Anisotropic multivariate deconvolution using projection on the Laguerre basis

Florian Dussap

To cite this version:

Florian Dussap. Anisotropic multivariate deconvolution using projection on the Laguerre basis. Journal of Statistical Planning and Inference, Elsevier, In press. hal-02400684v3

HAL Id: hal-02400684
https://hal.archives-ouvertes.fr/hal-02400684v3
Submitted on 10 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Anisotropic multivariate deconvolution using projection on the Laguerre basis

Florian Dussap
Université de Paris, CNRS, MAP5 UMR 8145, F-75006 Paris, France

Abstract
We investigate adaptive density estimation in the additive model $Z = X + Y$, where $X$ and $Y$ are independent $d$-dimensional random vectors with non-negative coordinates. Our goal is to recover the density of $X$ from independent observations of $Z$, assuming the density of $Y$ is known. In the $d = 1$ case, an estimation procedure using projection on the Laguerre basis have already been studied. We generalize this procedure in the multivariate case: we establish non-asymptotic upper bounds on the mean integrated squared error of the estimator and we derive convergence rates on anisotropic functional spaces. Moreover, we provide data-driven strategies for selecting the right projection space (for $d = 1$, we improve the previous projection procedure). We illustrate these procedures on simulated data, and in dimension $d = 1$ we compare our procedure with the previous adaptive projection procedure.

Keywords: anisotropic multivariate projection estimator, laguerre basis, model selection, nonparametric density estimation

2010 MSC: 62G07, 62H12

1. Statistical model and motivations

In this article, we study the problem of recovering the distribution of a random vector $X$ when we only observe its sum with a noise vector $Y$ with known distribution. This is a classical problem in nonparametric statistics (see references below), but we focus on the particular case where both $X$ and $Y$ have non-negative coordinates.

This assumption is quite unusual in deconvolution problems but is relevant for instance in reliability fields: we observe the failure times of several components in a system, each failure time being the sum of the lifetimes of two sub-components. In survival analysis, for $d = 1$, $X$ would be the time of infection of a disease and $Y$ the incubation time. The multivariate case can then be used to study the time of infection for multiple diseases.

More precisely, we consider the following statistical model:

$$Z_i = X_i + Y_i, \quad i = 1, \ldots, n,$$

(1)

where the $X_i$’s, the $Y_i$’s and $Z_i$’s are random vectors in $\mathbb{R}^d$ with non-negative coordinates. We assume that the $X_i$’s are i.i.d. with unknown density $f$ on $\mathbb{R}_+^d$, and that the $Y_i$’s are i.i.d. with known density $g$ on $\mathbb{R}_+^d$. Moreover, we assume that the $X_i$’s and the $Y_i$’s are independent. Our goal is to provide an adaptive procedure to estimate the density $f$ from the observations $Z_1, \ldots, Z_n$. 
In the univariate case, there are a lot of papers about recovering the density of a random variable when it is observed with an additive known noise. Many authors use a kernel estimator introduced by Stefanski and Carroll (1990). Fan (1991) first introduced the notion of ordinary smooth and supersmooth noise (rate of decay of the characteristic function) to study optimal rates of convergence on Hölder spaces. When \( f \) is supersmooth and the noise is ordinary smooth, Butucea (2004) showed that the kernel estimator achieves a good rate of convergence and that this rate is optimal. When both \( f \) and the noise are supersmooth, the problem is more complicated and have been studied by Butucea and Tsybakov (2008a,b) from lower bound point of view.

To provide an adaptive estimator, different procedures were proposed. When \( f \) belongs to a Sobolev space, Pensky and Vidakovic (1999) proposed a wavelet strategy that is adaptive and achieves optimal rates of convergence. For kernel estimators, Delaigle and Gijbels (2004) estimated the optimal bandwidth with a bootstrap procedure and showed its consistency. Hazelon and Turlach (2009, 2010) proposed a weighted kernel estimator with a data-driven way to choose the weights. Moreover, their weighted kernel estimator can be used in a multivariate setting. For projection with penalization strategies, Comte et al. (2006) used a Shannon type basis to construct an adaptive estimator that is minimax in most cases. More recently, non-compact supported bases were used by Mabon (2017) (Laguerre basis), by Comte and Genon-Catalot (2018) (Laguerre basis and Hermite basis), and by Sacko (2019) (Hermite basis) to construct adaptive estimators on suitable functional spaces.

The Laguerre basis was also used in a regression setting to study the problem of Laplace deconvolution, see Comte et al. (2017) and Vareschi (2015), or more recently Benhaddou et al. (2019).

The multivariate deconvolution literature is more sparse. Masry (1991) generalizes the kernel estimator for stationary random processes, with a dependence structure between the variables. The noise is assumed to have i.i.d. coordinates (isotropic noise) and no adaptive strategy is proposed, the author focuses on the problem of dependency between the variables. Youndjé and Wells (2008) propose a cross-validation strategy to estimate the optimal bandwidth of the kernel estimator in the multivariate setting, and show it is asymptotically optimal under the assumption that the noise is isotropic and ordinary smooth. Comte and Lacour (2013) use a bandwidth selection procedure inspired by Goldenshluger and Lepski (2011). Their procedure allows anisotropic noises, with both ordinary smooth and supersmooth components, and they derive rates of convergence for the pointwise risk and the \( L^2 \) risk, when \( f \) belongs to anisotropic Hölder, Nikol’skii or Sobolev classes. For ordinary smooth noise and when \( f \) belongs to anisotropic Nikol’skii classes, Rebelles (2016) provides an adaptive kernel estimator which is minimax for the \( L^p \)-loss. Recently, Lepski and Willer (2019) studied a more general model (with direct and indirect observations of \( X \)) and provided an adaptive kernel estimator on anisotropic Nikol’skii classes, under the \( L^p \)-loss.

Concerning our specific case of deconvolution with non-negative noise, the case \( d = 1 \) has already been studied by Mabon (2017) using a projection strategy on the Laguerre basis. We also use a projection strategy in the multivariate case. The main tool we use to construct our estimator is the theory of hypermatrices. Using the contraction product of hypermatrices, we show that it is possible to recover the coefficients of \( f \) from the observations. We recall the definitions of these objects in Section 2.

We provide rates of convergence for the MISE of our estimator on anisotropic functional spaces: Sobolev–Laguerre spaces and smooth Laguerre spaces. We propose a model selection procedure to produce an adaptive estimator, under mild assumptions on the noise density \( g \). This procedure is inspired by the work of Goldenshuger and Lepski (2011) concerning bandwidth selection. It was introduced for model selection by Chagny (2013b) for estimation of
conditional density, in a two-dimensional setting. We show this procedure can be applied to our deconvolution problem, in a d-dimensional setting, and we establish an asymptotic oracle inequality for this procedure. Moreover, the proof is written to provide general steps that can be applied to other contexts.

Outline of the paper. In Section 2, we provide the notations and the key facts about hypermatrices we will use in the rest of the paper. In Section 3 we construct the estimator. In section 4, we provide non asymptotic MISE bounds and we derive convergence rates on Sobolev–Laguerre balls and smooth Laguerre balls. In Section 5 we give model selection procedures to construct an adaptive estimator and we establish oracle inequalities. In Section 6 we illustrate the procedures on simulated data. All the proofs are gathered in Section 7.

2. Preliminary on hypermatrices

Let \( \alpha, \beta \in \mathbb{N}^d \) be multi-indices and \( x \in \mathbb{C}^d \). We will use the following notations:

\[
\alpha + \beta := (\alpha_1 + \beta_1, \ldots, \alpha_d + \beta_d), \quad \alpha \times \beta := (\alpha_1 \beta_1, \ldots, \alpha_d \beta_d),
\]

\[
\alpha \beta := \left( \frac{\alpha_1}{\beta_1}, \ldots, \frac{\alpha_d}{\beta_d} \right), \quad x^\alpha := x_1^{\alpha_1} \cdots x_d^{\alpha_d},
\]

\[
|\alpha| := \alpha_1 + \cdots + \alpha_d, \quad \beta \leq \alpha \iff \forall q \in \{1, \ldots, d\}, \beta_q \leq \alpha_q.
\]

\( m \land m' := (m_1 \land m'_1, \ldots, m_d \land m'_d) \) where \( m_i \land m'_i := \min(m_i, m'_i) \).

We denote \( 0 := (0, \ldots, 0) \) and \( 1 := (1, \ldots, 1) \). If \( k \in \mathbb{N}^d \) and \( \ell \in \mathbb{N}^p \) are two multi-indices, we denote \( (k, \ell) \in \mathbb{N}^{d+p} \) their concatenation.

**Proposition 2.1** (Multinomial theorem). If \( \alpha \in \mathbb{N}^d \) is a multi-index and if \( x, y \in \mathbb{R}^d \) are vectors, then \( (x + y)^\alpha = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} x^\beta y^{\alpha - \beta} \) where \( \binom{\alpha}{\beta} := \left( \frac{\alpha_1}{\beta_1} \right) \times \cdots \times \left( \frac{\alpha_d}{\beta_d} \right) \).

We recall some facts about hypermatrices; see Hogben (2013) for more details. A hypermatrix is a multi-dimensional array of numbers. The order of a hypermatrix is its number of dimensions: an order-1 hypermatrix is a vector, an order-2 hypermatrix is a matrix, and so on. If \( m = (m_1, \ldots, m_d) \in (\mathbb{N}^+)^d \) is multi-index, we denote by \( \mathbb{R}^m = \mathbb{R}^{(m_1, \ldots, m_d)} \) the set of order-\( d \) hypermatrices with \( m_j \) components according to the \( j \)-th dimension. Moreover, if \( m \in (\mathbb{N}^+)^d \) and \( r \in (\mathbb{N}^+)^p \) are two multi-indices, since \( (m, r) \) denotes the concatenation of \( m \) and \( r \), we denote \( \mathbb{R}^{(m, r)} := \mathbb{R}^{(m_1, \ldots, m_d, r_1, \ldots, r_p)} \). We choose to index the hypermatrix components starting from 0. So if \( T \in \mathbb{R}^m \) is a hypermatrix, its components are \( T_k \) for \( k \in \mathbb{N}^d \) such that \( k \leq m - 1 \).

**Definition 2.2.** Let \( m \in (\mathbb{N}^+)^d \), we make \( \mathbb{R}^m \) a Euclidean vector space, defining the addition and the scalar multiplication by \( [T + U]_k := T_k + U_k \) and \( [\lambda T]_k := \lambda T_k \), and defining the scalar product and the associated norm by \( \langle T, U \rangle_{\mathbb{R}^m} := \sum_{k \leq m-1} T_k U_k \) and \( \|T\|_{\mathbb{R}^m}^2 := \sum_{k \leq m-1} T_k^2 \).

For vectors \( (d = 1) \), we get the usual \( \ell^2 \) norm and for matrices \( (d = 2) \), we get the Frobenius norm.

**Definition 2.3** (Contraction product). Let \( m, p \) and \( r \) be multi-indices (with possibly different lengths) and denote \( d \) the length of \( p \). Let \( T \in \mathbb{R}^{(m, p)} \) and \( U \in \mathbb{R}^{(p, r)} \) be two hypermatrices, their \( d \)-contraction product is the hypermatrix \( T \times_d U \in \mathbb{R}^{(m, r)} \) defined by:

\[
[T \times_d U]_{(k, \ell)} := \sum_{j \leq p-r} T_{kj} U_{j\ell}.
\]

The contraction product generalizes the matrix product: if \( T \), respectively \( U \), are matrices \( m \times p \), respectively \( p \times r \), their 1-contraction product is their matrix product.
Definition 2.4 (Frobenius norm and spectral norm). Let \( m \in (\mathbb{N}^*)^d \) be a multi-index. If \( T \in \mathbb{R}^{(m,m)} \) is a hypermatrix, then \( T \) induces a linear map (still denoted by \( T \)):

\[
T: \mathbb{R}^m \rightarrow \mathbb{R}^m \quad x \mapsto T \times_d x.
\]

We denote \( T^* \in \mathbb{R}^{(m,m)} \) the hypermatrix defined by \( [T^*]_{k\ell} = T_{\ell k} \). The linear map induced by \( T^* \) is the adjoint linear map of the one induced by \( T \):

\[
\forall x, y \in \mathbb{R}^m, \quad \langle y, Tx \rangle_{\mathbb{R}^m} = \langle T^* y, x \rangle_{\mathbb{R}^m}.
\]

We define the Frobenius norm and the spectral norm of \( T \) in the same way we define them for any endomorphism: \( \|T\|_F^2 := \text{Tr}(T^*T) \) and \( \rho^2(T) := \lambda_{\text{max}}(T^*T) \).

It is easy to check that:

\[
\|T\|_F^2 = \sum_{k,\ell \leq m-1} T_{k\ell}^2, \quad \rho(T) = \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|Tx\|_{\mathbb{R}^m}}{\|x\|_{\mathbb{R}^m}}.
\]

Lastly, we recall the classical inequalities between the Frobenius and the spectral norm:

\[
\frac{1}{m_1 \cdots m_d} \|T\|_F^2 \leq \rho^2(T) \leq \|T\|_F^2.
\] (3)

Finally, we introduce the following definition that generalizes the concept of triangular matrices.

Definition 2.5. A hypermatrix \( T \in \mathbb{R}^{(m,m)} \) is said to be lower triangular if apart from multi-indices \( \ell, k \leq m - 1 \) such that \( k \leq \ell \), we have \( T_{\ell k} = 0 \).

An infinite hypermatrix \( T \in \mathbb{R}^{\mathbb{N}^d \times \mathbb{N}^d} \) is said to be lower triangular if apart from multi-indices \( \ell, k \in \mathbb{N}^d \) such that \( k \leq \ell \), we have \( T_{\ell k} = 0 \).

3. The estimation procedure

In the model (1), the \( Z_i \)'s are i.i.d. random vectors on \( \mathbb{R}^d_+ \), and they admit a density we denote by \( h \). This density function is given by the convolution product of \( f \) and \( g \):

\[
(f * g)(x) := \int_{\mathbb{R}^d} f(u) g(x - u) \, du = \int_{[0,x_1]} \cdots \int_{[0,x_d]} f(u) g(x - u) \, du.
\]

We assume that \( f \), \( g \) and \( h \) belong to \( L^2(\mathbb{R}^d_+) \) and we expand them in the multivariate Laguerre basis. We recall that the one-dimensional Laguerre functions \( (\varphi_k)_{k \in \mathbb{N}} \) are defined by:

\[
\forall x \in \mathbb{R}_+, \quad \varphi_k(x) := \sqrt{2} L_k(2x) e^{-x}, \text{ where } L_k(x) := \sum_{j=0}^{k} \binom{k}{j} \frac{(-x)^j}{j!},
\]

and that they form an orthonormal basis of \( L^2(\mathbb{R}_+) \). In the multivariate case, we tensorize the Laguerre basis. For \( k = (k_1, \ldots, k_d) \in \mathbb{N}^d \) a multi-index, we define the multivariate Laguerre function \( \varphi_k \) on \( \mathbb{R}^d_+ \) as the tensor product of one-dimensional Laguerre functions:

\[
\varphi_k(x_1, \ldots, x_d) := (\varphi_{k_1} \otimes \cdots \otimes \varphi_{k_d})(x_1, \ldots, x_d) = \varphi_{k_1}(x_1) \times \cdots \times \varphi_{k_d}(x_d).
\] (4)
The multivariate Laguerre functions \((\varphi_k)_{k \in \mathbb{N}^d}\) form a basis of \(L^2(\mathbb{R}_+^d)\), and we expand the functions \(f\), \(g\) and \(h\) in this basis:

\[
f = \sum_{k \in \mathbb{N}^d} a_k \varphi_k, \quad g = \sum_{j \in \mathbb{N}^d} b_j \varphi_j, \quad h = \sum_{\ell \in \mathbb{N}^d} c_{\ell} \varphi_{\ell}. \tag{5}
\]

The use of the Laguerre basis is relevant in this context, as the one-dimensional Laguerre functions verify the relation:

\[
\forall k, j \in \mathbb{N}, \quad \varphi_k \ast \varphi_j = 2^{-1/2}(\varphi_{k+j} - \varphi_{k+j+1}), \tag{6}
\]

see (Abramowitz and Stegun, 1972, formula 22.13.14). Using this relation and \(h = f \ast g\), by expanding the functions \(f\), \(g\) and \(h\) on the Laguerre basis, we get a relation between their coefficients.

**Proposition 3.1.** If \(a, b\) and \(c\) are the coefficients defined in \((5)\), then the following relation holds:

\[
\forall \ell \in \mathbb{N}^d, \quad c_{\ell} = 2^{-d/2} \sum_{\varepsilon \in \{0, 1\}^d} (-1)^{|\varepsilon|}(a \ast b)_{\ell-\varepsilon},
\]

where \(a \ast b\) is the discrete convolution product of \(a\) and \(b\) defined by \((a \ast b)_k := \sum_{j \leq k} a_j b_{k-j}\) if \(k \in \mathbb{N}^d\) and \((a \ast b)_k = 0\) if \(k \notin \mathbb{N}^d\).

This relation can be written as a discrete convolution product \(c = \beta \ast a\) with \(\beta \in \mathbb{N}^d\) defined by:

\[
\beta_k := 2^{-d/2} \sum_{\varepsilon \in \{0, 1\}^d} (-1)^{|\varepsilon|} b_{k-\varepsilon}, \tag{7}
\]

where by convention \(b_j = 0\) if \(j \notin \mathbb{N}^d\). Thus, we have a linear relation between the coefficients \(c\) and \(a\):

\[
\forall \ell \in \mathbb{N}^d, \quad c_{\ell} = \sum_{k \leq \ell} G_{\ell k} a_k, \quad \text{where} \quad G_{\ell k} := \begin{cases} \beta_{\ell-k} & \text{if } k \leq \ell, \\ 0 & \text{else.} \end{cases} \tag{8}
\]

If we consider \(G\) as an infinite hypermatrix \([G_{\ell k}]_{\ell, k \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d \times \mathbb{N}^d}\), then it is lower triangular according to Definition 2.5.

In the next proposition, we show that the linear relation between \(a\) and \(c\) is invertible.

**Proposition 3.2.** Let \(G\) be the infinite hypermatrix defined in \((8)\). For every \(k \in \mathbb{N}^d\), there exists \([H_{\ell k}]_{\ell \leq k}\) such that for every \(a, c \in \mathbb{N}^d\) satisfying the relation \((8)\), we have:

\[
a_{\ell} = \sum_{\ell \leq k} H_{\ell k} c_k. \tag{9}
\]

We denote \(H_{\ell k} := (G^{-1})_{\ell k}\).

We can write the linear relations between \(a\) and \(c\) using hypermatrices and contraction products. For \(m \in (\mathbb{N}^*)^d\), we denote by \(a_m\) (resp. \(c_m\)) the hypermatrix \([a_k]_{k \leq m-1} \in \mathbb{R}^m\) (resp. \([c_\ell]_{\ell \leq m-1} \in \mathbb{R}^m\), and we denote by \(G_m\) and \(G_m^{-1}\) the hypermatrices \([G_{\ell k}]_{\ell, k \leq m-1}\) and \([(G^{-1})_{\ell k}]_{\ell, k \leq m-1} \in \mathbb{R}^{(m,m)}\). Then, we have:

\[
c_m = G_m \times_d a_m, \quad a_m = G_m^{-1} \times_d c_m, \tag{10}
\]

where \(\times_d\) stands for the contraction product defined by \((2)\).
**Estimation procedure.** For \( m \in (\mathbb{N}^*)^d \), let \( S_m \) be the vector space spanned by the functions \( \varphi_k \) for \( k \leq m - 1 \), and let \( D_m := m_1 \cdots m_d \) be its dimension. We estimate \( f \) by estimating \( f_m \) the projection of \( f \) on \( S_m \). This projection is given by \( f_m = \sum_{k \leq m-1} a_k \varphi_k \), so the problem reduces to the estimation of \( a_m \). Because \( a_m \) is related to \( c_m \) by [10] and since \( c_k = \mathbb{E}[\varphi_k(Z_1)] \), we estimate \( f \) by:

\[
\hat{f}_m := \sum_{k \leq m-1} \hat{a}_k \varphi_k \text{ where } \hat{a}_m := G_m^{-1} \times_d \hat{c}_m \text{ and } \hat{c}_k := -\frac{1}{n} \sum_{i=1}^n \varphi_k(Z_i).
\]

**4. Non-asymptotic error bounds**

We quantify the quality of the estimator \( \hat{f}_m \) by its MISE (Mean Integrated Squared Error): \( \mathbb{E}\|f - \hat{f}_m\|_{L^2}^2 \). In the next proposition, we decompose the MISE in a bias term and a variance term and give a bound on the MISE of \( \hat{f}_m \).

**Proposition 4.1.** If \( f \) and \( g \) are \( L^2(\mathbb{R}^d_+) \) functions, then we have the inequality:

\[
\forall m \in \mathbb{N}^d, \quad \mathbb{E}\|f - \hat{f}_m\|_{L^2}^2 \leq \|f - f_m\|_{L^2}^2 + \frac{2^d D_m \rho^2(G_m^{-1})}{n}\|h\|_{L^\infty}\|G_m^{-1}\|_F^2.
\]

**Remark 4.2.** Using the Cauchy–Schwarz inequality, it holds \( \|h\|_{L^\infty} \leq \|f\|_{L^2}\|g\|_{L^2} \) which is finite by assumption. Moreover, if \( g \) is bounded, we also have \( \|h\|_{L^\infty} \leq \|g\|_{L^\infty} \).

**Remark 4.3.** The norm equivalence \( [3] \) implies that \( \|G_m^{-1}\|_F^2 \leq D_m \rho^2(G_m^{-1}) \), so the order of magnitude of the variance term is given by \( \|G_m^{-1}\|_F^2 \). The minimum is important only for the small values of \( m \) (because of the constants).

Under an additional assumption, [Comte and Genon-Catalot, 2018] improved the variance bound in the one-dimensional case. We generalize their result to the multivariate case.

**Proposition 4.4.** We denote \( Y_1^{(j)} \) the \( j \)-th coordinate of \( Y_1 \). We assume that for any nonempty subset \( J \) of \( \{1, \ldots, d\} \), we have:

\[
M_J(g) := \mathbb{E}\left[ \prod_{j \in J} \frac{1}{\sqrt{Y_1^{(j)}}} \right] < +\infty.
\]

Then we have the following inequality:

\[
\forall m \in (\mathbb{N}^*)^d, \quad \mathbb{E}\|f - \hat{f}_m\|_{L^2}^2 \leq \|f - f_m\|_{L^2}^2 + \frac{c(g)\sqrt{D_m \rho^2(G_m^{-1})}}{n}\|h\|_{L^\infty}\|G_m^{-1}\|_F^2,
\]

where \( c(g) \) is a positive constant depending on \( M_J(g) \) for \( J \subseteq \{1, \ldots, d\} \).

To study the bias term, we assume that \( f \) belongs to a Sobolev–Laguerre 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 coefficients was established later by [Comte and Genon-Catalot, 2015]. Following the same idea, we define Sobolev–Laguerre balls on \( \mathbb{R}^d_+ \).

**Definition 4.5 (Sobolev–Laguerre ball).** Let \( L > 0 \) and \( s \in (0, +\infty)^d \), we define the Sobolev–Laguerre ball of order \( s \) and radius \( L \) by:

\[
W^s(\mathbb{R}^d_+, L) := \left\{ f \in L^2(\mathbb{R}^d_+) \left| \sum_{k \in \mathbb{N}^d} a_k^2(f) k^s \leq L \right. \right\},
\]

with \( a_k(f) := (f, \varphi_k)_{L^2} \) the Laguerre coefficients of \( f \).
Assuming \( f \) to belong to \( W^s(\mathbb{R}^d_+, L) \), the bias term decreases to 0 with polynomial rate. Indeed, for \( m \in (\mathbb{N}^*)^d \), we have:

\[
\|f - f_m\|_{L^2}^2 = \sum_{k \in \mathbb{N}^d} a_k^2(f) \leq \sum_{q=1}^{d} \sum_{k \in \mathbb{N}^d} a_k^2(f) k_q^r \cdot k_q^{-s_q} \leq L \sum_{q=1}^{d} m_q^{-s_q}. \tag{12}
\]

The case where the Laguerre coefficients of \( f \) decrease with exponential rate is also interesting. We define new functional spaces, “smooth Laguerre spaces”, in the following way.

**Definition 4.6** (Smooth Laguerre ball). Let \( L > 0 \) and \( r \in (0, +\infty)^d \), we define the smooth Laguerre ball of order \( r \) and radius \( L \) as:

\[
S^r(\mathbb{R}^d_+, L) := \left \{ f \in L^2(\mathbb{R}^d_+) \mid \sum_{k \in \mathbb{N}^d} a_k^2(f) e^{-rk} \leq L \right \}.
\]

By the same argument as previously, if \( f \) belongs to \( S^r(\mathbb{R}^d_+, L) \), the bias term decreases to 0 with exponential rate:

\[
\forall m \in (\mathbb{N}^*)^d, \quad \|f - f_m\|_{L^2}^2 \leq L \sum_{q=1}^{d} e^{-r_m s_q}. \tag{13}
\]

Now, we need to control the variance term in Proposition 4.4. In the one-dimensional case, this control can be provided under assumptions on the Laplace transform \( G \) of \( g \) and under assumptions on the derivatives of \( g \), see lemma 3.6 in [Comte et al. 2017]. In the next proposition, we extend this result to the multivariate case. Moreover, we make assumptions only on the behavior of the Laplace transform of \( g \), we do not need to study its differentials.

We recall that the Laplace transform of \( g \) is the function \( G \) defined on the domain \( P_+^d \) by:

\[
G(s) := \int_{\mathbb{R}^d_+} e^{-s \cdot x} g(x) \, dx,
\]

where \( P_+ \) stands for the set of complex numbers with non-negative real part. In addition, we extend the set \( C \) of complex numbers by adding a point at infinity denoted by \( \infty \). The control of the Frobenius norm of \( G_m^{-1} \) relies on the behavior of \( G \) when some of its arguments take the \( \infty \) value.

**Proposition 4.7.** We assume that \( \beta \in \ell^1(\mathbb{N}^d) \), with \( \beta \) defined in [7]. We assume that \( G \) is non-zero on \( P_+^d \) and that there exists \( \alpha \in (\mathbb{N}^*)^d \) such that the function:

\[
K_\alpha(s) := (1 + s)^\alpha G(s), \quad s \in P_+^d,
\]

can be extended as a non-zero function on \( (P_+ \cup \{\infty\})^d \) such that the restriction of \( K_\alpha \) on \( (i\mathbb{R} \cup \{\infty\})^d \) is continuous. Then for \( m \in \mathbb{N}^d \) satisfying \( m \geq 4 \), there exists a constant \( C > 0 \) depending on \( \beta \) such that \( \|G_m^{-1}\|^2_F \leq C m^{2\alpha} \).

**Remark 4.8.** If \( g \in W^s(\mathbb{R}^d_+, L) \) with \( L > 0 \) and \( s \in (1, +\infty)^d \), the Laguerre coefficients of \( g \) belong to \( \ell^1(\mathbb{N}^d) \). Indeed, using Cauchy–Schwarz inequality,

\[
\sum_{k \in \mathbb{N}^d} |b_k| = \sum_{k \in \mathbb{N}^d} |b_k| k^{{1 \over 2} r} k^{-r} \leq \left( \sum_{k \in \mathbb{N}^d} b_k^2 k^r \right)^{1 \over 2} \left( \sum_{k \in \mathbb{N}^d} k^{-s} \right)^{1 \over 2} < +\infty,
\]

because for every \( q \in \{1, \ldots, d\}, s_q > 1 \).
Remark 4.9. We notice that in dimension $d = 1$, these assumptions simply become:

1. The Laplace transform $G$ does not vanish on $\mathbb{P}_+$.
2. The Fourier transform of $g$ admits an asymptotic expansion $g^\ast(\omega) = \omega^{-\alpha} (K_\alpha + o(1))$ when $|\omega|$ goes to $+\infty$, for some $\alpha \in \mathbb{N}^*$ and some non-zero constant $K_\alpha$.

It is easy to see that this second assumption is a consequence of the assumptions on the derivatives of $g$ made in [Comte et al., 2017, subsection 2.5].

Remark 4.10. If the distribution of $Y$ is a product of gamma distributions $\otimes_{q=1}^d \Gamma(\alpha_q, \lambda_q)$ with $\alpha \in (\mathbb{N}^*)^d$, then $G$ and $K_\alpha$ are given by:

$$G(s) = \prod_{q=1}^d \left( 1 + \frac{s_q}{\lambda_q} \right)^{-\alpha_q} \quad \text{and} \quad K_\alpha(s) = \lambda^\alpha \prod_{q=1}^d \left( \frac{1 + s_q}{\lambda_q + s_q} \right)^{\alpha_q},$$

so the assumptions of Proposition 4.7 are fulfilled.

Gathering (12) or (13) with Proposition 4.7, we obtain convergence rates for our estimator.

Theorem 4.11. Let $s, r \in (0, +\infty)^d$ and $L > 0$. Assume that $g$ satisfies the assumptions of Proposition 4.7 with $\alpha \in (\mathbb{N}^*)^d$.

1. For $m_{\text{opt}} \in (\mathbb{N}^*)^d$ given by $m_{\text{opt}, j} = n^{1/(s_j + s)} \sum_{i=1}^d \frac{2\alpha_i}{\alpha_j}$, there exists a constant $C > 0$ depending on $s, L$ and $g$ such that:

$$\sup_{f \in W^s(\mathbb{R}_+^d, L)} \mathbb{E}\|f - \hat{f}_{m_{\text{opt}}}\|_{L^2}^2 \leq C n^{-1/(1 + \sum_{i=1}^d \frac{2\alpha_i}{\alpha_j})}.$$

2. For $m_{\text{opt}} \in (\mathbb{N}^*)^d$ given by $m_{\text{opt}, j} = \frac{\log n}{r_j}$, there exists a constant $C > 0$ depending on $r$, $L$ and $g$ such that:

$$\sup_{f \in S^r(\mathbb{R}_+^d, L)} \mathbb{E}\|f - \hat{f}_{m_{\text{opt}}}\|_{L^2}^2 \leq C \frac{(\log n) \sum_{i=1}^d 2\alpha_i}{n}.$$

Remark 4.12. Our convergence rates on Sobolev–Laguerre balls are similar to convergence rates found by [Comte and Lacour, 2013] on anisotropic Sobolev balls, in the context of deconvolution using a kernel estimator with an ordinary smooth noise.

Remark 4.13. We could have considered mixed regularities for $f$. For instance, if the coefficients of $f$ satisfy:

$$\sum_{k \in \mathbb{N}^d} a_k^2 k_1^{s_1} \cdots k_{j_0}^{s_{j_0}} \exp(r_{j_0 + 1} k_{j_0 + 1} + \cdots + r_d k_d) \leq L,$$

for some $1 \leq j_0 < d$, then we can give the following upper-bound on the bias term:

$$\|f - \hat{f}_m\|_{L^2}^2 \leq L \left( \sum_{j=1}^{j_0} m_{j}^{-s_j} + \sum_{j=j_0+1}^d e^{-r_j m_j} \right).$$

By choosing $m_j \propto \frac{\log n}{r_j}$ for $j > j_0$, the MISE of $\hat{f}_m$ is then:

$$\mathbb{E}\|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{j=1}^{j_0} m_{j}^{-s_j} + \frac{\|h\|_{\infty}}{N_n} \prod_{j=1}^{j_0} m_{2\alpha_j} + \frac{L (d - j_0)}{n},$$
with \( N_n := \prod_{j > j_0}^{n} \frac{(\log(n)/r_j)^{\alpha_j}}{r_j} \). We can now choose the \( j_0 \)-first components of \( m \) like in the case 1 of Theorem 4.11, but with a sample size \( N_n \), and we get:

\[
E\|f - \hat{f}_{m_{opt}}\|_{L^2}^2 \leq C \left( \frac{n}{\prod_{j > j_0}^{n} (\log(n)/r_j)^{2\alpha_j}} \right)^{-1/2} \left( 1 + \sum_{j=1}^{j_0} \frac{2\alpha_j}{r_j} \right).
\]

It is the rate we would find on a Sobolev–Laguerre ball in dimension \( j_0 \), up to a logarithmic factor.

5. Adaptive estimation and oracle inequalities

In practice, we do not know the underlying regularity of \( f \), so we cannot compute the model \( m_{opt} \) of Theorem 4.11. We want a data-driven procedure that automatically provides the bias-variance compromise, without assumptions on \( f \). More precisely, we want to select \( \hat{m} \) from the data such that the MISE of \( \hat{f}_m \) is close to the oracle MISE:

\[
\inf_{m} E\|f - \hat{f}_m\|_{L^2}^2.
\]

Let \( m^* \in (\mathbb{N}^*)^d \) fixed such that \( D_{m^*} \leq n \). We only look at sub-models of \( m^* \) that are not “too big”. More precisely, we consider the models that belong to the following model collection:

\[
M_n := \left\{ m \leq m^* \mid D_m \rho^2(G_m^{-1}) \leq \frac{n}{\log n} \right\}.
\]

We assume that the \( Y_i \)'s distribution satisfies the following assumptions.

(A1) The noise density \( g \) is bounded.

(A2) For every nonempty \( J \subset \{1, \ldots, d\} \), \( M_J(g) \) defined by (11) is finite.

(A3) For every \( b > 0 \), we have \( \sum_{m \in M_n} \rho^2(G_m^{-1}) e^{-bD_m} \leq K(b) \), with \( K(b) \) a positive constant not depending on \( n \).

Under assumptions (A1) and (A2), we can apply Proposition 4.4 to control the variance term \( E\|\hat{f}_m - f_m\|_{L^2}^2 \) by:

\[
V(m) := \frac{c(g)\sqrt{D_m \rho^2(G_m^{-1})}}{n} \wedge \left( \|g\|_{\infty} \vee 1 \right) \|G_m^{-1}\|_F^2 \log n.
\]

Moreover, under assumption (A3), we can control the right deviation of \( \|\hat{f}_m - f_m\|_{L^2}^2 \) when \( m \) belongs to \( M_n \).

Lemma 5.1. Under assumptions (A1) to (A3), there exists a numerical constant \( a_0(d) > 0 \) depending on the dimension \( d \) such that for every \( a > a_0(d) \):

\[
\sum_{m \in M_n} E\left( \|\hat{f}_m - f_m\|_{L^2}^2 - a V(m) \right) \leq \frac{C}{n},
\]

with \( C \) a positive constant depending on \( g \) and \( a \).
The one-dimensional case. When \( d = 1 \), we consider two procedures, indexed by \( i \in \{1, 2\} \), the first procedure being the one studied by Mabon (2017). In both procedures, the projection space is chosen by minimizing the penalized criterion:

\[
\hat{m}_i := \arg \min_{m \in \mathcal{M}_n} \left[ -\|f_m\|_{L^2} + \kappa_i \text{pen}_i(m) \right], \quad i \in \{1, 2\}
\]

where \( \mathcal{M}_n \) is defined by [14]. \( \kappa_i \) is a numerical constant to be adjusted and \( \text{pen}_i \) is the penalty term. Mabon (2017) proposes the following penalty term:

\[
\text{pen}_1(m) := \frac{2m \rho^2(G_{m}^{-1})}{n} \wedge \left( \|g\|_{\infty} \vee 1 \right) \|G_{m}^{-1}\|_{F}^{2} \log n,
\]

and provides an oracle inequality for the estimator \( \hat{f}_{\hat{m}_1} \) under assumptions (A1) and (A3), see Theorem 4.1 in Mabon’s article. This choice is based on the bound of the variance term of Proposition 4.1. However if \( g \) is bounded, assumption (A2) holds automatically in dimension \( d = 1 \), so we can apply Proposition 4.4 to get a better bound on the variance term. Thus, we propose the penalty term:

\[
\text{pen}_2(m) := V(m) = \frac{c(g) \sqrt{n} \rho^2(G_{m}^{-1})}{n} \wedge \left( \|g\|_{\infty} \vee 1 \right) \|G_{m}^{-1}\|_{F}^{2} \log n.
\]

We show that our estimator \( \hat{f}_{\hat{m}_2} \) also satisfies an oracle inequality.

**Theorem 5.2.** We assume \( d = 1 \). Under assumptions (A1) to (A3), there exists a numerical constant \( \kappa_0 > 0 \) such that for every choice of \( \kappa > \kappa_0 \), we have the following oracle inequality:

\[
\mathbb{E}\left[\|f - \hat{f}_{\hat{m}_2}\|_{L^2}^2\right] \leq 4 \inf_{m \in \mathcal{M}_n} \left( \|f - f_m\|_{L^2}^2 + \kappa \text{pen}_2(m) \right) + \frac{C}{n},
\]

with \( C \) a positive constant depending on \( \kappa \) and \( g \).

The multivariate case. We use the procedure similar to Goldenshluger and Lepski (2011) for model selection introduced by Chagny (2013b) for the estimation of a conditional density function. We apply this procedure to our multivariate deconvolution problem, and we establish an oracle inequality.

We choose \( \hat{m} \) in the model collection \( \mathcal{M}_n \) defined in [14], minimizing:

\[
\hat{m} := \arg \min_{m \in \mathcal{M}_n} \left[ A(m) + \kappa_3 V(m) \right],
\]

where \( V(m) \) is defined by [15] and \( A(m) \) is a term which has the order of the bias term (see the proof):

\[
A(m) := \max_{m' \in \mathcal{M}_n} \left( \|\hat{f}_{m'} - f_{m \land m'}\|_{L^2}^2 - \kappa_1 V(m') \right),
\]

and where \( \kappa_1, \kappa_2 \) are two numerical constants to be adjusted.

**Theorem 5.3** (Oracle inequality). Under assumptions (A1) to (A3), there exists a numerical constant \( \kappa_0(d) > 0 \) depending on the dimension \( d \) such that for every choice of \( \kappa_1, \kappa_2 \) satisfying \( \kappa_0(d) < \kappa_1 \leq \kappa_2 \), we have the following oracle inequality:

\[
\mathbb{E}\left[\|f - \hat{f}_{\hat{m}}\|_{L^2}^2\right] \leq C \inf_{m \in \mathcal{M}_n} \left( \|f - f_m\|_{L^2}^2 + V(m) \right) + \frac{C'}{n},
\]

with \( C \) a positive constant depending on \( \kappa_1 \) and \( \kappa_2 \), and \( C' \) a positive constants depending on \( g \), \( d \) and \( \kappa_1 \).
Remark 5.4. It is important to notice that the proof of Theorem 5.3 only depends on the deconvolution structure through Lemma 5.1 to obtain a concentration bound on the variance term. The rest of the proof does not use the deconvolution structure again. Thus, the procedure can be applied to more general model selection problems.

6. Numerical illustrations

6.1. Estimators comparison in one-dimensional case

We want to compare the two estimators \( \hat{f}_{m_1} \) and \( \hat{f}_{m_2} \) defined in Section 5, with different distributions for \( X \) and \( Y \), and with different sample size \( n \). We compare their MISE:

\[
R_i = R_i(f, g, n) := \mathbb{E} \| f - \hat{f}_{m_i} \|^2_{L^2}, \quad i \in \{1, 2\}.
\]

We also compute the oracle risk:

\[
R_o = R_o(f, g, n) := \min_{1 \leq m \leq m^*} \mathbb{E} \| f - \hat{f}_m \|^2_{L^2},
\]

where \( m^* \) is the maximal element of \( M_n \). We compute an approximation of \( \| \cdot \|^2_{L^2} \) with Simpson’s rule on a bounded interval \( I \) of \( \mathbb{R}_+ \) with 1000 points. We compute the expectations with an empirical mean over 500 samples (we use the same samples for \( R_1, R_2 \) and \( R_o \)).

Remark 6.1. Since \( f \) is non-negative, \( \hat{f}_m \) is replaced by \( (\hat{f}_m)^+ := \max(\hat{f}_m, 0) \) in the following.

Distributions for \( X \). We consider several distributions for \( X \). In each case, we normalize the distribution for the variance to be 1. We use the same examples as Mabon (2017).

- Exponential \( \mathcal{E}(1), I = [0, 5] \).
- Gamma \( \Gamma(20, \sqrt{20}), I = [0, 10] \).
- Rayleigh, \( f(x) = \frac{x}{\sigma^2} \exp(-\frac{x^2}{2\sigma^2}) \) with \( \sigma^2 = \frac{2}{4-\pi}, I = [0, 6] \).
- Weibull, \( f(x) = \frac{k}{\lambda} \left( \frac{x}{\lambda} \right)^{k-1} e^{-(x/\lambda)^k} \), with \( k = \frac{3}{2} \) and \( \lambda = 1/\sqrt{\Gamma(1 + \frac{4}{3}) - \Gamma(1 + \frac{2}{3})^2}, I = [0, 5] \).
- Beta \( B(4, 5) \) with normalisation \( \frac{9}{\sqrt{2}}, I = [0, 8] \).
- Gamma mixture \( 0.4 \Gamma(2, 2) + 0.6 \Gamma(16, 4) \) with normalisation \( \frac{1}{\sqrt{2.96}}, I = [0, 5] \).

Distributions for \( Y \). We choose gamma distributions for \( Y \). We recall the density of the gamma distribution with parameters \( \alpha \geq 1 \) and \( \lambda > 0 \):

\[
g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0.
\]

with \( \Gamma \) the gamma function. These distributions satisfy assumptions \( (A1) \) to \( (A3) \), so we can use the procedure described in Section 5. We can compute the Laguerre coefficients of \( g \) exactly:

\[
\forall k \in \mathbb{N}, \quad b_k = \sqrt{2} \left( \frac{\lambda}{\lambda+1} \right)^\alpha \sum_{j=0}^{k} \binom{k}{j} \binom{\alpha-1+j}{j} \left( \frac{-2}{1+\lambda} \right)^j.
\]
We choose two distributions for $X$. Therefore, we can compute easily and efficiently the matrix $\Sigma(I)$. For every distribution of $X$, we considered earlier and for every $X$ we computed:

$$\mathbf{b}_m = \sqrt{2} \left( \frac{\lambda}{\lambda + 1} \right)^{\alpha} \mathbf{P}_m \mathbf{v}_m,$$

where $\mathbf{P}_m$ is the $m \times m$ matrix with components $P_{ij} := \binom{i}{j}$ for $0 \leq i, j \leq m - 1$, and $\mathbf{v}_m$ is the vector of size $m$:

$$\mathbf{v}_m := \left[ \binom{\alpha - 1 + j}{j} \left( \frac{-2}{1 + \lambda} \right)^j \right]_{0 \leq j \leq m - 1}.$$

Therefore, we can compute easily and efficiently the matrix $\mathbf{G}_m$ and its inverse.

We choose two distributions for $Y$: the $\Gamma(2, \sqrt{20})$ distribution which has variance $\frac{10}{2}$ and the $\Gamma(2, \sqrt{8})$ distribution which has variance $\frac{1}{4}$.

The constant $c(g)$. This constant in the penalty $\text{pen}_2$ is not known, so we have to evaluate it numerically. Following the proof of the Proposition 4.4, we see this constant appears in the upper bound:

$$\frac{\rho^2 (\mathbf{G}_m^{-1})}{n} \sum_{k=0}^{m-1} \mathbb{E}[\varphi_k(Z_1)^2] \leq c(g) \sqrt{m} \rho^2 (\mathbf{G}_m^{-1}) \frac{1}{n},$$

which is true for every density of $X_1$ and for every $m$. Therefore,

$$c(g) = \sup_{X_1 \sim f} \sup_{m \in \mathbb{N}} \frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} \mathbb{E}[\varphi_k(X_1 + Y_1)^2].$$

For every distribution of $X$ we considered earlier and for every $m$ from 1 to 50, we computed:

$$\frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} \mathbb{E}[\varphi_k(X_1 + Y_1)^2],$$

using an empirical mean over 1000 realizations of $(X_1, Y_1)$ to compute the expectation, then we took the maximum. For the two distributions we considered for $Y_1$, the worst constant was about 0.4 and was reached for $X_1 \sim \mathcal{E}(1)$. For the other distributions of $X_1$, we found constants between 0.2 and 0.3. To be safe, we chose a bigger constant $c(g) = 0.5$ for both $\Gamma(2, \sqrt{20})$ and $\Gamma(2, \sqrt{8})$ distributions.

The constants $\kappa$. We use the same constant $\kappa_1 = 0.03$ as [Mabon (2017)]. Concerning the constant $\kappa_2$, we made several simulations to calibrate it and we chose $\kappa_2 = 0.04$. During the calibration of $\kappa_2$, we realized that the collection $\mathcal{M}_n$ did not have enough models. The selected dimensions were too small and lead to bad performances. Therefore, we considered the following collection:

$$\mathcal{M}_n' := \left\{ 1 \leq m \leq m^* \mid m \rho^2 (\mathbf{G}_m^{-1}) \leq \frac{n}{\log n} \right\}.$$

Procedure. We take the maximal dimension to be $m^* = 20$. We use two sample sizes: $n = 200$ and $n = 2000$. Given a distribution for $X$, a distribution for $Y$, a sample size $n$, and a subdivision $\Sigma(I)$ of $I$, do:

1. Compute the matrix $\mathbf{G}_{20}$ and its inverse.
2. Compute the collection $\mathcal{M}_n'$.
3. Compute $f(x)$ for $x \in \Sigma(I)$.
4. Repeat 500 times:
   
   (a) Generate a sample $Z_i = X_i + Y_i$, $i = 1, \ldots, n$.
   
   (b) Compute the Laguerre coefficients vector $\hat{c}_{20}$ and compute $\hat{a}_{20} = G_{20}^{-1} \hat{c}_{20}$.
   
   (c) Compute $\hat{m}_1$ and $\hat{m}_2$ minimizing:
   
   $$\hat{m}_i = \arg\min_{m \in M'_n} \left( -\sum_{k=0}^{m-1} \hat{a}_k^2 + \kappa_i \text{pen}_i(m) \right), \quad i \in \{1, 2\},$$

   and compute $(\hat{f}_{\hat{m}_1}(x))_+$ and $(\hat{f}_{\hat{m}_2}(x))_+$ for $x \in \Sigma(I)$.
   
   (d) Compute $J_1 = \| f - (\hat{f}_{\hat{m}_1})_+ \|_{L^2}$ and $J_2 = \| f - (\hat{f}_{\hat{m}_2})_+ \|_{L^2}$ using Simpson’s rule.
   
   (e) For $m$ from 1 to 20, compute $(\hat{f}_m(x))_+$ for $x \in \Sigma(I)$; then compute $J(m) = \| f - (\hat{f}_m)_+ \|_{L^2}$ using Simpson’s rule.

5. Compute $R_1$ (resp. $R_2$) as the mean of $J_1$ (resp. $J_2$) over the 500 samples.

6. For each $m$, compute the mean of $J(m)$ over the 500 samples, then compute $R_o$ as the minimum of these quantities.

Results. We show our results in Table 1. We note that in the cases $X \sim \mathcal{E}(1)$ and $X \sim \Gamma(20, \sqrt{20})$, both estimators performed badly compared to the oracle. We see that our estimator $\hat{f}_{\hat{m}_2}$ is better when $X$ has gamma, Rayleigh, beta and mixture gamma distribution. The estimator $\hat{f}_{\hat{m}_1}$ is better when $X$ has Weibull and exponential distribution.

For illustration, Figure 1 shows several estimations when the distribution of $X$ is a mixture gamma. It’s a bimodal distribution, so it’s interesting to see if the estimators are able to recover the two “peaks” and the “hollow” of the true density. For small samples ($n = 200$), the
estimator \( \hat{f}_{m_1} \) seems to fail more often than the estimator \( \hat{f}_{m_2} \). For large samples (\( n = 2000 \)), both estimators locate well the two pikes (by overestimating the first one and underestimating the second one), but the estimator \( \hat{f}_{m_1} \) locates less well the hollow and underestimates more the second pike.

6.2. Model selection in two-dimensional case

In this subsection, our goal is to illustrate the procedure similar to Goldenshluger and Lepski (2011) on two examples, in the case \( d = 2 \).

**Distributions for X.** To generate a random vector \( X = (X^{(1)}, X^{(2)}) \), we do the following: we generate a random variable \( W^{(1)} \) with some distribution and we generate \( W^{(2)} \) independently with some other distribution, and we compute:

\[
\begin{bmatrix}
X^{(1)} \\
X^{(2)}
\end{bmatrix} = \begin{bmatrix} 1 & 0.1 \\ 0.2 & 1 \end{bmatrix} \begin{bmatrix} W^{(1)} \\ W^{(2)} \end{bmatrix},
\]

such that the coordinates of \( X \) are not independent. We consider the following distributions for \( W^{(1)} \) and \( W^{(2)} \):

- Gamma \( \Gamma(3, 1) \)
- Beta \( B(4, 5) \) with normalisation \( \frac{9}{\sqrt{2}} \)
- Rayleigh, \( f(x) = \frac{x}{\sigma^2} \exp(-\frac{x^2}{2\sigma^2}) \) with \( \sigma^2 = \frac{2}{\sqrt{\pi}} \)
- Weibull, \( f(x) = \frac{k}{\lambda} (\frac{x}{\lambda})^{k-1} \exp(-x/\lambda) \), with \( k = \frac{3}{2} \) and \( \lambda = 1/\sqrt{\Gamma(1 + \frac{3}{2}) - \Gamma(1 + \frac{2}{3})^2} \)

**Distribution for Y.** We chose distributions of the form \( \Gamma(\alpha_1, \lambda_1) \otimes \Gamma(\alpha_2, \lambda_2) \). These distributions satisfy the assumptions (A1) to (A3) of the Section 5. Moreover, the Laguerre coefficients can be computed easily: if \( \gamma_i \) is the density of the distribution \( \Gamma(\alpha_i, \lambda_i) \), then we have \( g = \gamma_1 \otimes \gamma_2 \), so:

\[
\forall k \in \mathbb{N}^2, \quad b_k = \langle \gamma_1, \varphi_{k_1} \rangle_{L^2} \times \langle \gamma_2, \varphi_{k_2} \rangle_{L^2}.
\]
In columns: distribution of $W^{(1)}$. In each cell, we show the MISE (multiplied by $10^4$) and the mean selected model.

Refer to Subsection 6.1 for the computation of $\langle \gamma_i, \varphi_{k_i} \rangle_{L^2}$. We chose the distribution $\Gamma(2, \sqrt{20}) \otimes \Gamma(2, \sqrt{20})$ for $Y$.

The $\kappa$ constants. After several simulations, we chose $\kappa_1 = \kappa_2 = 10^{-5}$. However, this calibration is rough. Like in the one-dimensional case, the model collection $\mathcal{M}_n$ defined by (14) was too small, so we considered the following collection:

$$\mathcal{M}'_n = \left\{ m \leq m^* \left| D_m \rho^2(G_m^{-1}) \leq 10^4 \frac{n}{\log n} \right. \right\}.$$  

Results. We took the maximal model to be $m^* = (12, 12)$. In each case, we simulated a sample of size $n = 500$ and $n = 5000$, then we computed the model $\hat{m}$ defined by (19). We computed $\mathbb{E} \| f - \hat{f}_{\hat{m}} \|_{L^2}$ by approximating the double integral by a Riemann sum:

$$\int_{\mathbb{R}_+^2} \left( f(x, y) - \hat{f}_{\hat{m}}(x, y) \right)^2 \, dx \, dy \approx \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \left( f(x_i, y_j) - \hat{f}_{\hat{m}}(x_i, y_j) \right)^2 \Delta x_i \Delta y_j$$

on a grid with step size 0.01, and by computing the expectation with an empirical mean over 100 samples.

We show our results in Table 2. In each case, we compute the MISE and the mean selected model. We emphasize that in the majority of the samples, the selected model is anisotropic: the components of $\hat{m}$ are not equal. It is a property we expected from a procedure similar to Goldenshluger and Lepski (2011). The estimator adapts to the regularity of $f$.

For illustration, Figure 2 shows the result of the estimation when $W^{(1)}$ has a Gamma distribution and $W^{(2)}$ has a Weibull distribution, for a sample of size 5000. We show both the adaptive estimator $\hat{f}_{\hat{m}}$ and the estimator with no selection procedure $\hat{f}_{(12,12)}$, where we simply choose the maximum model. We see that the procedure selected the model $\hat{m} = (5, 8)$, which has a smaller dimension than the maximum model.
Figure 2: Density estimation when $W^{(1)}$ has a Gamma distribution and $W^{(2)}$ has a Weibull distribution, for a sample of size $n = 5000$. First column: true density, second column: adaptive estimator $\hat{f}_m$, third column: max model estimator $\hat{f}_{(12,12)}$. The selected model is $\hat{m} = (5, 8)$.

7. Proofs

7.1. Proofs of Sections 3 and 4

Proof of Proposition 3.1. Using the relation (6), we have:

$$f \ast g = \sum_{k \in \mathbb{N}^d} \sum_{j \in \mathbb{N}^d} a_k b_j (\varphi_k \ast \varphi_j) = \sum_{k \in \mathbb{N}^d} \sum_{j \in \mathbb{N}^d} a_k b_j \bigotimes_{q=1}^d 2^{-1/2}(\varphi_{k_q+j_q} - \varphi_{k_q+j_q+1}).$$

By setting $\ell_q = k_q + j_q$, we get:

$$f \ast g = 2^{-d/2} \sum_{\ell \in \mathbb{N}^d} \sum_{k \in \mathbb{N}^d} a_k b_{\ell-k} \bigotimes_{q=1}^d (\varphi_{\ell_q} - \varphi_{\ell_q+1}) = 2^{-d/2} \sum_{\ell \in \mathbb{N}^d} (a \ast b)_{\ell} \bigotimes_{q=1}^d (\varphi_{\ell_q} - \varphi_{\ell_q+1}).$$

Using the tensor product multi-linearity:

$$\bigotimes_{q=1}^d (\varphi_{\ell_q} - \varphi_{\ell_q+1}) = \bigotimes_{q=1}^d \sum_{\varepsilon_q=0}^{1} (-1)^{\varepsilon_q} \varphi_{\ell_q+\varepsilon_q} = \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} \varphi_{\ell + \varepsilon}.$$ 

Thus, we get:

$$f \ast g = \sum_{\ell \in \mathbb{N}^d} (a \ast b)_{\ell} \left[ 2^{-d/2} \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} \varphi_{\ell + \varepsilon} \right] = \sum_{\ell \in \mathbb{N}^d} \left[ 2^{-d/2} \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} (a \ast b)_{\ell - \varepsilon} \right] \varphi_{\ell}.$$ 

Since $h = f \ast g$, by uniqueness of the Laguerre coefficients of $h$, we obtain the desired relation. \qed
Proof of Proposition 3.2 First, we notice that $\forall \ell$, $G_{\ell \ell} = E[e^{-Y_i^{(1)} + \cdots + Y_i^{(d)}}] > 0$. We proceed by induction on $|k|$.

- If $|k| = 0$, then $k = 0$ and $a_0 = (G_{0,0})^{-1} c_0$.
- Let $r \in \mathbb{N}$, we suppose \(\square\) is true for every $k \in \mathbb{N}^d$ such that $|k| \leq r$. Let $k \in \mathbb{N}^d$ with $|k| = r + 1$. From (8),
  \[
  c_k = \sum_{\ell \leq k \atop \ell \neq k} G_{k \ell} a_\ell + G_{kk} a_k.
  \]
  If $\ell \leq k$ with $\ell \neq k$ then $|\ell| < |k|$, so we can use the induction assumption:
  \[
  a_k = (G_{kk})^{-1} \left( c_k - \sum_{\ell \leq k \atop \ell \neq k} G_{k \ell} \sum_{j \leq \ell} H_{\ell j} c_j \right).
  \]
  Thus, by setting $H_{kk} := (G_{kk})^{-1}$ and $H_{kj} := (G_{kk})^{-1} \sum_{j \leq k, j \neq k} G_{k \ell} H_{\ell j}$ for every $j \leq k, j \neq k$, we’ve just proved (9) for all $k$ such that $|k| = r + 1$. \(\square\)

Proof of Proposition 4.1 By Pythagorean theorem, $\|f - \hat{f}_m\|_{L^2}^2 = \|f - f_m\|_{L^2}^2 + \|f_m - \hat{f}_m\|_{L^2}^2$.

We decompose the second term on the Laguerre basis:

\[
\|f_m - \hat{f}_m\|_{L^2}^2 = \sum_{k \leq m-1} (a_k - \hat{a}_k)^2 = \|\mathbf{a}_m - \hat{\mathbf{a}}_m\|_{\mathbb{R}^m}^2.
\]

We now give an upper bound on the last quantity in two different ways.

1. The first way is a bound using the spectral norm.

\[
E\|f_m - \hat{f}_m\|_{L^2}^2 = E\|G_m^{-1}(c_m - \hat{c}_m)\|_{\mathbb{R}^m}^2 \\
\leq \rho^2(G_m^{-1}) E\|c_m - \hat{c}_m\|_{\mathbb{R}^m}^2 \\
= \rho^2(G_m^{-1}) E \left[ \sum_{k \leq m-1} \left( \frac{1}{n} \sum_{i=1}^{n} \varphi_k(Z_i) - E[\varphi_k(Z_1)] \right) \right]^2 \\
= \frac{\rho^2(G_m^{-1})}{n} \sum_{k \leq m-1} \text{Var} (\varphi_k(Z_1)) \leq \frac{\rho^2(G_m^{-1})}{n} \sum_{k \leq m-1} E[\varphi_k(Z_1)^2].
\]

Yet for every $x \in \mathbb{R}_+^d$, $\varphi_k(x)^2 \leq 2^d$ and $\text{card}\{k \in \mathbb{N}^d \atop k \leq m-1\} = \prod_{q=1}^{d} m_q$, so:

\[
E\|f_m - \hat{f}_m\|_{L^2}^2 \leq \frac{\rho^2(G_m^{-1})}{n} \sum_{k \leq m-1} E[\varphi_k(Z_1)^2] \leq \frac{2^d (\prod_{q=1}^{d} m_q) \rho^2(G_m^{-1})}{n}.
\]
2. For the second way, we develop the square.

\[
E\|G_m^{-1}(c_m - \hat{c}_m)\|_{\mathbb{R}^n}^2 = E \left[ \sum_{\ell \leq m-1} \left( \sum_{k \leq m-1} |G_m^{-1}|_{\ell k}(c_k - \hat{c}_k) \right)^2 \right]
\]

\[
= E \left[ \sum_{\ell \leq m-1} \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} |G_m^{-1}|_{\ell k'} \text{Cov}(\hat{c}_k, \hat{c}_{k'}) \right]
\]

\[
= \sum_{\ell \leq m-1} \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} |G_m^{-1}|_{\ell k'} \frac{\text{Cov}(\varphi_k(Z_1), \varphi_{k'}(Z_1))}{n}
\]

\[
= \frac{1}{n} \sum_{\ell \leq m-1} \text{Var} \left( \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} \varphi_k(Z_1) \right).
\]

We control the variance by the expectation of the square:

\[
\text{Var} \left( \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} \varphi_k(Z_1) \right) \leq E \left[ \left( \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} \varphi_k(Z_1) \right)^2 \right]
\]

\[
= \int_{\mathbb{R}^+} \left( \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} \varphi_k(z) \right)^2 h(z) \, dz
\]

\[
\leq \|h\|_{\infty} \left\| \sum_{k \leq m-1} |G_m^{-1}|_{\ell k} \varphi_k \right\|_{L^2}^2 = \|h\|_{\infty} \sum_{k \leq m-1} |G_m^{-1}|_{\ell k}^2.
\]

In the end, we obtain:

\[
E\|f_m - \hat{f}_m\|_{L^2}^2 \leq \frac{\|h\|_{\infty}}{n} \sum_{\ell \leq m-1} \sum_{k \leq m-1} |G_m^{-1}|_{\ell k}^2 = \frac{||h||_{\infty} \|G_m^{-1}\|_{L^2}^2}{n}.
\]

To prove Proposition 4.3, we will use the following lemma about the order of magnitude of the functions \(\varphi_k\) for \(k \in \mathbb{N}\).

**Lemma 7.1** *(Askey and Wainger 1965)*. There exists \(k_0 \in \mathbb{N}\) and \(C > 0\) such that for every \(k \geq k_0\),

\[
\left| \varphi_k \left( \frac{x}{2} \right) \right| \leq C \begin{cases} 
1 & \text{if } 0 \leq x \leq \frac{1}{k}, \\
k^{-\frac{1}{2}}x^{-\frac{1}{2}} & \text{if } \frac{1}{k} \leq x \leq \delta k, \\
k^{-\frac{1}{2}}(\nu - x)^{-\frac{1}{2}} & \text{if } \delta k \leq x \leq \nu - \nu^\frac{1}{2}, \\
k^{-\frac{1}{2}} & \text{if } \nu - \nu^\frac{1}{2} \leq x \leq \nu + \nu^\frac{1}{2}, \\
k^{-\frac{1}{2}}(x - \nu)^{-\frac{1}{2}} \exp \left( -\eta(x - \nu)^{\frac{3}{2}} \nu^{-\frac{1}{2}} \right) & \text{if } \nu + \nu^\frac{1}{2} \leq x \leq (1 + \lambda) \nu, \\
\exp(\xi x) & \text{if } (1 + \lambda) \nu \leq x,
\end{cases}
\]
where $\nu := 4k + 2$, where $\delta$ et $\lambda$ are small enough positive constants, and where $\eta$ and $\xi$ are fixed positive constants.

**Proof of Proposition 4.4.** Following the proof of Proposition 4.1, we see we need to improve the upper bound in (20). Using Lemma 7.1, we have for all $k \in \mathbb{N}$,

$$|\varphi_k(x)|^2 \leq C \begin{cases} 1 & \text{if } 0 \leq 2x \leq \frac{1}{k}, \\ 1/\sqrt{kx} & \text{if } \frac{1}{k} \leq 2x \leq \delta k, \\ R_k & \text{if } 2x \geq \delta k. \end{cases}$$

where $R_k = o(1/\sqrt{k})$ does not depend on $x$. Since “$0 \leq x \leq \frac{1}{k}$” is equivalent to “$1 \leq \frac{1}{\sqrt{kx}}$”, we get $|\varphi_k(x)|^2 \leq C(\frac{1}{\sqrt{kx}} + R_k)$. Thus for $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$, we get:

$$E[\varphi_k(Z)^2] = E\left[\prod_{j=1}^d \varphi_{k_j}(Z^{(j)})^2\right] \leq E\left[\prod_{j=1}^d C\left(\frac{1}{\sqrt{k_jZ^{(j)}}} + R_{k_j}\right)\right]$$

$$\leq E\left[\prod_{j=1}^d C\left(\frac{1}{\sqrt{k_jY^{(j)}}} + R_{k_j}\right)\right]$$

$$= \frac{C^d}{\sqrt{k_1 \cdots k_d}} E\left[\frac{1}{\sqrt{Y^{(1)} \cdots Y^{(d)}}}\right] + \tilde{C} o\left(\frac{1}{\sqrt{k_1 \cdots k_d}}\right),$$

where $\tilde{C}$ depends on $M_J$ for $J \subset \{1, \ldots, d\}$. Therefore:

$$\sum_{k \leq m-1} E[\varphi_k(Z)^2] \leq c(g)\sqrt{D_m},$$

with $c(g)$ a positive constant depending on $M_J$ for $J \subset \{1, \ldots, d\}$. 

**Proof of Theorem 4.11.** Using Proposition 4.1, the bound (12) on the bias term, and Proposition 4.7 if $f$ belongs to $W^s(\mathbb{R}_+^d, L)$, we have:

$$\forall m \in (\mathbb{N}^*)^d, \quad E\|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{i=1}^d m_i^{-s_i} + \frac{\|h\|_\infty}{n} \prod_{i=1}^d m_i^{2s_i}.$$  

By Remark 4.2, $\|h\|_\infty$ is controlled by $\|f\|_{L^2} \|g\|_{L^2}$. Because $f \in W^s(\mathbb{R}_+^d, L)$, we have $\|f\|_{L^2}^2 \leq L$, so $\|h\|_\infty \leq \sqrt{L} \|g\|_{L^2}$. Thus,

$$\forall m \in (\mathbb{N}^*)^d, \quad E\|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{i=1}^d m_i^{-s_i} + \frac{\sqrt{L} \|g\|_{L^2}^2}{n} \prod_{i=1}^d m_i^{2s_i} =: \psi_n(m).$$

Let $m_{\text{opt}}$ minimizing $\psi_n$, then the gradient of $\psi_n$ vanishes on $m_{\text{opt}}$, so we have:

$$\forall j \in \{1, \ldots, d\}, \quad \frac{L s_j}{2\alpha_j} m_{\text{opt}, j}^{-s_j} = \frac{\sqrt{L} \|g\|_{L^2}}{n} \prod_{i=1}^d m_{\text{opt}, i}^{2s_i}.$$  

(21)

Therefore, the coordinates of $m_{\text{opt}}$ must satisfy $\frac{s_j}{2\alpha_j} m_{\text{opt}, i}^{-s_i} = \frac{s_j}{2\alpha_j} m_{\text{opt}, j}^{-s_j}$ for every $i, j$ in $\{1, \ldots, d\}$. Using these relations in (21), we obtain:

$$\forall j \in \{1, \ldots, d\}, \quad m_{\text{opt}, j} = C_j(s, L, g) n^{1/(s_j + s_j \sum_{i=1}^d \frac{\alpha_i}{s_i})},$$
where $C_j(s, L, g)$ are constants depending on $s$, $L$ and $g$. The minimum value of $\psi_n$ is then:

$$\psi_n(m_{\text{opt}}) = C(s, L, g) n^{-1/(1 + \sum_{i=1}^{d} 2\alpha_i)},$$

where $C(s, L, g)$ is a constant depending in $s$, $L$ and $g$.

Now, if $f$ belongs to $\mathcal{S}'(\mathbb{R}^d, L)$, we use the bound on the bias term:

$$\forall m \in (\mathbb{N}^*)^d, \quad E\|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{i=1}^{d} c^{-r_i m_i} + \frac{\sqrt{L} \|g\|_\infty}{n} \prod_{i=1}^{d} m_i^{2\alpha_i} =: \phi_n(m).$$

We minimize the function $\phi_n(m)$ as we did above, and we find the following relations:

$$\forall j \in \{1, \ldots, d\}, \quad r_j m_{\text{opt},j} e^{-r_j m_{\text{opt},j}} = C_j(r, L, g) \frac{(r_j m_{\text{opt},j}) \sum_{i=1}^{d} 2\alpha_i}{n},$$

where $C_j(r, L, g)$ are constants depending on $r$, $L$ and $g$. Taking the log, we find:

$$r_j m_{\text{opt},j} + \left( \sum_{i=1}^{d} 2\alpha_i - 1 \right) \log(r_j m_{\text{opt},j}) = \log n - \log C_j(r, L, g).$$

Thus, when $n$ goes to $+\infty$, we have $r_j m_{\text{opt},j} \sim \log n$. Taking $m_{\text{opt}} \in (\mathbb{N}^*)^d$ such that $m_{\text{opt},j} \propto \frac{\log n}{r_j}$, we find:

$$\phi_n(m_{\text{opt}}) \leq C(r, L, g) \frac{(\log n) \sum_{i=1}^{d} 2\alpha_i}{n},$$

where $C(r, L, g)$ is a constant depending on $r$, $L$ and $g$. \hfill \Box

7.2. Proposition 4.17

7.2.1. Preliminary results

To prove this proposition, we first need to extend the theory of Toeplitz matrices to hypermatrices. For more details about Toeplitz matrices, see Böttcher and Grudsky (2005). We say that $T$ is an infinite lower triangular Toeplitz hypermatrix if it is lower triangular according to Definition 2.5 and if the value of $T_{\ell k}$ depends only on the difference $\ell - k$:

$$T = [T_{\ell k}]_{\ell, k \in \mathbb{N}^*}, \quad T_{\ell k} = a_{\ell - k},$$

with $a = [a_k]_{k \in \mathbb{Z}^d}$ and $a_k = 0$ if $k \notin \mathbb{N}^d$. So there is a bijection that takes $a \in \mathbb{R}^{\mathbb{N}^d}$ and returns the corresponding Toeplitz hypermatrix $T(a)$. We can see $T(a)$ as a linear map on $\mathbb{R}^{\mathbb{N}^d}$:

$$x = [x_k]_{k \in \mathbb{N}^d} \mapsto \left[ \sum_{k \in \mathbb{N}^d} T_{\ell k} x_k \right]_{\ell \in \mathbb{N}^d} = \left[ \sum_{k \leq \ell} a_{\ell - k} x_k \right]_{\ell \in \mathbb{N}^d} = a * x.$$ 

The associativity of the convolution product gives that for every $a$ and $b$, we have $T(a) \times T(b) = T(a * b)$.

**Notation.** If $m \in (\mathbb{N}^*)^d$, we denote by $T_m(a) \in \mathbb{R}^{(m, m)}$ the sub-hypermatrix of $T(a)$ constructed by taking only the coefficients $[T(a)]_{\ell, k}$ for $\ell, k \leq m - 1$. Note that as a linear map on $\mathbb{R}^m$, we have $T_m(a) = T(a)|_{\mathbb{R}^m}$ because of the triangular structure.
**Operator on \( \ell^p(\mathbb{N}^d) \).** If \( a \in \ell^1(\mathbb{N}^d) \) and if \( x \) is in \( \ell^p(\mathbb{N}^d) \) with \( p \in [1, +\infty] \), then \( a \ast x \) belongs to \( \ell^p(\mathbb{N}^d) \) we have \( \|a \ast x\|_p \leq \|a\|_1 \|x\|_p \). In other words, \( T(a) \) is an operator from \( \ell^p(\mathbb{N}^d) \) to \( \ell^p(\mathbb{N}^d) \) and its operator norm is bounded by \( \|a\|_1 \).

The \( \ell^2 \) case. We will need to study \( T(a) \) as an operator on \( \ell^2(\mathbb{N}^d) \) and give a bound on its norm with milder assumptions than \( a \in \ell^1(\mathbb{N}^d) \). We define a subspace of \( L^2(\mathbb{T}^d) \) that plays an important role in this matter.

**Definition 7.2** (Hardy space). We define the Hardy space as the following subset of \( L^2(\mathbb{T}^d) \):

\[
H^2(\mathbb{T}^d) := \{ u \in L^2(\mathbb{T}^d) \mid \forall k \notin \mathbb{N}^d, c_k(u) = 0 \},
\]

where \( c_k(u) \) denotes the \( k \)-th Fourier coefficient of \( u \) and \( \mathbb{T} \) is the set of complex numbers with unitary module.

The map that takes a function and returns its Fourier coefficient is then an isometric bijection between \( H^2(\mathbb{T}^d) \) and \( \ell^2(\mathbb{N}^d) \).

\[
\mathcal{F}: \quad H^2(\mathbb{T}^d) \quad \mapsto \quad \ell^2(\mathbb{N}^d)
\]

\[
f \quad \mapsto \quad [c_k(f)]_{k \in \mathbb{N}^d}
\]

We see that if \( u \) and \( v \) are \( H^2(\mathbb{T}^d) \) functions, the identity \( \mathcal{F}[u] \ast \mathcal{F}[v] = \mathcal{F}[u \times v] \) translates for Toeplitz hypermatrices into \( T(\mathcal{F}[u]) \times T(\mathcal{F}[v]) = T(\mathcal{F}[u \times v]) \).

Under the additional assumption that \( \mathcal{F}^{-1}[a] \) belongs to \( L^\infty(\mathbb{T}^d) \), we show in the next proposition that \( T(a) \) defines an operator from \( \ell^2(\mathbb{N}^d) \) to \( \ell^2(\mathbb{N}^d) \).

**Proposition 7.3.** Let \( a \in \ell^2(\mathbb{N}^d) \) such that \( u := \mathcal{F}^{-1}[a] \in L^\infty(\mathbb{T}^d) \). Then \( T(a) \) defines an operator on \( \ell^2(\mathbb{N}^d) \) and its operator norm is bounded by \( \|u\|_{\infty} \).

**Proof.** Let \( x \in \ell^2(\mathbb{N}^d) \) and let \( v := \mathcal{F}^{-1}[x] \). Then:

\[
\|T(a)x\|_\ell^2 = \|\mathcal{F}[u] \ast \mathcal{F}[v]\|_\ell^2 = \|\mathcal{F}[u \times v]\|_\ell^2 = \|u \times v\|_{L^2} \leq \|u\|_{\infty} \|v\|_{L^2} = \|u\|_{\infty} \|x\|_{\ell^2}. \quad \Box
\]

Before we can prove Proposition 7.3, some facts need to be established. We denote by \( \hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\} \) the Riemann sphere. We will use the following functions:

- We make the assumption that \( \beta \) belongs to \( \ell^1(\mathbb{N}^d) \) so that the power series \( B(z) := \sum_{k \in \mathbb{N}^d} \beta_k z^k \) is normally convergent on \( \overline{\mathbb{D}}^d \) and defines a function which is continuous on \( \overline{\mathbb{D}}^d \) and holomorphic on \( \mathbb{D}^d \).

- We denote by \( G \) the Laplace transform of \( g \). This function is defined on \( \mathbb{P}^d_+ \), continuous on \( \mathbb{P}^d_+ \), and holomorphic on \{ \( s \in \mathbb{C}^d \mid \forall q, \Re(s_q) > 0 \} \).

- If \( u \) is a \( L^1(\mathbb{R}^d_+) \) or \( L^2(\mathbb{R}^d_+) \) function, we denote by:

\[
u^*(\omega) := \int_{\mathbb{R}^d_+} e^{i\omega \cdot x} u(x) \, dx, \quad \omega \in \mathbb{R}^d_+,\]

its Fourier transform, where \( \omega \cdot x := \sum \omega_i x_i \). The Fourier transform of the Laguerre functions \( \varphi_k, k \in \mathbb{N}^d \) can be computed from the case \( d = 1 \):

\[
\varphi_k^*(\omega) = \prod_{q=1}^d \varphi_{k_q}^*(\omega_q) = \prod_{q=1}^d (-1)^{k_q} \sqrt{2 \frac{(1 + i\omega_q)^{k_q}}{(1 - i\omega_q)^{k_q + 1}}}.
\]
We will need to understand the behaviour of the map $z \mapsto \frac{1+z}{1-z}$.

**Lemma 7.4.** Let $\eta: \hat{C} \to \hat{C}$ be the homographic function $\eta(s) := \frac{s-1}{s+1}$ (with $\eta(\infty) = 1$ and $\eta(-1) = \infty$).

1. The function $\eta$ is invertible and its inverse is $\eta^{-1}(z) = \frac{1+z}{1-z}$.
2. The image of $\{ s \in \mathbb{C} \mid \Re(s) > 0 \}$ by $\eta$ is $\mathbb{D}$.
3. The image of $\{ s \in \mathbb{C} \mid \Re(s) = 0 \}$ by $\eta$ is $\mathbb{T} \setminus \{1\}$.

In the next proposition that generalizes lemma C.1 from Comte et al. (2017), we show that the functions $G$ and $B$ are linked through the transformation $\eta$.

**Proposition 7.5.** If $\theta \in \mathbb{R}^d$, we denote $e^{i\theta} := (e^{i\theta_1}, \ldots, e^{i\theta_d})$. If $\beta \in \ell^1(\mathbb{N}^d)$, then $[\beta_k]_{k \in \mathbb{N}^d}$ are the Fourier coefficients of $\theta \mapsto G(\frac{1+e^{i\theta}}{1-e^{i\theta}})$ and we have:

$$\forall \theta \in \mathbb{R}^d, \ G\left(\frac{1+e^{i\theta}}{1-e^{i\theta}}\right) = \sum_{k \in \mathbb{N}^d} \beta_k e^{ik \cdot \theta} = B(e^{i\theta}),$$

with normal convergence of the series.

**Remark 7.6.** If $g \in W^s(\mathbb{R}^d_+, L)$ with $s \in (1, +\infty)^d$, then $\beta$ is $\ell^1(\mathbb{N}^d)$, see Remark 4.8.

**Proof of Proposition 7.5.** We start from the expression of $\beta_k$ and use the Plancherel isometry:

$$\beta_k = 2^{-d/2} \sum_{k \leq 1} (-1)^{|\varepsilon|} b_{k-\varepsilon} = 2^{-d/2} \sum_{k \leq 1} (-1)^{|\varepsilon|} \langle g, \varphi_{k-\varepsilon} \rangle_{L^2}$$

$$= 2^{-d/2} \sum_{k \leq 1} (-1)^{|\varepsilon|} \frac{1}{(2\pi)^d} \langle g^*, \varphi_{k-\varepsilon} \rangle_{L^2} = \langle g^*, \frac{2^{-d/2}}{(2\pi)^d} \sum_{k \leq 1} (-1)^{|\varepsilon|} \varphi_{k-\varepsilon} \rangle_{L^2}.$$

Let us compute the second factor in the scalar product. The reader is referred to Section 2 for the notations governing vectors and multi-indices computation.

$$\frac{2^{-d/2}}{(2\pi)^d} \sum_{\varepsilon \leq 1} (-1)^{|\varepsilon|} \varphi_{k-\varepsilon}(\omega) = \frac{2^{-d/2}}{(2\pi)^d} \sum_{\varepsilon \leq 1} (-1)^{|\varepsilon|} \prod_{q=1}^d (-1)^{k_q - \varepsilon_q} \sqrt{2} e^{i\omega_q k_q - \varepsilon_q (1 - i\omega_q) k_q - \varepsilon_q + 1}$$

$$= (-1)^{|k|} \sum_{\varepsilon \leq 1} \left( \frac{1 + i\omega}{1 - i\omega} \right)^{k - \varepsilon} \left( \frac{1}{1 - i\omega} \right)^{\varepsilon}.$$

Using the multibinomial theorem (Proposition 2.1),

$$= \left( \frac{i\omega + 1}{i\omega - 1} \right)^k \frac{1}{(2\pi)^d} \left( \frac{1 - i\omega}{1 + i\omega} + 1 \right)^1 \left( \frac{1}{1 - i\omega} \right)^1$$

$$= \left( \frac{i\omega + 1}{i\omega - 1} \right)^k \frac{1}{\pi^d} \prod_{q=1}^d \frac{1}{1 + \omega_q^2}.$$

We derive the following expression:

$$\beta_k = \frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\omega) \left( \frac{i\omega - 1}{i\omega + 1} \right)^k \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)}.$$
The change of variable $e^{-i\theta} = \frac{i\omega_q - 1}{i\omega_q + 1} = \eta(i\omega_q)$ gives then:

$$
\beta_k = \frac{1}{(2\pi)^d} \int_{[0,2\pi]^d} g^* \left( \frac{e^{i\theta} + 1}{i(e^{i\theta} - 1)} \right) e^{-ik\cdot\theta} \, d\theta
= \frac{1}{(2\pi)^d} \int_{[0,2\pi]^d} G \left( \frac{1 + e^{i\theta}}{1 - e^{i\theta}} \right) e^{-ik\cdot\theta} \, d\theta.
$$

(22)

Therefore, if $k \in \mathbb{N}^d$, the $k$-th Fourier coefficient of the function $\theta \mapsto G(\frac{1 + e^{i\theta}}{1 - e^{i\theta}})$ is $\beta_k$. On the other hand, if $k \notin \mathbb{N}^d$, let us show that the Fourier coefficients vanish. Without loss of generality, we assume that $k_1 < 0$, and we compute the $k$-th Fourier coefficient:

$$
\frac{1}{(2\pi)^d} \int_{[0,2\pi]^d} G \left( \frac{1 + e^{i\theta}}{1 - e^{i\theta}} \right) e^{-ik\cdot\theta} \, d\theta
= \frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\omega) \left( \frac{i\omega_1 - 1}{i\omega_1 + 1} \right)^{k_1} \prod_{q=2}^{d} \left( \frac{i\omega_q - 1}{i\omega_q + 1} \right)^{k_q} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)}
= \frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\omega) \left( \frac{i(-\omega_1) - 1}{i(-\omega_1) + 1} \right)^{-k_1} \prod_{q=2}^{d} \left( \frac{i\omega_q - 1}{i\omega_q + 1} \right)^{k_q} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)}
= \frac{-1}{\pi^d} \int_{\mathbb{R}^d} g^*(-\omega_1, \omega_2, \ldots, \omega_d) \left( \frac{i\omega - 1}{i\omega + 1} \right)^{k'} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)},
$$

where $k' := (-k_1, k_2, \ldots, k_d)$. Let $\delta$ be the map on $\mathbb{R}^d$ defined by $\delta(\omega) := (-\omega_1, \omega_2, \ldots, \omega_d)$, then we have:

$$
-\frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\delta(\omega)) \left( \frac{i\omega - 1}{i\omega + 1} \right)^{k'} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)} = \left\langle -g^* \circ \delta, \frac{2^{-d/2}}{(2\pi)^d} \sum_{\epsilon \leq 1} (-1)^{|\epsilon|} \varphi_{k'-\epsilon} \right\rangle_{L^2}
= 2^{-d/2} \sum_{k \leq 1} (-1)^{|\epsilon|} \frac{1}{(2\pi)^d} \left\langle -g^* \circ \delta, \varphi_{k'-\epsilon}^* \right\rangle_{L^2}
= 2^{-d/2} \sum_{k \leq 1} (-1)^{|\epsilon|} \left\langle g \circ \delta, \varphi_{k'-\epsilon} \right\rangle_{L^2},
$$

because $(g \circ \delta)^* = -g^* \circ \delta$. This last expression is always zero:

$$
\left\langle g \circ \delta, \varphi_{k'-\epsilon} \right\rangle_{L^2} = \int_{\mathbb{R}^d} g(-x_1, x_2, \ldots, x_d) \varphi_{k'-\epsilon}(x) \, dx = 0,
$$

since the function $g$ is zero of $x_1 > 0$ and $\varphi_{k'-\epsilon}$ is zero if $x_1 < 0$. Thus, the Fourier coefficients of $\theta \mapsto G(\frac{1 + e^{i\theta}}{1 - e^{i\theta}})$ are $[\beta_k]_{k \in \mathbb{N}^d}$.

The function $G$ is continuous on $(i\mathbb{R} \cup \{\infty\})^d$, so the function $G \circ (\eta^{-1}) \otimes^d$ is continuous on $\mathbb{T}^d$ by Lemma 7.4. Therefore, since the Fourier series $\theta \mapsto G(\frac{1 + e^{i\theta}}{1 - e^{i\theta}})$ is normally convergent, this function is equal to its Fourier series at each point:

$$
\forall \theta \in \mathbb{R}^d, \quad G \left( \frac{1 + e^{i\theta}}{1 - e^{i\theta}} \right) = \sum_{k \in \mathbb{N}^d} \beta_k \, e^{ik\cdot\theta}.
$$

We need a last technical lemma before we start the proof of Proposition 4.7.
Lemma 7.7. Let $\alpha \in (\mathbb{N}^*)^d$, then $(1 - z)^{-\alpha}$ admits a power series expansion on $\mathbb{D}^d$ given by:

$$\forall z \in \mathbb{D}^d, \quad (1 - z)^{-\alpha} = \sum_{j \in \mathbb{N}^d} \left( \frac{\alpha - 1 + j}{j} \right) z^j.$$  

Moreover, denoting $\zeta_j$ the $j$-th coefficient in the power series above, for $m \geq 4$ we have $\|T_m(\zeta)\|_F^2 \leq m^{2\alpha}$ where $T(\zeta)$ is the Toeplitz hypermatrix contructed from the coefficients $\zeta_j$.

Proof of Lemma 7.7. We recall the following identity: for $z \in \mathbb{D}$ and $r \in \mathbb{N}^*$, we have $(1 - z)^{-r} = \sum_{j=0}^{\infty} \binom{r - 1 + j}{j} z^j$. Thus, for $z \in \mathbb{D}^d$ and $\alpha \in (\mathbb{N}^*)^d$, we have:

$$(1 - z)^{-\alpha} = \prod_{q=1}^{d} (1 - z_q)^{-\alpha_q} = \prod_{q=1}^{d} \sum_{j_q=0}^{\infty} \left( \frac{\alpha_q - 1 + j_q}{j_q} \right) z_q^{j_q} = \sum_{j \in \mathbb{N}^d} \left( \frac{\alpha - 1 + j}{j} \right) z^j.$$  

Therefore,

$$\|T_m(\zeta)\|_F^2 = \sum_{j \leq m - 1} \left( \frac{\alpha - 1 + j}{j} \right)^2 \# \{(k, \ell) \in \mathbb{N}^d \times \mathbb{N}^d | k \leq \ell \leq m - 1, \ell - k = j\}$$

$$= \prod_{q=1}^{d} \sum_{j_q=0}^{\infty - 1} \left( \frac{\alpha - 1 + j_q}{j_q} \right)^2 (m_q - j_q),$$

and it reduces to the case $d = 1$ which was already solved in [Comte et al., 2017, appendix C]. So if $m_q \geq 4$ for every $q$, then $\|T_m(\zeta)\|_F^2 \leq \prod_{q=1}^{d} m_q^{2\alpha_q} = m^{2\alpha}$. \qed

7.2.2. Proof of Proposition 4.7.

From Proposition 7.5 we get $G = B \circ \eta^\otimes d$. We define a function $w$ on $(\mathbb{D} \setminus \{1\})^d$ by:

$$\forall z \in (\mathbb{D} \setminus \{1\})^d, \quad w(z) := (1 - z)^{-\alpha} B(z).$$

This function is related to $K_\alpha$ by the identity:

$$\forall z \in (\mathbb{D} \setminus \{1\})^d, \quad w(z) = 2^{-\alpha} K_\alpha((\eta^{-1})^\otimes d(z)).$$

Thus, the function $w$ can be extended as a function on $\mathbb{D}^d$, still denoted $w$, and according to our assumptions on $K_\alpha$ and Lemma 7.4, $w$ satisfies:

- $w|_{\mathbb{T}^d}$ is continuous;
- $w$ is continuous on $(\mathbb{D} \setminus \{1\})^d$;
- $w$ is holomorphic on $\mathbb{D}^d$;
- $w$ doesn’t vanish on $\mathbb{D}^d$.

Thus, the function $w^{-1} = 1/w$ is well defined on $\mathbb{D}^d$ and has the same properties. In particular, since it is holomorphic on $\mathbb{D}^d$, it admits a power series expansion:

$$\forall z \in \mathbb{D}^d, \quad w^{-1}(z) = \sum_{k \in \mathbb{N}^d} d_k z^k.$$
Let us have look on what is happening on $\mathbb{T}^d$. For $r \in [0, 1[$, we denote $w_{r}^{-1}$ the function on $\mathbb{T}^d$ defined by:

$$\forall t \in \mathbb{T}^d, \quad w_{r}^{-1} = w^{-1}(rt) = \sum_{k \in \mathbb{N}^d} d_k r^{|k|} t^k.$$  

On the one hand, the Fourier coefficients of $w_{r}^{-1}$ are $d_k r^{|k|}$ (we set $d_k = 0$ if $k \not\in \mathbb{N}^d$). On the other hand, we can compute the Fourier coefficients and we get:

$$\forall k \in \mathbb{Z}^d, \quad d_k r^{|k|} = \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} w^{-1}(re^{i\theta}) e^{-ik \cdot \theta} \, d\theta.$$  

Since $w^{-1}$ is continuous on $(\mathbb{T} \setminus \{1\})^d$, we have $w_{r}^{-1} \to w_{|\mathbb{T}^d|}^{-1}$ a.e. on $\mathbb{T}^d$. By dominated convergence, we obtain:

$$\forall k \in \mathbb{Z}^d, \quad d_k = \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} w^{-1}(e^{i\theta}) e^{-ik \cdot \theta} \, d\theta.$$  

Therefore, $(d_k)$ are the Fourier coefficients of $w_{|\mathbb{T}^d|}^{-1}$. Thus, we have shown that $w_{|\mathbb{T}^d|}^{-1} \in H^2(\mathbb{T}^d)$ and:

$$w^{-1}(t) = \sum_{k \in \mathbb{N}^d} d_k t^k,$$

with $L^2(\mathbb{T}^d)$-convergence of the series.

By Lemma 7.7, the function $(1 - z)^{-\alpha}$ admits a power series expansion on $\mathbb{D}$ given by $(1 - z)^{-\alpha} = \sum_{k \in \mathbb{N}^d} \zeta_k z^k$. Thus, the power series equality $B(z) \times (1 - z)^{-\alpha} \times w^{-1}(z) = 1$ on the domain $\mathbb{D}$ translates to their coefficients into the relation $\beta \ast \zeta \ast d = \delta_0$, where $\delta_0$ is the element of $\mathbb{R}^{N^d}$ defined by:

$$[\delta_0]_k = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{else}. \end{cases}$$

Taking the corresponding Toeplitz hypermatrices, we get $G \times T(\zeta) \times T(d) = I_{N^d}$, where $I_{N^d}$ is the infinite hypermatrix given by:

$$[I_{N^d}]_{\ell k} = \begin{cases} 1 & \text{if } \ell = k, \\ 0 & \text{else}. \end{cases}$$

Thus, for $m \in \mathbb{N}^*$, we get $G_m^{-1} = T_m(\zeta) \times T_m(d)$. Taking the Frobenius norm, we obtain the following inequality:

$$\|G_m^{-1}\|_F^2 \leq \|T_m(\zeta)\|_F^2 \times \rho^2(T_m(d)).$$

From Proposition 7.3 we have:

$$\rho^2(T_m(d)) = \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|T_m(d)x\|_F^2}{\|x\|_F^2} = \sup_{x \in \mathbb{C}^m \setminus \{0\}} \frac{\|T_m(d)x\|_F^2}{\|x\|_F^2} = \|w_{|\mathbb{T}^d|}^{-1}\|_\infty^2,$$

and by Lemma 7.7 we have $\|T_m(\zeta)\|_F^2 \leq m^{2\alpha}$ if $m \geq 4$. \qed

### 7.3. Proofs of Section 5

**Proof of Theorem 5.2** The proof is identical to the one in Mabon (2017), but we use our Lemma 5.1 instead of Mabon’s proposition 7.1. \qed
Proof of Lemma 5.1 We will prove that \( \| \hat{f}_m - f_m \|_{L^2}^2 \) can be written as the supremum of an empirical process, then we will use Talagrand’s inequality (Theorem A.1). Let \( m \in (\mathbb{N}^*)^d \), for every \( t \in S_m \) we set:

\[
G_m^{-1}(t) := \sum_{k \leq m-1} [G_m^{-1} \times_d t_m]_k \varphi_k,
\]

where \( t_m \in \mathbb{R}^m \) is the matrix of Laguerre coefficients of \( t \). We define the empirical process \( \nu_n \) on the space \( S_{m^*} \) by:

\[
\nu_n(t) := \frac{1}{n} \sum_{i=1}^n (G_m^{-1}(t)(Z_i) - E[G_m^{-1}(t)(Z_i)]).
\]

Note that \( S_{m^*} \) contains every \( S_m \) as a subspace for \( m \in \mathcal{M}_n \) and because of the triangular structure of \( G^{-1} \), we have for every \( t \in S_m \), \( G_m^{-1}(t) = G_m^{-1}(t) \). For \( m \in \mathcal{M}_n \), we have then:

\[
\| \hat{f}_m - f_m \|_{L^2} = \sum_{k \leq m-1} (\hat{a}_k - a_k)^2 = \sum_{k \leq m-1} \nu_n^2(\varphi_k).
\]

The map \( t \mapsto \nu_n(t) \) being linear on \( S_m \) and \( (\varphi_k)_{k \leq m-1} \) being a basis of \( S_m \), we have by Cauchy–Schwarz inequality:

\[
\sum_{k \leq m-1} \nu_n^2(\varphi_k) = \sup_{t \in B_m} \nu_n^2(t), \quad \text{where } B_m := \{ t \in S_m \mid \| t \|_{L^2} = 1 \}.
\]

Thus, we have:

\[
\| \hat{f}_m - f_m \|_{L^2}^2 = \sup_{t \in B_m} \nu_n^2(t).
\]

We apply Talagrand’s inequality\(^1\) We need the constants \( M, H \) and \( \nu \) to apply this inequality.

- Computation of \( M \):

\[
\sup_{t \in B_m} \sup_{x \in \mathbb{R}^d} |G_m^{-1}(t)(x)| = \sup_{t \in B_m} \sup_{x \in \mathbb{R}^d} \left| \sum_{k \leq m-1} [G_m^{-1} \times_d t_m]_k \varphi_k(x) \right|
\leq 2^{d/2} \sup_{t \in B_m} \sum_{k \leq m-1} \left| [G_m^{-1} \times_d t_m]_k \right|
\leq 2^{d/2} D_m \sup_{t \in B_m} \| G_m^{-1} \times_d t_m \|_{\mathbb{R}^d} = 2^{d/2} \sqrt{D_m} \rho(G_m^{-1}).
\]

Thus, \( M^2 := 2^d D_m \rho^2(G_m^{-1}) \).

- Computation of \( H \):

\[
E\left[ \sup_{t \in B_m} |\nu_n(t)| \right]^2 = E\left[ \| \hat{f}_m - f_m \|_{L^2}^2 \right] \leq E\left[ \| \hat{f}_m - f_m \|_{L^2}^2 \right]
\leq \frac{c(g) \sqrt{D_m} \rho^2(G_m^{-1})}{n} + \frac{\| g \|_\infty \| G_m^{-1} \|_F^2}{n},
\]

by Proposition 4.4 Thus, we take:

\[
H^2 := \frac{c(g) \sqrt{D_m} \rho^2(G_m^{-1})}{n} + \frac{\| g \|_\infty \| G_m^{-1} \|_F^2}{n}.
\]

\(^1\)this inequality concerns countable families of functions, but it’s not a problem here since \( \nu_n \) is continuous on \( S_{m^*} \) and \( B_m \) is separable.
We consider two cases.

1. \( c(g)\sqrt{D_m} \rho^2(G_m^{-1}) \leq (\|g\|_\infty \vee 1)\|G_m^{-1}\|_F^2 \) case, then we have:

\[
\frac{nH^2}{v} = \frac{c(g)}{\|g\|_\infty} \sqrt{D_m}, \quad \frac{nH}{M} = \frac{\sqrt{c(g)}}{2^{d/2}} \sqrt{nD_m^{-1/4}} \geq \frac{\sqrt{c(g)}}{2^{d/2}} n^{1/4},
\]

since \( D_m \leq n \). For \( \delta > 0 \), Talagrand’s inequality gives:

\[
E\left[ \left( \|f_m - f_m\|_{L^2}^2 - 2(1 + 2\delta)H^2 \right)^+ \right] \leq \frac{\|g\|_\infty \rho^2(G_m^{-1})}{n} \exp\left( -K\delta \frac{c(g)}{\|g\|_\infty} \sqrt{D_m} \right) + \frac{D_m \rho^2(G_m^{-1})}{C(\delta)^2 n^2} \exp\left( -\frac{KC(\delta)\sqrt{2}\sqrt{\sqrt{c(g)}}}{7} \frac{n^{1/4}}{2^{d/2}} \right).
\]

We control the first term using assumption (A3’):

\[
\sum_{m \in \mathcal{M}_n} \frac{\|g\|_\infty \rho^2(G_m^{-1})}{n} \exp\left( -K\delta \frac{c(g)}{\|g\|_\infty} \sqrt{D_m} \right) \leq \frac{\|g\|_\infty K(\delta)}{n},
\]

where \( K(\delta) \) is a constant not depending on \( n \). To control the second term, we use that \( D_m \rho^2(G_m^{-1}) \leq n \) when \( m \in \mathcal{M}_n \):

\[
\sum_{m \in \mathcal{M}_n} \frac{D_m \rho^2(G_m^{-1})}{C(\delta)^2 n^2} \exp\left( -\frac{KC(\delta)\sqrt{2}\sqrt{\sqrt{c(g)}}}{7} \frac{n^{1/4}}{2^{d/2}} \right) \leq \exp\left( -\tilde{K}(\delta) n^{1/4} \right) = o\left( \frac{1}{n} \right),
\]

so we have our result.

2. \( c(g)\sqrt{D_m} \rho^2(G_m^{-1}) \geq (\|g\|_\infty \vee 1)\|G_m^{-1}\|_F^2 \) case, then we have:

\[
\frac{nH^2}{v} = \frac{\|G_m^{-1}\|_F^2}{\rho^2(G_m^{-1})} \geq 1, \quad \frac{nH}{M} = \frac{\sqrt{n} (\|g\|_\infty \vee 1)^{1/2} \|G_m^{-1}\|_F}{2^{d/2} \sqrt{D_m} \rho(G_m^{-1})} \geq \frac{\sqrt{n}}{2^{d/2}} D_m^{-1/2}.
\]

Let us show that \( \rho(G_m^{-1}) \geq 1 \). We notice that \( \frac{2^{d/2}}{b_0} \) is an eigenvalue of \( G_m^{-1} \): indeed, if \( E_0 \) is the elementary hypermatrix with 1 in position \((0, \ldots, 0)\) and zeros elsewhere, then:

\[
G_m^{-1} \times_d E_0 = \frac{2^{d/2}}{b_0} E_0.
\]

Since the operator norm is bigger than the spectral radius, we have:

\[
\rho(G_m^{-1}) \geq \frac{2^{d/2}}{b_0} = \frac{1}{E\left[ e^{-Y_i^{(1)} - \ldots - Y_i^{(d)}} \right]} \geq 1.
\]
Thus, if $m \in \mathcal{M}_n$ we have $D_m \leq D_m \rho^2(\mathbf{G}_m^{-1}) \leq \frac{n}{\log n}$. Therefore, $\frac{nH}{\mathcal{M}} \geq \frac{\log n}{2^{d/2}}$. We apply Talagrand’s inequality with $\delta = a \log n$ and $a > 0$ to be chosen later:

$$E \left[ \left\| f - m - f' \right\|_{L^2}^2 - 2(1 + 2a \log n)H^2 \right] \leq$$

$$\frac{\|g\|_{\infty}^2 \rho^2(\mathbf{G}_m^{-1})}{n} \frac{1}{n^{1 + Ka}} + \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{n^2 \exp \left( - \frac{KC(a \log n) \sqrt{2a \log n}}{7} \right)} \times$$

$$+ \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{n^2 \exp \left( - \frac{K \sqrt{2a}}{7 \times 2^{d/2}} \log n \right)} .$$

We assume that $\frac{n}{\log n} \geq \exp(3/a)$, such that $C(a \log n) = 1$. Then:

$$E \left[ \left\| f - m - f' \right\|_{L^2}^2 - 2(1 + 2a \log n)H^2 \right] \leq \frac{\|g\|_{\infty}^2 \rho^2(\mathbf{G}_m^{-1})}{n^{1 + Ka}} \leq \frac{\|g\|_{\infty}^2}{n} ,$$

if $a \geq 2/K = 12$. To control the second term, we use that if $m \in \mathcal{M}_n$ then $D_m \rho^2(\mathbf{G}_m^{-1}) \leq n$:

$$\sum_{m \in \mathcal{M}_n} \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{n^2} \exp \left( - \frac{K \sqrt{2a}}{7 \times 2^{d/2}} \log n \right) \leq n - \frac{\sqrt{\frac{a}{7}}} {2^{d/2}} \leq \frac{1}{n}$$

if $a \geq \frac{2d}{2^{K+2}}$.

**Proof of Theorem 5.3** Let $m \in \mathcal{M}_n$,

$$\left\| f - m \right\|_{L^2}^2 \leq 3 \left\| f - m \right\|_{L^2}^2 + 3 \left\| f - m \right\|_{L^2}^2 + 3 \left\| f - m \right\|_{L^2}^2 .$$

By definition of $A(m)$ and $\hat{m}$, and since $\kappa_1 \leq \kappa_2$, we get:

$$\left\| f - m \right\|_{L^2}^2 + \left\| f - m \right\|_{L^2}^2 \leq A(m) + \kappa_1 V(\hat{m}) + \kappa_1 V(m) \leq$$

$$\leq A(m) + (\kappa_1 + \kappa_2) V(m) .$$

Thus, by taking the expectation in (23) and using Proposition 4.4, we obtain:

$$E \left[ \left\| f - m \right\|_{L^2}^2 \right] \leq 6E[A(m)] + 3(\kappa_1 + \kappa_2) V(m) + 3 \left( \left\| f - m \right\|_{L^2}^2 + V(m) \right) .$$

Now, we need to control $E[A(m)]$. Let $m' \in \mathcal{M}_n$.

$$\left\| f - m \right\|_{L^2}^2 \leq 3 \left\| f - m \right\|_{L^2}^2 + 3 \left\| f - m \right\|_{L^2}^2 .$$

Thus,

$$A(m) \leq 3 \left( \max_{m' \in \mathcal{M}_n} \left( \left\| f - m \right\|_{L^2}^2 - \frac{\kappa_1}{6} V(m') \right) + \right. + \left. \max_{m' \in \mathcal{M}_n} \left( \left\| f - m \right\|_{L^2}^2 - \frac{\kappa_1}{6} V(m') \right) + \max_{m' \in \mathcal{M}_n} \left\| f - m \right\|_{L^2}^2 \right) .$$

We denote these three terms $T_1$, $T_2(m)$ and $T_3(m)$, in that order.
Bound on $\mathbb{E}[T_1]$. We control the maximum by the sum on all terms:
\[
T_1 \leq \sum_{m' \in \mathcal{M}_n} \left( \| f_{m'} - f_{m'} \|_{L^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+.
\]

By Lemma 5.1 there exist $\kappa_0$ such that for every choice of $\kappa_1 > \kappa_0$, $\mathbb{E}[T_1] \leq C/n$ for some constant $C$ depending on $\kappa_1$ and $g$.

Bound on $\mathbb{E}[T_2(m)]$. We decompose the maximum by looking at which indices $j$ we have $m'_j < m_j$:
\[
T_2(m) = \max_{J \subseteq \{1, \ldots, d\}} \max_{m' \in \mathcal{M}_n} \max_{\{j \mid m'_j < m_j\} = J} \left( \| f_{m \land m'} - \hat{f}_{m \land m'} \|_{L^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+ \leq \max_{J \subseteq \{1, \ldots, d\}} \max_{m' \in \mathcal{M}_n} \max_{\{j \mid m'_j < m_j\} = J} \left( \| f_{m \land m'} - \hat{f}_{m \land m'} \|_{L^2}^2 - \frac{\kappa_1}{6} V(m \land m') \right)_+, \tag{26}
\]

since $V$ is non-decreasing in its arguments. For $J \subseteq \{1, \ldots, d\}$ fixed, we introduce the collection:
\[
\mathcal{M}_n(m, J) := \{ m' \in \mathcal{M}_n \mid \forall j \in J, m'_j < m_j \text{ and } \forall j \notin J, m'_j = m_j \}.
\]

Then we notice that $\{j \mid m'_j < m_j\} = J$ if and only if $m \land m' \in \mathcal{M}_n(m, J)$. So (26) becomes:
\[
T_2(m) \leq \max_{J \subseteq \{1, \ldots, d\}} \max_{m' \in \mathcal{M}_n(m, J)} \left( \| f_{m'} - \hat{f}_{m'} \|_{L^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+ \leq 2^d \sum_{m' \in \mathcal{M}_n} \left( \| f_{m'} - \hat{f}_{m'} \|_{L^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+.
\]

Then, as we did with $T_1$, we have for every choice of $\kappa_1 > \kappa_0$, $\mathbb{E}[T_2(m)] \leq 2^d \times C/n$.

Bound on $T_3(m)$. For $m' \in \mathcal{M}_n$ fixed, let $J_{m,m'} := \{ j \mid m_j < m'_j \}$ and let $S_{m,m'}$ be the following closed subspace of $L^2(\mathbb{R}_d)$:
\[
S_{m,m'} := \text{Span}\{ \varphi_k : k \in \mathbb{N}^d \text{ and } \forall j \in J_{m,m'}, k_j \leq m_j - 1 \}.
\]

Then we notice that $S_{m'} \cap S_{m,m'} = S_{m,m'}$ and that $S_m \subseteq S_{m,m'}$. We denote by $\Pi_{m'}$ the projection on $S_{m'}$ and $\Pi_{m,m'}$ the projection on $S_{m,m'}$. Then, because $\Pi_{m'}$ is 1-lipschitz,
\[
\| f_{m'} - f_{m \land m'} \|_{L^2}^2 = \| \Pi_{m'}(f - \Pi_{m,m'} f) \|_{L^2}^2 \leq \| f - \Pi_{m,m'} f \|_{L^2}^2 \leq \| f - f_m \|_{L^2}^2,
\]

where the last inequality comes from the definition of the projection on $S_{m,m'}$. So $T_3(m)$ is controlled by the bias term $\| f - f_m \|_{L^2}^2$.

End of the proof. We use the bounds on $\mathbb{E}[T_1]$, $\mathbb{E}[T_2(m)]$ and $T_3(m)$ to control $\mathbb{E}[A(m)]$ in (24):
\[
\mathbb{E}\left[ \| f - \hat{f}_m \|_{L^2}^2 \right] \leq 9 \| f - f_m \|_{L^2}^2 + 3(1 + \kappa_1 + \kappa_2) V(m) + \frac{6(1 + 2^d)C}{n}.
\]

Since this bound holds for every $m \in \mathcal{M}_n$, we get:
\[
\mathbb{E}\left[ \| f - \hat{f}_m \|_{L^2}^2 \right] \leq \max\{9, 3(1 + \kappa_1 + \kappa_2)\} \inf_{m \in \mathcal{M}_n} (\| f - f_m \|_{L^2}^2 + V(m)) + \frac{6(1 + 2^d)C}{n}. \quad \square
\]
Acknowledgments
The author wants to thank Fabienne Comte and Céline Duval for their helpful advice.

Funding
This work was supported by a grant from Région Île-de-France.

Appendix A. Talagrand’s inequality
This concentration inequality is derived from the work of [Klein and Rio (2005)]. A proof can be found in the PhD thesis of [Chagny (2013a) chapter 2].

Theorem A.1 (Talagrand’s inequality). Let \( n \in \mathbb{N}^* \) and let \( F \) be a set at most countable of measurable functions on a polish space \( E \). Let \( \xi_1, \dots, \xi_n \) be independent random variables with values in \( E \). We consider \( \nu_n \) the centred empirical process defined by:

\[
\nu_n(f) := \frac{1}{n} \sum_{i=1}^{n} (f(\xi_i) - \mathbb{E}[f(\xi_i)]), \quad f \in F.
\]

We assume that there exists three positive constants \( M, H \) and \( v \) such that:

\[
\sup_{f \in F} \|f\|_{\infty} \leq M, \quad \mathbb{E}\left[ \sup_{f \in F} |\nu_n(f)| \right] \leq H, \quad \sup_{f \in F} \frac{1}{n} \sum_{i=1}^{n} \text{Var}(f(\xi_i)) \leq v.
\]

Then for every \( \delta > 0 \), we have the following inequality:

\[
\mathbb{E} \left[ \sup_{f \in F} \nu_n^2(f) - 2(1 + 2\delta)H^2 \right] \leq \frac{4}{K} \left( \frac{v}{n} e^{-K \frac{\delta^2 H^2}{2}} + \frac{49 M^2}{K C(\delta)^2 n^2} e^{-K C(\delta)^2 \frac{\delta^2 H^2}{2}} \right),
\]

with \( C(\delta) := (\sqrt{1 + \delta} - 1) \wedge 1 \) and \( K = 1/6 \).

References


