ADAPTIVE DECONVOLUTION OF LINEAR FUNCTIONALS ON THE NONNEGATIVE REAL LINE

Gwennaëlle Mabon

ADAPTIVE DECONVOLUTION OF LINEAR FUNCTIONALS ON THE NONNEGATIVE REAL LINE

GWENNAËLLE MABON

CREST - ENSAE
3 avenue Pierre Larousse
92245 Malakoff, France

MAP5*, Université Paris Descartes*
45 rue des Saints-Pères
75006 Paris, France

ABSTRACT. In this paper we consider the convolution model $Z = X + Y$ with X of unknown density f, independent of Y, when both random variables are nonnegative. Our goal is to estimate linear functionals of f such as $\langle \psi, f \rangle$ for a known function ψ assuming that the distribution of Y is known and only Z is observed. We propose an estimator of $\langle \psi, f \rangle$ based on a projection estimator of f on Laguerre spaces, present upper bounds on the quadratic risk and derive the rate of convergence in function of the smoothness of f, g and ψ. Then we propose a nonparametric data driven strategy, inspired Goldenshluger and Lepski (2011) method to select a relevant projection space. This methodology permits to estimate the cumulative distribution function of X for instance. In addition it is adapted to the pointwise estimation of f. We illustrate the good performance of the new method through simulations. We also test a new approach for choosing the tuning parameter in Goldenshluger-Lepski data driven estimators following ideas developed in Lacour and Massart (2015).

Keywords. Deconvolution. Nonparametric density estimation. Adaptive estimation. Linear functionals. Laguerre basis. Mean squared risk.

AMS Subject Classification 2010: Primary 62G07; secondary 62G99, 62J99.

1. INTRODUCTION

In many experiments statisticians do not observe directly the variable of interest X ; instead they have at hand observations of Z, equal to the sum of X and another random variable Y. In various situations, Y can modelize a measurement error, and as such, is assumed to be symmetric or centered. But we can also, in reliability fields, observe the sum of the lifetimes of two components, the second one being well known. In survival analysis, X can be the time of infection of a disease and Y the incubation time, and this happens in the so called back calculation problems in AIDS research. In these last two cases, distributions of X and Y are $\mathbb{R}^+$-supported. In this situation, Y is not considered as a noise but as an additional nuisance process. Indeed a noise distribution is assumed to be centered, which is not the case anymore. Thus we consider the following model

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

where the X_i's are independent identically distributed (i.i.d.) nonnegative random variables (r.v.) with unknown density f. The Y_i's are also i.i.d. nonnegative variables with known density g. We denote by h the density of the Z_i's. Moreover the X_i's and the Y_i's are assumed to be independent, they are not observed. Our goal is to estimate linear functionals of f defined by

$\vartheta(f) = \langle \psi, f \rangle = \mathbb{E}[\psi(X_1)]$ with ψ a known function, from observations $Z_1, \ldots, Z_n$.

Those assumptions imply that, in Model (1), $h(x) = (f \star g)(x)$ where $(\varphi \star \psi)(x) = \int \varphi(x - u)\psi(u) \, \mathrm{d}u$ denotes the convolution product. This setting matches convolution models which is a classical topic in nonparametric statistics. The problem of recovering the signal distribution f when it is observed with an additive noise with known error distribution on the real line, has been extensively studied, see for rates of convergence and their optimality for kernel estimators Carroll and Hall (1988) and Fan (1991), for wavelets strategy Pensky and Vidakovic (1999), for projection strategies with penalization Comte et al. (2006) and Butucea and Tsybakov (2008a,b) for the study of sharp asymptotic optimality ; with an unknown error density, see Neumann (1997) for kernel estimator and minimax optimality, Johannes (2009) for minimax optimality under various regularity conditions on f, Comte and Lacour (2011) and Kappus and Mabon (2014) for projection strategies with penalization.

The problem of one-sided error in the convolution model has been first introduced by Groeneboom and Wellner (1992) under a constraint of monotonicity of the cumulative distribution function (c.d.f.). They concentrate on deriving nonparametric maximum likelihood estimators (NPMLE) of the c.d.f. Some particular cases have been tackled as uniform or exponential deconvolution by Groeneboom and Jongbloed (2003) and Jongbloed (1998) who propose NPMLE of the c.d.f. of the X_i's, which have explicit expressions. For other cases van Es et al. (1998) circumvent the lack of explicit expression for the NPMLE by proposing an isotonic inverse estimator. In this paper, we do not use the approach of the NPLME.

Moreover Model (1) is also related to the field of mixture models, see Roueff and Rydén (2005) and Rebafka and Roueff (2010) who study in particular mixtures of Exponential and Gamma distributions, which are contained in our framework. These models play a major role in natural sciences phenomena of discharge or disexcitation as in radioactive decays, the electric discharge of a capacitor or the temperature difference between two objects.

On one hand the problem of estimating linear functionals in linear models has been widely studied especially in the setting of the white noise model, see Cai and Low (2003, 2005) and Laurent et al. (2008) for instance. On the other hand adaptive estimation of linear functionals has not been much studied in the convolution model. Butucea and Comte (2009), in the context of Model (1) when the variables X_i's and Y_i's are $\mathbb{R}$-supported, propose a general estimator of $\vartheta(f)$ using a spectal cut-off in the Fourier domain when the random variables are distributed on the real line. They apply it to the pointwise estimation of the density on the real line and prove that their losses in their adaptive procedure is optimal in the minimax sense. They do not prove it for their general estimator. They also apply their adaptive procedure to the pointwise Laplace transform estimation and the stochastic volatility model. Recently Pensky (2015) has improved their results by deriving minimax lower bounds for estimators of a general linear functional of the deconvolution density. The author even extends the techniques when ψ is not integrable or square integrable and considers the possibility that the vector of observations is sparse (i.e. has a lot of zeros).

In this paper, our goal is to establish a specific method for the estimation of linear functional when we know that the random variables are $\mathbb{R}^+$-supported. In that way we can cite the work of Mabon (2014) who proposes a specific estimation for nonnegative variables in Model (1). The methodology, based on a penalized projection strategy in a Laguerre basis, allows to estimate the density and survival function of X. Moreover this work has showed that for certain classes of density, the Laguerre estimation in the convolution model (1) gives faster rates of convergence of the estimators than with the classical method based on Fourier inversion as obtained in Comte et al. (2006) for instance. In particular, it is verified for mixed Gamma distributions. The contribution of this paper is to extend the particular methodology developed for nonnegative variables to estimate linear functionals of f. Thanks to this new methodology, we can derive the pointwise estimation of the probability density function (p.d.f.) f, of the c.d.f of X and also its Laplace transform.

In Section 2, we explain how projection coefficient estimators in the Laguerre basis can be used to define estimators of linear functionals. Next we lead a theoretical study of estimators of $\vartheta(f)$ and derive upper bound on the quadratic risk and rates of convergence in function of the smoothness of f, g and ψ when these three functions belong to $\mathbb{L}^2(\mathbb{R}^+)$. We show that under some assumptions on g or ψ, the parametric rate of estimation can be achieved. In Section 3 we propose a nonparametric data driven strategy, following Goldenshluger and Lepski (2011) method, for selecting a relevant projection space. In Section 4, when f and g belong to $\mathbb{L}^2(\mathbb{R}^+)$ we adapt the procedure to pointwise estimation of f in the setting of Model (1) in this case ψ is specified and do not belong to $\mathbb{L}^2(\mathbb{R}^+)$ anymore. The method is then illustrated through simulations. Following Lacour and Massart (2015), we apply a new procedure to choose the tuning parameter appearing in the penalization term of the data-driven estimator. This procedure is promising and shows good results.

To sum up the paper is organized as follows. In Section 2, we give the notations, specify the statistical model and estimation procedures for projection estimators of f and $\vartheta(f)$, upper bound on the pointwise mean squared error. In Section 3, we propose a new data-driven procedure for choosing the tuning parameter linear functionals of f which allows in particular to estimate the cumulative distribution function of X. In Section 4, we derive an adaptive procedure for the pointwise estimation of f and provide an empirical study on simulations. All the proofs are postponed to Section 5.

2. Statistical model and estimation procedure

2.1. **Notations.** For two real numbers a and b, we denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For two functions $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ belonging to $\mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the $\mathbb{L}^2$ norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ defined by $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)dx$.

Let d be an integer, for two vectors $\vec{u}$ and $\vec{v}$ belonging to $\mathbb{R}^d$, we denote $\|\vec{u}\|_2$ the Euclidean norm defined by $\|\vec{u}\|_2^2 = {}^t\vec{u}\vec{u}$ where ${}^t\vec{u}$ is the transpose of $\vec{u}$. The scalar product between $\vec{u}$ and $\vec{v}$ is $\langle \vec{u}, \vec{v} \rangle_2 = {}^t\vec{u}\vec{v} = {}^t\vec{v}\vec{u}$. We introduce the square of the spectral norm of a matrix $\mathbf{A}$: $\varrho^2(\mathbf{A}) = \lambda_{\max}({}^t\mathbf{A}\mathbf{A})$ where $\lambda_{\max}({}^t\mathbf{A}\mathbf{A})$ is the largest eigenvalue of ${}^t\mathbf{A}\mathbf{A}$ in absolute value. We also introduce the following operator norm: $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq d} \sum_{i=1}^{d} |a_{ij}|$.

2.2. **Laguerre basis.** We define the Laguerre basis as

$$\forall k \in \mathbb{N}, \forall x \geq 0, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^{k}(-1)^j \binom{k}{j}\frac{x^j}{j!}. \tag{2}$$

The Laguerre polynomials L_k defined by Equation (2) are orthonormal with respect to the weight function $x \mapsto e^{-x}$ on $\mathbb{R}^+$. In other words, $\int_{\mathbb{R}^+} L_k(x)L_{k'}(x)e^{-x}\, dx = \delta_{k,k'}$ where $\delta_{k,k'}$ is the Kronecker symbol. Thus $(\varphi_k)_{k \geq 0}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$. We can also notice that the Laguerre basis verifies the following inequality for any integer k

$$\sup_{x \in \mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_\infty \leq \sqrt{2}. \tag{3}$$

We also introduce the space $\mathcal{S}_m = \mathrm{Span}\{\varphi_0, \ldots, \varphi_{m-1}\}$. For a function p in $\mathbb{L}^2(\mathbb{R}^+)$, we note

$$p(x) = \sum_{k \geq 0} a_k(p)\varphi_k(x) \quad \text{where} \quad a_k(p) = \int_{\mathbb{R}^+} p(u)\varphi_k(u)\, du.$$

According to formula 22.13.14 in Abramowitz and Stegun (1964), what makes the Laguerre basis relevant in our deconvolution setting is the relation

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u)\, du = 2^{-1/2}\left(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\right) \tag{4}$$

where $\star$ stands for the convolution product.

2.3. **Projection estimator of the linear functional.**

2.3.1. *Statistical procedure.* The goal is to estimate $\vartheta(f) = \langle \psi, f \rangle = \mathbb{E}[\psi(X_1)]$ when only the Z_i's are observed for $i = 1, \ldots, n$. If f and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$ then they can be decomposed on the Laguerre basis, it yields that

$$\vartheta(f) = \langle \psi, f \rangle = \sum_{k=0}^{\infty} a_k(\psi) a_k(f).$$

Since ψ is known and the φ_k's are known, coefficients $a_k(\psi)$ can be explicitly computed. Therefore we have to define estimators $\hat{a}_k$ of $a_k(f)$ but an estimator of $\vartheta(f)$ would not be $\widehat{\vartheta}(f) = \sum_{k=0}^{\infty} a_k(\psi) \hat{a}_k$. Indeed the last sum is not necessarily convergent nor computable in practice, so we need to truncate it. Let us consider $\vartheta_m(f)$ defined by

$$\vartheta_m = \vartheta_m(f) = \sum_{k=0}^{m-1} a_k(\psi) a_k(f).$$

Let us define ψ_m and f_m respectively the projection of ψ and f on the space $\mathcal{S}_m$. It yields that

$$\langle \psi_m, f_m \rangle = \vartheta_m = \langle \psi, f_m \rangle = \langle \psi_m, f \rangle.$$

Then we can estimate ϑ_m by

$$\hat{\vartheta}_m = \vartheta(\hat{f}_m) = \langle \psi_m, \hat{f}_m \rangle = \sum_{k=0}^{m-1} a_k(\psi) \hat{a}_k, \tag{5}$$

with $\hat{f}_m = \sum_{k=0}^{m-1} \hat{a}_k \varphi_k$. The coefficients $\hat{a}_k$ are computed as follows

$${}^t(\hat{a}_0 \ldots \hat{a}_{m-1}) := \vec{\hat{f}}_m = \mathbf{G}_m^{-1} \vec{\hat{h}}_m, \quad \text{with} \quad \vec{\hat{h}}_m = {}^t(\hat{a}_0(Z) \ldots \hat{a}_{m-1}(Z)) \tag{6}$$

where $\hat{a}_k(Z) = (1/n) \sum_{i=1}^{n} \varphi_k(Z_i)$ for $k = 0, \ldots, m-1$ are estimators of $a_k(h)$ and

$$[\mathbf{G}_m] := \left([\mathbf{G}_m]_{i,j} \right)_{0 \le i,j \le m-1} = \begin{cases} 2^{-1/2} a_0(g) & \text{if } i = j, \\ 2^{-1/2} \left(a_{i-j}(g) - a_{i-j-1}(g) \right) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

We know from Mabon (2014) that formula (4) implies that $\hat{\vartheta}_m$ is an unbiased estimator of ϑ_m, see details therein, i.e. $\mathbb{E}\hat{\vartheta}_m = \vartheta_m$.

Before deriving the statistical properties of our estimator let us give some examples to show that the method is feasible. Since the vector $\vec{\hat{h}}_m$ is computed directly from the data, let us concentrate on the matrix $\mathbf{G}_m$ and derive, if possible, an explicit expression of its coefficients.

Example 1. If g is an Exponential distribution of parameter $\lambda > 0$, then for all integer k we have

$$a_k(g) = \int_{\mathbb{R}^+} g(u) \varphi_k(u) \, du = \sqrt{2} \lambda \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^j}{j!} \int_{\mathbb{R}^+} (2u)^j e^{-(1+\lambda)u} \, du$$

$$= \frac{\sqrt{2}\lambda}{\lambda + 1} \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^j}{(1+\lambda)^j} = \frac{\sqrt{2}\lambda}{\lambda + 1} \frac{(\lambda - 1)^k}{(\lambda + 1)^k}.$$

We can compute the coefficients of the matrix as

$$a_0(g) = \frac{\sqrt{2}\lambda}{1 + \lambda} \quad \text{and} \quad a_{i-j}(g) - a_{i-j-1}(g) = -2\sqrt{2}\lambda \frac{(\lambda - 1)^{i-j-1}}{(\lambda + 1)^{i-j+1}} \quad \text{if} \quad j < i.$$

In that case, we can also inverse the matrix as follows

$$[\mathbf{G}_m^{-1}]_{i,j} = \frac{\lambda + 1}{\lambda} \begin{cases} 1 & \text{if } i = j, \\ \dfrac{2}{\lambda + 1} & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Thus we can explicitly compute the coefficients $\hat{a}_k$ with formula (6) and $\hat{\vartheta}_m$ is easy to obtain by (5). Extension to the case of a Gamma distribution of parameters p an integer, $p \geq 1$ and $\lambda > 0$, can be obtained by using the more general formula

$$a_k(g) = \int_{\mathbb{R}^+} \lambda^p u^{p-1} e^{-\lambda u} \sqrt{2} e^{-u} L_k(2u) \, \mathrm{d}u = \sqrt{2} \lambda^p \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^j}{j!} \int_{\mathbb{R}^+} u^{p+j-1} e^{-(1+\lambda)u} \, \mathrm{d}u$$

$$= \frac{\sqrt{2}\lambda^p}{(1+\lambda)^p} \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^j}{(1+\lambda)^j} \frac{(p+j-1)!}{j!} = \frac{\sqrt{2}\lambda^p}{(1+\lambda)^p} S_{p-1,k}\left(\frac{2}{1+\lambda}\right). \qquad (9)$$

with

$$S_{p-1,k}(x) = \frac{1}{(p-1)!} \frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}} \left[x^{p-1}(1-x)^k \right].$$

2.3.2. *Discussion: noticeable linear functionals.* We propose to point out some interesting linear functionals. Some can be directly obtained with the previous methodology since they can be written as an inner product of f with a square integrable function. Others do not fit this particular assumption but can nonetheless be estimated. As said in the introduction, Mabon (2014) studies the problem of global estimation in Model (1) and proposes adaptive estimators of the probability density and survival functions. Thanks to the particular procedure for linear functionals, we can derive the pointwise estimation of the c.d.f., p.d.f and Laplace transform.

a) Cumulative distribution function.
Let us define $\psi(x) = \mathbb{1}_{[0,c]}(x)$ for $c > 0$. The function ψ belongs to $\mathbb{L}^2(\mathbb{R}^+)$. This definition of ψ enables us to consider the estimation of the c.d.f. of f denoted by F. Indeed we can write the c.d.f. as

$$F(c) = \int_0^c f(u) \, \mathrm{d}u = \int_0^\infty \mathbb{1}_{[0,c]}(u) f(u) \, \mathrm{d}u = \langle \psi, f \rangle.$$

The methodology of Subsection 2.3.1 can therefore be applied.

b) Probability density function.
Let x_0 be a nonnegative real number, we have

$$f(x_0) = \int_0^\infty f(u) \delta_{\{x_0\}}(u) \, \mathrm{d}u = \langle \delta_{\{x_0\}}(.), f \rangle = \langle \psi, f \rangle,$$

with $\psi(x) = \delta_{\{x_0\}}(x)$ and δ stands for the Dirac measure. Obviously in that case ψ is not square integrable. This case is studied in details in Section 4.

c) Laplace transform.
Let t be a postive real number and consider the function $\psi_t(x) = e^{-tx}$ which is square integrable, we can write the Laplace transform of X as follows

$$\mathbb{E}e^{-tX} = \int_0^\infty e^{-tu} f(u) \, \mathrm{d}u = \int_0^\infty e^{-tu} \sum_{k \geq 0} a_k(f) \varphi_k(u) \, \mathrm{d}u$$

$$= \sum_{k \geq 0} a_k(f) \int_0^\infty e^{-tu} \varphi_k(u) \, \mathrm{d}u = \sum_{k \geq 0} a_k(f) a_k(\psi_t).$$

We can apply the procedure of Subsection 2.3.1. Moreover we explain in Subsection 2.4.4 how to estimate the Laplace transform with a parametric rate.

2.4. **Upper bound on the mean squared error and rates of convergence.**

2.4.1. *Risk bound.* Let us notice that if $f, g \in \mathbb{L}^2(\mathbb{R}^+)$ then h is bounded. Indeed

$$\|h\|_\infty = \sup_{x \geq 0} |f \star g(x)| \leq \|f\|\|g\|.$$

Then we can state the following Proposition:

Proposition 2.1. *For f, g and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, for $\mathbf{G}_m$ defined by Equation (7) and $\hat{\vartheta}_m$ defined by Equation (5), the following result holds*

$$\mathbb{E}\left[(\vartheta - \hat{\vartheta}_m)^2\right] \leq (\vartheta - \vartheta_m)^2 + \|h\|_\infty \frac{\|{}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2}{n}. \tag{10}$$

The terms of the right-hand side of Equation (10) correspond to a squared bias variance decomposition. We can notice by applying the Cauchy-Schwarz inequality that

$$(\vartheta - \vartheta_m)^2 = \left(\sum_{k \geq m} a_k(\psi) a_k(f)\right)^2 \leq \|\psi - \psi_m\|^2 \|f - f_m\|^2 = \sum_{k \geq m} a_k^2(\psi) \sum_{k \geq m} a_k^2(f). \tag{11}$$

Therefore the first term gets smaller when m increases. On the contrary noticing that

$$\|{}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 \leq \|\psi\|^2 \varrho^2(\mathbf{G}_m^{-1}),$$

we obtain a bound on the variance which is nondecreasing with m (see Lemma 3.4 in Mabon (2014)). In that case we have to tackle the usual bias variance trade-off.

We can note that if there exist some function ψ and/or some density function g such that $\|{}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2$ is bounded by a constant not depending on m, then we can estimate the resulting linear functionals with a parametric rate. First let us consider the general rates of convergence.

2.4.2. *General rates of convergence.* In order to derive general rates of convergence of estimators $\hat{\vartheta}_m$ defined by Equation (5), we need to evaluate the smoothness of the signal along with the order of $\varrho^2(\mathbf{G}_m^{-1})$. In the first place, we assume that f belongs to a Laguerre-Sobolev space defined as

$$W^s(\mathbb{R}^+, L) = \left\{p : \mathbb{R}^+ \to \mathbb{R}, \, p \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s a_k^2(p) \leq L < +\infty\right\} \quad \text{with} \quad s \geq 0 \tag{12}$$

where we recall that $a_k(p) = \langle p, \varphi_k \rangle$. Bongioanni and Torrea (2009) have introduced Laguerre-Sobolev spaces but the link with the coefficients of a function on a Laguerre basis was done by Comte and Genon-Catalot (2015) when s is an integer. Indeed, let s be an integer, $f \in W^s(\mathbb{R}^+, L)$ is equivalent to the fact that f admits derivatives up to order $s - 1$ with $f^{(s-1)}$ absolutely continuous and for $0 \leq k \leq s - 1$, $x^{(k+1)/2} \sum_{j=0}^{k+1} \binom{k+1}{j} f^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$. For more details we refer to section 7 of Comte and Genon-Catalot (2015).

Thanks to those spaces we can evaluate the bias order. Let $f \in W^s(\mathbb{R}^+, L)$ and $\psi \in W^\alpha(\mathbb{R}^+, L')$ with $\alpha, s > 1$, starting from Equation (11) we have

$$(\vartheta - \vartheta_m)^2 \leq \sum_{k \geq m} a_k^2(\psi) k^{-\alpha} k^\alpha \sum_{k \geq m} a_k^2(f) k^{-s} k^s \leq LL' m^{-(s+\alpha)}.$$

Then we must evaluate the variance term of Equation (10) which means assess the order of $\varrho^2(\mathbf{G}_m^{-1})$. Comte et al. (2013) show that under the following conditions on the density g, we can recover the order of the spectral norm of $\mathbf{G}_m^{-1}$. First we define an integer $r \geq 1$ such that

$$\frac{\mathrm{d}^j}{\mathrm{d}x^j} g(x) \big|_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \dots, r-2 \\ B_r \neq 0 & \text{if } j = r-1. \end{cases}$$

And we make the two following assumptions:

(C1) $g \in \mathbb{L}^1(\mathbb{R}^+)$ is r times differentiable and $g^{(r)} \in \mathbb{L}^1(\mathbb{R}^+)$.
(C2) The Laplace transform of g, defined by $G(z) = \mathbb{E}[e^{-zY}]$, has no zero with non negative real parts except for the zeros of the form $\infty + ib$.

For instance if Y follows a Gamma distribution of parameter p and λ, then its density g verifies these three conditions for $r = p$. Especially an Exponential distribution satisfies those assumptions for $r = 1$. According to Comte et al. (2013), under Assumptions **(C1)-(C2)** there exist some positive constants C_ϱ and C_ϱ' such that

$$C_\varrho' m^{2r} \leq \varrho^2\left(\mathbf{G}_m^{-1}\right) \leq C_\varrho m^{2r}.$$

Then the order of the variance can be derived as follows

$$\| {}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 \leq \|{}^t\vec{\psi}_m\|_2^2 \varrho^2(\mathbf{G}_m^{-1}) \leq \|\psi\|^2 \varrho^2(\mathbf{G}_m^{-1}) \leq \|\psi\|^2 C_\varrho m^{2r}. \tag{13}$$

Proposition 2.2. *Assume that $\psi \in W^\alpha(\mathbb{R}^+, L')$ and $f \in W^s(\mathbb{R}^+, L)$ with $\alpha, s > 1$, that g satisfies Assumptions **(C1)-(C2)** and let $m_{opt} \propto n^{1/(s+\alpha+2r)}$, then there exists a positive constant c_1 depending on $L, L', s, r, \alpha, \|\psi\|, C_\varrho$ such that*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}\left[(\vartheta - \hat\vartheta_{m_{opt}})^2\right] \leq c_1 n^{-(s+\alpha)/(s+\alpha+2r)}.$$

Thus the choice $m = m_{opt}$ enables us to compute the rate of convergence of the estimator. However this choice depends on the regularity of the unknown function f and cannot be used in practice. This is why we will look for another way of performing the compromise between the squared bias and the variance.

2.4.3. *Almost parametric rate: mixed Gamma densities.* We want to point out that the Laguerre method enables us to obtain fast rates of convergence on the class of mixed Gamma densities noted $\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})$ and defined by

$$\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda}) = \left\{ f = \sum_{i=1}^q \alpha_i \gamma_i \geq 0, \ \gamma_i \sim \Gamma(p_i, \lambda_i), \ \lambda_i > 0, \ p_i \in \mathbb{N}, \right.$$
$$\left. q \in \mathbb{N} \cup \{\infty\}, \ \alpha_i \in \mathbb{R}, \ \sum_{i=1}^q |\alpha_i| < \infty, \ \sum_{i=1}^q \alpha_i = 1 \right\}, \tag{14}$$

Let us assume that f belongs to $\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})$. We obtain from (9) (see details in Mabon (2014)):

$$(\vartheta - \vartheta_m)^2 = \left(\sum_{k \geq m} a_k(\psi) \sum_{i=1}^q \alpha_i a_k(\gamma_i) \right)^2 \leq \sum_{k \geq m} a_k^2(\psi) \sum_{k \geq m} \sum_{i=1}^q |\alpha_i| a_k^2(\gamma_i)$$
$$\leq \|\psi\|^2 \sum_{i=1}^q |\alpha_i| C(p_i, \lambda_i) \left(\frac{\lambda_i - 1}{\lambda_i + 1} \right)^{2m} m^{2(p_i - 1)} \leq C\|\psi\|^2 m^{2(p^* - 1)} \rho^{2m}.$$

with $\rho = \max_i |(\lambda_i - 1)/(\lambda_i + 1)| \in (0, 1)$ and $p^* = \max_i p_i$. Thus the squared bias decays exponentially. And if we assume that g verifies **(C1)-(C2)** according to Equation (13), the order of the variance is less than $C_\varrho \|\psi\|^2 m^{2r}$. Therefore we obtain the following result

Proposition 2.3. *Assume that $f \in \mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})$, that g satisfies Assumptions **(C1)-(C2)** and let $m_{opt} = c \log n / |\log \rho|$ with $c \geq 1$, then there exists a positive constant c_2 depending on $q, \vec{\alpha}, \vec{p}, \vec{\lambda}, \alpha, \|\psi\|, C_\varrho$ such that*

$$\sup_{f \in \mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})} \mathbb{E}\left[(\vartheta - \hat\vartheta_{m_{opt}})^2\right] \leq c_2 \frac{(\log n)^{2r}}{n}.$$

So for the classes of mixed Gamma densities, the Laguerre procedure enables us to obtain almost parametric rates of convergence, which is remarkable in the deconvolution model. As previously, the choice m_{opt} cannot be performed in practice, since f is unknown.

2.4.4. *Parametric rate.* In this section, we propose to find conditions ensuring that the parametric rate can be achieved.

Proposition 2.4. *If $\psi(x) = \alpha e^{-x}$ with $\alpha \in \mathbb{R}$, then*

$$\mathbb{E}\left[(\vartheta - \hat{\vartheta}_1)^2\right] = \frac{\alpha^2}{a_0^2(g)} \frac{\mathbb{V}\mathrm{ar}[\varphi_0(Z_1)]}{n}. \tag{15}$$

This result is interesting in the sense that it does not require any additional assumptions on the density g. It allows us to recover the Laplace transform of X at the point $t = 1$ with a parametric rate. Note that, as Model (1) implies that $\mathbb{E}[e^{-Z_i}] = \mathbb{E}[e^{-X_i}]\mathbb{E}[e^{-Y_i}]$ and the distribution of Y is assumed to be known, we can estimate the Laplace transform with a parametric rate with $(1/n)\sum_{i=1}^{n} e^{-X_i}/\mathbb{E}[e^{-Y_1}]$, without knowing anything about Laguerre basis. This formula shows that the Laplace transform can be estimated at a parametric rate at any point $t > 0$. It can also be achieved with the Laguerre basis but the procedure is a little bit more complicated than with a plug-in estimator.

In fact, the Laguerre basis can be parametrized. In the present work for sake of clarity we consider that this scale parameter is equal to 1. More generally we can write the basis as follows

$$\forall k \in \mathbb{N}, \forall x \geq 0, \quad \varphi_k^\tau(x) = \sqrt{2\tau} L_k(2\tau x)e^{-\tau x} \quad \text{with} \quad \tau > 0.$$

For $\tau > 0$, if we take the Laguerre basis $(\varphi_k^\tau)_{k \geq 0}$ and proceed exactly as in the proof of Proposition 2.4, we find that the functions proportional to $\exp(-\tau x)$ satisfy Equation (15) for the corresponding scale parameter τ of the basis. The extension is straightforward. So for $\psi(x) = \alpha \exp(-\tau x)$, the parametric rate is still achieved. We can use this methodology to estimate the Laplace transform of X at any point.

The following two results are more elaborate.

Proposition 2.5. *If $\sum_{j \geq 2} |a_{j-1}(g) - a_{j-2}(g)| / a_0(g) < 1$ then $\varrho\left(\mathbf{G}_m^{-1}\right)$ is upper bounded by a constant independent of m and the parametric rate is achieved.*

We can notice that Exponential distributions with parameter $\lambda = 1$ satisfies this Proposition. From Example 1, we have the coefficients of $g \sim \mathcal{E}(\lambda)$, which yields for $\lambda = 1$

$$\sum_{j \geq 2} \frac{|a_{j-1}(g) - a_{j-2}(g)|}{a_0(g)} = 0,$$

and for $\lambda \neq 1$

$$\sum_{j \geq 2} \frac{|a_{j-1}(g) - a_{j-2}(g)|}{a_0(g)} = \sum_{j \geq 2} 2\frac{|\lambda - 1|^{j-2}}{|\lambda + 1|^{j-1}} = \frac{2}{\lambda + 1}\sum_{j \geq 2} \frac{|\lambda - 1|^{j-2}}{|\lambda + 1|^{j-2}}$$

$$= \frac{2}{\lambda + 1}\frac{1}{1 - (\lambda - 1)/(\lambda + 1)} = 1.$$

Proposition 2.6. *If $g \sim \mathcal{E}(\lambda)$ with $\lambda > 0$ and $\sum_{k=0}^{m-1}(\sum_{j=k}^{m-1} |a_j(\psi)|)^2 < \infty$, then there exists a positive constant C depending on λ such that*

$$\mathbb{E}\left[(\vartheta - \hat{\vartheta}_m)^2\right] \leq (\vartheta - \vartheta_m)^2 + \frac{C}{n}.$$

Propositions 2.5 and 2.6 give conditions ensuring that the variance can be upper bounded by C/n whatever the value of m. Then to derive a parametric rate in this case we need to choose m_n large enough such that $(\vartheta - \vartheta_{m_n})^2 = \mathrm{O}(n^{-1})$.

Let us give some examples of functions ψ to illustrate Proposition 2.6.

Example 2. Let us consider here that ψ is the density $\Gamma(p, \mu)$. According to Equation (9) we have

$$a_j(\psi) = \frac{\sqrt{2}\mu^p}{(1+\mu)^p} S_{p-1,j}\left(\frac{2}{1+\mu}\right)$$

The term $S_{p-1,j}(x)$ has the following order for in $x \in [0,1]$

$$|S_{p-1,j}(x)| \leq 2^p |1-x|^{j-p+1}, \quad \text{for} \quad j \geq p-1$$

(see details in Appendix B), it yields

$$\sum_{k=0}^{m-1} \left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2 \leq \frac{2^{2p+1}\mu^{2p}}{(\mu+1)^{2p}} \sum_{k=p-1}^{m-1} \left(\sum_{j=k}^{m-1} \left| \frac{\mu-1}{\mu+1} \right|^{j-p+1} \right)^2$$

$$\leq \frac{2^{2p+1}\mu^{2p}}{(\mu+1)^{2p}} \sum_{k=p-1}^{m-1} \left| \frac{\mu-1}{\mu+1} \right|^{2k} \left(\sum_{\ell=0}^{m-1-k} \left| \frac{\mu-1}{\mu+1} \right|^{\ell} \right)^2$$

$$\leq \frac{2^{2p+1}\mu^{2p}}{(\mu+1)^{2p}} \frac{(\mu+1)^4}{2\mu(\mu+1-|\mu-1|)^2}$$

Thus condition $\sum_{k=0}^{m-1}(\sum_{j=k}^{m-1}|a_j(\psi)|)^2 < \infty$ of Proposition 2.6 is fulfilled. Therefore we conclude that if g is $\mathcal{E}(\lambda)$ and ψ a finite mixture of Gamma distributions, then the parametric rate is achieved.

Example 3. Let α an integer such that $\alpha > 1$ and assume $\psi \in W^\alpha(\mathbb{R}^+, L')$, it yields

$$\left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2 = \left(\sum_{j=k}^{m-1} j^{\alpha/2}|a_j(\psi)|j^{-\alpha/2} \right)^2 \leq \sum_{j=k}^{m-1} a_j^2(\psi)j^\alpha \sum_{j=k}^{m-1} j^{-\alpha} \leq L'\frac{\alpha k^{-\alpha+1}}{\alpha-1},$$

then

$$\sum_{k=0}^{m-1} \left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2 \leq \frac{L'\alpha}{\alpha-1} \sum_{k=0}^{m-1} k^{-(\alpha-1)} < \infty \quad \text{if} \quad \alpha > 1.$$

In conclusion any function ψ such that $\sum_{k\geq 0} k^\alpha a_k^2(\psi) \leq L' < +\infty$ with $\alpha > 1$ satisfies the assumption of Proposition 2.6.

For instance $\psi(x) = (1+x)^{-\beta}\mathbb{1}_{x\geq 0}$ with $\beta > (\alpha+1)/2$ belongs to $W^\alpha(\mathbb{R}^+, L')$ (see details in Appendix C).

3. Model selection and adaptive estimation

To build data-driven estimators of linear functionals of the density f, we propose a new selection strategy in the spirit of Goldenshluger and Lepski (2011). First we add the following assumption on the model collection:

(A) $\mathcal{M} = \left\{ m \in \{1, \ldots, n\}, \ m \leq n/\log n \ \text{and} \ \max_{1\leq k\leq m} \| {}^t\vec{\psi}_k \mathbf{G}_k^{-1} \|_2^2 \leq n \right\}.$

The model selection is done by minimizing the following quantity

$$\widehat{m} = \operatorname*{argmin}_{m\in\mathcal{M}} \{A(m) + V(m)\} \tag{16}$$

where $V(m)$ has the order of the variance in the right-hand side of Equation (10) and is defined by

$$V(m) = \kappa(\|h\|_\infty \vee 1) \max_{1\leq k\leq m} \frac{\| {}^t\vec{\psi}_k \mathbf{G}_k^{-1} \|_2^2 \log n}{n},$$

with κ a numerical constant which will be chosen later (see comment below). The term of variance is slightly modified to ensure that V is nondecreasing with m. The term $A(m)$ has the order of the bias and is based on the comparison of the estimators built in the previous section as follows

$$A(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - V(m') \right\}_+.$$

We can now state the following oracle inequality:

Theorem 3.1. *For f, g and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption **(A)** is true. Let $\hat{\vartheta}_{\hat{m}}$ be defined by Equations* (5) *and* (16). *Then there exists a constant κ_0 such that for $\kappa \geq \kappa_0$ such that*

$$\mathbb{E}\left[(\vartheta - \hat{\vartheta}_{\hat{m}})^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + V(m) \right\} + \frac{C}{n}. \tag{17}$$

where C^{ad} is a positive numerical constant and C depends only on $\|h\|_\infty$.

The oracle inequality (17) establishes a non asymptotic oracle bound. It shows that the squared bias variance tradeoff is automatically made up to a loss of logarithmic order and a multiplicative constant. Let us notice that $C^{ad} = 7$ and $\kappa_0 = 384$ would suit. Let us comment the specific squared bias term which differs from the squared bias established in Proposition 2.1. The term $\sup_{k>m}\left\{(\vartheta_k - \vartheta_m)^2\right\}$ is defined such that it is non-decreasing in m and we take for convention $\vartheta_\infty = \vartheta$. Usually in Lepski's method, it is assumed that the p.d.f. belongs to some regularity space indexed by β and the squared bias is expressed as a function of β and m which is non-decreasing in m. Nevertheless it a strong assumption since we do not know the coefficient β depending on f. Our approach enables us to circumvent this assumption.

Theorem 3.1 is derived under mild assumptions. Indeed we do not need to assume that ψ or f belong to some semi-parametric space to derive the theorem. Concerning the model selection procedure, it is simpler in its formulation than those of Laurent et al. (2008) and Butucea and Comte (2009) thanks to the Goldenshluger and Lepski (2011) method. Butucea and Comte (2009) obtain as a consequence of their general adaptive procedure the pointwise estimation of the density on the real line, the stochastic volatility model and the pointwise Laplace transform estimation when X is a positive random variable, even when the Laplace transform of the noise is infinite. In the next section, we also show that a procedure for the pointwise estimation on the nonnegative real line is possible. A direct application of Theorem 3.1 is the estimation of the c.d.f. of X as pointed out in Subsection 2.4, which is a novelty compared to Butucea and Comte (2009). Pensky (2015) also develops an adaptive procedure based on Lepski's method. This paper clearly improves previous papers and covers a lot of cases (c.d.f, pointwise, moments) for which lower bounds are provided.

Some comments for practical use are in order. Indeed in the variance term $V(m)$, there are two quantities which deserve some explanations: κ and $\|h\|_\infty$. First note that when κ gets larger, so do $V(m)$ and thus the bound (17). It follows from the proof that $\kappa = 384$ would suit. But in practice, values obtained from the theory are generally too large and the constant is calibrated by simulations. Once chosen, it remains fixed for all simulation experiments. There is still an unknown term in the penalty, $\|h\|_\infty$, that needs to be estimated. We have to check that we can derive an oracle inequality when this term is estimated, which is done in the following Corollary.

Beforehand let

$$\hat{h}_D(x) = \sum_{k=0}^{D-1} \hat{a}_k(Z) \varphi_k(x), \tag{18}$$

with $\hat{a}_k(Z) = (1/n) \sum_{i=1}^{n} \varphi_k(Z_i)$. We can see that $\hat{h}_D$ is an unbiased estimator of $h_D(x) = \sum_{k=0}^{D-1} a_k(h) \varphi_k(x)$.

Corollary 3.2. *For f, g and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption **(A)** is true. Let $\hat{\vartheta}_{\widetilde{m}}$ be defined by* (5) *and*

$$\widetilde{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ \widetilde{A}(m) + \widetilde{V}(m) \right\} \tag{19}$$

with $\widetilde{A}(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - \widetilde{V}(m') \right\}_+$ and $\widetilde{V}(m) = \kappa (2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \| {}^t\vec{\psi}_k \mathbf{G}_k^{-1} \|_2^2 \frac{\log n}{n}$

where $\hat{h}_D$ is given by (18) *and D satisfies $\log n \leq D \leq \dfrac{\|h\|_\infty}{128\sqrt{2}} \dfrac{n}{(\log n)^3}$. Then there exists a constant κ_0 such that for $\kappa \geq \kappa_0$ such that*

$$\mathbb{E}\left[(\vartheta - \hat{\vartheta}_{\widetilde{m}})^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + V(m) \right\} + \frac{C}{n}.$$

where C^{ad} is a positive numerical constant and C depends only on $\|h\|_\infty$.

Note that the constraint on D is fulfilled for n large enough as soon as $D \simeq \sqrt{n}$ for instance. In this sense Corollary 3.2 has rather an asymptotic flavor.

4. PARTICULAR CASE OF POINTWISE ESTIMATION

4.1. **Theoretical results.** For pointwise estimation of f, we take $\psi(x) = \delta_{\{x_0\}}(x)$ for any nonnegative x_0, where δ stands for the Dirac measure. It yields that $\vartheta = f(x_0)$. In this case Equation (5) becomes

$$\hat{f}_m(x_0) = \sum_{k=0}^{m-1} \hat{a}_k \varphi_k(x_0), \tag{20}$$

since $a_k(\psi) = \langle \delta_{\{x_0\}}, \varphi_k \rangle = \varphi_k(x_0)$. We can notice that ψ is not square integrable.

Proposition 4.1. *For $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, for $\mathbf{G}_m$ defined by Equation* (7) *and $\hat{f}_m(x_0)$ defined by Equation* (20) *for any nonnegative x_0, the following result holds*

$$\mathbb{E}\left[(f(x_0) - \hat{f}_m(x_0))^2 \right] \leq (f(x_0) - f_m(x_0))^2 + \|h\|_\infty \frac{\| {}^t\vec{\varphi}_m(x_0)\mathbf{G}_m^{-1} \|_2^2}{n},$$

where ${}^t\vec{\varphi}_m(x_0) = (\varphi_0(x_0) \dots \varphi_{m-1}(x_0))$.

We can notice that the squared bias and the variance have not the same order as in the case where ψ is square integrable. Let us assume as in Section 2.4.2 that f belongs to $W^s(\mathbb{R}^+, L)$, applying Equation (3) and the Cauchy-Schwarz inequality we get the following order for the squared bias

$$(f(x_0) - f_m(x_0))^2 = \left(\sum_{k \geq m} a_k(f)\varphi_k(x_0) \right)^2 \leq 2 \left(\sum_{k \geq m} |a_k(f)| \right)^2 = 2 \left(\sum_{k \geq m} |a_k(f)| k^{s/2} k^{-s/2} \right)^2$$

$$\leq 2 \sum_{k \geq m} |a_k(f)|^2 k^s \sum_{k \geq m} k^{-s} \leq \frac{2sL}{s-1} m^{-s+1}$$

for $s > 1$. Moreover applying Equation (3), we get

$$\| {}^t\vec{\varphi}_m(x_0)\mathbf{G}_m^{-1} \|_2^2 \leq \| {}^t\vec{\varphi}_m(x_0) \|_2^2 \varrho^2(\mathbf{G}_m^{-1}) \leq \sum_{k=0}^{m-1} |\varphi_k(x_0)|^2 \varrho^2(\mathbf{G}_m^{-1}) \leq 2m\varrho^2(\mathbf{G}_m^{-1}). \tag{21}$$

Proposition 4.2. *Assume that $f \in W^s(\mathbb{R}^+, L)$ with $s > 1$, that Assumptions **(C1)-(C2)** are true and let $m_{opt} \propto n^{1/(s+2r)}$, then there exists a positive constant c_2 depending on L, s, r, C_ϱ such that*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}\left[(f(x_0) - \hat{f}_{m_{opt}}(x_0))^2 \right] \leq c_2 n^{-(s-1)/(s+2r)}.$$

Note that this rate is different from the global one: according to Mabon (2014) for the global estimation of the density f with the $\mathbb{L}^2$-risk, the rate of convergence is of order $n^{-s/(s+2r+1)}$. As in Section 2.4.2, we can point out that faster rates of convergence of order $\log n/n$ can be obtained for functions of Gamma type defined by Equation (14).

The data-driven strategy is done according to the one developed in Section 3 for the particular case of pointwise estimation, with the specific variance term according to this model. It yields that

$$\widetilde{A}(m, x_0) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{f}_{m'}(x_0) - \hat{f}_{m' \wedge m}(x_0))^2 - \widetilde{V}(m', x_0) \right\}_+ \quad \text{with}$$

$$\widetilde{V}(m, x_0) = \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \le k \le m} \frac{\| {}^t\vec{\varphi}_k(x_0)\mathbf{G}_k^{-1}\|_2^2 \log n}{n} \tag{22}$$

where κ is a numerical constant, $\hat{h}_D$ is the estimator of the projection of h on $\mathcal{S}_D$ given by (18) and D satisfies $\log n \le D \le \dfrac{\|h\|_\infty}{128\sqrt{2}} \dfrac{n}{(\log n)^2}$ and

$$\widetilde{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ \widetilde{A}(m, x_0) + \widetilde{V}(m, x_0) \right\}. \tag{23}$$

Theorem 4.3. *For $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption **(A)** is true. Let $\hat{f}_{\widetilde{m}}(x_0)$ be defined by Equations (20)-(23). Then there exists a constant κ_0 such that for $\kappa \ge \kappa_0$ such that*

$$\mathbb{E}\left[(f(x_0) - \hat{f}_{\widetilde{m}}(x_0))^2\right] \le C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} (f_k(x_0) - f_m(x_0))^2 + V(m, x_0) \right\} + \frac{C}{n}.$$

where C^{ad} is a positive numerical constant, C depends only on $\|h\|_\infty$ and $V(m, x_0) = \kappa(\|h\|_\infty \vee 1) \max_{1 \le k \le m} \| {}^t\vec{\varphi}_k(x_0)\mathbf{G}_k^{-1}\|_2^2 \log n/n$.

Applying Equation (3) note that

$$\sup_{k > m} |f_k(x_0) - f_m(x_0)|^2 = \sup_{k > m} \left| \sum_{j=m}^{k-1} a_j(f)\varphi_k(x_0) \right|^2 \le 2 \left(\sum_{j \ge m} |a_j(f)| \right)^2$$

and using Equation (21) yields

$$V(m, x_0) \le 2\kappa\|h\|_\infty m \varrho^2(\mathbf{G}_m^{-1})\frac{\log n}{n},$$

since $m \mapsto \varrho^2(\mathbf{G}_m^{-1})$ is nondecreasing with m (see Lemma 3.4 in Mabon (2014)). These two bounds are independent of x_0 which proves that the risk, may tend uniformly to 0 for an adequate choice $m = m(n)$, when n tends to infinity.

4.2. **Illustrations.** The whole implementation is conducted using R software. The mean squared error (MSE) $\mathbb{E}(f(x_0) - \hat{f}_{\widetilde{m}}(x_0))^2$ is computed as the empirical mean over 500 simulation samples.

4.2.1. *Simulation setting.* The performance of the procedure is studied for the three following distributions for X. Moreover all the densities are normalized with unit variance. Each density is computed over an interval I.

    ▷ Gamma distribution : $\Gamma(4, \frac{1}{2})$, $I = [0, 8]$.
    ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I = [0, 13]$.
    ▷ Log-normal distribution : $L\mathcal{N}(1, 1)$, $I = [0, 15]$.

We then choose a Gamma distribution for the error distribution which verifies **(C1)-(C2)** for $r = 2$.

    ▷ Gamma noise: $\Gamma(2, \frac{1}{\sqrt{20}})$ and $\Gamma(2, \frac{1}{\sqrt{8}})$.

Thus the first Gamma distribution has a variance $1/10$ and the second $1/4$. We refer to Example 1 for the computation of the matrix $\mathbf{G}_m$.

We also compute estimators where there is no additional noise, i.e. $\sigma^2 = 0$. The procedure is applied with $\mathbf{G}_m = \mathbf{I}_m$.

4.2.2. *Implementation of the estimation procedure.* We consider the following model collection $\mathcal{M}_n = \{m \in \mathbb{N}, \quad 0 \le m \le \lfloor n/\log n \rfloor\}$. The pointwise estimation at a point x_0 is computed according to the following adaptive procedure:

> ▷ For $m \in \mathcal{M}_n$, compute $\widetilde{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left\{ \widetilde{A}(m, x_0) + \widetilde{V}(m, x_0) \right\}$, with
>
> ○ $\widetilde{A}(m, x_0) = \sup_{m' \in \mathcal{M}_n} \left\{ (\hat{f}_{m'}(x_0) - \hat{f}_{m' \wedge m}(x_0))^2 - \widetilde{V}(m', x_0) \right\}_+$
>
> ○ $\widetilde{V}(m, x_0) = \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{0 \le k \le m-1} \| {}^t\vec{\varphi}_k(x_0) \mathbf{G}_k^{-1} \|_2^2 \log n / n$ with $\hat{h}_D$ defined by (18).
>
> ▷ Compute $\hat{f}_{\widetilde{m}}(x_0) = \sum_{k=0}^{\widetilde{m}-1} \hat{a}_k \varphi_k(x_0)$.

In addition we compute the integrated squared risk over an interval I of 300 equidistant points, by computing

$$\frac{\text{length}(I)}{300} \sum_{k=1}^{300} (f(x_k) - \hat{f}_{\widetilde{m}_k}(x_k))^2$$

and $(x_k)_k$ a subdivision of I, which gives an approximation of the integrated risk for our method. Thus we can measure the performances of the pointwise estimation by comparing our method with the method of Mabon (2014) who proposes adaptive estimators of the density f associated with the $\mathbb{L}^2$-risk.

One of the major issues in data-driven strategies is the choice of the constant κ. To detect this value in the model selection paradigm, we can use the slope heuristics for instance based on the works of Birgé and Massart (2007). Baudry et al. (2012) developed a method based on slope heuristics or on dimension jump. We cannot use those methods in the present work since we use Goldenshluger-Lepski methodology for the automatic selection of an estimator among a collection.

The results of Goldenshluger and Lepski (2011) ensure that if the variance term is large enough, the data-driven estimator is almost as efficient as the best one in the collection. In the proof we derive a theoretical value of κ defined by Equation (22). Nonetheless it is well known that theoretical constants in $\widetilde{V}(m, x_0)$ are often too large in practice. Thus for practical implementation, the good behavior of the estimator is based on the calibration of the constant κ. Few studies have been conducted on the constant calibration in Goldenshluger-Lepski methodology to see if there exists a critical value such that, if $V(m) > V_0$ the risk is quasi optimal.

To find this value in the setting of Goldenshluger-Lepski method, we follow the recent work of Lacour and Massart (2015) who show that, for density estimation based on kernel estimators with Goldenshluger-Lepski type bandwith selection, if the variance term is chosen too small then the procedure fails. They are the first to give a minimal penalty for Goldenshluger-Lepski methodology. In our case, we do not necessarily expect the same behavior of the variance term since we are in the particular case of the deconvolution model but we expect to find a critical value of κ which characterizes a failure of the procedure.

To detect such a value, we propose to observe the evolution of the choice $\widetilde{m}$ for different values of κ. It is clear that to challenge the global estimation, our pointwise procedure must be able to adapt to the point and to choose different $\widetilde{m}$ for different points.

To avoid over fitting, we choose two distributions that we will not estimate in the following: $\beta(4, 5) \times 5$ and $\chi^2(10)/\sqrt{20}$. The first one has variance $50/81$ and has a compact support $[0, 5]$. The other one is normalized with unit variance and its support is $\mathbb{R}^+$. Only here in our procedure, we replace $\widetilde{V}(m)$ by its uniform bound $2\kappa m \|\hat{h}_D\|_\infty \varrho^2(\mathbf{G}_m^{-1}) \log n / n$. Indeed with this approach it

is easier to detect stable ranges where the choice of $\widetilde{m}$ does not change. Otherwise we observe stronger oscillations of the choice $\widetilde{m}$ and it is difficult to isolate the good values of κ from the bad ones.

In Figure 1, we apply this methodology for a wide range of κ and in Figure 2 we concentrate on smaller range of κ which corresponds to the neighborhood of the κ of the black thin line with circles. In Figure 1, we also draw the data driven choice for $\kappa = 0$ and $\kappa = 20$ which respectively corresponds to the bold black and the bold red lines. For $\kappa = 0$, the selected dimension is systematically the largest as possible while for $\kappa = 20$ it is mainly the smallest. We choose to concentrate on the black thin line with circles because it is the curve which moves the most and is the farthest from the two curves $\kappa = 0$ and $\kappa = 20$ which represent extreme situations. In Figure 2, we see that for both distributions the curves have the same behavior. Finally according to this methodology we choose a small value of $\kappa = 3 \cdot 10^{-4}$.
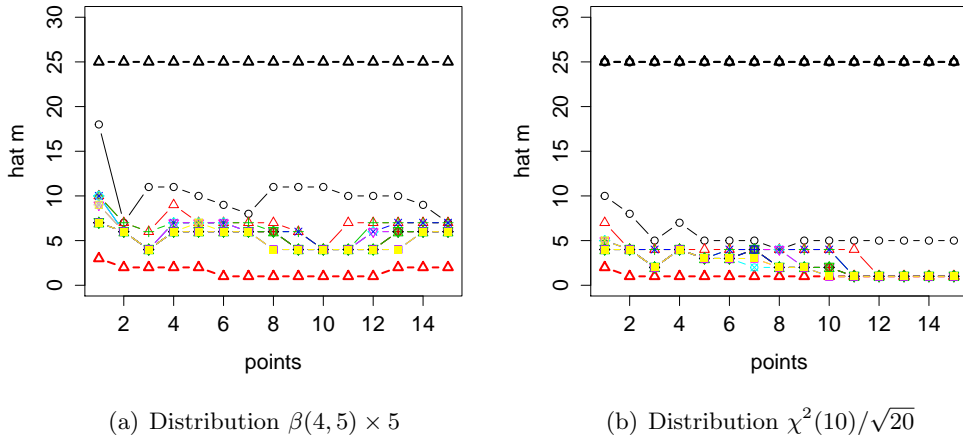

(a) Distribution $\beta(4,5) \times 5$ (b) Distribution $\chi^2(10)/\sqrt{20}$

FIGURE 1. $\widetilde{m}$ in function of $\kappa \in [10^{-4}, 10]$ (15 equidistant calibration constants, some are superposed) plus $\kappa = 0$ (top) and $\kappa = 20$ (bottom), $n = 5000$ and $\sigma^2 = 1/10$ and 15 equidistant x_0 in the interval $[0,5]$ for the β and $[0,10]$ for the χ^2.

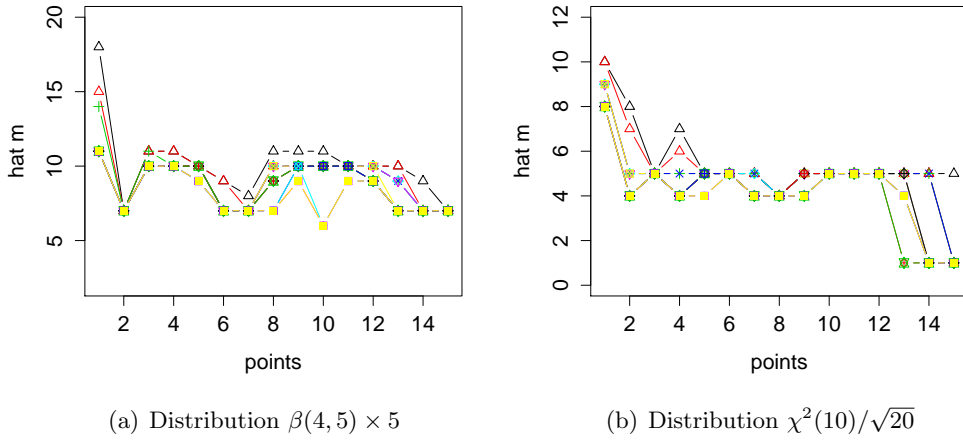(a) Distribution $\beta(4,5) \times 5$ (b) Distribution $\chi^2(10)/\sqrt{20}$

FIGURE 2. $\widetilde{m}$ in function of $\kappa \in [10^{-4}, 10^{-3}]$ (15 equidistant calibration constants, some are superposed), $n = 5000$ and $\sigma^2 = 1/10$ and 15 equidistant x_0 in the interval $[0,5]$ for the β and $[0,10]$ for the χ^2.

In a second step we also propose to test another calibration suggested by Lacour and Massart (2015) in their concluding remark section. The idea is to consider

$$\overline{A}(m, x_0) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \| {}^t\vec{\varphi}_k(x_0)\mathbf{G}_k^{-1}\|^2 \log n/n \right\}_+$$

$$\overline{m} = \operatorname*{argmin}_{m \in \mathcal{M}} \left\{ \overline{A}(m, x_0) + \kappa'(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \| {}^t\vec{\varphi}_k(x_0)\mathbf{G}_k^{-1}\|_2^2 \log n/n \right\}.$$

Then, it is easy to prove that Theorem 3.1 holds for $\widehat{\vartheta_{\overline{m}}}$ as soon as $\kappa' \geq \kappa$. The authors suggest to find an adequate value of κ for $\kappa = \kappa'$ and then to put $\kappa' = 2\kappa$. This method is illustrated in the next section.

4.2.3. *Simulation results.* In Table 1 we report the results of the simulation (averaged values of the integrated squared errors multiplied by 100) for the global estimator computed according to Mabon (2014). The results of the pointwise estimation are presented in Table 2. It reports averaged values of the approximation of the integrated risk for our method of pointwise estimation. First remarks: increasing the sample size improves the quality of the estimation and increasing the variance degrades the estimation but in an acceptable way for both methods.

If we compare the methods, we can see for the mixed Gamma distribution that the pointwise estimation is better than the global estimation. The risk is approximately divided by 2. For the Gamma and the log-normal, the estimation is clearly better with the global method where there is no additional noise. Otherwise the two methods are equivalent. We can note for the pointwise estimation that when the sample size increases, the results are very close to those without noise ($\sigma^2 = 0$).

In Table 2 we compute the risk associated to the oracles. We see that the oracles are very good, approximately divided by 30 compared to data driven estimators. It shows that the pointwise estimation could be excellent if the data driven strategy could perform adequately. Nonetheless in Section 3, we have seen that the bias variance compromise was made up to a logarithmic loss. It may also explain the gap between oracles and data driven procedure results. We also compare the quality of the estimation when $\kappa' = \kappa$ or $\kappa' = 2\kappa$. The results are always and slightly better when $\kappa' = \kappa$. Our examples do not permit us to highlight the idea of dissociation of the constant presented in Lacour and Massart (2015) for the Goldenshluger-Lepski strategy. Nonetheless this detection method of the constant κ leads to very good results.

We also illustrate the results with some figures. Figures 3-5 display the results of the data driven estimation respectively for the Gamma and the mixed Gamma. For each Figure we present the Goldenshluger-Lepksi estimator with the calibration proposed by Lacour and Massart (2015) for $\kappa' = \kappa$. First we see that the risk decreases when the sample size increases. For the Gamma, we see that data driven estimators are very close to oracle estimators which is not really the case for the mixed Gamma especially for $n = 2000$. We can observe some oscillations near the origin. Thus the pointwise estimation does not seem to radically improve the estimation near the origin.

	$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
	200	2000	200	2000	200	2000
Gamma	0.313	0.045	0.625	0.037	0.752	0.289
Mixed Gamma	0.334	0.031	1.304	0.342	2.738	0.348
Log normal	0.347	0.043	0.684	0.129	0.829	0.246

TABLE 1. Results of simulation for the global estimation based on Mabon (2014). MISE $\mathbb{E}(\|f - \hat{f}_{\hat{m}}\|^2) \times 100$ averaged over 500 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$.

	$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
	200	2000	200	2000	200	2000
oracle						
Gamma	0.064	0.004	0.016	0.008	0.056	0.016
Mixed Gamma	0.007	0.005	0.026	0.013	0.003	0.013
Log normal	0.090	0.0006	0.045	0.002	0.105	0.045
$\kappa' = 2\kappa$						
Gamma	0.464	0.080	0.504	0.064	1.176	0.144
Mixed Gamma	0.442	0.299	0.455	0.286	0.624	0.299
Log normal	0.390	0.165	0.480	0.165	0.960	0.180
$\kappa' = \kappa$						
Gamma	0.352	0.064	0.416	0.064	0.976	0.088
Mixed Gamma	0.429	0.338	0.442	0.286	0.572	0.286
Log normal	0.345	0.150	0.495	0.165	0.885	0.180

TABLE 2. Results of simulation for the pointwise estimation. Approximation of the integrated risk multiplied by 100 averaged over 500 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$. *oracle* = length$(I)/300 \sum_{k=1}^{300} \min_{m \in \mathcal{M}} \mathbb{E}|f(x_0) - f_m(x_0)|^2$
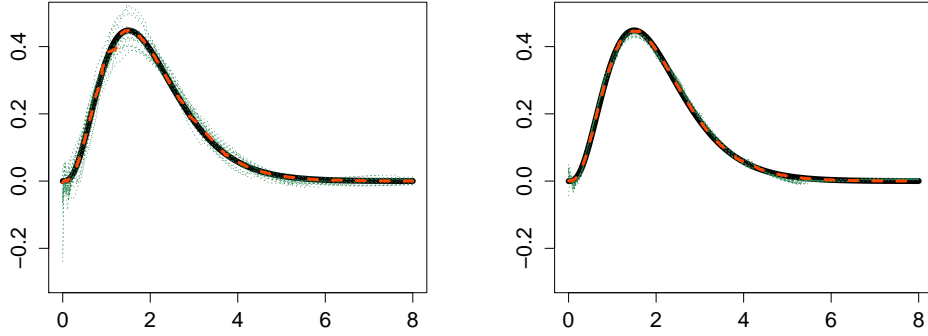

FIGURE 3. Estimation of the Gamma density: left for $n = 200$ and right $n = 2000$; $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.

5. PROOFS

5.1. **Proofs of Section 2.**

5.1.1. *Proof of Proposition 2.1.* We have the following bias variance decomposition:

$$\mathbb{E}(\vartheta - \hat{\vartheta}_m)^2 = (\vartheta - \vartheta_m)^2 + \mathbb{E}(\vartheta_m - \hat{\vartheta}_m)^2.$$

The first term corresponds to the squared bias term of Equation (10). Let us study the second term, by definition we have

$$(\vartheta_m - \hat{\vartheta}_m)^2 = \langle \psi, f_m - \hat{f}_m \rangle^2 = \langle \vec{\psi_m}, \vec{f_m} - \hat{\vec{f}}_m \rangle_2^2 = \langle \vec{\psi_m}, \mathbf{G}_m^{-1}(\vec{h}_m - \hat{\vec{h}}_m) \rangle_2^2.$$
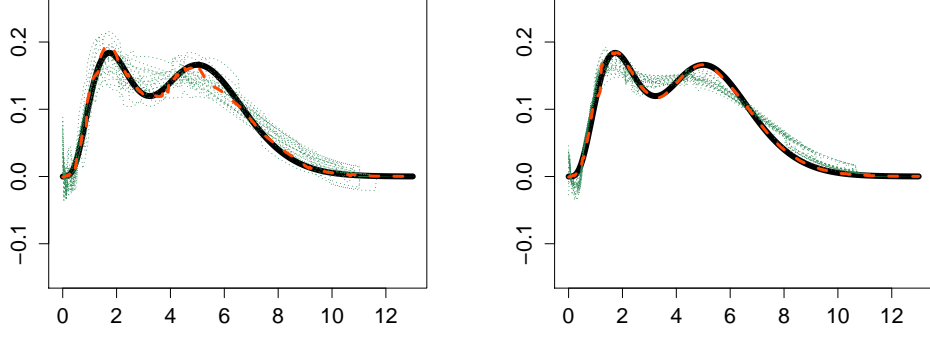
FIGURE 4. Estimation of the Mixed Gamma density: left for $n = 200$ and right $n = 2000$, $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.
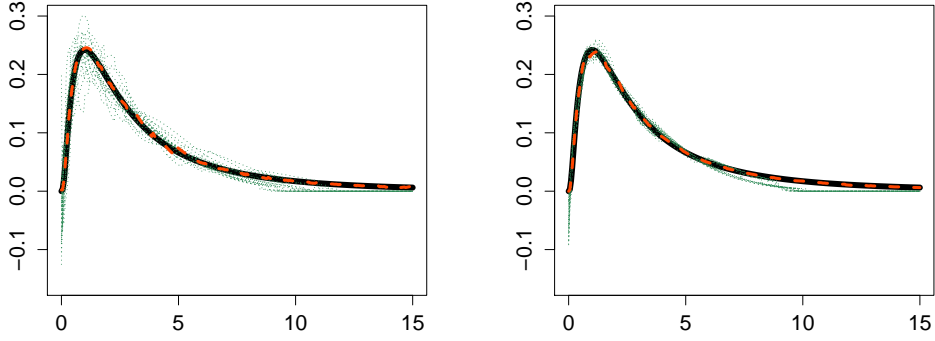

FIGURE 5. Estimation of the log Normal density: left for $n = 200$ and right $n = 2000$, $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.

Then we get

$$
\begin{aligned}
\mathbb{E}(\vartheta_m - \hat{\vartheta}_m)^2 = \mathbb{V}\mathrm{ar}[\hat{\vartheta}_m] &= \mathbb{V}\mathrm{ar}\left[\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} a_k(\psi)\left[\mathbf{G}_m^{-1}\right]_{k,l}\hat{a}_l(Z)\right] \\
&= \mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} a_k(\psi)\left[\mathbf{G}_m^{-1}\right]_{k,l}\varphi_l(Z_i)\right] \\
&= \frac{1}{n}\mathbb{V}\mathrm{ar}\left[\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} a_k(\psi)\left[\mathbf{G}_m^{-1}\right]_{k,l}\varphi_l(Z_1)\right] \leq \frac{1}{n}\mathbb{E}\left[\left(\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} a_k(\psi)\left[\mathbf{G}_m^{-1}\right]_{k,l}\varphi_l(Z_1)\right)^2\right] \\
&\leq \frac{1}{n}\int_{\mathbb{R}^+}\left(\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} a_k(\psi)\left[\mathbf{G}_m^{-1}\right]_{k,l}\varphi_l(u)\right)^2 h(u)\,\mathrm{d}u \leq \frac{\|h\|_\infty}{n}\|{}^t\vec{\psi}_m\mathbf{G}_m^{-1}\|_2^2. \quad \square
\end{aligned}
$$

5.1.2. *Proof of Proposition 2.4.* According to Equation (10), we just have to find the eigenvectors of the matrix $\mathbf{G}_m^{-1}$. Indeed if ${}^t\vec{\psi}_m$ is an eigenvector of $\mathbf{G}_m^{-1}$ associated with eigenvalue λ_m, it yields that $\|{}^t\vec{\psi}_m\mathbf{G}_m^{-1}\|_2^2 = \lambda_m^2\|{}^t\vec{\psi}_m\|_2^2$ and since we assumed that $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, the variance is upper bounded $\lambda_m^2\|\psi\|^2$. First note that

- For any vector $\vec{u}$, $\|{}^t\mathbf{G}_m^{-1}\vec{u}\|_2^2 = \|{}^t\vec{u}\mathbf{G}_m^{-1}\|_2^2$.
- If we have ${}^t\mathbf{G}_m\vec{u} = \alpha\vec{u}$ with $\alpha \neq 0$, then $(1/\alpha)\vec{u} = {}^t\mathbf{G}_m^{-1}\vec{u}$.

Therefore let us look for the eigenvalues of ${}^t\mathbf{G}_m$.

Since ${}^t\mathbf{G}_m$ is a triangular matrix, its eigenvalues are its diagonal elements. Thus the only eigenvalue of ${}^t\mathbf{G}_m$ is $a_0(g)/\sqrt{2}$ associated with the eigenvectors belonging to $\mathcal{S}_0 = \mathrm{Span}\{\varphi_0\}$. Then the matrix ${}^t\mathbf{G}_m^{-1}$ has for eigenvalue $\sqrt{2}/a_0(g)$ associated with the same space of eigenvectors.

Let $\psi \in \mathcal{S}_0$, i.e. $\psi(x) = \alpha e^{-x}\mathbb{1}_{\mathbb{R}^+}(x)$ with $\alpha \in \mathbb{R}$, it yields that ${}^t\vec{\psi}_m = (\alpha/\sqrt{2}\ 0\ \ldots\ 0)$ and ${}^t\vec{\psi}_m\mathbf{G}_m^{-1} = (\alpha/a_0(g)\ 0\ \ldots\ 0)$ which implies $\|{}^t\mathbf{G}_m^{-1}\vec{\psi}_m\|_2^2 = \|{}^t\vec{\psi}_m\mathbf{G}_m^{-1}\|_2^2 = \alpha^2/a_0^2(g) = \vartheta_1$. Thus the variance term does not depend on the dimension m of the model as announced.

Now for the bias term, let us notice that $\vartheta = \langle\psi, f\rangle = \sum_{k\geq 0} a_k(\psi)a_k(f) = \alpha a_0(f)/\sqrt{2}$ which implies that $\mathbb{E}[\hat{\vartheta}_1] = \vartheta_1 = \vartheta$. So the estimator $\hat{\vartheta}_1$ of ϑ, in that case, is unbiased.

Finally, we have that

$$\mathbb{E}(\vartheta - \hat{\vartheta}_1)^2 = \mathbb{E}(\vartheta_1 - \hat{\vartheta}_1)^2 = \mathbb{V}\mathrm{ar}[\hat{\vartheta}_1] = \frac{1}{n}\mathbb{V}\mathrm{ar}\left[a_0(\psi)[\mathbf{G}_m^{-1}]_{1,1}\varphi_0(Z_1)\right] = \frac{\alpha^2}{a_0^2(g)}\frac{\mathbb{V}\mathrm{ar}[\varphi_0(Z_1)]}{n}.$$

$\square$

5.1.3. *Proof of Proposition 2.5.* The matrix $\mathbf{G}_m$ admits only one eigenvalue $a_0(g)/\sqrt{2}$ with multiplicity m, it yields that

$$\left(\mathbf{G}_m - \frac{a_0(g)}{\sqrt{2}}\mathbf{I}_m\right)^m = 0$$

which implies that

$$\left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right)^m = 0.$$

Writing that

$$\mathbf{I}_m = \mathbf{I}_m - \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right)^m = \left(\mathbf{I}_m - \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right)\right)\sum_{k=0}^{m-1}\left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right)^k,$$

therefore

$$\mathbf{G}_m^{-1} = \frac{\sqrt{2}}{a_0(g)}\sum_{k=0}^{m-1}\left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right)^k.$$

It yields that

$$\varrho\left(\mathbf{G}_m^{-1}\right) \leq \frac{\sqrt{2}}{a_0(g)}\sum_{k=0}^{m-1}\left(\varrho\left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right)\right)^k.$$

Yet using the fact that for any matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$ (see Chapter 5 in Horn and Johnson (1990)),

$$\varrho(\mathbf{A}) \leq \sqrt{\|\mathbf{A}\|_1\|\mathbf{A}\|_\infty}$$

where $\|.\|_1$ and $\|.\|_\infty$ are respectively the maximum column sum matrix and the maximum row sum matrix. Besides noticing that since $\mathbf{G}_m^{-1}$ is a triangular Toeplitz matrix we have that $\|\mathbf{G}_m^{-1}\|_1 = \|{}^t\mathbf{G}_m^{-1}\|_\infty$. So we get

$$\varrho\left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\right) \leq \|\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\|_1$$

Moreover let us notice that

$$\|\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)}\mathbf{G}_m\|_1 = \sum_{j \geq 2}^{m} \left| \frac{a_{j-1}(g) - a_{j-2}(g)}{a_0(g)} \right| \leq \sum_{j \geq 2} \left| \frac{a_{j-1}(g) - a_{j-2}(g)}{a_0(g)} \right|.$$

Thus if

$$q = \sum_{j \geq 2} \left| \frac{a_{j-1}(g) - a_{j-2}(g)}{a_0(g)} \right| < 1,$$

it yields that

$$\varrho\left(\mathbf{G}_m^{-1}\right) \leq \frac{\sqrt{2}}{a_0(g)} \sum_{k=0}^{m-1} q^k \leq \frac{\sqrt{2}}{a_0(g)} \sum_{k=0}^{\infty} q^k = \frac{\sqrt{2}}{a_0(g)} \frac{1}{1-q}$$

then $\varrho\left(\mathbf{G}_m^{-1}\right)$ is upper bounded by a constant independent of m. □

5.1.4. *Proof of Proposition 2.6.* Let the coefficients of $\mathbf{G}_m^{-1}$ be defined by Equation (8), it yields

$$\| {}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 = \left(\frac{\lambda+1}{\lambda}\right)^2 \sum_{k=0}^{m-1} \left(a_k(\psi) + \frac{2}{\lambda+1} \sum_{j=k+1}^{m-1} a_j(\psi) \right)^2$$

$$\leq \left(\frac{\lambda+1}{\lambda}\right)^2 \left(1 \vee \frac{2}{\lambda+1}\right)^2 \sum_{k=0}^{m-1} \left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2.$$

Then if $\sum_{k=0}^{m-1} (\sum_{j=k}^{m-1} |a_j(\psi)|)^2 < \infty$, we get that $\| {}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 \leq C$ where C does not depend on m. □

5.2. **Proofs of Section 3.**

5.2.1. *Proof of Theorem 3.1.* By definition of $\widehat{m}$, the following inequalities hold

$$(\vartheta - \hat{\vartheta}_{\widehat{m}})^2 \leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(\hat{\vartheta}_m - \hat{\vartheta}_{m \wedge \widehat{m}})^2 + 3(\hat{\vartheta}_{m \wedge \widehat{m}} - \hat{\vartheta}_{\widehat{m}})^2$$

$$\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(A(\widehat{m}) + V(m) + A(m) + V(\widehat{m}))$$

$$\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(A(m) + V(m))$$

Taking expectation, we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\widehat{m}})^2 \leq 3\mathbb{E}(\vartheta - \hat{\vartheta}_m)^2 + 6\mathbb{E}[A(m)] + 6V(m) \tag{24}$$

with

$$A(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - V(m') \right\}_+ \tag{25}$$

Proposition 5.1. *Under Assumptions of Theorem 3.1, it holds true that*

$$\mathbb{E}[A(m)] \leq \sup_{\substack{k > m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \frac{C}{n}.$$

Starting from Equation (24), applying Proposition 2.1 and Proposition 5.1 we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\widehat{m}})^2 \leq 3(\vartheta - \vartheta_m)^2 + \|h\|_\infty \frac{\| {}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2}{n} + 6 \sup_{\substack{k > m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \frac{C}{n} + 6V(m)$$

$$\leq 3(\vartheta - \vartheta_m)^2 + 6 \sup_{\substack{k > m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \left(\frac{1}{\kappa} + 6 \right) V(m) + \frac{C}{n}$$

$$\leq 6 \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + 7V(m) + \frac{C}{n},$$

since $\kappa \geq 1$ and given that $\vartheta = \vartheta_\infty = \lim_{m\to\infty} \vartheta_m$ if $f, g \in \mathbb{L}^2(\mathbb{R}^+)$ Finally by taking the infimum over all $m \in \mathcal{M}$ we get the desired result

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\hat{m}})^2 \leq C^{ad} \inf_{m\in\mathcal{M}} \left\{ \sup_{\substack{k>m \\ k\in\mathbb{N}\cup\{+\infty\}}} \{(\vartheta_k - \vartheta_m)^2\} + V(m) \right\} + \frac{C}{n}.$$

$\square$

Proof of Proposition 5.1. First let us notice the following upper bound

$$A(m) = \sup_{m'\in\mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m'\wedge m})^2 - V(m') \right\}_+ \leq 3(D_1 + D_2 + D_3)$$

where

$$D_1 = \sup_{m'\in\mathcal{M}} \left\{ (\vartheta_{m'} - \vartheta_{m\wedge m'})^2 \right\}$$

$$D_2 = \sup_{m'\in\mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+$$

$$D_3 = \sup_{m'\in\mathcal{M}} \left\{ (\hat{\vartheta}_{m\wedge m'} - \vartheta_{m\wedge m'})^2 - \frac{V(m')}{6} \right\}_+.$$

- First consider D_1, we have

$$D_1 = \sup_{m'\in\mathcal{M}} \left\{ (\vartheta_{m'} - \vartheta_{m\wedge m'})^2 \right\} = \sup_{\substack{k>m \\ k\in\mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\}. \qquad (26)$$

- Now consider D_2

$$\mathbb{E}[D_2] \leq \sum_{m'\in\mathcal{M}} \mathbb{E}\left[\left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+ \right] \leq \sum_{m'\in\mathcal{M}} \int_{\mathbb{R}^+} \mathbb{P}\left[\left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+ \geq u \right] du$$

$$\leq \sum_{m'\in\mathcal{M}} \int_{\mathbb{R}^+} \mathbb{P}\left[|\hat{\vartheta}_{m'} - \vartheta_{m'}| \geq \sqrt{\frac{V(m')}{6} + u} \right] du. \qquad (27)$$

To handle this probability, we will apply Bernstein's inequality. First we need to identify the empirical process, by definition we have

$$\hat{\vartheta}_{m'} = \sum_{k=0}^{m'-1} a_k(\psi)\hat{a}_k = \sum_{k=0}^{m'-1} a_k(\psi) \left(\sum_{i=0}^{m'-1} [\mathbf{G}_{m'}^{-1}]_{k,i} \frac{1}{n}\sum_{j=1}^{n} \varphi_i(Z_j) \right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} \sum_{k=0}^{m'-1} \sum_{i=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \varphi_i(Z_j) = \frac{1}{n}\sum_{j=1}^{n} \omega_{m'}(Z_j),$$

which yields $\hat{\vartheta}_{m'} - \vartheta_{m'} = \frac{1}{n}\sum_{j=1}^{n} (\omega_{m'}(Z_j) - \mathbb{E}[\omega_{m'}(Z_j)])$. We need to ensure that there exist b and v such that $|\omega_{m'}(Z_j)| \leq b$ a.s. and $\mathbb{V}\text{ar}[\omega_{m'}(Z_j)] \leq v$. First notice

$$|\omega_{m'}(Z_j)| = \left| \sum_{0\leq i,k\leq m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \varphi_i(Z_j) \right| \leq \sum_{i=0}^{m'-1} \left| \varphi_i(Z_j) \sum_{k=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \right|$$

$$\leq \sqrt{2} \sum_{i=0}^{m'-1} \left| \sum_{k=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \right| \leq \sqrt{2}\| {}^t\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1} \|_1 := b.$$

Then we have

$$\mathbb{V}\mathrm{ar}[\omega_{m'}(Z_j)] \leq \mathbb{E}[\omega_{m'}^2(Z_j)] \leq \mathbb{E}\left[\left(\sum_{k=0}^{m'-1}\sum_{i=0}^{m'-1} a_k(\psi)\left[\mathbf{G}_{m'}^{-1}\right]_{k,i}\varphi_i(Z_j)\right)^2\right]$$

$$\leq \int_{\mathbb{R}^+}\left(\sum_{k=0}^{m'-1}\sum_{i=0}^{m'-1} a_k(\psi)\left[\mathbf{G}_{m'}^{-1}\right]_{k,i}\varphi_i(u)\right)^2 h(u)\,\mathrm{d}u \leq \|h\|_\infty\|\,{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_2^2 := v.$$

At last applying Bernstein inequality, we get

$$\mathbb{P}\left[|\hat{\vartheta}_{m'} - \vartheta_{m'}| \geq \sqrt{\frac{V(m')}{6} + u}\right]$$

$$\leq 2\left\{\exp\left(-\frac{n}{4v}\left(\frac{V(m')}{6} + u\right)\right) \vee \exp\left(-\frac{n}{4b}\sqrt{\frac{V(m')}{6} + u}\right)\right\}$$

$$\leq 2\left\{\exp\left(-\frac{n}{4v}\left(\frac{V(m')}{6} + u\right)\right) \vee \exp\left(-\frac{n\alpha}{4b}\sqrt{\frac{V(m')}{6} + u} - \frac{n(1-\alpha)}{4b}\sqrt{\frac{V(m')}{6} + u}\right)\right\}$$

$$\leq 2\left\{\exp\left(-\frac{n}{4v}\left(\frac{V(m')}{6} + u\right)\right) \vee \exp\left(-\frac{n\alpha}{4b}\sqrt{\frac{V(m')}{6}}\right)\exp\left(-\frac{n(1-\alpha)}{4b}\sqrt{u}\right)\right\},$$

with $\alpha \in [0,1]$. Yet we get

$$\frac{n}{4v}\frac{V(m')}{6} \geq \frac{\kappa \log n}{24} \geq p\log n$$

as soon as $\kappa \geq 24p$.

By definition we have

$$\frac{n\alpha}{4b}\sqrt{\frac{V(m')}{6}} = \frac{n\alpha}{4\sqrt{2}\|\,{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_1}\sqrt{\frac{\kappa(\|h\|_\infty \vee 1)\max_{1\leq k\leq m'}\|\,{}^t\vec{\psi}_k\mathbf{G}_k^{-1}\|_2^2\log n}{6n}}$$

$$= \alpha\sqrt{\frac{\kappa(\|h\|_\infty \vee 1)}{96}}\sqrt{\frac{\max_{1\leq k\leq m'}\|\,{}^t\vec{\psi}_k\mathbf{G}_k^{-1}\|_2^2 n\log n}{\|\,{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_1^2}}$$

And as $m' \in \mathcal{M}$, $m' \leq n/\log n$ and according to the following inequalites for any $\vec{u} \in \mathbb{R}^k$

$$\|\vec{u}\|_2 \leq \|\vec{u}\|_1 \leq \sqrt{k}\|\vec{u}\|_2,$$

it yields using **(A)**

$$\frac{n\alpha}{4b}\sqrt{\frac{V(m')}{6}} \geq \alpha\sqrt{\frac{\kappa(\|h\|_\infty \vee 1)}{96}}\sqrt{\frac{n\log n}{m'}} \geq \alpha\sqrt{\frac{(\|h\|_\infty \vee 1)}{96}}\log n \geq \alpha\sqrt{\frac{\|h\|_\infty}{96}}\log n.$$

Thus

$$\frac{n\alpha}{4b}\sqrt{\frac{V(m')}{6}} \geq p\log n$$

as soon as $\kappa \geq 96p^2/\alpha^2$.

Now putting these results into Equation (27), we get

$$\mathbb{E}[D_2] \leq 2 \sum_{m' \in \mathcal{M}} \int_{\mathbb{R}^+} \exp\left(-\frac{nu}{4v} - p\log n\right) \vee \exp\left(-\frac{n(1-\alpha)}{4b}\sqrt{u} - p\log n\right) \mathrm{d}u$$

$$\leq 2 \sum_{m' \in \mathcal{M}} n^{-p} \int_{\mathbb{R}^+} \exp\left(-\frac{nu}{4v}\right) \vee \exp\left(-\frac{n(1-\alpha)}{4b}\sqrt{u}\right) \mathrm{d}u$$

$$\leq 8n^{-p} \sum_{m' \in \mathcal{M}} \max\left(\|h\|_\infty \frac{\|{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_2^2}{n}, 8\frac{\|{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_1^2}{n^2(1-\alpha)^2}\right)$$

$$\leq C(\alpha, \|h\|_\infty)n^{-p} \sum_{m' \in \mathcal{M}} \max\left(\frac{\|{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_2^2}{n}, \frac{\|{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_1^2}{n^2}\right).$$

Since $m' \in \mathcal{M}$, we get

$$\mathbb{E}[D_2] \leq C(\alpha, \|h\|_\infty)n^{-p} \sum_{m' \in \mathcal{M}} \frac{\|{}^t\vec{\psi}_{m'}\mathbf{G}_{m'}^{-1}\|_2^2}{n} \leq C(\alpha, \|h\|_\infty)n^{-p}\mathrm{Card}(\mathcal{M}).$$

Thus for $p = 2$, we have

$$\mathbb{E}[D_2] \leq \frac{C}{n}. \tag{28}$$

• Finally noticing that

$$D_3 = \max\left(\sup_{m'>m}\left\{(\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m')}{6}\right\}_+, \sup_{m' \leq m}\left\{(\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6}\right\}_+\right)$$

$$\leq \max\left(\sup_{m'>m}\left\{(\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m')}{6}\right\}_+, D_2\right)$$

$$\leq \max\left(\sup_{m'>m}\left\{(\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m)}{6}\right\}_+, D_2\right)$$

$$\leq \left\{(\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m)}{6}\right\}_+ + D_2 \leq 2D_2,$$

where we used the fact that $m \mapsto V(m)$ is nondecreasing. Now applying the same reasoning as for D_2 it yields that

$$\mathbb{E}[D_3] \leq \frac{C}{n}. \tag{29}$$

Now gathering Equations (26), (28) and (29) into Equation (25), we get

$$\mathbb{E}[A(m)] \leq \sup_{\substack{k>m \\ k \in \mathcal{M}}}\left\{(\vartheta_k - \vartheta_m)^2\right\} + \frac{C}{n}.$$

$\square$

5.2.2. *Proof of Corollary 3.2.* By definition of $\widetilde{m}$, the following inequalities hold

$$(\vartheta - \hat{\vartheta}_{\widetilde{m}})^2 \leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(\hat{\vartheta}_m - \hat{\vartheta}_{m \wedge \widetilde{m}})^2 + 3(\hat{\vartheta}_{m \wedge \widetilde{m}} - \hat{\vartheta}_{\widetilde{m}})^2$$

$$\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(\widetilde{A}(\widetilde{m}) + \widetilde{V}(m) + \widetilde{A}(m) + \widetilde{V}(\widetilde{m}))$$

$$\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(\widetilde{A}(m) + \widetilde{V}(m))$$

$$\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(A(m) + V(m)) + 6(\widetilde{V}(m) - V(m)) + 6\sup_{\substack{m'>m \\ m,m' \in \mathcal{M}}}\left\{V(m') - \widetilde{V}(m')\right\}_+$$

$$\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(A(m) + V(m)) + 12\sup_{\substack{m' \geq m \\ m,m' \in \mathcal{M}}}\left|V(m') - \widetilde{V}(m')\right|.$$

Taking expectation, we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\widetilde{m}})^2 \le 3\mathbb{E}(\vartheta - \hat{\vartheta}_m)^2 + 6\left(\mathbb{E}[A(m)] + V(m)\right) + 12\mathbb{E}\left[\sup_{\substack{m' \ge m \\ m,m' \in \mathcal{M}}} \left|V(m') - \widetilde{V}(m')\right|\right].$$

To bound the first two terms on the r.h.s. of the previous inequality we apply Proposition 2.1 and Proposition 5.1 we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\widetilde{m}})^2 \le 3\sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{(\vartheta_k - \vartheta_m)^2\right\} + 48V(m) + \frac{C}{n} + 12\mathbb{E}\left[\sup_{\substack{m' \ge m \\ m,m' \in \mathcal{M}}} \left|V(m') - \widetilde{V}(m')\right|\right] \quad (30)$$

To control the fluctuation of the random penalty $\widetilde{V}(m')$ around the true one, we need to introduce the following set

$$\Lambda = \left\{\left|\|\hat{h}_D\|_\infty - \|h\|_\infty\right| < \frac{\|h\|_\infty}{2}\right\}.$$

On one hand, on the set Λ we have that $\|\hat{h}_D\|_\infty \le (3/2)\|h\|_\infty$, it yields that

$$\mathbb{E}\left[\sup_{\substack{m' > m \\ m,m' \in \mathcal{M}}} \left|V(m') - \widetilde{V}(m')\right| \mathbb{1}_\Lambda\right]$$

$$= \mathbb{E}\left[\left|(\|h\|_\infty \vee 1) - (2\|\hat{h}_D\|_\infty \vee 1)\right| \mathbb{1}_\Lambda\right] \kappa \max_{1 \le k \le m} \|{}^t\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n}$$

$$\le 4\kappa(\|h\|_\infty \vee 1) \max_{1 \le k \le m} \|{}^t\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n} = 4V(m).$$

On the set Λ^c, on the other hand using that $m' \in \mathcal{M}$ and Assumption **(A)**

$$\mathbb{E}\left[\sup_{\substack{m' > m \\ m,m' \in \mathcal{M}}} \left|V(m') - \widetilde{V}(m')\right| \mathbb{1}_{\Lambda^c}\right]$$

$$= \mathbb{E}\left[\left|(\|h\|_\infty \vee 1) - (2\|\hat{h}_D\|_\infty \vee 1)\right| \mathbb{1}_{\Lambda^c}\right] \kappa \max_{1 \le k \le m} \|{}^t\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n}$$

$$\le \kappa \log n \, \mathbb{E}\left[\left|(\|h\|_\infty \vee 1) - (2\|\hat{h}_D\|_\infty \vee 1)\right| \mathbb{1}_{\Lambda^c}\right]$$

$$\le \kappa \log n \, \mathbb{E}\left[(\|h\|_\infty \vee 1)\mathbb{1}_{\Lambda^c}\right] + 2\kappa \log n \, \mathbb{E}\left[(\|\hat{h}_D\|_\infty \vee 1)\mathbb{1}_{\Lambda^c}\right].$$

Noticing that $\|\hat{h}_D\|_\infty \le \left\|\sum_{k=0}^{D-1} \varphi_k\right\|_\infty \le 2D \le 2n$, we get

$$\mathbb{E}\left[\sup_{\substack{m' > m \\ m,m' \in \mathcal{M}}} \left|V(m') - \widetilde{V}(m')\right| \mathbb{1}_{\Lambda^c}\right] \le \kappa \log n(\|h\|_\infty \vee 1)\mathbb{E}\left[\mathbb{1}_{\Lambda^c}\right] + 2\kappa n \log n \, \mathbb{E}\left[\mathbb{1}_{\Lambda^c}\right].$$

We now apply the following Lemma proved just after

Lemma 5.2. *For $p > 0$, $\log n \le D \le \dfrac{\|h\|_\infty}{128\sqrt{2}} \dfrac{n}{(\log n)^p}$ and n large enough*

$$\mathbb{P}\left[\left|\|\hat{h}_D\|_\infty - \|h\|_\infty\right| \ge \frac{\|h\|_\infty}{2}\right] \le \frac{2D}{n^p}.$$

It yields for $p = 2$

$$\log n \, \mathbb{E}\left[\mathbb{1}_{\Lambda^c}\right] \leq 2 \log n \, \frac{D}{n^2} \leq 2 \log n \, \frac{\|h\|_\infty}{128\sqrt{2}} \frac{n}{(\log n)^2} \frac{1}{n^2} \leq \frac{\sqrt{2}}{128} \frac{\|h\|_\infty}{n}$$

and similarly for $p = 3$

$$n \log n \, \mathbb{E}\left[\mathbb{1}_{\Lambda^c}\right] \leq \frac{\sqrt{2}}{128} \frac{\|h\|_\infty}{n}.$$

Finally, we have

$$\mathbb{E}\left[\sup_{\substack{m' \geq m \\ m,m' \in \mathcal{M}}} \left|V(m') - \widetilde{V}(m')\right|\right] \leq 4V(m) + \frac{C}{n}. \tag{31}$$

In conclusion, gathering Equations (30) and (31) gives the desired result. □

Proof of Lemma 5.2. Let us notice that

$$\mathbb{P}\left[\left|\|\hat{h}_D\|_\infty - \|h\|_\infty\right| \geq \|h\|_\infty/2\right] \leq \mathbb{P}\left[\|\hat{h}_D - h\|_\infty \geq \|h\|_\infty/2\right]$$

$$\leq \mathbb{P}\left[\|\hat{h}_D - h_D\|_\infty \geq \|h\|_\infty/4\right]$$

$$+ \mathbb{P}\left[\|h_D - h\|_\infty \geq \|h\|_\infty/4\right] := P_1 + P_2. \tag{32}$$

• First consider P_1, the following inequalities hold true.

$$P_1 = \mathbb{P}\left[\sup_{x \in \mathbb{R}^+} \left|\sum_{k=0}^{D-1} (\hat{a}_k(Z) - a_k(h))\varphi_k(x)\right| \geq \|h\|_\infty/4\right] \leq \mathbb{P}\left[\sum_{k=0}^{D-1} |\hat{a}_k(Z) - a_k(h)| \geq \|h\|_\infty/(4\sqrt{2})\right]$$

$$\leq \sum_{k=0}^{D-1} \mathbb{P}\left[|\hat{a}_k(Z) - a_k(h)| \geq \|h\|_\infty/(4\sqrt{2}D)\right]$$

$$\leq \sum_{k=0}^{D-1} \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} (\varphi_k(Z_i) - \mathbb{E}[\varphi_k(Z_i)])\right| \geq \|h\|_\infty/(4\sqrt{2}D)\right]$$

In order to apply Bernstein's inequality, let us notice that $|\varphi_k(Z_1)| \leq \sqrt{2}$ and $\mathbb{V}\text{ar}[\varphi_k(Z_1)] \leq 2$ which yields that

$$P_1 \leq 2D\left(\exp\left(-\frac{n}{8}\frac{\|h\|_\infty}{4\sqrt{2}D}\right) \vee \exp\left(-\frac{n}{4\sqrt{2}}\sqrt{\frac{\|h\|_\infty}{4\sqrt{2}D}}\right)\right).$$

Taking $D \leq \frac{\|h\|_\infty}{128\sqrt{2}} \frac{n}{(\log n)^p}$, we get that

$$P_1 \leq \frac{2D}{n^p}.$$

• Now let us prove that for n large enough $P_2 = 0$. According to Mabon (2014) we can write the following series of equalities,

$$h_D(x) = \sum_{k=0}^{D-1} a_k(h)\varphi_k(x) = \sum_{k=0}^{D-1} \varphi_k(x)\left(2^{-1/2}a_k(f)a_0(g) + \sum_{l=0}^{k-1} 2^{-1/2}\left(a_{k-l}(g) - a_{k-l-1}(g)\right)a_l(f)\right)$$

$$= \sum_{k=0}^{D-1}\sum_{j=0}^{D-1} a_k(f)a_j(g)\int_0^x \varphi_k(u)\varphi_j(x-u)\,\mathrm{d}u = \int_0^x f_D(x-u)g_D(u)\,\mathrm{d}u.$$

Thus we have

$$|h_D(x) - h(x)| = \left| \int_0^x f_D(x-u)g_D(u)\,\mathrm{d}u - \int_0^x f(x-u)g(u)\,\mathrm{d}u \right|$$

$$= \left| \int_0^x (f_D - f)(x-u)g_D(u)\,\mathrm{d}u - \int_0^x f(x-u)(g - g_D)(u)\,\mathrm{d}u \right|$$

$$\leq \left| \int_0^x (f_D - f)(x-u)g_D(u)\,\mathrm{d}u \right| + \left| \int_0^x f(x-u)r_D(u)\,\mathrm{d}u \right|$$

$$\leq \|f_D - f\|\|g\| + \|f\|\|g - g_D\|.$$

Since $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, we get

$$\|g - g_D\|^2 = \sum_{k \geq D} a_k^2(g) \xrightarrow[D \to \infty]{} 0 \qquad \text{and} \qquad \|f - f_D\|^2 = \sum_{k \geq D} a_k^2(f) \xrightarrow[D \to \infty]{} 0$$

Then there exists a certain n_0 such that $\|f_D - f\|\|g\| + \|f\|\|g - g_D\| \leq \varepsilon$. Besides $D \geq \log n$ and $n \geq n_0$ ensure that $\|f_D - f\|\|g\| + \|f\|\|g - g_D\| \leq \varepsilon$. Then starting from Equation (32), we get

$$P_2 \leq \mathbb{P}\left[\varepsilon \geq \frac{\|h\|_\infty}{4}\right] = 0$$

In the end

$$\mathbb{P}\left[\left|\|\hat{h}_D\|_\infty - \|h\|_\infty\right| \geq \frac{\|h\|_\infty}{2}\right] \leq \frac{2D}{n^p}.$$

$\square$

APPENDIX A.

Lemma A.1. *(Bernstein's inequality) Let $X_1, \ldots, X_n$ be i.i.d.random variables such that $\mathbb{V}\mathrm{ar}(X_1) \leq v^2$ and $|X_1| \leq b$ a.s. Let $S_n = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, then for $\varepsilon > 0$*

$$\mathbb{P}\left[|S_n - \mathbb{E}[S_n]| \geq n\varepsilon\right] \leq 2\max\left(\exp\left(-\frac{n\varepsilon^2}{4v^2}\right), \exp\left(-\frac{n\varepsilon}{4b}\right)\right).$$

APPENDIX B.

Let $x \in [0, 1]$, we have

$$S_{p,j}(x) = \frac{1}{p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}\left[x^p(1-x)^j\right]$$

$$= \frac{1}{p!}\sum_{k=0}^p \left(\binom{p}{k}\frac{\mathrm{d}^{p-k}}{\mathrm{d}x^{p-k}}[x^p]\frac{\mathrm{d}^k}{\mathrm{d}x^k}[(1-x)^j]\right)$$

$$= \frac{1}{p!}\sum_{k=0}^p \binom{p}{k}\frac{p!}{k!}x^k(-1)^k\frac{j!}{(j-k)!}(1-x)^{j-k}$$

$$= (1-x)^{j-p}\sum_{k=0}^p \binom{p}{k}\frac{x^k}{k!}(-1)^k\frac{j!}{(j-k)!}(1-x)^{p-k}$$

$$= (1-x)^{j-p}\sum_{k=0}^p \binom{p}{k}\binom{j}{k}\frac{x^k}{k!}(-1)^k(1-x)^{p-k}$$

$$\leq |1-x|^{j-p}\max_{0 \leq k \leq p}\binom{j}{k}(|x| + |1-x|)^p$$

$$\leq 2^p|1-x|^{j-p}.$$

Appendix C.

According to Comte and Genon-Catalot (2015), proving that a function ψ belongs to a Laguerre-Sobolev defined by (12), with α an integer, is equivalent to prove that

$$\|x^{\alpha/2}[\psi(x)e^x]^{(\alpha)}e^{-x}\|^2 = \sum_{k \geq \alpha} k(k-1)\ldots(k-\alpha+1)a_k^2(\psi) \leq L.$$

And we have for ψ defined as in Example 3 that

$$\|x^{\alpha/2}[\psi(x)e^x]^{(\alpha)}e^{-x}\|^2 = \left\| x^{\alpha/2} \sum_{j=0}^{\alpha} \binom{\alpha}{j} (-1)^j \beta(\beta+1)\ldots(\beta+j-1)(1+x)^{-\beta-j} \right\|^2$$

$$= \int_0^{\infty} x^{\alpha} \left| \sum_{j=0}^{\alpha} \binom{\alpha}{j} \beta(\beta+1)\ldots(\beta+j-1)(1+x)^{-\beta-j} \right|^2 dx$$

$$= \sum_{i=0}^{\alpha} \sum_{j=0}^{\alpha} \binom{\alpha}{i}\binom{\alpha}{j} \beta(\beta+1)\ldots(\beta+j-1)\beta(\beta+1)\ldots(\beta+i-1)$$

$$\int_0^{\infty} x^{\alpha}(1+x)^{-2\beta-i-j} dx.$$

Yet $x \mapsto x^{\alpha}(1+x)^{-2\beta-i-j}$ is integrable for all i,j if $\beta > (\alpha+1)/2$.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition.

Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470.

Birgé, L. and Massart, P. (2007). Minimal penalties for gaussian model selection. *Probability Theory and Related Fields*, 138(1-2):33–73.

Bongioanni, B. and Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.*, 192(2):147–172.

Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1):69–98.

Butucea, C. and Tsybakov, A. (2008a). Sharp optimality in density deconvolution with dominating bias I. *Theory Proba. Appl.*, 52(1):24–39.

Butucea, C. and Tsybakov, A. (2008b). Sharp optimality in density deconvolution with dominating bias II. *Theory Proba. Appl.*, 52(2):237–249.

Cai, T. T. and Low, M. G. (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.*, 31(4):1140–1153.

Cai, T. T. and Low, M. G. (2005). Adaptive estimation of linear functionals under different performance measures. *Bernoulli*, 11(2):341–358.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.

Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2013). Laplace deconvolution and its application to dynamic contrast enhanced imaging. Prépublication MAP5 2012-17.

Comte, F. and Genon-Catalot, V. (2015). Adaptive Laguerre density estimation for mixed Poisson models. *Electronic Journal of Statistics*, 9:1112–1148.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B*, 73:601–627.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *The Canadian Journal of Statistics*, 3(34):431–452.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.

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

Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Statistica Neerlandica*, 57(1):136–157.

Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel.

Horn, R. and Johnson, C. (1990). *Matrix Analysis*. Cambridge University Press.

Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301–2323.

Jongbloed, G. (1998). Exponential deconvolution: two asymptotically equivalent estimators. *Statistica Neerlandica*, 52(1):6–17.

Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electronic Journal of Statistics*, 8(2):2879–2904.

Lacour, C. and Massart, P. (2015). Minimal penalty for Goldenshluger-Lepski method. *ArXiv e-prints 1503.00946*.

Laurent, B., Ludeña, C., and Prieur, C. (2008). Adaptive estimation of linear functionals by model selection. *Electronic Journal of Statistics*, 2:993–1020.

Mabon, G. (2014). Adaptive deconvolution on the nonnegative real line. preprint MAP5 2014-33.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330.

Pensky, M. (2015). Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity. *ArXiv e-prints 1411.1660*.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053.

Rebafka, T. and Roueff, F. (2010). Nonparametric estimation of the mixing density using polynomials. *ArXiv e-prints 1002.4516*.

Roueff, F. and Rydén, T. (2005). Nonparametric estimation of mixing densities for discrete distributions. *Ann. Statist.*, 33(5):2066–2108.

van Es, B., Jongbloed, G., and van Zuijlen, M. (1998). Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.*, 26(6):2395–2406.