Laguerre basis for inverse problems

Fabienne Comte, Valentine Genon-Catalot

LAGUERRE BASIS FOR INVERSE PROBLEMS

F. COMTE[(1)] & V. GENON-CATALOT[(2)]

Abstract. We present a series of inverse problems of nonparametric statistics which have an easy solution using projection estimators on a Laguerre basis. The models are $Y_i = X_iU_i,\ Z_i = X_i+V_i,\ W_i = (X_i+V_i)U_i,\ T_i = X_iU_i+V_i,\ i = 1,\dots,n$ where the X_i's and V_i's are nonnegative, the X_i's are *i.i.d.* with unknown density f, the V_i's are *i.i.d.* with known density f_V, the U_i's are *i.i.d.* with uniform density on $[0,1]$. The sequences $(X_i), (U_i), (V_i)$ are independent. We aim at estimating f on $\mathbb{R}^+$ in the four cases of indirect observations of $(X_1,\dots,X_n)$. We propose projection estimators using a Laguerre basis and give upper bounds of their $\mathbb{L}^2$-risks on specific Sobolev-Laguerre spaces. In each case, a data-driven procedure is described and proved to perform automatically the bias variance compromise.

(1) Université Paris Descartes, MAP5, UMR CNRS 8145, email: fabienne.comte@parisdescartes.fr
(2) Université Paris Descartes, MAP5, UMR CNRS 8145, valentine.genon-catalot@parisdescartes.fr.

Keywords. Adaptive estimation. Additive noise. Inverse problem. Model selection. Multiplicative censoring. Projection estimator.
MSC2010. 62G07

1. Introduction

The aim of this paper is to present a series of inverse problems of nonparametric statistics which have an easy solution using projection estimators on a Laguerre basis. The paper is partly a review of some recent results but also contains new aspects.
Consider $X_1,\dots,X_n$ n *i.i.d.* non negative random variables with unknown density f. If the X_i's are observed and if f belongs to $\mathbb{L}^2(\mathbb{R}^+)$, nonparametric estimators of f can be built by using a projection method on an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$. The basis of normalized Laguerre functions is a possibility and has the advantage of being composed of $\mathbb{R}^+$-supported functions. If the X_i's are not directly observed, the estimation of f is an inverse problem. Depending on the kind of observations, the estimation of f can be difficult. In what follows, we consider four cases of indirect observations of the X_i's and show that the use of a projection method on a Laguerre basis leads to an explicit and implementable solution.
First, we assume that observations are

$$Y_i = X_iU_i, \quad i = 1,\dots,n \tag{1}$$

where the sequences $(X_i), (U_i)$ are independent and (U_i) are *i.i.d.* with uniform distribution on $[0,1]$. The model $Y_i = X_iU_i$ is called multiplicative censoring model and covers several important statistical problems, in particular estimation under monotonicity constraints (see *e.g.* Vardi(1989)). Numerous papers deal with the estimation of f for model (1) whether by nonparametric maximum likelihood (Vardi (1989), Vardi and Zhang (1992), Asgharian *et al.* (2012)), by projection methods (Andersen and Hansen (2001), Abbaszadeh *et al.* (2012,2013)) or kernel methods (Brunel *et al.* (2015)). In Belomestny *et al.* (2016), the estimation of f by projection estimators on a Laguerre basis is investigated in the more general situation where U_i

Date: January 30, 2017.

has $beta(r,k)$-distribution. An adaptive procedure is proposed. We recall the results therein in the case where U_i has uniform distribution. Moreover, under a slight additional assumption, an improvement of the risk bound is provided in Comte and Dion (2017) and the adaptive procedure is modified accordingly.
Second, we consider observations $Z_1, \dots, Z_n$ such that

$$Z_i = X_i + V_i, \quad i = 1, \dots, n. \tag{2}$$

where X_i, V_i are nonnegative random variables, (V_i) are *i.i.d.* with known density f_V and the sequences $(X_i), (V_i)$ are independent. Density estimation from noisy observations is also the subject of a huge number of contributions. For real-valued random variables, this deconvolution problem is classically solved by Fourier methods. However, recently, the study of one-sided errors, *i.e.* $V_i \geq 0$, was motivated by applications in the field of finance (see Jirak *et al.* (2014)) or in survival models, (see van Es *et al.* (1998), Jongbloed (1998)). In particular, Mabon (2016) proposes for model (2) projection estimators of f using a Laguerre basis whose properties allow deconvolution of densities on $\mathbb{R}^+$. We detail this approach.
Finally, we combine the two previous situations. This can be done in two ways which are not equivalent. On one hand, we assume that observations are:

$$W_i = (X_i + V_i)U_i, \quad i = 1, \dots, n. \tag{3}$$

On the other hand, we assume that observations are

$$T_i = X_i U_i + V_i, \quad i = 1, \dots, n. \tag{4}$$

The sequences $(X_i), (U_i), (V_i)$ are supposed to be independent. In each case, we show how to build projections estimators of f on a Laguerre basis and propose a data-driven choice of the dimension of the projection space.
The Laguerre basis is related to specific function spaces, the Sobolev-Laguerre spaces (see *e.g.* Shen (2000) and Bongioanni and Torrea (2007)). The link between projection coefficients and regularity conditions in these spaces has been described in Comte and Genon-Catalot (2015). In each of the above models, we exhibit explicit relations between the projection coefficients of the density of the observed variables in the Laguerre basis and the projection coefficients of the unknown density f. This allows to build projection estimators of f. We provide risk bounds for the estimators, allowing to compute upper bounds for the rates of convergence. Afterwards, we propose a data-driven procedure leading to an adaptive estimator performing automatically the bias variance compromise.
In Section 2, we describe the basis and the Sobolev-Laguerre spaces. In Section 3, for the purpose of comparison with the other models, we study the case of direct observations of $X_1, \dots, X_n$. Sections 4-6 deal with the four successive models. In Section 7, we review some extensions and other inverse problems that can be solved by the Lagurre appproach. Section 8 contains a recap of useful formulae for Laguerre functions and all proofs. In the Appendix, we give the Talagrand inequality used for proving the adaptation results.

2. About Laguerre bases and spaces

We start by presenting the Laguerre basis that we have chosen and the Sobolev-Laguerre spaces. More details on Laguerre functions are given in Section 8.1.

2.1. Laguerre basis.

Below we denote the scalar product and the $\mathbb{L}^2$-norm on $\mathbb{L}^2(\mathbb{R}^+)$ by:

$$\forall s, t \in \mathbb{L}^2(\mathbb{R}^+),\ \langle s, t\rangle = \int_0^{+\infty} s(x)t(x)dx,\ \|t\|^2 = \int_0^{+\infty} t^2(x)dx.$$

Consider the Laguerre polynomials (L_j) and the Laguerre functions (φ_j) given by

$$L_j(x) = \sum_{k=0}^{j}(-1)^k \binom{j}{k}\frac{x^k}{k!}, \qquad \varphi_j(x) = \sqrt{2}L_j(2x)e^{-x}\mathbb{1}_{x\geq 0}, \quad j \geq 0. \tag{5}$$

The collection $(\varphi_j)_{j\geq 0}$ constitutes a complete orthonormal system on $\mathbb{L}^2(\mathbb{R}^+)$, and is such that (see Abramowitz and Stegun (1964)):

$$\forall j \geq 1,\ \ \forall x \in \mathbb{R}^+,\ \ |\varphi_j(x)| \leq \sqrt{2}. \tag{6}$$

For $h \in \mathbb{L}^2(\mathbb{R}^+)$, we can develop h on the Laguerre basis with:

$$h = \sum_{j\geq 0} a_j(h)\varphi_j, \qquad a_j(h) = \langle h, \varphi_j\rangle.$$

When h is a density, $a_0(h) = \langle h, \varphi_0\rangle = \sqrt{2}\int_0^{+\infty} h(x)e^{-x}dx > 0$.
By convention, we set $\varphi_j = 0$ if $j \leq -1$ and define the vector of coefficients of h on $(\varphi_0, \dots, \varphi_{m-1})$:

$$\vec{a}_{m-1}(h) :=\ {}^t(a_j(h))_{0\leq j\leq m-1}.$$

We define the m-dimensional space $S_m = \text{span}(\varphi_0, \varphi_1, \dots, \varphi_{m-1})$. The function

$$h_m = \sum_{j=0}^{m-1} a_j(h)\varphi_j$$

is the orthogonal projection of h on S_m.

2.2. **Sobolev-Laguerre spaces.** For $s \geq 0$, the Sobolev-Laguerre space with index s (see Bongioanni and Torrea (2007)) is defined by:

$$W^s = \{h : \mathbb{R}^+ \to \mathbb{R},\ h \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k\geq 0} k^s a_k^2(h) < +\infty\}. \tag{7}$$

The following results have been proved in Section 7 of Comte and Genon-Catalot (2015) and Section 7.2 of Belomestny *et al.* (2016).
For s integer, if $h : \mathbb{R}^+ \to \mathbb{R}$ belongs to $\mathbb{L}^2(\mathbb{R}^+)$, then

$$|h|_s^2 := \sum_{k\geq 0} k^s a_k^2(h) < +\infty. \tag{8}$$

is equivalent to the property that h admits derivatives up to order $s-1$, with $h^{(s-1)}$ being absolutely continuous and for $m = 0, \dots, s-1$, the functions

$$\xi_{m+1}(x) := x^{(m+1)/2}(h(x)e^x)^{(m+1)}e^{-x} = x^{(m+1)/2}\sum_{j=0}^{m+1}\binom{m+1}{j}h^{(j)}(x)$$

belong to $\mathbb{L}^2(\mathbb{R}^+)$. Moreover, for $m = 0, 1, \dots, s-1$,

$$\|\xi_{m+1}\|^2 = \sum_{k\geq m+1} k(k-1)\dots(k-m)a_k^2(h).$$

For $h \in W^s$ with s integer, we set $\|h\|_0^2 = \|h\|^2$ and for $s \geq 1$

$$\|h\|_s = \|\xi_s\| = [\sum_{k\geq s} k(k-1)\dots(k-s+1)a_k^2(h)]^{1/2}. \tag{9}$$

Now we set

$$|||h|||_s^2 := \sum_{j=0}^{s} \|h\|_j^2.$$

Then it holds that, when s is integer, the two norms $|||h|||_s$ and $|h|_s$ are equivalent.
We define the ball $W^s(D)$ by (see (7)-(8)):

$$W^s(D) \doteq \left\{ f \in W^s, |f|_s^2 = \sum_{k=0}^{\infty} k^s a_k^2(f) \leq D \right\}.$$

3. Projection estimators of f in the Laguerre basis when X_i's are observed

We assume that f belongs to $\mathbb{L}^2(\mathbb{R}^+)$ and provide for each $m \geq 1$, a projection estimator of f by estimating the coefficients $a_j(f), j = 0, \ldots, m-1$. In the case where the X_i's are observed, we define the empirical and unbiased estimator of $a_j(f)$ by

$$\hat{a}_j(X) = \frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) \quad \text{and the projection estimator} \quad \hat{f}_m^X = \sum_{j=0}^{m-1} \hat{a}_j(X) \varphi_j.$$

Clearly, $\hat{f}_m^X$ an unbiased estimator of $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$, the orthogonal projection of f on S_m. By the Pythagoras Theorem, we have $\|\hat{f}_m^X - f\|^2 = \|f - f_m\|^2 + \|\hat{f}_m^X - f_m\|^2$. As $(\varphi_j)_j$ is orthonormal, we get $\|\hat{f}_m^X - f_m\|^2 = \sum_{j=0}^{m-1} (\hat{a}_j(X) - a_j(f))^2$ and

$$\mathbb{E}[(\hat{a}_j(X) - a_j(f))^2] = \frac{1}{n} \mathrm{Var}(\varphi_j(X)) \leq \frac{1}{n} \mathbb{E}(\varphi_j^2(X)).$$

Therefore, with (6), we obtain the risk bound:

$$\mathbb{E}(\|\hat{f}_m^X - f\|^2) \leq \|f - f_m\|^2 + 2\frac{m}{n}. \tag{10}$$

Remark 3.1. The risk bound decomposition (10) classically involves a bias term $\|f - f_m\|^2 = \sum_{j \geq m} a_j^2(f)$ which is decreasing with m and a variance term of order m/n which is increasing with m. Therefore, to evaluate the rate of convergence, we have to perform a compromise to select relevantly m.

For $f \in W^s(D)$, $\|f - f_m\|^2 = \sum_{j \geq m} a_j^2(f) \leq Dm^{-s}$. Choosing $m_{\text{opt}} = [n^{1/(s+1)}]$ in the r.h.s. of (10) implies

$$\mathbb{E}(\|\hat{f}_{m_{\text{opt}}}^X - f\|^2) \leq C_0(s, D) n^{-s/(s+1)}$$

where $C_0(s, D)$, is a constant depending on s and D only. The following lower bound result is proved in Belomestny *et al.* (2016) implying that the above rate is minimax optimal on Sobolev-Laguerre balls (up to a logarithmic term).

Theorem 3.1. *Assume that s is an integer, $s > 1$.*
Then for any estimator $\hat{f}_n$ built as a measurable function of $X_1, \ldots, X_n$, for any $\epsilon > 0$ and for n large enough,

$$\sup_{f \in W^s(D)} \mathbb{E}_f \left[\|\hat{f}_n - f\|^2 \right] \gtrsim \psi_n, \quad \psi_n = n^{-s/(s+1)} / \log^{(1+\epsilon)/(s+1)}(n).$$

Remark 3.2. On some concrete examples, faster rates of convergence of the $\mathbb{L}^2$-risk may be obtained. Exponential distributions provide examples of such a case. If X has exponential distribution $\mathcal{E}(\theta)$, $\theta > 0$, then the projection coefficients are given by $a_k(f) = \sqrt{2}[\theta/(\theta+1)]\,((\theta-1)/(\theta+1))^k$ and the bias can be explicitly computed,

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} a_k^2(f) = \frac{\theta}{2}\left|\frac{\theta-1}{\theta+1}\right|^{2m}.$$

Therefore, the bias is exponentially decreasing. Consequently, for $m_{\rm opt} = \log(n)/\rho$, with $\rho = |\log[|(\theta-1)/(\theta+1)|]|$, the rate of the $\mathbb{L}^2$-risk of $\hat{f}^X_{m_{\rm opt}}$ is of order $[\log(n)]/n$. This kind of result can be generalized to the case of a density f defined as a mixture of exponential densities and to Gamma distributions $\Gamma(p,\theta)$, with p an integer (see Comte and Genon-Catalot (2015), Mabon (2015)). More precisely, if f_p is the density $\Gamma(p,\theta)$,

$$a_k(f_p) = \frac{\sqrt{2}}{\Gamma(p)}\left(\frac{\theta}{\theta+1}\right)^p S_{p,k}\left(\frac{2}{\theta+1}\right), \quad \text{with} \quad S_{p,k}(x) = \frac{d^{p-1}}{dx^{p-1}}\left[x^{p-1}(1-x)^k\right].$$

This term can be computed explicitly and we get the bound, for $p \geq 2$ and $C_0(p,\theta)$ a constant depending on p and θ only,

$$|a_k(f_p)| \leq C_0(p,\theta)k^{p-1}\left|\frac{\theta-1}{\theta+1}\right|^k.$$

Thus for $m \geq p-1$,

$$\sum_{k\geq m}[a_k(f_p)]^2 \leq C(p,\theta)m^{2(p-1)}\left(\frac{\theta-1}{\theta+1}\right)^{2m}, \quad \text{with} \quad 0 < C(p,\theta) < +\infty.$$

Note that the bias is null for $\theta = 1$ and $m > p-1$, which is expected since $f_p \in S_{p-1}$. Moreover, the bias order depends on θ.

Of course, if we know that f belongs to some parametric model, it is better to use a parametric method. But, in our framework, f is unknown, so we have to face all situations. This is why a data-driven choice of the dimension of the projection space has to be done.
The interest of the adaptive procedure is that it realizes automatically the finite sample bias-variance compromise and also *automatically* reaches the best possible asymptotic rate without requiring any knowledge on the bias order. The data-driven choice of m mimicks the minimization of the squared bias-variance bound using estimators of the risk bound terms. As $\|f-f_m\|^2 = \|f\|^2 - \|f_m\|^2$, the squared bias is estimated by $-\|\hat{f}^X_m\|^2$, getting rid of $\|f\|^2$ which is unknown but constant (not depending on m). Thus we set, for κ a numerical constant,

$$\hat{m}_X = \arg\min_{m\in\{1,\ldots,n\}}\left(-\|\hat{f}^X_m\|^2 + \text{pen}_X(m)\right), \quad \text{pen}_X(m) = \kappa\frac{m}{n}.$$

It follows from Massart (2007), Chapter 7, Theorem 7.5 that there exists a numerical value κ_0 such that for all $\kappa \geq \kappa_0$,

$$\mathbb{E}(\|\hat{f}^X_{\hat{m}_X} - f\|^2) \leq 3\inf_{m\in\{1,\ldots,n\}}\left(\|f-f_m\|^2 + \text{pen}_X(m)\right) + \frac{C}{n},$$

where C is a constant depending on $\|f\|$.

4. Projection estimator of f in the Laguerre basis when Y_i's are observed

Now, our aim is to build an estimator of f from the observations $Y_1,\ldots,Y_n$, still taking into account that all variables are nonnegative.

4.1. **Preliminary properties and formulas.** The construction relies on the following steps. We have

$$f_Y(y) = \int_y^{+\infty} \frac{f(u)}{u} du 1_{y\geq 0}.$$

Let $F(x) = \int_0^x f(t)\,dt$, $F_Y(y) = \int_0^y f_Y(t)\,dt$. Elementary computations yield that, for any $y \geq 0$,

$$f(y) = -y f_Y'(y), \qquad F(y) = F_Y(y) - y f_Y(y). \tag{11}$$

The second equality implies

$$\lim_{y\to 0} y\ f_Y(y) = \lim_{y\to +\infty} y\ f_Y(y) = 0. \tag{12}$$

Lemma 4.1. (1) *Let $t : \mathbb{R} \to \mathbb{R}$ be bounded, derivable, then*

$$\mathbb{E}(t(Y) + Yt'(Y)) = \mathbb{E}t(X). \tag{13}$$

(2) *Assume that $\mathbb{E}(X) < +\infty$. Let $t \in \mathbb{L}^2(\mathbb{R}^+)$, then $\mathbb{E}(Y^2 t^2(Y)) \leq \mathbb{E}(X)\|t\|^2$.*

Equality (13) is the basement of the estimation procedure. Using it, we can link the coefficients of f and f_Y on the Laguerre basis and these relations are used for building the projection estimators.

Proposition 4.1. *For all $j \geq 0$,*

$$a_j(f) = \langle f, \varphi_j \rangle = \langle f_Y, (y\varphi_j)' \rangle \tag{14}$$

Relation (14) is just an application of (13).
Moreover, using formula (30) (see Section 8.1), we get $a_0(f) = (1/2)a_0(f_Y) + (1/2)a_1(f_Y)$ and for $j \geq 1$,

$$a_j(f) = -\frac{j}{2} a_{j-1}(f_Y) + \frac{1}{2} a_j(f_Y) + \frac{j+1}{2} a_{j+1}(f_Y).$$

Introducing the matrix $\mathbf{H}_m = ([\mathbf{H}_m]_{k,\ell})_{1\leq k,\ell\leq m}$ with size $m \times (m+1)$ given by $[\mathbf{H}_m]_{k,\ell} = 0$ if $\ell \neq k-1, k, k+1$ and $[\mathbf{H}_m]_{1,1} = 1/2$, $[\mathbf{H}_m]_{1,2} = 1/2$ and for $k \geq 2$,

$$[\mathbf{H}_m]_{k,k-1} = -\frac{k-1}{2}, \quad [\mathbf{H}_m]_{k,k} = \frac{1}{2}, \quad [\mathbf{H}_m]_{k,k+1} = \frac{k}{2}, \tag{15}$$

we obtain the linear relation between the vectors of coefficients of f and f_Y:

$$\vec{a}_{m-1}(f) = \mathbf{H}_m \vec{a}_m(f_Y).$$

4.2. **Projection estimator and upper risk bound.** Consequently, we define a collection of projection estimators of f by:

$$\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \qquad \text{with } \hat{a}_j = \frac{1}{n}\sum_{i=1}^n [Y_i \varphi_j'(Y_i) + \varphi_j(Y_i)]. \tag{16}$$

We also have for $m \geq 1$, setting $\vec{\hat{a}}_{m-1} = {}^t(\hat{a}_j)_{0\leq j\leq m-1}$, $\vec{\hat{a}}_m(Y) = {}^t(\hat{a}_j(Y))_{0\leq j\leq m}$, the following relation which is convenient to compute the estimator

$$\vec{\hat{a}}_{m-1} = \mathbf{H}_m \vec{\hat{a}}_m(Y) \quad \text{with} \qquad \hat{a}_j(Y) := \frac{1}{n}\sum_{i=1}^n \varphi_j(Y_i). \tag{17}$$

Proposition 4.2. *Let $\hat{f}_m$ be given by (16).*
If $\mathbb{E}(X_1) < +\infty$, then we have

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \le \|f - f_m\|^2 + 4\mathbb{E}(Y_1)\frac{m^2}{n} + 2\frac{m}{n}.$$

Otherwise, we have

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \le \|f - f_m\|^2 + \frac{2m^3}{n} + \frac{3m}{2n}.$$

These bounds are given in Belomenstny *et al.* (2016) and in Comte and Dion (2017). We deduce from Proposition 4.2 rates of convergence of the estimator on the Sobolev-Laguerre spaces described in Section 2.2.

Corollary 4.1. *Assume that $f \in W^s(D)$. Let $\hat{f}_m$ be given by (16).*
If $\mathbb{E}(X_1) < +\infty$, then choosing $m_{opt} = [n^{s+2}]$ gives

$$\mathbb{E}(\|\hat{f}_{m_{opt}} - f\|^2) \le C_1(s, D)n^{-s/(s+2)}$$

where $C_1(D, s)$ is a constant depending on D and s.
Otherwise, choosing $m_{opt} = [n^{s+3}]$ gives

$$\mathbb{E}(\|\hat{f}_{m_{opt}} - f\|^2) \le C_2(s, D)n^{-s/(s+3)}$$

where $C_2(D, s)$ is a constant depending on D and s.

Remark 3.2 applies here. For exponential, Gamma or mixed Gamma densities f, the bias is exponentially decreasing. Thus, the same choice m_{opt} yields a rate of order $[\log(n)]^2/n$.

4.3. **Adaptive estimation.** We propose a penalization method to select m automatically. For κ a numerical constant, let

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} \left(-\|\hat{f}_m\|^2 + \text{pen}(m)\right) \quad \text{with} \quad \text{pen}(m) = \kappa\frac{m\log(m+2)}{n}(1 + 2\mathbb{E}(Y_1)m) \tag{18}$$

and

$$\mathcal{M}_n = \{m \in \{1, \dots, n\},\ m \le \sqrt{n}\}.$$

Theorem 4.1. *Assume that $\mathbb{E}(X_1) < +\infty$. Let $\hat{f}_m$ be given by (16)and $\hat{m}$ by (18). There exists a constant κ_0 such that for any $\kappa \ge \kappa_0$, we have*

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \le C_1 \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \text{pen}(m)\right) + \frac{C_2}{n} \tag{19}$$

where C_1 is a numerical constant ($C_1 = 4$ suits) and C_2 is a positive constant depending on $\mathbb{E}(Y_1)$.

The estimator $\hat{f}_{\hat{m}}$ realizes an automatic trade-off between the squared bias $\|f - f_m\|^2$, and the variance, increased by a logarithmic term. The penalty contains the unknown quantity $\mathbb{E}(Y_1)$: to compute the estimator, this term is replaced by the empirical mean $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ and it is possible to prove that the bound (19) still holds (see Comte and Dion (2017)).

5. Projection estimator of f when $Z_i = X_i + V_i$ are observed

5.1. Projection estimator and risk bound. A consequence of Model (2) is

$$f_Z(x) = f \star f_V(x) = \int_0^x f(u) f_V(x-u) du.$$

By using (34), this convolution equation can be rewritten:

$$\begin{aligned}\sum_{k=0}^{\infty} a_k(f_Z)\varphi_k(x) &= \sum_{j=0}^{+\infty}\sum_{k=0}^{+\infty} a_j(f)a_k(f_V)\varphi_j \star \varphi_k(x) \\ &= \sum_{k=0}^{\infty} \varphi_k(x) \sum_{\ell=0}^{k} 2^{-1/2}\left(a_{k-\ell}(f_V) - a_{k-\ell-1}(f_V)\right)a_\ell(f).\end{aligned}$$

Define the $m \times m$ triangular matrix $\mathbf{V}_m = (v_{i,j})_{0 \le i,j \le m-1}$ where

$$v_{i,j} = 2^{-1/2}(\langle f_V, \varphi_{i-j}\rangle \mathbf{I}_{i-j \ge 0} - \langle f_V, \varphi_{i-j-1}\rangle \mathbf{I}_{i-j-1 \ge 0}). \tag{20}$$

As $v_{i,j} = v(i-j)\mathbf{I}_{i-j\ge 0}$, $\mathbf{V}_m$ is a Toeplitz triangular matrix with diagonal elements $v_{i,i} = 2^{-1/2}\langle f_V, \varphi_0\rangle > 0$. It is thus invertible and for all $m \ge 1$,

$$\vec{a}_{m-1}(f_Y) = {}^t(a_j(f))_{0\le j\le m-1} = \mathbf{V}_m^{-1}[(a_j(f_Z))_{0\le j\le m-1}] = \mathbf{V}_m^{-1}\vec{a}_{m-1}(f_Z), \tag{21}$$

Formula (20) relies on a convolution property of the Laguerre functions (φ_j) (Formula (34), Section 8.1) which can be used in $\mathbb{R}^+$-deconvolution. The projection estimator of f on S_m based on $(Z_1, \dots, Z_n)$ is given by

$$\tilde{f}_m = \sum_{j=0}^{m-1} \tilde{a}_j \varphi_j, \quad \vec{\tilde{a}}_{m-1} = {}^t(\tilde{a}_j)_{0\le j\le m-1} = \mathbf{V}_m^{-1}\vec{\hat{a}}_{m-1}(Z), \qquad m \ge 1 \tag{22}$$

where $\vec{\hat{a}}_{m-1}(Z) = [(\hat{a}_j(Z))_{0\le j\le m-1}]$ and $\hat{a}_j(Z)$ is defined by

$$\hat{a}_j(Z) := \frac{1}{n}\sum_{i=1}^{n} \varphi_j(Z_i). \tag{23}$$

The following risk bound holds (Mabon (2016)):

Proposition 5.1. *Assume that* $\|f_V\|_\infty < +\infty$. *Let* $\tilde{f}_m$ *be given by (22). Then we have*

$$\mathbb{E}(\|\tilde{f}_m - f\|^2) \le \|f - f_m\|^2 + (2 \vee \|f_V\|_\infty)\frac{\|\mathbf{V}_m^{-1}\|_F^2}{n}$$

where $\|\mathbf{A}\|_F^2 = \mathrm{Tr}({}^t\mathbf{A}\mathbf{A})$.

Remark 3.1 still applies here. The bias term is unchanged. For the variance term, it is of order $\|\mathbf{V}_m^{-1}\|_F^2/n$, which is increasing in m because of the special form of $\mathbf{V}_m^{-1}$ (lower triangular and Toeplitz, see Mabon (2016)).

We can deduce from Proposition 5.1 rates of convergence of the estimator on Sobolev-Laguerre spaces. In Comte et *al.* (2017), the order of $\|\mathbf{V}_m^{-1}\|_F^2$ in function of m is studied. In particular, if V_i has a Gamma distribution $\Gamma(r, \lambda)$, $r \in \mathbb{N}$, $r \ge 1$, there exist constants c, C such that

$$cm^{2r} \le \|\mathbf{V}_m^{-1}\|_F^2 \le Cm^{2r}.$$

Note that the case $V_i = 0$ and $\mathbf{V}_m = Id$ is excluded from this context.

Therefore the following corollary holds:

Corollary 5.1. *Assume that $f \in W^s(D)$, and that V_i has a Gamma distribution $\Gamma(r,\lambda)$, r integer, $r \geq 1$. Then $\tilde{f}_m$ given by (22) satisfies, for $m_{opt} = [n^{2r+s}]$*

$$\mathbb{E}(\|\tilde{f}_{m_{opt}} - f\|^2) \leq C(s,D) n^{-s/(2r+s)}.$$

Remark 5.1. • For $V \sim \mathcal{E}(\lambda) = \gamma(1,\lambda)$, we have $[\mathbf{V}_m]_{i,i} = \lambda/(1+\lambda)$ and

$$[\mathbf{V}_m]_{i,j} = -2\lambda \frac{(\lambda-1)^{i-j-1}}{(\lambda+1)^{(i-j+1)}} \text{ if } j < i \tag{24}$$

and $[\mathbf{V}_m]_{i,j} = 0$ otherwise. We can compute $[\mathbf{V}_m^{-1}]_{i,j} = (\lambda+1)/\lambda$ if $i = j$, $2/\lambda$ if $i > j$ and 0 otherwise. Note that

$$\|\mathbf{V}_m^{-1}\|_F^2 = 2\frac{m^2}{\lambda^2} + m(1 + \frac{2}{\lambda} - \frac{1}{\lambda^2}).$$

• For $V \sim \Gamma(2,\mu)$, we have $[\mathbf{V}_m]_{i,i} = (\mu/(1+\mu))^2$, $[\mathbf{V}_m]_{i+1,i} = -4\mu^2/(1+\mu)^3$ and

$$[\mathbf{V}_m]_{i,j} = 4(i-j-\mu)\mu^2 \frac{(\mu-1)^{i-j-2}}{(\mu+1)^{i-j+2}} \text{ if } i > j+1 \tag{25}$$

and $[\mathbf{V}_m]_{i,j} = 0$ otherwise.

5.2. **Adaptive estimation.** A data driven method to relevantly select m can be proposed and yields an automatic bias variance compromise. Let us define, for κ a numerical constant,

$$\tilde{m} = \arg\min_{m \in \mathcal{M}_n} \left(-\|\tilde{f}_m\|^2 + \widetilde{\text{pen}}(m)\right) \quad \text{with } \widetilde{\text{pen}}(m) = \kappa \frac{\log(2 + \|\mathbf{V}_m^{-1}\|_F^2)\|\mathbf{V}_m^{-1}\|_F^2}{n} \tag{26}$$

where

$$\mathcal{M}_n = \{m \in \mathbb{N}^*,\ m \leq n/\log(2+n),\ \|\mathbf{V}_m^{-1}\|_F^2 \leq n\}.$$

Theorem 5.1. *Let $\tilde{f}_m$ be given by (22) and $\tilde{m}$ by (26). There exists a numerical constant κ_0 such that for any $\kappa \geq \kappa_0$, we have*

$$\mathbb{E}(\|\tilde{f}_{\tilde{m}} - f\|^2) \leq C_1 \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \widetilde{\text{pen}}(m)\right) + \frac{C_2}{n}.$$

6. Combining the models

6.1. **Projection estimator when $W_i = (X_i + V_i)U_i$ are observed.** Now we combine the two previous procedures. We define the projection estimator $\check{f}_m$ by

$$\check{f}_m = \sum_{j=0}^{m-1} \check{a}_j \varphi_j, \quad \text{with} \quad \vec{\check{a}}_{m-1} = \mathbf{V}_m^{-1} \vec{\check{a}}_{m-1}(W) \tag{27}$$

and

$$\check{a}_j(W) = \frac{1}{n}\sum_{i=1}^n [W_i \varphi_j'(W_i) + \varphi_j(W_i)], \quad \vec{\check{a}}_{m-1}(W) = {}^t(\check{a}_0(W), \ldots, \check{a}_{m-1}(W)).$$

Note that, as previously, with $\mathbf{H}_m$ is defined by (15),

$$\vec{\check{a}}_{m-1}(W) = \mathbf{H}_m \vec{\hat{a}}_m(W), \quad \vec{\hat{a}}_m(W) = {}^t(\hat{a}_0(W), \ldots, \hat{a}_m(W)), \quad \hat{a}_j(W) = \frac{1}{n}\sum_{i=1}^n \varphi_j(W_i).$$

Proposition 6.1. *Let $\check{f}_m$ be given by (27). If $\mathbb{E}(W_1) < +\infty$, then we have*

$$\mathbb{E}(\|\check{f}_m - f\|^2) \le \|f - f_m\|^2 + 2\|\mathbf{V}_m^{-1}\|_{\text{op}}^2 \left(2\mathbb{E}(W_1)\frac{m^2}{n} + \frac{m}{n}\right),$$

where $\|\mathbf{A}\|_{\text{op}}^2 = \lambda_{\max}({}^t\mathbf{A}\,\mathbf{A})$, the largest eigenvalue of the matrix ${}^t\mathbf{A}\,\mathbf{A}$.

If $V_i = 0$, *i.e.* $\mathbf{V}_m = Id$, then $\|\mathbf{V}_m^{-1}\|_{\text{op}}^2 = 1$, so we recover the first result of Proposition 4.2. Let

$$\check{\mathcal{M}}_n = \{m \in \mathbb{N}, m^2 \le n/\log(n+2), \quad m^2\|\mathbf{V}_m^{-1}\|_{\text{op}}^2 \le n\}$$

and, for κ a numerical constant,

$$\check{\text{pen}}(m) = \kappa \log(2 + m^2\|\mathbf{V}_m^{-1}\|_{\text{op}}^2)\|\mathbf{V}_m^{-1}\|_{\text{op}}^2 \left(2\mathbb{E}(W_1)\frac{m^2}{n} + \frac{m}{n}\right).$$

Then setting

$$\check{m} = \arg\min_{m \in \check{\mathcal{M}}_n} \left(-\|\check{f}_m\|^2 + \check{\text{pen}}(m)\right),$$

we can prove an adaptation result for $\check{f}_{\check{m}}$ analogous to the ones of Theorems 4.1 and 5.1. The proof being analogous is omitted.

6.2. **Projection estimator when $T_i = X_iU_i + V_i$ are observed.** Define $\mathbf{K}_m$ with size $m \times (m+1)$ by

$$\mathbf{K}_m := \mathbf{H}_m\mathbf{V}_{m+1}^{-1}.$$

Then, we have

$$\vec{a}_{m-1}(f) = \mathbf{K}_m\vec{a}_m(f_T).$$

Thus we can define the estimator of f by

$$\breve{f}_m = \sum_{j=0}^{m-1} \breve{a}_j\varphi_j, \quad \vec{\breve{a}}_{m-1} = \mathbf{K}_m\vec{\hat{a}}_m(T) \tag{28}$$

where

$$\hat{a}_j(T) = \frac{1}{n}\sum_{i=1}^n \varphi_j(T_i), \quad \text{and} \quad \vec{\hat{a}}_m = {}^t(\hat{a}_0, \dots, \hat{a}_m), \quad \vec{\breve{a}}_{m-1} = {}^t(\breve{a}_0, \dots, \breve{a}_{m-1}).$$

Proposition 6.2. *Assume that $\|f_V\|_\infty < +\infty$. Then,*

$$\mathbb{E}(\|\breve{f}_m - f\|^2) \le \|f - f_m\|^2 + (2\|f_V\|_\infty)\|\mathbf{H}_m\|_{\text{op}}^2 \frac{\|\mathbf{V}_{m+1}^{-1}\|_F^2}{n}.$$

It follows from Corollary 2.1 in Belomestny *et al.* (2016) and its proof (for $k = 1$), that $\|\mathbf{H}_m\|_{\text{op}}^2 \le 3(m+1)^2$. Therefore, in this case, if $V_i = 0$, then $\mathbf{V}_{m+1}^{-1} = Id$ and $\|\mathbf{V}_{m+1}^{-1}\|_F^2 = m+1$ and we recover the variance order of the second Inequality of Proposition 4.2.

In this case too, we can propose a data-driven selection of m leading to an adaptive estimator.

7. Extensions and concluding remarks

In this paper, the use of a Laguerre basis to estimate a function $f \in \mathbb{L}^2(\mathbb{R}^+)$ is illustrated in several examples of inverse problems. Projection estimators which are easy to implement are built and studied. Data-driven choices of the projection dimension can be proposed leading to adaptive estimators.

In Mabon (2016), Belomestny *et al.* (2016), Comte and Dion (2017), the adaptive estimators for additive and multiplicative censoring models are studied and implemented. The choice of the constant κ in the penalty is a specific difficulty of the method: indeed, the theoretical constant κ_0 obtained in proofs is not optimal and generally much too large. Finding the optimal theoretical value is hard but in practice, the constant κ is calibrated by preliminary simulations.

For Model (1) and for Model (2), the estimation of the survival function $S(x) = \mathbb{P}(X_1 > x)$ can be done, still using a Laguerre basis, under the assumption that S is in $\mathbb{L}^2(\mathbb{R}^+)$, which holds if $\mathbb{E}(X_1) = \int_0^{+\infty} S(x)dx < +\infty$. A natural idea would be to assume that f belongs to $\mathbb{L}^2(\mathbb{R}^+)$ and integrate the development of f under maybe additional assumptions, using Formulae (31) and (33). However there is a wiser approach and we need not assume that f belongs to $\mathbb{L}^2(\mathbb{R}^+)$ to build estimators of S in Models (1) and (2). Indeed, in Model (1), it follows from formula (11) that $S(x) = S_Y(x) + xf_Y(x)$, where $S_Y(x)$ is the survival function of Y_1. The estimation procedure, developed in Comte and Dion (2017), is based on

$$\langle \varphi_j, S\rangle = \int_0^{+\infty} \mathbb{E}(\mathbf{1}_{Y_1>x})\varphi_j(x)dx + \mathbb{E}[Y_1\varphi_j(Y_1)] = \mathbb{E}[\Phi_j(Y_1) + Y_1\varphi_j(Y_1)].$$

In Model (2), Mabon (2016) proves the basic relation

$$S_Z(x) = S_X \star f_V(x) + S_V(x)$$

and deduce the estimation procedure.

Extensions of the results presented here are possible. First, the case of multiplicative censoring is investigated in Belomestny *et al.* (2016) when the multiplicative noise U_i has $beta(r, k)$ distribution with r, k integers, $r, k \geq 1$. Second, as in Chesneau (2013), we can consider that $U_i = U_i^{(1)} \dots U_i^{(\ell)}$ with $U_i^{(j)}$'s *i.i.d.* and uniform. Third, the case of noisy observations with unknown distribution of the noise V is studied in Comte and Mabon (2016). A preliminary sample of the V_i's is then required for identifiability.

Other models and inverse problems have been investigated with the use of a Laguerre basis. First, the estimation of f from the observations

$$y(t_i) = \int_0^{t_i} g(t_i - \tau)f(\tau)d\tau + \sigma\varepsilon_i, \quad i = 1, \dots, n$$

where $f,\ g\ :\ \mathbb{R}^+ \to \mathbb{R}^+$, g is known and (ε_i) are *i.i.d.* centered with unit variance is studied in Comte *et al.*. (2017). The solution is provided in two steps: estimation of the regression function $f \star g$ and deconvolution.

The Laguerre basis $(\varphi_j)_j$ is especially well-fitted for estimating the mixture density from *i.i.d.* mixed Poisson observations. In Comte an Genon-Catalot (2015), upper and lower risk bounds are obtained.

The estimation of the density of interarrival times $D_i = T_i - T_{i-1}$ in the renewal model

$$R_t = \sum_{i\geq 0} \mathbf{1}_{T_i \leq t}, \quad t \in [0, T]$$

is a difficult problem. The Laguerre basis is a convenient tool for continuous time observations of R_t, which gets even crucial for discrete time observations, see Comte and Duval (2016).

All these examples show that, as long as nonnegative random variables are involved in the model, the Laguerre basis is a powerful and adequate tool for solving estimation problems.

8. Proofs

8.1. **Formulae for Laguerre functions.** The Laguerre polynomials (L_j) (see (5)) are orthogonal with respect to the weight function e^{-x}. Therefore, Laguerre bases can be defined with a parameter $\mathtt{a} > 0$ by setting

$$\varphi_j^{\mathtt{a}}(x) = \sqrt{\mathtt{a}} L_j(\mathtt{a}x) e^{-\mathtt{a}x/2} \mathbf{1}_{\mathbb{R}^+}(x).$$

We focus on the choice $\mathtt{a} = 2$ which is especially convenient for computing derivatives or integrals of the basis functions and denote $\varphi_j^{\mathtt{a}}$ by φ_j for $\mathtt{a} = 2$.
We give several formulae which are used in the present paper and others which are not used but are necessary for the extensions mentionned in Section 7. Formula (22.7.12) in Abramowitz and Stegun (1964) states that

$$xL_j(x) = -(j+1)L_{j+1}(x) + (2j+1)L_j(x) - jL_{j-1}(x). \tag{29}$$

implying

$$(y\varphi_j(y))' = \varphi_j(y) + y\varphi_j'(y) = -\frac{j}{2}\varphi_{j-1}(y) + \frac{1}{2}\varphi_j(y) + \frac{j+1}{2}\varphi_{j+1}(y). \tag{30}$$

By elementary computations, we get

$$\int_0^{+\infty} \varphi_j(x)dx = \sqrt{2}(-1)^j.$$

The functions $\varphi_j, \varphi_j', \varphi_j''$ are uniformly bounded. We already mentioned that $\forall x \geq 0, |\varphi_j(x)| \leq \sqrt{2}$. By Lemma 6.1 in Comte and Genon-Catalot (2015), it holds that, for all $x \geq 0$,

$$|\varphi_j'(x)| \leq \sqrt{2}(2j+1) \leq 2\sqrt{2}(j+1), \quad |\varphi_j''(x)| \leq 2\sqrt{2}(j+1)^2.$$

Consequently, for all $x \geq 0$,

$$\sum_{j=0}^{\ell} \varphi_j^2(x) \leq 2(\ell+1), \quad \sum_{j=0}^{\ell} [\varphi_j'(x)]^2 \leq 8(\ell+1)^3, \quad \sum_{k=0}^{\ell} [\varphi_j''(x)]^2 \leq 8(\ell+1)^5.$$

The functions $\varphi_j'(x)$ and

$$\Phi_j(x) = \int_x^{+\infty} \varphi_j(u)du \tag{31}$$

belong the the space S_j spanned by $(\varphi_0, \ldots, \varphi_j)$ and we can compute their components on the Laguerre basis which allows to compute easily their $\mathbb{L}^2$-norms.

Proposition 8.1.

$$\varphi_0'(x) = -\varphi_0(x), \quad \varphi_j'(x) = -\varphi_j(x) - 2\sum_{k=0}^{j-1} \varphi_k(x), \quad j \geq 1. \tag{32}$$

$$\Phi_0(x) = \varphi_0(x), \quad \Phi_j(x) = \varphi_j(x) + 2(-1)^j \sum_{k=0}^{j-1} (-1)^k \varphi_k(x). \tag{33}$$

Proof. The following equality holds $\varphi_j'(x) = -\varphi_j(x) + 2\sqrt{2}e^{-x}L_j'(2x)$ which is a polynomial function of degree j multiplied by e^{-x}. Thus, it can be decomposed as $\varphi_j'(x) = \sum_{k=0}^{j} a_k^{(j)}\varphi_k(x)$ with

$$\begin{aligned} a_k^{(j)} &= \langle \varphi_j', \varphi_k\rangle = \int_0^{+\infty} \varphi_j'(x)\varphi_k(x)dx = [\varphi_j(x)\varphi_k(x)]_0^{+\infty} - \int_0^{+\infty} \varphi_j(x)\varphi_k'(x)dx \\ &= -\varphi_j(0)\varphi_k(0) - \int_0^{+\infty} \varphi_j(x)\varphi_k'(x)dx = -2 - \langle \varphi_j, \varphi_k'\rangle = -2 - a_j^{(k)} \end{aligned}$$

Notice that this formula is also true when $k = j$: $\langle \varphi_j', \varphi_j\rangle = \int_0^{+\infty} \varphi_j'(x)\varphi_j(x)dx = -(1/2)\varphi_j^2(0) = -2/2 = -1$. Thus we obtain:

$$\begin{aligned} \varphi_j'(x) &= \sum_{k=0}^{j} a_k^{(j)}\varphi_k(x) = -2\sum_{k=0}^{j}\varphi_k(x) - \sum_{k=0}^{j}\langle \varphi_j, \varphi_k'\rangle\varphi_k(x) \\ &= -\varphi_j(x) - 2\sum_{k=0}^{j-1}\varphi_k(x) - \sum_{k=0}^{j-1}\langle \varphi_j, \varphi_k'\rangle\varphi_k(x) \end{aligned}$$

Note that the $\langle \varphi_j, \varphi_k'\rangle$ are zero for $k \le j-1$. Thus we obtain (32).

Then integrating from x to $+\infty$ formula (32) for $j \ge 1$, we obtain $\varphi_j = \Phi_j + 2\sum_{k=0}^{j-1}\Phi_k$. Thus, $\Phi_j = \varphi_j - \varphi_{j-1} - \Phi_{j-1}$. Using that $\Phi_0 = \varphi_0$, we obtain by elementary induction $\Phi_j = \varphi_j + 2\sum_{k=1}^{j}(-1)^k\varphi_{j-k}$, which implies formula (33). □

Proposition 8.2. $\|\varphi_j^{(\ell)}\|^2 \le 2^{\ell+1}(j+1)^{2\ell-1}$ *for* $\ell \ge 1$ *and* $j \ge 0$.

Proof. It follows from formula (32) that

$$^t(\varphi_0^{(\ell)} \ \dots\ \varphi_{m-1}^{(\ell)}) = \mathbf{A}^\ell \quad ^t(\varphi_0 \ \dots\ \varphi_{m-1})$$

where $m \times m$ matrix $\mathbf{A}$ is a lower triangular Toeplitz matrix defined by $[\mathbf{A}]_{i,j} = a(i-j)1_{i-j\ge 0}$, $a(0) = -1$, $a(k) = -2$ for $k \ge 1$. We can write

$$\mathbf{A} = -Id_m - 2\sum_{k=1}^{m-1}\mathbf{J}^k,$$

where $\mathbf{J}$ is the lower triangular Jordan matrix of order m (sub-diagonal coefficients equal to 1, and all others null), which satisfies $\mathbf{J}^m = 0$. The matrix $\mathbf{A}^\ell$ is also lower triangular Toeplitz and we denote by $a^{(\ell)}(i-j)$ its coefficients. Using that $\mathbf{J}^m = 0$ and $\mathbf{A}^\ell = \mathbf{A} \times \mathbf{A}^{\ell-1}$, we get

$$\begin{cases} a^{(\ell)}(0) = -a^{(\ell-1)}(0), a^{(\ell)}(1) = -2a^{(\ell-1)}(0) - a^{(\ell-1)}(1), \\ a^{(\ell)}(k) = -2a^{(\ell-1)}(0) - a^{(\ell-1)}(k) - 2\sum_{p=1}^{k-1} a^{(\ell-1)}(p), \quad \text{for } k \ge 1. \end{cases}$$

Now we have

$$\begin{aligned} \|\varphi_j^{(\ell)}\|^2 &= \sum_{k=0}^{j-1}[a^{(\ell)}(k)]^2 \le [a^{(\ell-1)}(0)]^2 + \sum_{k=0}^{j-1}\sum_{p=0}^{k}[a^{(\ell-1)}(p)]^2(4+1+4(k-1)) \\ &\le \sum_{k=0}^{j-1}(4k+1)\sum_{p=0}^{k}[a^{(\ell-1)}(p)]^2 \\ &\le \|\varphi_j^{(\ell-1)}\|^2(4\frac{j(j-1)}{2}+1) \le 2(j+1)^2\|\varphi_j^{(\ell-1)}\|^2 \end{aligned}$$

As for $\ell=1$, we have $\|\varphi_j^{(1)}\|^2 = 1+4j \le 4(1+j)$, the result follows. □

The following convolution property (formula 22.13.14 in Abramowitz and Stegun (1964)) makes the Laguerre basis relevant in the deconvolution setting

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u)du = 2^{-1/2}\left(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\right) \tag{34}$$

where $\star$ stands for the convolution product.

8.2. **Proof of formula 11 and Lemma 4.1.** The first equality is elementary. For $y \ge 0$,

$$\begin{aligned}\bar{F}_Y(y) &= \int_y^{+\infty} f_Y(z)dz = \int_y^{+\infty}\int_z^{+\infty}\frac{f(x)}{x}dxdz = \int(\int_y^x dz)\frac{f(x)}{x}\mathbb{1}_{y\le x}dx\\ &= \int_y^{+\infty}(x-y)\frac{f(x)}{x}dx = \int_y^{+\infty} f(x)dx - y\int_y^{+\infty}\frac{f(x)}{x}dx = \bar{F}(y) - yf_Y(y).\end{aligned}$$

By (11), $yf_Y(y)$ tends to 0 as both y tends to $+\infty$ and 0. Therefore, integrating by parts yields as t is bounded,

$$\begin{aligned}\int_{\mathbb{R}^+} f_Y(y)(t(y)+yt'(y))dy &= [f_Y(y)yt(y)]_0^{+\infty} - \int_{\mathbb{R}^+} yt(y)(f_Y(y))'dy = -\int_0^{+\infty} yt(y)(-\frac{f(y)}{y})dy\\ &= \int_0^{+\infty} t(y)f(y)dy.\end{aligned}$$

Note that $\mathbb{E}Y^2t^2(Y) \le \mathbb{E}X^2t^2(UX)$. Then,

$$\mathbb{E}X^2t^2(UX) = \int_{x\ge 0, 0\le u\le 1} x^2 1_{[0,1]}(u)f(x)t^2(ux)dxdu = \int_0^{+\infty} xf(x)(\int_0^x t^2(v)dv)dx \le \mathbb{E}(X)\|t\|^2.$$

□

8.3. **Proof of Proposition 4.2.** We have $\|\hat{f}_m - f\|^2 = \|f-f_m\|^2 + \|\hat{f}_m - f_m\|^2$ by Pythagoras Theorem. Also, $\|\hat{f}_m - f_m\|^2 = \sum_{j=0}^{m-1}(\hat{a}_j - a_j)^2$ where $a_j = \mathbb{E}(\hat{a}_j) = \langle f, \varphi_j\rangle$. Now we bound in two different ways the expectation of this last term.
We first assume that $\mathbb{E}(X_1) < +\infty$. We have

$$\begin{aligned}\mathbb{E}(\|\hat{f}_m - f_m\|^2) &= \sum_{j=0}^{m-1}\mathrm{Var}(\hat{a}_j) = \frac{1}{n}\sum_{j=0}^{m-1}\mathrm{Var}[Y_1\varphi_j'(Y_1) + \varphi_j(Y_1)]\\ &\le \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left[(Y_1\varphi_j'(Y_1)+\varphi_j(Y_1))^2\right]\\ &= \frac{1}{n}\sum_{j=0}^{m-1}\left\{\mathbb{E}\left[(Y_1\varphi_j'(Y_1))^2\right] + \mathbb{E}\left[2Y_1\varphi_j'(Y_1)\varphi_j(Y_1) + \varphi_j^2(Y_1)\right]\right\}\end{aligned}$$

Formula (13) applied to $t = \varphi_j^2$ yields

$$\mathbb{E}\left[2Y_1\varphi_j'(Y_1)\varphi_j(Y_1) + \varphi_j^2(Y_1)\right] = \mathbb{E}(\varphi_j^2(X_1)) \le 2.$$

By Formula (32), $\|\varphi_j'\|^2 = 1+4j$, therefore, using Lemma 4.1,

$$\mathbb{E}\left[(Y_1\varphi_j'(Y_1))^2\right] \le \mathbb{E}(X_1)\|\varphi_j'\|^2 = (1+4j)\mathbb{E}(X_1).$$

It follows that

$$\begin{aligned}\mathbb{E}(\|\hat{f}_m - f_m\|^2) &\leq \frac{1}{n}\sum_{j=0}^{m-1}[(4j+1)\mathbb{E}(X_1)+2) = \frac{2m}{n}\left[1+(2m-1)\mathbb{E}(Y_1)\right] \\ &\leq 4\frac{m^2}{n}\mathbb{E}(Y_1) + 2\frac{m}{n},\end{aligned}$$

using that $\mathbb{E}(X_1) = 2\mathbb{E}(Y_1)$. This gives the first bound of Proposition 4.2.

Now, we no longer assume that $\mathbb{E}(X_1) < +\infty$. Relation (30) and the Cauchy-Schwarz inequality imply

$$\begin{aligned}\mathbb{E}[\|\hat{f}_m - f_m\|^2] &= \sum_{j=0}^{m-1}\mathrm{Var}(\hat{a}_j) \leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}[(Y_1\varphi_j'(Y_1)+\varphi_j(Y_1))^2] \\ &= \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\{[[(y\varphi_j(y))']](Y_1)]^2]\} \\ &\leq \frac{1}{n}\sum_{j=0}^{m-1}3\mathbb{E}\left[\left(\frac{j}{2}\varphi_{j-1}(Y_1)\right)^2 + \left(\frac{1}{2}\varphi_j(Y_1)\right)^2 + \left(\frac{j+1}{2}\varphi_{j+1}(Y_1)\right)^2\right].\end{aligned}$$

Then we use that

$$\mathbb{E}\left[\left(\frac{j}{2}\varphi_{j-1}(Y)\right)^2\right] = \int\left(\frac{j}{2}\varphi_{j-1}(y)\right)^2 f_Y(y)dy \leq \|\varphi_{j-1}\|_\infty^2\left(\frac{j}{2}\right)^2\int f_Y(y)dy \leq \frac{1}{2}j^2$$

and it yields

$$\begin{aligned}\mathbb{E}[\|\hat{f}_m - f_m\|^2] &\leq \frac{3}{n}\sum_{j=0}^{m-1}\left(\frac{j^2}{2}+\frac{1}{2}+\frac{(j+1)^2}{2}\right) \\ &\leq \frac{3}{2n}\left(\frac{m^3}{3}+m+m^3\right) \leq \frac{2m^3}{n}+\frac{3m}{2n}. \qquad (35)\end{aligned}$$

This gives the second bound stated in Proposition 4.2.□

8.4. **Proof of Theorem 4.1.** Let us define, for t a function from $\mathbb{R}^+$ into $\mathbb{R}$, the contrast

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n}[t(Y_i)+Y_it'(Y_i)]. \qquad (36)$$

For $t = \sum_{j=0}^{m-1}a_j\varphi_j$, $\gamma_n(t) = \sum_{j=0}^{m-1}(a_j-\hat{a}_j)^2 - \sum_{j=0}^{m-1}\hat{a}_j^2$. Thus, $\hat{f}_m = \underset{t\in\mathcal{S}_m}{\mathrm{argmin}}\ \gamma_n(t)$ and $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$. We notice that

$$\gamma_n(t)-\gamma_n(s) = \|t-f\|^2 - \|s-f\|^2 - 2\nu_n(t-s) \qquad (37)$$

where

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^{n}[\phi_t(Y_i)-\mathbb{E}\phi_t(Y_i)], \quad \text{with} \quad \phi_t(y) = (yt(y))' = t(y)+yt'(y) \qquad (38)$$

and $\mathbb{E}\phi_t(Y_i) = \langle t, f\rangle$. By definition of $\hat{f}_{\hat{m}}$, for all $m \in \mathcal{M}_n$, we have $\gamma_n(\hat{f}_{\hat{m}}) + \text{pen}(\hat{m}) \le \gamma_n(f_m) + \text{pen}(m)$. Set for simplicity of notations $m \vee m' = m^*$ and

$$\mathcal{B}_{m,m'} = \{t \in \mathcal{S}_{m^*}, \|t\| = 1\}. \tag{39}$$

Using (37) yields

$$\begin{aligned}\|\hat{f}_{\hat{m}} - f\|^2 &\le \|f - f_m\|^2 + 2\nu_n(\hat{f}_{\hat{m}} - f_m) + \text{pen}(m) - \text{pen}(\hat{m})\\ &\le \|f - f_m\|^2 + \frac{1}{4}\|\hat{f}_{\hat{m}} - f_m\|^2 + 4\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) + \text{pen}(m) - \text{pen}(\hat{m})\\ &\le \|f - f_m\|^2 + \frac{1}{2}\|\hat{f}_{\hat{m}} - f\|^2 + \frac{1}{2}\|f_m - f\|^2 + 4\sup_{t\in\mathcal{B}_{m,\hat{m}}} \nu_n^2(t) + \text{pen}(m) - \text{pen}(\hat{m})\end{aligned}$$

Therefore,

$$\begin{aligned}\|\hat{f}_{\hat{m}} - f\|^2 &\le 3\|f - f_m\|^2 + 8\sup_{t\in\mathcal{B}_{m,\hat{m}}} \nu_n^2(t) + 2\text{pen}(m) - 2\text{pen}(\hat{m})\\ &\le 3\|f - f_m\|^2 + 2\text{pen}(m) + 8(\sup_{t\in\mathcal{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\hat{m})) + 8p(m,\hat{m}) - 2\text{pen}(\hat{m}). \qquad (40)\end{aligned}$$

Now, $p(m, m')$ must be determined such that

$$\exists\kappa_0,\ \text{a numerical constant, such that}\quad \forall\kappa \ge \kappa_0,\quad 4\,p(m,m') \le \text{pen}(m) + \text{pen}(m'). \tag{41}$$

With such a choice, we obtain

$$\|\hat{f}_{\hat{m}} - f\|^2 \le 3\|f - f_m\|^2 + 4\text{pen}(m) + 8(\sup_{t\in\mathcal{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\hat{m})). \tag{42}$$

The next lemma gives $p(m, m')$. Recall $m^* = m \vee m'$.

Lemma 8.1. *Under the assumption of Theorem 4.1, for $\nu_n(t)$ given by (38) and $p(m, m') = 4(1 + 48\log(2 + m^*))m^*(1 + 2\mathbb{E}(Y_1)m^*)/n \le 4 \times 50\log(2 + m^*)m^*(1 + 2\mathbb{E}(Y_1)m^*)/n$, (41) holds and moreover,*

$$\mathbb{E}\left[(\sup_{t\in\mathcal{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\widehat{m}))_+\right] \le K/n.$$

The proof of Theorem 4.1 is now achieved by taking the expectation of (42), and applying Lemma 8.1.

8.5. **Proof of Lemma 8.1.** First notice that,

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\widehat{m})\right)_+\right] \le \sum_{m'\in\mathcal{M}_n} \mathbb{E}\left(\sup_{t\in\mathcal{B}_{m,m'}} \nu_n^2(t) - p(m,m')\right)_+.$$

We now apply Talagrand's inequality to bound the r.h.s. of the above term (see Theorem A.1). Consider the class $\mathcal{F} = \{\phi_t(x) = t(x) + xt'(x), t \in \mathcal{B}_{m,m'}\}$. We compute the corresponding terms denoted by H^2, v and M in Theorem A.1.
To obtain H^2, we bound $\mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} \nu_n^2(t)]$. For $t \in \mathcal{B}_{m,m'}$, using that $t \mapsto \nu_n(t)$ is linear and $t = \sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle\varphi_j$ with $\sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle^2 = 1$, we get

$$\nu_n^2(t) = \left(\nu_n\left(\sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle\varphi_j\right)\right)^2 = \left(\sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle\nu_n(\varphi_j)\right)^2 \le \sum_{j=0}^{m^*-1}\nu_n^2(\varphi_j).$$

Thus it follows from Proposition 4.2 (see the bound (35)),

$$\begin{aligned}\mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} \nu_n^2(t)] &\le \sum_{j=0}^{m^*-1} \mathbb{E}[\nu_n^2(\varphi_j)] = \sum_{j=0}^{m^*-1} \frac{1}{n}\mathrm{Var}((Y_1\varphi_j'(Y_1)+\varphi_j(Y_1)) \\ &\le \frac{2m^*}{n}(1+2m^*\mathbb{E}(Y_1)) := H^2.\end{aligned}$$

Now, to obtain v, we note that

$$\mathrm{Var}(Y_1t'(Y_1)+t(Y_1)) \le \mathbb{E}[(Y_1t'(Y_1)+t(Y_1))^2] \le nH^2 := v.$$

Finally, using Formula (30) and the fact that the φ_j's are bounded by $\sqrt{2}$, we get

$$\begin{aligned}\sup_{t\in\mathcal{B}_{m,m'}} \sup_y |(yt(y))'| &\le \left(\sum_{j=0}^{m^*-1} (\sup_y (y\varphi_j(y))')^2\right)^{1/2} \\ &\le \left(\sum_{j=0}^{m^*-1} (\sqrt{2}(j+1))^2\right)^{1/2} \le \sqrt{2/3}(m^*)^{3/2} := M.\end{aligned}$$

We set $\alpha = \alpha(m^*) = 24\log(m^*+2)$, $C(\alpha)=1$, $p(m,m') = 2(1+2\alpha(m^*))H^2$. Applying Theorem A.1 yields:

$$\begin{aligned}&\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,m'}} \nu_n^2(t) - 2(1+2\alpha(m^*))\frac{2m^*}{n}(1+2m^*\mathbb{E}(Y_1))\right)_+\right] \\ &\le \frac{C}{n}\left(\frac{2m^*(1+2m^*\mathbb{E}(Y_1))}{(m^*+2)^4} + \frac{(m^*)^3}{n}e^{-C_2\sqrt{\mathbb{E}(Y_1)}n^{1/4}}\right),\end{aligned}$$

for some constants C, C_2, using that any $m\in\mathcal{M}_n$ satisfies $m\le\sqrt{n}$. Consequently,

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,m'}} \nu_n^2(t) - 2(1+2\alpha(m^*))\frac{2m^*}{n}(1+2m^*\mathbb{E}(Y_1))\right)_+\right] \le \frac{K_1}{n}$$

where K_1 is a constant and $p(m,m')$ satisfies $4p(m,m') \le \mathrm{pen(m)} + \mathrm{pen(m')}$ for all $\kappa \ge 2^6.5^2$. □

8.6. **Proof of Proposition 5.1.** We prove that

$$\mathbb{E}(\|\tilde{f}_m - f_m\|^2) \le \|f-f_m\|^2 + \frac{2m\|\mathbf{V}_m^{-1}\|_{\mathrm{op}}^2 \wedge \|f_V\|_\infty \|\mathbf{V}_m^{-1}\|_F^2}{n} \tag{43}$$

where we recall that $\|\mathbf{A}\|_F^2 = \mathrm{Tr}(\,{}^t\mathbf{A}\mathbf{A})$ and $\|\mathbf{A}\|_{\mathrm{op}}^2 = \lambda_{\max}(\,{}^t\mathbf{A}\mathbf{A})$ is the maximal eigenvalue of ${}^t\mathbf{A}\mathbf{A}$. The risk of the estimator can be written as usual

$$\|\tilde{f}_m - f\|^2 = \|f-f_m\|^2 + \|\tilde{f}_m - f_m\|^2$$

where $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$ is the projection of f on $S_m = \mathrm{span}(\varphi_0,\dots,\varphi_{m-1})$ and $\|f-f_m\|^2$ is the square bias term. Next we have

$$\|\tilde{f}_m - f_m\|^2 = \sum_{j=0}^{m-1} (\tilde{a}_j - a_j(f))^2 = \|\mathbf{V}_m^{-1}(\vec{\hat{a}}(Z)_{m-1} - \mathbb{E}(\vec{\hat{a}}(Z)_{m-1}))\|_2^2,$$

where $\|\vec{x}\|_2$ denotes the Euclidean norm of the m-vector $\vec{x}$. So,

$$\begin{aligned}\mathbb{E}(\|\tilde{f}_m - f_m\|^2) &\leq \|\mathbf{V}_m^{-1}\|_{\rm op}^2 \mathbb{E}(\|\vec{\hat{a}}(Z)_{m-1} - \mathbb{E}(\vec{\hat{a}}(Z)_{m-1})\|_2^2) \\ &\leq \|\mathbf{V}_m^{-1}\|_{\rm op}^2 \sum_{j=0}^{m-1} {\rm Var}(\hat{a}_j(Z)) = \frac{1}{n}\|\mathbf{V}_m^{-1}\|_{\rm op}^2 \sum_{j=0}^{m-1} {\rm Var}(\varphi_j(Z_1)) \\ &\leq \frac{1}{n}\|\mathbf{V}_m^{-1}\|_{\rm op}^2 \sum_{j=0}^{m-1} \mathbb{E}(\varphi_j^2(Z_1)) \leq \frac{2m\|\mathbf{V}_m^{-1}\|_{\rm op}^2}{n},\end{aligned}$$

as $\sum_{j=0}^{m-1} \varphi_j^2(x) \leq 2m$, $\forall x \in \mathbb{R}^+$. Therefore we get

$$\mathbb{E}(\|\tilde{f}_m - f\|^2) \leq \|f - f_m\|^2 + 2\frac{m\|\mathbf{V}_m^{-1}\|_{\rm op}^2}{n}.$$

On the other hand,

$$\begin{aligned}\mathbb{E}(\|\tilde{f}_m - f_m\|^2) &= \frac{1}{n}\sum_{\ell} {\rm Var}\left(\sum_j [\mathbf{V}_m^{-1}]_{\ell,j}\varphi_j(Z_1)\right) \leq \frac{1}{n}\sum_{\ell}\mathbb{E}\left[\left(\sum_j [\mathbf{V}_m^{-1}]_{\ell,j}\varphi_j(Z_1)\right)^2\right] \\ &\leq \frac{\|f_Z\|_\infty}{n}\sum_{\ell}\int\left(\sum_j [\mathbf{V}_m^{-1}]_{\ell,j}\varphi_j(z)\right)^2 dz \\ &= \frac{\|f_Z\|_\infty}{n}\sum_{\ell}\sum_j [\mathbf{V}_m^{-1}]_{\ell,j}^2 \leq \frac{\|f_V\|_\infty}{n}\|\mathbf{V}_m^{-1}\|_F^2.\end{aligned}$$

Combining the previous bounds implies (43). As $\|\mathbf{V}_m^{-1}\|_F^2 \leq m\|\mathbf{V}_m^{-1}\|_{\rm op}^2$, we get the result. □

8.7. **Proof of Theorem 5.1.** Let $\mathfrak{M} = \max \mathcal{M}_n$ denote the maximal element of the collection. We follow the lines of the proof of Theorem 4.1, with (36) replaced by

$$\tilde{\gamma}_n(t) = \|t\|^2 - 2\langle t, \tilde{f}_{\mathfrak{M}}\rangle,$$

and (38) by

$$\tilde{\nu}_n(t) = \langle t, \tilde{f}_{\mathfrak{M}} - f_{\mathfrak{M}}\rangle.$$

Note that for $t \in S_m$, then $\tilde{\nu}_n(t) = \langle t, \tilde{f}_m - f_m\rangle$. Thus we get

$$\|\tilde{f}_{\tilde{m}} - f\|^2 \leq 3\|f - f_m\|^2 + 2\widetilde{\rm pen}(m) + 8(\sup_{t\in\mathcal{B}_{m,\tilde{m}}} \tilde{\nu}_n^2(t) - \tilde{p}(m,\tilde{m})) + 8\tilde{p}(m,\tilde{m}) - 2\widetilde{\rm pen}(\tilde{m}).$$

We must determine $\tilde{p}(m, m')$ such that there exists a numerical constant κ_0 for which $4\tilde{p}(m, m') \leq \widetilde{\rm pen}(m) + \widetilde{\rm pen}(m')$ for all $\kappa \geq \kappa_0$. The next lemma gives $\tilde{p}(m, m')$ and allows to deduce κ_0.

Lemma 8.2. *Under the assumptions of Theorem 5.1, for*

$$\tilde{p}(m, m') = 2(2 \vee \|f_V\|_\infty)(1 + 2c\log(2 + \|\mathbf{V}_{m^*}^{-1}\|_F^2))\frac{\|\mathbf{V}_{m^*}^{-1}\|_F^2}{n}, \quad c \geq \max(3/b, 21^2/2b^2)$$

where b is the constant given in Theorem A.1, we have

$$\mathbb{E}\left[\left(\sup_{\vec{t}\in B(\tilde{m},m)} \tilde{\nu}_n^2(t) - \tilde{p}(m,\tilde{m})\right)_+\right] \leq \frac{K}{n} \tag{44}$$

Finally, we obtain that $\forall m \in \mathcal{M}_n$,

$$\mathbb{E}(\|\tilde{f}_{\tilde{m}} - f\|^2) \quad \leq \quad 3\|f - f_m\|^2 + 4\widetilde{\text{pen}}(m) + 8\frac{K}{n},$$

which ends the proof of Theorem 5.1. □

8.8. **Proof of Lemma 8.2.** The proof of (44) follows the line of the proof of Proposition 7.1 in Mabon (2015). We detail it and proceed as in the proof of Lemma 8.1. For $t \in \mathcal{B}(m, m')$ and $m^* = m \vee m'$, we have

$$\tilde{\nu}_n(t) = \langle t, \tilde{f}_{m^*} - f_{m^*}\rangle = \frac{1}{n}\sum_{i=1}^{n}[\psi_t(Z_i) - \mathbb{E}(\psi_t(Z_i))], \quad \psi_t(x) = \sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle[\mathbf{V}_{m^*}^{-1}\vec{\varphi}_{m^*-1}(x)]_j$$

where $\vec{\varphi}_{m-1}(x) = {}^t(\varphi_0(x), \ldots, \varphi_{m-1}(x))$ and $[\vec{x}]_j$ denotes the jth coordinate of vector $\vec{x}$.

We compute the terms H^2, v and M of Theorem A.1 for the class $\mathcal{F} = \{\psi_t(.), t \in \mathcal{B}(m, m')\}$. For H^2, we bound

$$\mathbb{E}\left(\sup_{t\in\mathcal{B}(m,m')}\tilde{\nu}_n^2(t)\right) \leq \sum_{j=0}^{m^*-1}\mathbb{E}\left(\tilde{\nu}_n^2(\varphi_j)\right) \leq \sum_{j=0}^{m^*-1}\mathbb{E}\left(\langle\varphi_j, \tilde{f}_{m^*} - f_{m^*}\rangle^2\right) = \mathbb{E}(\|\tilde{f}_{m^*} - f_{m^*}\|^2).$$

From Proposition 5.1, we deduce $H^2 = (2 \vee \|f_V\|_\infty)\|\mathbf{V}_{m^*}^{-1}\|_F^2/n$. Clearly, $v = nH^2$. To obtain M, we compute

$$\sup_{t\in\mathcal{B}(m',m)}\sup_x|\psi_t(x)| \leq \sup_x\|\mathbf{V}_{m^*}^{-1}\vec{\varphi}_{m^*-1}(x)\|_2 \leq \|\mathbf{V}_{m^*}^{-1}\|_{\text{op}}\sqrt{2m^*} := M.$$

Let $\alpha(m^*) = c\log(2 + \|\mathbf{V}_{m^*}^{1}\|_F^2)$, and let us apply Theorem A.1:

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\in B(m',m)}\tilde{\nu}_n^2(t) - 2(1 + 2\alpha(m^*))H^2\right)_+ \\
\leq\ & \frac{C}{n}\left(\|\mathbf{V}_{m^*}^{-1}\|_F^2\exp(-b\alpha(m^*)) + \frac{m^*\|\mathbf{V}_{m^*}^{-1}\|_{\text{op}}^2}{n}\exp\left(-\frac{\sqrt{2}b}{7}\frac{\sqrt{\alpha(m^*)n}\|\mathbf{V}_{m^*}^{-1}\|_F}{\sqrt{m^*}\|\mathbf{V}_{m^*}^{-1}\|_{\text{op}}}\right)\right) \\
\leq\ & \frac{C}{n}\left(\frac{1}{\|\mathbf{V}_{m^*}^{-1}\|_F^{2bc-2}} + \|V_{m^*}^{-1}\|_F^2\exp\left(-\frac{\sqrt{2}b}{7}\sqrt{\alpha(m^*)\log(n+2)}\right)\right),
\end{aligned}$$

where we have used that $m^* \leq n/\log(n+2)$ and $\|\mathbf{V}_m^{-1}\|_{\text{op}}^2 \leq \|\mathbf{V}_m^{-1}\|_F^2$. Therefore

$$\mathbb{E}\left(\sup_{t\in\mathcal{B}(m,m')}\tilde{\nu}_n^2(t) - 2(1 + 2\alpha(m^*))H^2\right)_+ \leq \frac{C}{n}\left(\frac{1}{\|\mathbf{V}_{m^*}^{-1}\|_F^{2bc-2}} + \frac{1}{\|\mathbf{V}_{m^*}^{-1}\|_F^{\sqrt{2c}b/7-2}}\right).$$

For $c \geq \max(3/b, 21^2/2b^2)$ and as $\|\mathbf{V}_m^{-1}\|_F^2 \geq 2m^*/a_0^2(f_V)$, which is the sum of squares of diagonal terms of $\mathbf{V}_m^{-1}$, we get

$$\mathbb{E}\left(\sup_{t\in\mathcal{B}(m,m')}\tilde{\nu}_n^2(t) - 2(1 + 2\alpha(m^*))H^2\right)_+ \leq \frac{C'}{n}\frac{1}{(m^*)^4}$$

so that

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in\mathcal{B}(m,m')}\tilde{\nu}_n^2(t) - 2(1 + 2\alpha(m^*))H^2\right)_+ \leq C\text{"}/n.$$

This concludes of Lemma 8.2. □

8.9. **Proof of Proposition 6.1.** The bias variance decomposition is $\mathbb{E}(\|\check{f}_m - f\|^2) = \|f - f_m\|^2 + \mathbb{E}(\|\check{f}_m - f_m\|^2)$ as clearly $\mathbb{E}(\check{f}_m) = f_m$. Next,

$$\mathbb{E}(\|\check{f}_m - f_m\|^2) = \mathbb{E}\left(\left\|\mathbf{V}_m^{-1}(\vec{\check{a}}_{m-1}(W) - \vec{a}_{m-1}(f_{X+V}))\right\|_2^2\right)$$

where $f_{X+V} = f \star f_V$ is the density of $X_1 + V_1$. Thus

$$\begin{aligned}\mathbb{E}(\|\check{f}_m - f_m\|^2) &\leq \|\mathbf{V}_m^{-1}\|_{\mathrm{op}}^2 \mathbb{E}\left(\|\vec{\check{a}}_{m-1}(W) - \vec{a}_{m-1}(f_{X+V})\|_2^2\right) \\ &= \|\mathbf{V}_m^{-1}\|_{\mathrm{op}}^2 \frac{1}{n}\sum_{j=1}^{m-1} \mathrm{Var}\left(W_1\varphi_j'(W_1) + \varphi_j(W_1)\right).\end{aligned}$$

The proof is ended as the proof of Proposition 4.2. □

8.10. **Proof of Proposition 6.2.** As in the proof of Proposition 5.1 (see Inequality(43), we have

$$\begin{aligned}\mathbb{E}(\|\check{f}_m - f_m\|^2) &\leq \|f - f_m\|^2 + \frac{2(m+1)\|\mathbf{K}_m\|_{\mathrm{op}}^2 \wedge \|f_V\|_\infty \|\mathbf{K}_m\|_F^2}{n} \\ &\leq \|f - f_m\|^2 + (2 \vee \|f_V\|_\infty)\frac{(m+1)\|\mathbf{K}_m\|_{\mathrm{op}}^2 \wedge \|\mathbf{K}_m\|_F^2}{n} \\ &= \|f - f_m\|^2 + (2 \vee \|f_V\|_\infty)\frac{\|\mathbf{K}_m\|_F^2}{n} \\ &\leq \|f - f_m\|^2 + (2 \vee \|f_V\|_\infty)\frac{\|\mathbf{H}_m\|_{\mathrm{op}}^2 \|\mathbf{V}_m^{-1}\|_F^2}{n}, \qquad (45)\end{aligned}$$

by using that $\|\mathbf{K}_m\|_F^2 \leq \|\mathbf{H}_m\|_{\mathrm{op}}^2 \|\mathbf{V}_m^{-1}\|_F^2$, see Magnus and Neudecker (1988), sec.6 p.231. □

References

[1] Abbaszadeh, M., Chesneau, C. and Doosti, H. (2012) Nonparametric estimation of density under bias and multiplicative censoring via wavelet methods. *Statist. Probab. Lett.* **82**, 932-941.

[2] Abbaszadeh, M, Chesneau, C. and Doosti, H. (2013) Multiplicative censoring: estimation of a density and its derivatives under the Lp-risk. *REVSTAT* **11**, 255-276.

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

[4] Asgharian, M., Carone, M., Fakoor, V. (2012) Large-sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.* **40**, 159-187.

[5] Andersen, K. E. and Hansen, M. B. (2001) Multiplicative censoring: density estimation by a series expansion approach. *J. Statist. Plann. Inference* **98**, 137-155.

[6] D. Belomestny, F. Comte and V. Genon-Catalot. (2016) Nonparametric Laguerre estimation in the multiplicative censoring model. *Electron. J. Stat.* **10**, 3114-3152.

[7] Bongioanni, B. and Torrea, J.L. (2009). What is a Sobolev space for the Laguerre function system? *Studia Mathematica* **192** (2), 147-172.

[8] Brunel, E., Comte, F. and Genon-Catalot, V. (2016) Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST* **25**, 570-590.

[9] Chesneau, C. (2013) Wavelet estimation of a density in a GARCH-type model. *Comm. Statist. Theory Methods* **42**, 98-117.

[10] Comte, F., Cuenod, C.-A., Pensky, M. and Rozenholc, Y. (2017). Laplace deconvolution and its application to Dynamic Contrast Enhanced imaging. *J. R. Stat. Soc.* Ser. **B**, **79**, 69-94.

[11] Comte, F. and Dion, C. (2017) Laguerre estimation under constraint at a single point. Preprint MAP5 2017-04.

[12] Comte, F. and Duval, C. (2016). Statistical inference for renewal processes. Preprint Hal and Preprint MAP5 2016-20.

[13] Comte, F. and Genon-Catalot, V. (2015) Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.* **9**, 1113-1149.
[14] Comte, F. and Mabon, G. (2016). Laguerre deconvolution with unkown matrix operator. Preprint HAL-01416412 and Preprint MAP5 2016-33.
[15] van Es, B., Spreij, P. and van Zanten, H. (2003). Nonparametric volatility density estimation. *Bernoulli* **9**, 451-465.
[16] van Es, B., Jongbloed, G. and van Zuijlen, M. (1998) Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.* **26**, 2395-2406.
[17] Jirak, M., Meister, A. and Reiss, M. (2014) Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42**, 1970-2002.
[18] Jongbloed, G. (1998) Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica* **52**, 6-17.
[19] Mabon, G. (2016). Adaptive deconvolution on the nonnegative real line. Preprint HAL hal-01076927, version 2. To appear in *Scandinavian Journal of Statistics*.
[20] Magnus, J.R. and Neudecker, H. (1988)*Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.
[21] Shen, J. (2000). Stable and efficient spectral methods in unbounded domains using Laguerre functions. *SIAM J. Numer. Anal.* **38**, 1113-1133.
[22] Vardi, Y. (1989) Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76**, 751-761.
[23] Vardi, Y. and Zhang, C.-H. (1992) Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022-1039.

Appendix A. Talagrand's inequality

The following result follows from the Talagrand concentration inequality.

Theorem A.1. *Consider $n \in \mathbb{N}^*$, $\mathcal{F}$ a class at most countable of measurable functions, and $(X_i)_{i\in\{1,\dots,n\}}$ a family of real independent random variables. Define, for $f \in \mathcal{F}$, $\nu_n(f) = (1/n)\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)])$, and assume that there are three positive constants M, H and v such that $\sup_{f\in\mathcal{F}} \|f\|_\infty \leq M$,*
$\mathbb{E}[\sup_{f\in\mathcal{F}} |\nu_n(f)|] \leq H$, and $\sup_{f\in\mathcal{F}} (1/n)\sum_{i=1}^n \mathrm{Var}(f(X_i)) \leq v$. Then for all $\alpha > 0$,

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{f\in\mathcal{F}}|\nu_n(f)|^2 - 2(1+2\alpha)H^2\right)_+\right] &\leq \frac{4}{b}\left(\frac{v}{n}\exp\left(-b\alpha\frac{nH^2}{v}\right)\right.\\ &+ \left.\frac{49M^2}{bC^2(\alpha)n^2}\exp\left(-\frac{\sqrt{2}bC(\alpha)\sqrt{\alpha}}{7}\frac{nH}{M}\right)\right)\end{aligned}$$

with $C(\alpha) = (\sqrt{1+\alpha} - 1) \wedge 1$, and $b = \frac{1}{6}$.

By density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \to \nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.