frac{u}{\sqrt{m_2}}\right).$$

Using the properties of $\psi$, we can write

$$I_{2j} \lesssim m_2^{-1/2} \left| \sum_{q=0}^{r} \binom{r}{q} D_j\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)^{(q)} \cdot m_2^{(q-r)/2}(\psi^*)^{(r-q)}\left(\frac{u}{\sqrt{m_2}}\right) \right|^2 du$$

$$\lesssim C_\psi m_2^{-1/2} \sum_{q=0}^{r} \binom{r}{q} m_2^{q-r} \int_{-\infty}^{\infty} \left| \left[ D_j\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)\right]^{(q)} \right|^2 du$$

where $S = [-8\pi/3, -2\pi/3] \cup [2\pi/3, 8\pi/3]$ is the support of $\psi^*$. Using again Leibniz rule, observe that

$$\left[ D_j\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)\right]^{(q)} = \sum_{p=0}^{q} \binom{q}{p} (f_\epsilon^*)^{(p)}(u)m_2^{(p-q)/2}D_j^{(q-p)}\left(\frac{u}{\sqrt{m_2}}\right).$$

Then with the variable change $v = u/\sqrt{m_2}$

$$\int_{S} \left| \left[ D_j\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)\right]^{(q)} \right|^2 dv \lesssim 2^q \sum_{p=0}^{q} \binom{q}{p} \int_{S} m_2^{p-q+1/2} \left| (f_\epsilon^*)^{(p)}(v\sqrt{m_2}) \right|^2 \left| D_j^{(q-p)}(v) \right|^2 dv.$$  

Now we use the assumptions $A3, A4$ on $f_\epsilon$, for $|v| \geq 2\pi/3$:

$$|(f_\epsilon^*)^{(p)}(v\sqrt{m_2})| \leq k_2 (v^2 m_2)^{-\alpha_p/2} \exp(-\beta|v\sqrt{m_2}|^\gamma) \lesssim m_2^{-\alpha_p/2} \exp(-\beta' m_2^{\gamma/2})$$

denoting $\alpha_0 = \alpha$ and $\beta' = \beta(2\pi/3)^\gamma$. Thus

$$\int_{S} \left| \left[ D_j\left(\frac{u}{\sqrt{m_2}}\right)f_\epsilon^*(u)\right]^{(q)} \right|^2 dv \lesssim \sum_{p=0}^{q} m_2^{p-q+1/2-\alpha_p} \exp(-2\beta' m_2^{\gamma/2}) \int_{S} \left| D_j^{(q-p)}(v) \right|^2 dv$$

and, coming back to (46).

$$I_{2j} \lesssim \sum_{q=0}^{r} \sum_{p=0}^{q} m_2^{-r+p-\alpha_p} \exp(-2\beta' m_2^{\gamma/2}) \int_{S} \left| D_j^{(q-p)}(v) \right|^2 dv.$$  

Since $D_j(v) = \sum_k \theta_{j,k}e^{ivk}$, we compute

$$D_j^{(q-p)}(v) = \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k}(ik)^{q-p}e^{ivk}.$$
which is a trigonometric polynomial with Fourier coefficients \( \theta_{j,k}(ik)q^{-p}1_{0 \leq k \leq \sqrt{m_2}-1} \). Parseval’s theorem gives

\[
\int_{-3\pi}^{3\pi} |D_j^{(q-p)}(v)|^2 dv = 6\pi \sum_{k=0}^{\sqrt{m_2}-1} k^{2q-2p} \theta_{j,k}^2 \lesssim (\sqrt{m_2})^{2q-2p+1}.
\]

Finally, denoting \( \alpha' = \min(\alpha_p) \)

\[
I_{2j} \lesssim \sum_{q=0}^{r} \sum_{p=0}^{q} m_2^{-r+p-\alpha'} \exp(-2\beta' m_2^{\gamma/2}) m_2^{q-p+1/2} \lesssim m_2^{1/2-\alpha'} e^{-2\beta' m_2^{\gamma/2}}.
\]

We bound \( I_{1j} \) in the same way, that provides \( I_{1j} \lesssim C_\psi^2 m_2^{1/2-\alpha} e^{-2\beta' m_2^{\gamma/2}} \). Thus

\[
\chi^2(P_0, P_0) = \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} (I_{1j} + I_{2j}) \lesssim C_\psi^2 \frac{\delta^2}{n} m_1 m_2^{1/2-\alpha} e^{-2\beta' m_2^{\gamma/2}}.
\]

- Let \( \theta \) and \( \theta' \) in \( \{0,1\}^{m_1 \sqrt{m_2}} \). Denoting \( A' = \Psi_{m_1}^{-1/2} \Theta \),

\[
\|\pi_\theta - \pi_{\theta'}\|_f^2 = \frac{\delta^2}{2 \pi} \sum_{j,k,k'} (\theta_{j,k} - \theta'_{j,k})(\theta'_{j,k'} - \theta_{j,k'}) \int \psi^*_{m_2,k} \psi_{m_2,k'}^* \int f(x) dx dy \\
= \frac{1}{2 \pi} \sum_{j,k,k'} (\theta_{j,k} - \theta'_{j,k})(\theta_{j,k'} - \theta'_{j,k'}) \int m_2^{-1/2} e^{iu(k-k')/\sqrt{m_2}} |\psi^*(u/\sqrt{m_2})|^2 du \\
= \frac{1}{2 \pi} \sum_{j,k,k'} (\theta_{j,k} - \theta'_{j,k})(\theta_{j,k'} - \theta'_{j,k'}) \int e^{iu(k-k')} |\psi^*(v)|^2 dv \\
= \frac{1}{2 \pi} \sum_{j} \left| \sum_k (\theta_{j,k} - \theta'_{j,k}) e^{ik} \right|^2 dv
\]

(47) where we have exploited the formula of \( \psi^*_{m_2,k} \) and the variable change \( v = u/\sqrt{m_2} \). Recall that \( |\psi^*| \) is lowerbounded on \([-2\pi, -\pi] \cup [\pi, 2\pi] \), thus

\[
\|\pi_\theta - \pi_{\theta'}\|_f^2 \geq \frac{c_\psi^2 n}{2 \pi} \sum_{j} \left| \sum_k (\theta_{j,k} - \theta'_{j,k}) e^{ik} \right|^2 dv \\
\geq \frac{c_\psi^2 n}{2 \pi} \sum_{j} \sum_k (\theta_{j,k} - \theta'_{j,k})^2 = \frac{c_\psi^2 n}{2 \pi} \rho(\theta, \theta').
\]
8.7.3. Proof of Lemma

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, \( \pi_0 \) is a conditional density. Now we have to prove that the function \( \pi_0 \) is in \( W_0^s(\mathbb{R}^2, L) \) for some \( L > 0 \). We first explain why the univariate function \( y \mapsto \pi_0(y) \) is in \( W_{s_2}(\mathbb{R}, L_2) \) for some \( L_2 > 0 \). According to \cite{Belomestny2019} it is sufficient to see that \( \pi_0, \pi_0', \ldots, \pi_0^{(s_2)} \) and \( y^{s_2-\ell} \pi_0^{(\ell)}(y), \ell = 0, \ldots, s_2-1 \), belong to \( L_2^2(\mathbb{R}) \). It is true since \( \pi_0^{(\ell)}(y) \sim cy^{-2r-\ell} \) when \( y \to \infty \) and \( r > s/2 + 1/4 \).

Now we want to prove that \( (x, y) \mapsto \pi_0(x, y) = \pi_0(y) \) belongs to the weighted bivariate space \( W_0^s(\mathbb{R}^2, L) \) for some \( L > 0 \), for \( s = (s_1, s_2) \). We have

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

with

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

Thus, it holds

\[
\|\pi_0 - (\pi_0)_{\ell,f}\|_f^2 = \iint \left( \sum_{j \geq \ell_1 \text{ or } k \geq \ell_2} a^{(f)}_{j,k} \varphi_j(x) \varphi_k(y) \right)^2 f(x)dx dy
\]

\[
\leq 2 \int \left( \sum_{j \geq 0} \langle \varphi_j, 1 \rangle_f \varphi_j(x) \right)^2 f(x)dx \int \left( \sum_{k \geq \ell_2} \langle \pi_0, \varphi_k \rangle \varphi_k(y) \right)^2 dy
\]

\[
+ 2 \int \left( \sum_{j \geq \ell_1} \langle \varphi_j, 1 \rangle_f \varphi_j(x) \right)^2 f(x)dx \int \left( \sum_{k \geq 0} \langle \pi_0, \varphi_k \rangle \varphi_k(y) \right)^2 dy.
\]

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 \( \pi_0 \) 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 \( f \in W_{s_1}(\mathbb{R}, R) \) 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

\[
\|\pi_0 - (\pi_0)_{\ell,f}\|_f^2 \leq L \ell_1^{-s_1} + \ell_2^{-s_2} \] and thus \( \pi_0 \in W_0^s(\mathbb{R}^2, L) \) for some \( L > 0 \) depending on \( \|f\|_\infty, R, s_2 \).

(b) Since \( \int \psi = 0 \), we have \( \int \pi_0(x, y)dy = 1 \) and we prove hereafter that \( \pi_0(x, y) \geq 0 \). It is sufficient to prove that

\[
|\pi_0(x, y) - \pi_0(x, y)| \leq \pi_0(x, y)
\]
i.e. to bound
\[
\frac{\pi_0(x, y) - \pi_0(y)}{\pi_0(y)} = \frac{\delta m_2^{1/4} \sqrt{m_2^{-1}} \sum_{k=0}^{m_1-1} \left( \sum_{j=0}^{m_1-1} A_{j,k} \varphi_j(x) \right) \psi(\sqrt{m_2 y} - k)}{\pi_0(y)}.
\]

Denoting \( \varphi(x) = (\varphi_0(x), \ldots, \varphi_{m_1-1}(x)) \) and \( \|\cdot\| \) the Euclidean norm on \( \mathbb{R}^{m_1} \), we have for any \( k_0 \)
\[
\left| \sum_{j=0}^{m_1-1} A_{j,k_0} \varphi_j(x) \right| \leq \| [\mathbf{t} \varphi(x)]_{k_0} \| = \| [\mathbf{t} \mathbf{\Psi}^{-1/2}_m \varphi(x)]_{k_0} \|
\leq \| \mathbf{t} e_{k_0} \| \mathbf{\Psi}^{-1/2}_m \varphi(x) \| \leq \| [\mathbf{\Theta e}_{k_0}] \| \mathbf{\Psi}^{-1/2}_m \varphi(x) \|
\leq \sqrt{L(m_1)} \mathbf{\Psi}^{-1}_m \|_{\text{op}} \sum_{j=1}^{m_1} \theta^2_{j,k_0} \leq \sqrt{m_1} L(m_1) \mathbf{\Psi}^{-1}_m \|_{\text{op}}^{1/2}
\]
(48)

Using the definition of \( \pi_0 \) and the property \( |\psi(y)| \leq C_\psi (1 + |y|)^{-2} \), we see that it is sufficient to bound
\[
S_0 := (1 + y^2) \sum_{k=0}^{\sqrt{m_2}-1} (1 + |\sqrt{m_2} y - k|)^{-2}.
\]
To do this we consider separately the cases \(|y| \) small and \(|y| \) large:

- If \(|y| \geq 2, 1 + y^2 \leq 5y^2/4 \) and for any \( 0 \leq k \leq \sqrt{m_2} \),
\[
|\sqrt{m_2} y - k| = \sqrt{m_2} |y| - k \geq \sqrt{m_2} (|y| - 1) > 0.
\]

Then \( |\sqrt{m_2} y - k| \geq \sqrt{m_2} (|y| - 1) \) and
\[
S_0 \leq 5 \sum_{k=0}^{\sqrt{m_2}-1} \frac{y^2}{m_2 (|y| - 1)^2} \leq \frac{5}{4} \frac{1}{\sqrt{m_2}} \sup_{|y| \geq 2} \left( \frac{|y|}{|y| - 1} \right)^2 \leq \frac{5}{\sqrt{m_2}}.
\]

- If \(|y| < 2, 1 + y^2 \leq 5 \) and
\[
S_0 \leq 5 \sum_{k=0}^{\sqrt{m_2}-1} \frac{1}{(|\sqrt{m_2} y - k| + 1)^2} \leq 10 \sum_{\ell=1}^{\infty} \frac{1}{\ell^2}.
\]

To prove the last inequality we can use that:
- if \( y \leq 0 \) then \( |\sqrt{m_2} y - k| + 1 \geq k + 1 \),
- if \( y > 0 \) and \( k \leq \lfloor \sqrt{m_2} y \rfloor \) then \( |\sqrt{m_2} y - k| + 1 \geq \lfloor \sqrt{m_2} y \rfloor - k + 1 \),
- if \( y > 0 \) and \( k > \lceil \sqrt{m_2} y \rceil \) then \( |\sqrt{m_2} y - k| + 1 \geq k - \lceil \sqrt{m_2} y \rceil \).

Gathering these bounds on \( S_0 \) with (48) entails
\[
\left| \frac{\pi_0(x, y) - \pi_0(y)}{\pi_0(y)} \right| \leq C_1 \sqrt{\frac{\delta^2 m_1 \sqrt{m_2} L(m_1) \mathbf{\Psi}^{-1}_m \|_{\text{op}}}{n}}
\]
where \( C_1 \) is a constant only depending on \( C_\psi \). This quantity is smaller than 1 thanks to our assumption.

(c) Next, it remains to prove that \( h := \pi_0 - \pi_0 \) belongs to \( W^r_{8,1}(\mathbb{R}^2, L) \); this will give \( \pi_0 \in W^r_{8,2}(\mathbb{R}^2, 4L) \). In the following, in order to see the bidimensional subscripts, we denote by \( h_{(k_1,k_2)} = \cdots = h_{(k_1,k_2)} = \cdots \).
where $\xi_j$ the orthogonal projection of $h$ on $S_\ell = S_{\ell_1} \otimes S_{\ell_2}$ in $L^2(\mathbb{R}^2, f(x)dx)$. We note that for any function $h$,

$$\|h - h_{\ell,\xi}^f\|_f = \|h - h_{\ell,\xi}^f + h_{\ell,\xi}^f - h_{\ell,\xi}^f\|_f = \|h - h_{\ell,\xi}^f + \left(h - h_{\ell,\xi}^f\right)_{\ell,\xi,\xi}\|_f \leq \|h - h_{\ell,\xi}^f\|_f + \|h - h_{\ell,\xi}^f\|_f.$$ 

So, to check that $\pi_{\ell} - \pi_0$ belongs to $W_\ell^f (\mathbb{R}^2, L)$, we prove

(i) $\ell_1^s (\pi_{\ell} - \pi_0) - (\pi_{\ell} - \pi_0)_{\ell,\xi,\xi}\|_f \leq L/2$ and (ii) $\ell_2^s (\pi_{\ell} - \pi_0) - (\pi_{\ell} - \pi_0)_{\ell,\xi,\xi}\|_f \leq L/2$.

Let us first check condition (i). For the case $\ell_1 \leq m_1$, we write, using the same computation as \[\|\|, \]

$$\|\pi_{\ell} - \pi_0\|_f \leq \|\pi_{\ell} - \pi_0\|_f \leq 3C_\ell^2 \frac{\delta^2}{n} \sum_{j,k} \theta_{j,k}^2 \leq 3C_\ell^2 \frac{\delta^2 m_1^2 \sqrt{m_2}}{n}$$

so that

$$\ell_1^s (\pi_{\ell} - \pi_0) - (\pi_{\ell} - \pi_0)_{\ell,\xi,\xi}\|_f \leq 3C_\ell^2 \frac{\delta^2 m_1^2 \sqrt{m_2}}{n} \leq L/2$$

by assumption. On the other hand, for $m_1 < \ell_1$, then $(\pi_{\ell} - \pi_0)_{\ell,\xi,\xi,\xi} = \pi_{\ell} - \pi_0$ and

$$\|\pi_{\ell} - \pi_0\|_f = \|\pi_{\ell} - \pi_0\|_f = 0.$$ 

Therefore, (i) is proved.

Now, we turn to condition (ii). Denoting $\psi_{m_2,k}(S_{\ell_2})$ the $L^2(dy)$-orthogonal projection on $S_{\ell_2}$ of $\psi_{m_2,k}$, we can prove that

$$(\pi_{\ell} - \pi_0)_{\ell,\xi,\xi}\|_f = \frac{\delta}{\sqrt{n}} \sum_{j,k} \sum_{k,j \neq k} A_{j,k} \varphi_j(x) \psi_{m_2,k}(y).$$

Thus, we obtain

$$\|\pi_{\ell} - \pi_0\|_f = \frac{\delta^2}{n} \sum_{j,k} \sum_{k,j \neq k} A_{j,k} \varphi_j(x) \psi_{m_2,k} - \psi_{m_2,k}(y)^2$$

$$= \frac{\delta^2}{n} \sum_{j,k} \sum_{k,j \neq k} \left( \Theta \Theta \right)_{j,k,k'} \psi_{m_2,k} - \psi_{m_2,k'}$$

$$= \frac{\delta^2}{n} \sum_{j,k} \sum_{k,j \neq k} \left( \sum_{k,j \neq k} \theta_{j,k} \psi_{m_2,k} - \left( \sum_{k,j \neq k} \theta_{j,k} \psi_{m_2,k} \right) \right)^2$$

$$= \frac{\delta^2}{n} \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \left( \sum_{k,j \neq k} \theta_{j,k} \psi_{m_2,k} - \left( \sum_{k,j \neq k} \theta_{j,k} \psi_{m_2,k} \right) \right)^2$$

where $\xi = \sum_{k=0}^{m_2-1} \theta_{j,k} \psi_{m_2,k}$ and $\xi_{\ell_2}$ is the $L^2$-orthogonal projection of $\xi_j$ on $S_{\ell_2}$. For a function $h \in L^2(\mathbb{R})$ we denote by

$$|h|_x^2 := \sum_{k=0}^{m_1-1} a_k \delta^2 \left< \sum_{k=0}^{m_2-1} \theta_{j,k} \psi_{m_2,k} - \left( \sum_{k,j \neq k} \theta_{j,k} \psi_{m_2,k} \right) \right>^2$$
its Sobolev-Hermite squared norm. Then
\[ \|\xi_j - \xi_j^{(S)}\|_2^2 = \sum_{p \geq \ell_2} a_p^2(\xi_j) \leq \xi_2^{-\ell_2} \sum_{p \geq \ell_2} a_p^2(\xi_j) p^{\ell_2} \leq \xi_2^{-\ell_2} |\xi_j|_{S_2}. \]

Now we use the result in [Belomestny et al., 2019, Sec. 4.1] (see Proposition 4 and its proof, Sec. 7.4), which states that the squared norm \( |h|_{S}^2 \) is equivalent to the squared norm
\[ N_s^2(h) := \sum_{q=0}^{s} \|h(q)\|^2 + \sum_{\ell=0}^{s-\ell} \|x^{\ell}h(\ell)\|^2. \]

Lemma 10 below gives the bound \( N_{s}^2(\xi_j) \leq C(\psi, s_2) m_2^{s_2+1/2} \). Consequently, we obtain
\[ \ell_2^s \|\pi_\theta - \pi_0\|_{\infty, \ell_2}^2 \leq \frac{\delta^2 n}{s} \sum_{j=0}^{m_1-1} |\xi_j|_{S_2}^2 \leq C_2(\psi, s_2) \frac{\delta^2 n}{s} m_2 m_2^{s_2+1/2}, \]

and this quantity is bounded by \( L/2 \) using our assumption. \( \square \)

**Lemma 10.** Let \( q \) and \( a \) be some integers. Let \( \xi_j = \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} \psi_{m_2,k} \) with \( \theta_{j,k} \in \{0, 1\} \). Then there exists a positive constant \( C \) only depending on \( C_\psi, q \) and \( a \) such that
\[ \|\xi_j^{(q)}(y) y^a\|^2 \leq C m_2^{q+1/2}. \]

Consequently (taking \( a = 0, q = 0, \ldots, s \) and \( a = s - q, q = 0, \ldots, s \))
\[ N_s^2(\xi_j) \leq C(\psi, s_2) m_2^{s+1/2}. \]

**Proof.** Classical Fourier results give
\[ \|\xi_j^{(q)}(y) y^a\|^2 = \frac{1}{2\pi} \int \left| \left( \xi_j^{(q)} y^a \right)^* \right|^2 = \frac{1}{2\pi} \int \left| \left( \xi_j^{(q)}(y) y^a \right)^* \right|^2. \]

Since, \( \xi_j = \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} \psi_{m_2,k} \), we have \( \xi_j^{(q)}(y) = \sum_{k=0}^{\sqrt{m_2}-1} \theta_{j,k} m_2^{1/4+q/2} \psi_0(y-k) \) and
\[ \left( \xi_j^{(q)} \right)^*(u) = m_2^{-1/4+q/2} D_j \left( \frac{u}{\sqrt{m_2}} \right) \psi_0^*(\frac{u}{\sqrt{m_2}}) \]

where \( D_j(v) = \sum_k \theta_{j,k} e^{ivk} \). Then, the Leibniz rule provides
\[ \left( \xi_j^{(q)} \right)^*(a) (u) = m_2^{-1/4+q/2} \sum_{p=0}^{a} \binom{a}{p} m_2^{-q/p} D_j^{(p)} \left( \frac{u}{\sqrt{m_2}} \right) m_2^{-(a-p)/2} \psi_0^{*(a-p)} \left( \frac{u}{\sqrt{m_2}} \right). \]

Now we use that the function \( \psi^{(q)}(v) = \psi^*(v) (-iv)^q \) has support \( S \) and is bounded, as well as its derivatives. Hence, proceeding as for the study of \( I_{2j} \) previously, we get
\[ \|\xi_j^{(q)}(y) y^a\|^2 \lesssim m_2^{q-a} \sum_{p=0}^{a} \binom{a}{p} (\sqrt{m_2})^{2p+1} \lesssim m_2^{q+1/2} \]

\( \square \)
8.8. **Proof of Lemma 3** We check that the coefficients of the \( n \times m_2 \) matrix, \( \mathbb{E}(\hat{\Theta}_{m_2}(Z)|X) = (\mathbb{E}(\ell_j(Z_i)|X_i))_{1 \leq i \leq n, 0 \leq j \leq m_2-1} \) are the same as those of \( \mathbb{E}(\hat{\Theta}_{m_2}(Y)|X) \ 'G_{m_2} \). On the one hand, by using Formula (31), we have for \( i = 1, \ldots, n \) and \( j = 0, \ldots, m_2 - 1 \),

\[
\mathbb{E}(\ell_j(Z_i)|X_i) = \int \pi_{Z_i|X}(X_i, z) \ell_j(z) dz
\]

\[
= \sum_{j',k \geq 0} \left( \sum_{p=0}^k \langle \pi, \ell_{j'} \otimes \ell_p \rangle g_{k,p} \right) \ell_{j'}(X_i) \int \ell_k(z) \ell_j(z) dz
\]

\[
= \sum_{j' \geq 0} \left( \sum_{p=0}^j \langle \pi, \ell_{j'} \otimes \ell_p \rangle g_{j,p} \right) \ell_{j'}(X_i)
\]

(49)

as \( [(\langle \pi, \ell_{j'} \otimes \ell_p \rangle)_{0 \leq p \leq m_2-1} 'G_{m_2}]_j \ell_{j'}(X_i) \) for \( j = 0, \ldots, m_2 - 1 \). On the other hand,

\[
\mathbb{E}(\hat{\Theta}_{m_2}(Y)|X)_{i,j} = \mathbb{E}(\ell_j(Y_i)|X_i) = \int \pi(X_i, z) \ell_j(z) dz
\]

\[
= \int \sum_{j',k \geq 0} \langle \pi, \ell_{j'} \otimes \ell_k \rangle \ell_{j'}(X_i) \ell_k(z) \ell_j(z) dz = \sum_{j' \geq 0} \langle \pi, \ell_{j'} \otimes \ell_j \rangle \ell_{j'}(X_i).
\]

Therefore

\[
\left[ \mathbb{E}(\hat{\Theta}_{m_2}(Y)|X) 'G_{m_2} \right]_{i,j} = \sum_{p=0}^{m_2-1} \mathbb{E}(\ell_p(Y_i)|X_i) 'G_{m_2} = \sum_{p=0}^{m_2-1} \sum_{j' \geq 0} \langle \pi, \ell_{j'} \otimes \ell_p \rangle \ell_{j'}(X_i) 'G_{m_2}
\]

\[
= \sum_{j' \geq 0} \sum_{p=0}^{m_2-1} \langle \pi, \ell_{j'} \otimes \ell_p \rangle 'G_{m_2} \ell_{j'}(X_i)
\]

(50)

The equality of (49) and (50) gives the result. \( \square \)

8.9. **Proof of Proposition 5** It follows from Lemma 3 that \( \mathbb{E}((\hat{\pi}_m^{(L)}|X) = \pi_{m,n} \) and \( \|\hat{\pi}_m^{(L)} - \pi\|^2_n = \|\hat{\pi}_m^{(L)} - \pi_{m,n}\|^2_n + \|\pi_{m,n} - \pi\|^2_n \).

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

\[
\|\hat{\pi}_m^{(L)} - \pi_{m,n}\|^2_n = \mathbb{E}((\hat{\pi}_m^{(L)}|X)'^2_n
\]

\[
= \frac{1}{n^2} \text{Tr} \left[ G_{m_2}^{-1} t \left( \hat{\Theta}_{m_2}(Z) - \mathbb{E} \left( \hat{\Theta}_{m_2}(Z)|X \right) \right) \Phi_m \tilde{\Phi}_m^{-1} \Phi_m \left( \hat{\Theta}_{m_2}(Z) - \mathbb{E} \left( \hat{\Theta}_{m_2}(Z)|X \right) \right) 'G_{m_2} \right].
\]

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

\[
\text{Tr} \left[ G_{m_2}^{-1} t \left( \hat{\Theta}_{m_2}(Z) - \mathbb{E} \left( \hat{\Theta}_{m_2}(Z)|X \right) \right) \Phi_m \tilde{\Phi}_m^{-1} \Phi_m \left( \hat{\Theta}_{m_2}(Z) - \mathbb{E} \left( \hat{\Theta}_{m_2}(Z)|X \right) \right) 'G_{m_2} \right]
\]

\[
\leq \|G_{m_2}\|^2_{op} \text{Tr} \left[ t \left( \hat{\Theta}_{m_2}(Z) - \mathbb{E} \left( \hat{\Theta}_{m_2}(Z)|X \right) \right) \Phi_m \tilde{\Phi}_m^{-1} \Phi_m \left( \hat{\Theta}_{m_2}(Z) - \mathbb{E} \left( \hat{\Theta}_{m_2}(Z)|X \right) \right) \right].
\]
Therefore, we obtain, analogously to (44),
\[ E[\|\hat{\pi}_m^{(L)} - \pi_{m,n}\|_2^2 | X] \leq \frac{\|G_{m,n}^{-1}\|_{op}^2 L(m_2)m_1}{n}. \]
This gives the first result.

Next it is easy to see that similarly to (44), we have
\[ E \left[ \|\hat{\pi}_m^{(L)} - \pi_{m,n}\|_2^2 | X \right] \leq \frac{\|G_{m,n}^{-1}\|_{op}^2}{n^2} \sum_{i=1}^{n} \sum_{j=0}^{m_2-1} E \left[ (\varphi_j(Z_i) - \mathbb{E}(\varphi_j(Z_i)|X_i))^2 | X_i \right] \|\hat{\Phi}_{m_1} \tilde{\psi}_{m_1}^{-1} \tilde{\Phi}_{m_1}\|_{i,i}. \]

and condition (34) implies that \( \sum_{j=0}^{m_2-1} E \left[ (\varphi_j(Z_i) - \mathbb{E}(\varphi_j(Z_i)|X_i))^2 | X_i \right] \leq C \sqrt{m_2} \), with the same argument as for (7). As \( \|\hat{\phi}_{m_1} \tilde{\psi}_{m_1}^{-1} \tilde{\phi}_{m_1}\|_{i,i} \geq 0 \) a.s., this yields (35). As \( Y \geq 0 \) and \( \varepsilon \geq 0 \),
\[ E \left( \frac{1}{\sqrt{Z}} | X = x \right) \leq \min \left( E \left( \frac{1}{\sqrt{Y}} | X = x \right), E(1/\sqrt{Z}) \right), \]
which explains the comment.

8.10. Proof of Theorem 6. In this proof, we shall denote by \( \mathbb{E}_X[\cdot] = \mathbb{E}[\cdot | X] \).
Let \( m \) be an arbitrary element of \( \hat{M}_n^{(sup)} \). First, we write
\[ \|\hat{\pi}^{(sup)} - \pi\|_n^2 \leq 3 \left( \|\hat{\pi}^{(sup)} - \pi_{m,m}\|_n^2 + \|\hat{\pi}^{(sup)} - \pi_{m,m}\|_n^2 + \|\hat{\pi}^{(sup)} - \pi\|_n^2 \right) \]
\[ \leq 3 \left( \sup_{m' \in \hat{M}_n^{(sup)}} \|\hat{\pi}^{(sup)} - \pi_{m',n}\|_n^2 - V^{(sup)}(m')/6 \right) + 3 \sup_{m' \in \hat{M}_n^{(sup)}} \|\pi_{m',n} - \pi_{(m,m'),n}\|_n^2 \]
\[ + 3 \sup_{m' \in \hat{M}_n^{(sup)}} \|\hat{\pi}^{(sup)} - \pi_{(m,m'),n}\|_n^2 - V^{(sup)}(m')/6 \right) + \]
and Proposition 7 holds if we have
\[ (a) \quad \mathbb{E}_X \left[ \sup_{m' \in \hat{M}_n^{(sup)}} \left( \|\hat{\pi}^{(sup)}_{m',n} - \pi_{m',n}\|_n^2 - V^{(sup)}(m')/6 \right) \right] \leq \frac{C}{n} \]
\[ (b) \quad \mathbb{E}_X \left[ \sup_{m' \in \hat{M}_n^{(sup)}} \left( \|\pi_{m,m'} - \pi_{(m,m'),n}\|_n^2 - V^{(sup)}(m')/6 \right) \right] \leq \frac{C}{n} \]
\[ (c) \quad \sup_{m' \in \hat{M}_n^{(sup)}} \|\pi_{m',n} - \pi_{(m,m'),n}\|_n^2 \leq \|\pi - \pi_{m,n}\|_n^2. \]
We state here a Lemma proved in Section 8.11.
Lemma 11. \[ \|\hat{\pi}_m^{(\text{Sup})} - \pi_{m,n}\|_n = \sup_{T \in B_m} \|\hat{\pi}_m^{(\text{Sup})} - \pi_{m,n}\|_n = \sup_{T \in B_m} \nu^{(\text{Sup})}(T), \]

where \( B_m = \{T \in S_m, \|T\|_n = 1\} \) and

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

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

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

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

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

\[ \Psi_T(x, z) = \sum_{j=0}^{m_1 - 1} \sum_{k=0}^{m_2 - 1} [BG^{-1}_{m_2}]_{j,k} \varphi_j(x) \varphi_k(z). \]

Moreover, note that the following result holds.

Lemma 12. If \( T \in S_m \) then \( \|T\|_n^2 \leq L(m_1)L(m_2)\|T\|_2^2 \), and \( \|T\|_2^2 \leq \|\hat{\Psi}_m^{-1}\|_{\text{op}}\|T\|_n^2 \).

Proof of Lemma 12. If \( T(x, y) = \sum_{j,k} b_{j,k} \varphi_j(x) \varphi_k(y) \), then

\[ |T(x, y)|^2 \leq \sum_{j,k} b_{j,k}^2 \varphi_j^2(x) \varphi_k^2(y) \leq 2L(m_1)L(m_2)\|T\|_2^2, \]

which gives the first inequality. Moreover, we have \( \|T\|_2^2 = \text{Tr}(T^T B T) \), where \( B \) is the matrix \( (b_{j,k}) \), and \( \|T\|_n^2 = \text{Tr}(B \hat{\Psi}_m^{-1} B) \). Then

\[ \|T\|_2^2 \leq \|\hat{\Psi}_m^{-1}\|_{\text{op}}\|T\|_n^2 \leq \|\hat{\Psi}_m^{-1}\|_{\text{op}}\|T\|_n^2, \]

which is the second inequality.

Proof of (a).

First, using Lemma 11,

\[ T_n^{(\text{Sup})} := \mathbb{E}_X \left[ \sup_{m' \in \mathcal{M}_m^{(\text{Sup})}} \left( \|\hat{\pi}_m^{(\text{Sup})} - \pi_{m',n}\|_n^2 - \frac{V^{(\text{Sup})}(m')}{6} \right) \right] \]

\[ \leq \sum_{m' \in \mathcal{M}_m^{(\text{Sup})}} \mathbb{E}_X \left[ \left( \|\hat{\pi}_m^{(\text{Sup})} - \pi_{m',n}\|_n^2 - \frac{V^{(\text{Sup})}(m')}{6} \right) \right] \]

\[ \leq \sum_{m' \in \mathcal{M}_m^{(\text{Sup})}} \left( \sup_{T \in S_m, \|T\|_n = 1} \left( \nu_n^{(\text{Sup})} \right)^2 \right) \frac{V^{(\text{Sup})}(m')}{6} \]

Now we consider separately the three different cases.

Direct case (D). We use Talagrand inequality recalled in Lemma 13, conditionally to \( X \). Remember that we have already proved (see the proof of Proposition 1), that

\[ \mathbb{E} \left[ \left( \sup_{T \in B_m} \frac{1}{n} \sum_{i=1}^{n} [T(X_i, Y_i) - \mathbb{E}_X(T(X_i, Y_i))] \right)^2 | X \right] = \mathbb{E} \left[ \|\hat{\pi}_m - \pi_{m,n}\|_n^2 | X \right] \leq \frac{m_1L(m_2)}{n} := H^2. \]
Moreover
\[
\frac{1}{n} \sum_{i=1}^{n} E(T^2(X_i, Y_i)|X) \leq \|\pi\|_\infty \|T\|_n^2
\]
so that \(v = \|\pi\|_\infty\). To compute \(b\), we use Lemma 12
\[
\|T\|_n^2 \leq L(m_1)L(m_2)\|\hat{\Psi}^{-1}_{m_1}\|_{op} \leq \delta^* n \frac{L(m_2)}{\log^*(n)} =: b^2
\]
Thus we apply Lemma 13 (Talagrand) so that for \(K_0 \geq 12(1 + 2e^2)\) we get
\[
T_1^{(D)} \leq C_0 \left(\frac{\|\pi\|_\infty}{n}\right) \sum_{m \in \{1, \ldots, n\}^2} \left( e^{-c_1 m_1 L(m_2)} + \frac{L(m_2)}{\log^2(n)} e^{-c_2 \log(n) \sqrt{m_1}} \right) \leq \frac{1}{n},
\]
where \(c_2 = 2eC(e^2)K_1/(7\sqrt{\delta^*})\). Thus, use that \(\sum_{m_1 \geq 1} e^{-\kappa \sqrt{m_1}} \leq S e^{-\kappa} \leq \epsilon C(e^2)K_1/7\). So, using that for \(m \in \mathcal{M}_n^{(D)}\), \(L(m_2) \leq n/K_0\), the result holds for a well chosen constant \(\delta^* \) under condition (39).

Case Noisy-Hermite (H).
We now proceed to the application of Talagrand inequality to \(\nu_n^{(H)}(T)\) conditionally to \(X = (X_1, \ldots, X_n)\), where we already saw that
\[
E_X \sup_{T \in \mathcal{S}_m, \|T\|_1 = 1} [\nu_n^{(H)}(T)]^2 = E_X \|\hat{\Psi}_{m_1}^{(n)} - \xi_0\|_n^2 \leq m_1 \Delta(m_2) := H^2.
\]
Next we determine \(v\). Let \(T = \sum_{j,k} b_{j,k} \varphi_j \otimes \varphi_k \in \mathcal{S}_m\), \(B = \langle b_{j,k} \rangle, \) such that \(\|T\|_n = 1\).
\[
\frac{1}{n} \sum_{i=1}^{n} E_X \Phi^2_T(X_i, Z_i) = \frac{1}{n} \sum_{i=1}^{n} \int \Phi^2_T(X_i, z) \pi_{Z|X}(X_i, z) dz \leq \|\pi_{Z|X}\|_\infty \frac{1}{n} \sum_{i=1}^{n} \int \Phi^2_T(X_i, z) dz
\]
\[
\leq \|\pi\|_\infty \frac{1}{n} \sum_{j,k,j',k'} b_{j,k} b_{j',k'}(T) \varphi_j(X_i) \varphi_{j'}(X_i) \int v_{\varphi_k}(z) v_{\varphi_{k'}}(z) dz
\]
\[
= \|\pi\|_\infty \text{Tr} \left[ B \hat{\Psi}^{-1}_{m_1} B (\langle v_{\varphi_k}, v_{\varphi_{k'}} \rangle)_{0 \leq k,k' \leq m_2-1} \right].
\]
As \(\Sigma_0 := t^* B \hat{\Psi}^{-1}_{m_1} B\) is square symmetric positive definite and \(S_0 := \langle v_{\varphi_k}, v_{\varphi_{k'}} \rangle_{0 \leq k,k' \leq m_2-1}\) is symmetric, we can prove that \(\text{Tr}(\Sigma_0 S_0) \leq \|S_0\|_{op} \text{Tr}(\Sigma_0)\). Indeed, \(S_0 = t^* D S_0 P\) with \(D S_0 = \text{diag}(d_i(S_0))\), \(t\) diagonal and \(P\) orthogonal, and
\[
\text{Tr}[\Sigma_0 S_0] = \text{Tr}[\Sigma_0 t^* P D S_0 P] = \text{Tr}[P \Sigma_0 t^* P D S_0]
\]
\[
= \sum_{i=1}^{m_2} d_i(S_0) [P \Sigma_0 t^* P]_{i,i} \quad \text{with} \quad [P \Sigma_0 t^* P]_{i,i} = \|\Sigma_0^{1/2} P e_i\|^2 \geq 0
\]
\[
\leq \max_i (|d_i(S_0)|) \sum_{i=1}^{m_2} [P \Sigma_0 t^* P]_{i,i} = \max_i (|d_i(S_0)|) \text{Tr}(P \Sigma_0 t^* P) \leq \max_i (|d_i(S_0)|) \text{Tr}(\Sigma_0).
\]
Therefore
\[
\text{Tr} \left[ B \hat{\Psi}^{-1}_{m_1} B (\langle v_{\varphi_k}, v_{\varphi_{k'}} \rangle)_{0 \leq k,k' \leq m_2-1} \right] \leq \| (\langle v_{\varphi_k}, v_{\varphi_{k'}} \rangle)_{0 \leq k,k' \leq m_2-1} \|_{op} \text{Tr} \left[ B \hat{\Psi}^{-1}_{m_1} B \right].
\]
Then \(\text{Tr} \left[ B \hat{\Psi}^{-1}_{m_1} B \right] = \|T\|^2 = 1\) and we have to bound the operator norm. First
\[
\| (\langle v_{\varphi_k}, v_{\varphi_{k'}} \rangle)_{0 \leq k,k' \leq m_2-1} \|_{op} = \sup_{x \in \mathbb{R}^{m_2}, \|x\| = 1} t^* (\langle v_{\varphi_k}, v_{\varphi_{k'}} \rangle)_{0 \leq k,k' \leq m_2-1} x = \sup_{t \in \mathbb{R}^{m_2}, \|t\| = 1} ||t||^2.
\]
Next, as \( v_t = (1/2\pi)(t^*/f^*_x)^*(-) \), we have

\[
\|v_t\|^2 = \frac{1}{2\pi} \int \left| \frac{t^*(z)}{f^*_x(z)} \right|^2 dz \\
\leq \frac{1}{2\pi} \sup_{|z| \leq \sqrt{\frac{m_2}{2}}} \left| \frac{f^*_x(z)}{f^*_x(z)} \right|^2 \int \left| t^*(z) \right|^2 dz + \frac{1}{2\pi} \int_{|z| > \sqrt{\frac{m_2}{2}}} \left| t^*(z) \right|^2 dz \\
\leq \sup_{|z| \leq \sqrt{\frac{m_2}{2}}} \left| \frac{1}{f^*_x(z)} \right|^2 + \frac{1}{2\pi} \sum_{j=0}^{m_2-1} \frac{1}{2\pi} \int_{|z| > \sqrt{\frac{m_2}{2}}} \left| \varphi^*_j(z) \right|^2 dz \\
\leq \sup_{|z| \leq \sqrt{\frac{m_2}{2}}} \left| \frac{1}{f^*_x(z)} \right|^2 + \epsilon = \delta(m_2),
\]

by proceeding as in the proof of Proposition 3. As a consequence,

\[
\sup_{T \in S_m, \|T\|_n = 1} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_X \left[ \Phi_T^2(X_i, Z_i) \right] \right) \leq \sup_{|z| \leq \sqrt{\frac{m_2}{2}}} \frac{1}{|f^*_x(z)|^2} + \epsilon := \nu.
\]

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

\[
\sup_{T \in S_m, \|T\|_n = 1} \sup_{x,z} \left| \Phi_T(x, z) \right|^2 \leq \sup_{T \in S_m, \|T\|_n = 1} \sup_{x,z} \left| \sum_{j,k} b_{j,k} \varphi_j(x) \varphi_k(z) \right|^2 \\
\leq \sup_{T \in S_m, \|T\|_n = 1} \sup_{x,z} \text{Tr}[TB] \sum_j \varphi_j^2(x) \sum_k v_{\varphi_k}(z) \\
\leq \sup_{T \in S_m, \|T\|_n = 1} \sup_{x,z} \|ar{\Psi}_{m_1}^{-1}\|_{op} \text{Tr}[\bar{B} \Psi_{m_1}^{-1} B] L(m_1) \Delta(m_2) \\
= \|ar{\Psi}_{m_1}^{-1}\|_{op} L(m_1) \Delta(m_2) \leq \frac{\delta^* n}{\log(n)} \Delta(m_2) := b^2
\]
as \( \text{Tr}[B \Psi_{m_1}^{-1} B] = \|T\|_n^2 = 1 \) and using that \( \mathbf{m} \in \tilde{\mathcal{M}}_n^{(H)} \).

As a consequence, by Talagrand inequality,

\[
\sum_{\mathbf{m}} \mathbb{E} \left[ \left( \sup_{T \in S_m, \|T\|_n = 1} [\nu(H)(T)]^2 - V(H)(\mathbf{m}) \right) \right] \\
\leq \frac{c_2 n}{n} \sum_{\mathbf{m}} \left\{ \delta(m_2) \exp(-c_2^* m_1 \Delta(m_2)/\delta(m_2)) + \frac{\Delta(m_2)}{\log^2(n)} \exp(-c_2 \log(n) \sqrt{m_1}) \right\},
\]

where \( c_2 \) is the same as previously. As for \( \mathbf{m} \in \tilde{\mathcal{M}}_n^{(H)} \), we have \( \Delta(m_2) \leq n \), and the choice of \( \delta \) manage with the second sum. The first one is handled with condition (40).

Consequently

\[
\mathbb{E}_X \left[ \sup_{\mathbf{m} \in \bar{\mathcal{M}}_n^{(H)}} (\|\bar{\pi}_m - \pi_{m,n}\|_n^2 - V(H)(\mathbf{m})/6)_+ \right] \leq C/n.
\]

This ends the proof of case (H). \( \square \)
Noisy-Laguerre case (L).

Now we can apply Talagrand Inequality (Theorem 13) to

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

First, we get from the proof of Proposition 5

\[ \mathbb{E} \left( \sup_{T \in S_m, \|T\|_{\infty} = 1} \nu_n^{(L)}(T) \right) = \mathbb{E}(\|T\|_{\infty} - \pi_{m,n} X \|_{n}^{2}) \leq \frac{m_1 L(m_2) \|G_{m_2}^{-1}\|_{op}^2}{n} := H^2. \]

Next we have

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

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

\[ \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\Psi_T^2(X_i, Z_i)|X] \leq \frac{\|\pi\|_{\infty}}{n} \sum_{i=1}^{n} \int \left( \sum_{j,k} [BG_{m_2}^{-1}]_{j,k} (j \varphi_j(x_i) \varphi_k(z)) \right)^2 \]

\[ = \frac{\|\pi\|_{\infty}}{n} \sum_{j,k} [BG_{m_2}^{-1}]_{j,k} [BG_{m_2}^{-1}]_{j',k} \pi_{m_1} \varphi_j \varphi_k \leq \frac{\|\pi\|_{\infty}}{n} \text{Tr}[B \hat{\Psi}_m B] = \frac{\|\pi\|_{\infty}}{n} \|G_{m_2}^{-1}\|_{op}^2 \|T\|_n^2 \]

which implies that \( v = \|\pi\|_{\infty} \|G_{m_2}^{-1}\|_{op}^2 \). Lastly we write

\[ \|\Psi_T\|_{\infty} = \sup_{x,z} \sum_{j,k} [BG_{m_2}^{-1}]_{j,k} \varphi_j(x) \varphi_k(z) \leq \sqrt{L(m_1) L(m_2) \|BG_{m_2}^{-1}\|_F^2} \]

\[ \leq \sqrt{L(m_1) L(m_2) \|G_{m_2}^{-1}\|_{op}^2 \|B\|_F^2}, \]

and by Lemma 12

\[ \|B\|_F^2 = \text{Tr}[B^* B] = \|T\|_2^2 \leq \|\hat{\Psi}_{m_1}^{-1}\|_{op}^2 \|T\|_n^2. \]

Therefore, we get

\[ \|\Psi_T\|_{\infty}^2 \leq L(m_1) L(m_2) \|G_{m_2}^{-1}\|_{op}^2 \|\hat{\Psi}_{m_1}^{-1}\|_{op}^2 \|T\|_n^2 \leq \frac{\rho^* n}{\log^2(n)} L(m_2) \|G_{m_2}^{-1}\|_{op}^2 := b^2 \]

by using that \( m \in \hat{M}_n^{(L)} \) (second constraint).

Therefore, by applying Talagrand Inequality (Theorem 13) that for \( K_0 \geq 12(1 + 2e^2) \), we get

\[ T^{(L)}_1 \leq C_0 \|\pi\|_{\infty} \sum_{m \in \hat{M}_n} \left( \|G_{m_2}^{-1}\|_{op}^2 e^{-c_1 m_1 L(m_2)} + \frac{L(m_2) \|G_{m_2}^{-1}\|_{op}^2}{\log^2(n)} e^{-c_2 \log(n) \sqrt{m_1}} \right), \]

where \( c_2 \) is the same as in case (D). So, using that \( m \in \hat{M}_n^{(m)} \), \( L(m_2) \|G_{m_2}^{-1}\|_{op} \leq n/K_0 \) (first constraint), the result holds if (H1) holds.

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 m \land m' was still in the collection,

\[
\begin{align*}
&\sup_{m' \in \mathcal{M}_n^{(\text{up})}} \left( \|\hat{\pi}(\text{up})_{m',m} - \pi_{m,m'},n\|^2 - V(\text{up})(m')/6 \right) \\
&\leq \sup_{m' \in \mathcal{M}_n^{(\text{up})}} \left( \|\hat{\pi}(\text{up})_{m,m'} - \pi_{m,m'},n\|^2 - V(\text{up})(m\land m')/6 \right) \\
&\leq \sup_{m' \in \mathcal{M}_n^{(\text{up})}} \left( \|\hat{\pi}(\text{up})_{m'} - \pi_{m',n}\|^2 - V(\text{up})(m'')/6 \right).
\end{align*}
\]

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

Proof of (c)

We already noticed that \( \mathbb{E}(\hat{\pi}_m^{(D)}|X) = \mathbb{E}(\hat{\pi}_m^{(L)}|X) = \mathbb{E}(\hat{\pi}_m^{(H)}|X) = \pi_{m,n} \), so the bias terms are exactly the same in the three cases.

Let us define \( \text{Proj}_{S_m}^{(n)} \) denotes the empirical projection on \( S_m \) which associates to \( (x,y) \mapsto T(x,y) \) the function \( (x,y) \mapsto (\text{Proj}_{S_m}^{(n)}T)(x,y) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \left( \hat{\Phi}_{m_1} \hat{\Phi}_{m_1}^{-1} \left( \int \varphi_k(z)T(X,z)dz \right)_{1 \leq i \leq n, 0 \leq k \leq m_2-1} \right) j,k \varphi_j(x)\varphi_k(y). \)

For any bivariate function \( T \), the following holds:

\[ \text{Proj}_{S_m}^{(n)} T = \text{Proj}_{S_m}^{(n)} \left( \text{Proj}_{S_m}^{(n)} T \right). \]

Then

\[
\|\pi_{m',n} - \pi_{m,m'},n\| = \|\text{Proj}_{S_m}^{(n)} T - \text{Proj}_{S_m}^{(n)} T\| = \|\pi - \text{Proj}_{S_m}^{(n)} T\|.
\]

Thus we obtain (c).


\[
\|\hat{\pi}_m^{(\text{up})} - \mathbb{E}_X \hat{\pi}_m^{(\text{up})}\|_n = \sup_{T \in S_m, \|T\|_n=1} \langle \hat{\pi}_m^{(\text{up})} - \mathbb{E}_X \hat{\pi}_m^{(\text{up})}, T \rangle_n = \sup_{T \in S_m, \|T\|_n=1} \nu_n^{(\text{up})}(T).
\]

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

\[
\langle \hat{\pi}_m^{(\text{up})} - \pi_{m,n}, T \rangle_n = \sum_{j,j',k,k'} (\hat{\pi}_m^{(\text{up})} - \mathbb{E}_X \hat{\pi}_m^{(\text{up})})_{j,k} B_{j,k} \langle \varphi_j \otimes \varphi_k, \varphi_{j'} \otimes \varphi_{k'} \rangle_n > \sum_{j,j',k} (\hat{\pi}_m^{(\text{up})} - \mathbb{E}_X \hat{\pi}_m^{(\text{up})})_{j,k} B_{j,k} (\tilde{\varphi}_m^{(\text{up})})_{j,j'} = \text{Tr} \left[ (\hat{\pi}_m^{(\text{up})} - \mathbb{E}_X \hat{\pi}_m^{(\text{up})}) \tilde{\varphi}_m^{(\text{up})} B \right].
\]
For the rest of the proof, we study separately the three cases.

**Direct case.** Recall that \( \hat{A}_m^{(D)} = \frac{1}{n} \tilde{\Psi}_m^{-1} \hat{\Phi}_m \hat{\Theta}_{m_2}(Y) \). Then
\[
\langle \hat{\pi}_m^{(D)} - \pi_{m,n}, T \rangle_n = \frac{1}{n} \text{Tr} \left[ \langle \hat{\Theta}_{m_2}(Y) - \mathbb{E} \hat{\Theta}_{m_2}(Y) \rangle \hat{\Phi}_m B \right] \\
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j,k} (\varphi_k(Y_i) - \mathbb{E} \varphi_k(Y_i)) \varphi_j(X_i) B_{jk} \\
= \frac{1}{n} \sum_{i=1}^{n} [T(X_i, Y_i) - \mathbb{E} T(X_i, Y_i)].
\]

**Hermite case.** Here we use that \( \hat{A}_m^{(H)} = \frac{1}{n} \tilde{\Psi}_m^{-1} \hat{\Phi}_m \hat{\Phi}_m \hat{\Theta}_{m_2}(Z) \). Thus
\[
\langle \hat{\pi}_m^{(H)} - \pi_{m,n}, T \rangle_n = \frac{1}{n} \text{Tr} \left[ \langle \hat{\Theta}_{m_2}(Z) - \mathbb{E} \hat{\Theta}_{m_2}(Z) \rangle \hat{\Phi}_m B \right] \\
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j,k} (\varphi_k(Z_i) - \mathbb{E} \varphi_k(Z_i)) \varphi_j(X_i) B_{jk} \\
= \frac{1}{n} \sum_{i=1}^{n} [\Phi_T(X_i, Z_i) - \mathbb{E} \Phi_T(X_i, Z_i)].
\]

**Laguerre case.** In this case \( \hat{A}_m^{(L)} = \frac{1}{n} \tilde{\Psi}_m^{-1} \hat{\Phi}_m \hat{\Theta}_{m_2}(Z) \hat{G}_{m_2}^{-1} \), then
\[
\langle \hat{\pi}_m^{(L)} - \pi_{m,n}, T \rangle_n = \frac{1}{n} \text{Tr} \left[ \hat{G}_{m_2}^{-1} \langle \hat{\Theta}_{m_2}(Z) - \mathbb{E} \hat{\Theta}_{m_2}(Z) \rangle \hat{\Phi}_m B \right] \\
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j,k} (\varphi_k(Z_i) - \mathbb{E} \varphi_k(Z_i)) \varphi_j(X_i) (BG_{m_2})_{jk} \\
= \frac{1}{n} \sum_{i=1}^{n} [\Psi_T(X_i, Z_i) - \mathbb{E} \Psi_T(X_i, Z_i)].
\]

8.12. **Proof of Corollary**

De-conditionning is justified by Lemma stated and proved in Appendix.

Let \( \Lambda_n^{(\text{sup})} = \{ \mathcal{M}_n^{(\text{sup})} \subset \hat{M}_n^{(\text{sup})} \} \) and write
\[
\mathbb{E}[\| \hat{\pi}_m^{(\text{sup})} - \pi \|_n^2] = \mathbb{E}[\mathbb{E}[\| \hat{\pi}_m^{(\text{sup})} - \pi \|_n^2 | 1_{\Lambda_n^{(\text{sup})}}]] + \mathbb{E}[\| \hat{\pi}_m^{(\text{sup})} - \pi \|_n^2 1_{(\Lambda_n^{(\text{sup})})^c}] := T_1 + T_2.
\]

We first study \( T_1 \). On \( \Lambda_n^{(\text{sup})} \), we have
\[
\inf_{m \in \mathcal{M}_n^{(\text{sup})}} \left\{ \| \pi_{m,n} - \pi \|_n^2 + V^{(\text{sup})}(m) \right\} \leq \inf_{m \in \mathcal{M}_n^{(\text{sup})}} \left\{ \| \pi_{m,n} - \pi \|_n^2 + V^{(\text{sup})}(m) \right\} \\
\leq \inf_{m \in \mathcal{M}_n^{(\text{sup})}} \left\{ \| \pi - \pi \|_n^2 + V^{(\text{sup})}(m) \right\}.
\]

So, for the first term, we have, using Theorem and the definition of \( \Lambda_n^{(\text{sup})} \),
\[
\mathbb{E}[\| \hat{\pi}_m^{(\text{sup})} - \pi \|_n^2 1_{\Lambda_n^{(\text{sup})}}] \leq C \inf_{m \in \mathcal{M}_n^{(\text{sup})}} \left\{ \| \pi - \pi_{m,n} \|_n^2 + V^{(\text{sup})}(m) \right\} + C',
\]
and taking the expectation yields
\[
\mathbb{E}[\mathbb{E}[\| \hat{\pi}_m^{(\text{sup})} - \pi \|_n^2 | 1_{\Lambda_n^{(\text{sup})}}]] \leq C \inf_{m \in \mathcal{M}_n^{(\text{sup})}} \left\{ \| \pi - \pi_{m} \|_n^2 + V^{(\text{sup})}(m) \right\} + C'.
\]
Now, $T_2$ is bounded thanks to the two facts:

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

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

$$\|\hat{\pi}_m^{(D)}\|_n^2 = \frac{1}{n^2} \text{Tr} \left( \hat{\Phi}_{m_1} \hat{\Psi}^{-1}_{m_1} \hat{\Phi}_{m_1} \hat{\Theta}_{m_2}(\mathbf{Y}) \hat{\Theta}_{m_2}(\mathbf{Y}) \right) \leq \frac{1}{n^2} \|\hat{\Phi}_{m_1} \hat{\Psi}^{-1}_{m_1} \|_{\text{op}} \text{Tr} \left[ \hat{\Theta}_{m_2}(\mathbf{Y}) \hat{\Theta}_{m_2}(\mathbf{Y}) \right] \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{m_2-1} \varphi_k^2(Y_i) \leq L(m_2) \leq K_0 n$$

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

$$\|\hat{\pi}_m^{(L)}\|_n^2 = \frac{1}{n^2} \text{Tr} \left( \hat{\Phi}_{m_1} \hat{\Psi}^{-1}_{m_1} \hat{\Phi}_{m_1} \hat{G}_{m_2}^{-1} \hat{\Theta}_{m_2}(\mathbf{Z}) \hat{\Theta}_{m_2}(\mathbf{Z}) \hat{G}_{m_2}^{-1} \right) \leq L(m_2) \|\hat{G}_{m_2}^{-1}\|_{\text{op}} \leq K_0 n$$

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

$$\|\hat{\pi}_m^{(H)}\|_n^2 = \frac{1}{n^2} \text{Tr} \left( \hat{\Phi}_{m_1} \hat{\Psi}^{-1}_{m_1} \hat{\Phi}_{m_1} \hat{Y}_{m_2}(\mathbf{Z}) \hat{Y}_{m_2}(\mathbf{Z}) \right) \leq \Delta(m_2) \leq K_0 n.$$  

To bound $\|\pi\|_n^2$, as $\pi$ 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

$$\mathbb{P}((\Lambda_n^{(\sup)})^c) = \mathbb{P}\left( \left\{ \mathcal{M}_n^{(\sup)} \subset \hat{\mathcal{M}}_n^{(\sup)} \right\}^c \right) = \mathbb{P}(\exists \mathbf{m} \in \mathcal{M}_n^{(\sup)}, \text{such that } \mathbf{m} \notin \hat{\mathcal{M}}_n^{(\sup)})$$

$$\leq \sum_{\mathbf{m} \in \mathcal{M}_n^{(\sup)}} \mathbb{P}\left( L(m_1)\|\Psi^{-1}_{m_1}\|_{\text{op}} \leq \frac{\hat{\delta}^* \sqrt{n}}{2 \log^2(n)} \text{ and } L(m_1)\|\Psi^{-1}_{m_1}\|_{\text{op}} > \frac{\hat{\delta}^* \sqrt{n}}{2 \log^2(n)} \right)$$

$$\leq \sum_{\mathbf{m} \in \mathcal{M}_n^{(\sup)}} \mathbb{P}\left( L(m_1)(\|\Psi^{-1}_{m_1}\|_{\text{op}} - \|\Psi^{-1}_{m_1}\|_{\text{op}}) \geq \frac{\hat{\delta}^* \sqrt{n}}{2 \log^2(n)} \right)$$

$$\leq \sum_{\mathbf{m} \in \mathcal{M}_n^{(\sup)}} \mathbb{P}\left( (\|\Psi^{-1}_{m_1} - \Psi_{m_1}\|_{\text{op}}) > \|\Psi^{-1}_{m_1}\|_{\text{op}} \right)$$

$$\leq \sum_{\mathbf{m} \in \mathcal{M}_n^{(\sup)}} \mathbb{P}\left( \|\Psi^{-1/2}_{m_1} \Psi_{m_1}^{-1/2} - \text{Id}_{m_1}\|_{\text{op}} > \frac{1}{2} \right),$$
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 $0 \leq \delta \leq 1$,

$$
\mathbb{P}\left(\|\Psi_{m1}^{-1/2}\Psi_{m2}\Psi_{m1}^{-1/2} - \mathbb{I}_{m1}\|_\text{op} > \delta\right) \leq 2m_1 \exp\left(-c(\delta) \frac{n}{L(m_1)(\|\Psi_{m1}^{-1}\|_\text{op} \vee 1)}\right),
$$

where $c(\delta) = (1 + \delta) \log(1 + \delta) - \delta$, which for $\delta = 1/2$ yields $c(1/2) = (3/2) \log(3/2) - 1/2 = 5\delta$, $c(1/2) \sim 0.11$. Thus, under the condition $L(m_1)(\|\Psi_{m1}^{-1}\|_\text{op} \leq \delta^* n/\log^2(n) \leq \delta n/\log(n)$, for $n$ large enough, in the definition of $\mathcal{M}_n^{(\text{sup})}$, we get

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

This ends the proof. □

**APPENDIX**

**Lemma 13** (Talagrand Inequality). Let $Y_1, \ldots, Y_n$ be independent random variables and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\varepsilon^2 > 0$

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

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

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

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 $\mathcal{F} = B_m(0,1)$.

**Lemma 14.** 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 $(X_1, \ldots, X_n)$.

This Lemma legitimates the application of Talagrand inequality conditionally to $(X_1, \ldots, X_n)$.

**Proof of Lemma 14** First $Y_1, \ldots, Y_n$ are independent conditionally to $X_1, \ldots, X_n$ if, for all measurable (bounded or nonnegative) functions $f_i : \mathbb{R} \to \mathbb{R}$,

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

As collection of test functions of $X_1, \ldots, X_n$ for characterization of the conditional expectation, we consider $g(X_1, \ldots, X_n) = \prod_{i=1}^n g_i(X_i)$ 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 $\sigma$-algebra the cartesian product $A = A_1 \times \cdots \times A_n$ which are generators). Therefore (54) holds if

$$
\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i) \prod_{i=1}^n g_i(X_i)\right] = \mathbb{E}\left[\prod_{i=1}^n f_i(Y_i) | X_1, \ldots, X_n\right] \prod_{i=1}^n \mathbb{E}[g_i(X_i)].
$$
To check that this equality holds, let us start from the right-hand-side term.

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

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

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

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

where the last line follows by independence of the \((X_1, Y_1), \ldots, (X_n, Y_n)\). This ends the proof of Lemma 14. □

**References**


