



# Laguerre deconvolution with unknown matrix operator

Fabienne Comte, Gwennaëlle Mabon

► **To cite this version:**

Fabienne Comte, Gwennaëlle Mabon. Laguerre deconvolution with unknown matrix operator. MAP5 2016-33. 2016. <hal-01416412>

**HAL Id: hal-01416412**

**<https://hal.archives-ouvertes.fr/hal-01416412>**

Submitted on 14 Dec 2016

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# LAGUERRE DECONVOLUTION WITH UNKNOWN MATRIX OPERATOR

F. COMTE<sup>†</sup> AND G. MABON<sup>‡,\*</sup>

<sup>†</sup> MAP5, UMR CNRS 8145, Université Paris Descartes, Paris, France

<sup>‡</sup> Institute of Mathematics Humboldt-Universität zu Berlin, Berlin, Germany

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 the unknown density  $f$  of  $X$  from  $n$  independent, identically, distributed observations of  $Z$ , when the law of the additive process  $Y$  is unknown. When the density of  $Y$  is known, a solution to the problem has been proposed in Mabon (2016b). To make the problem identifiable for unknown density of  $Y$ , we assume that we have access to a preliminary sample of the nuisance process  $Y$ . The question is to propose a solution to an inverse problem with unknown operator. To that aim, we build a family of projection estimators of  $f$  on the Laguerre basis, particularly adapted to the non-negativeness of both random variables. The dimension of the projection space is chosen thanks to a model selection procedure by penalization. At last we prove that the final estimator satisfies an oracle inequality. It can be noted that the study of the mean integrated square risk is based on Bernstein's type concentration inequalities developed for random matrices in Tropp (2015). Finally we illustrate our method on some simulated data.

**Keywords.** Convolution model. Linear inverse problem. Nonnegative random variables. Laguerre basis. Nonparametric density estimation. Random matrix. Oracle inequalities. Adaptive estimation.

**AMS Subject Classification 2010:** Primary 62G07; secondary 62N02.

## 1. INTRODUCTION

**1.1. The model.** We consider in this work the following convolution model:  $Z_i = X_i + Y_i$ , for  $i = 1, \dots, n$  where the observation is the sequence  $(Z_i)_{1 \leq i \leq n}$  while the  $X_i$ 's are the independent and identically distributed variables (i.i.d.) of interest, with common density denoted by  $f$ . The random variables  $Y_i$ ,  $i = 1, \dots, n$  represent a nuisance process, they are also i.i.d. with common density  $g$ . The sequences  $(X_i)_{1 \leq i \leq n}$  and  $(Y_i)_{i \leq i \leq n}$  are assumed to be independent.

Our aim is to perform nonparametric estimation of  $f$ . The specificity of our framework is that all random variables are assumed to be nonnegative. Moreover, we do not suppose that the density  $g$  of the nuisance variables is known. Nevertheless, to make the problem identifiable, we assume that we have at hand an auxiliary nuisance sample  $(Y'_i)_{1 \leq i \leq n_0}$ , independent of  $(X_i, Y_i)_{1 \leq i \leq n}$ . To sum up, we have to solve an inverse problem with unknown operator.

**1.2. Bibliography for real-valued variables.** The literature studies the convolution model for real-valued random variables and for centered  $Y_i$ 's, which are then understood as a noise or a measurement error. Most solutions are based on Fourier methods, relying on the fact that the characteristic function of the observations is the product of the Fourier transforms of  $f$  and  $g$ : then, cautious Fourier inversion of a quotient should allow one to recover  $f$ .

---

\* Corresponding author

*E-mail address:* [fabienne.comte@parisdescartes.fr](mailto:fabienne.comte@parisdescartes.fr), [gwennaelle.mabon@hu-berlin.de](mailto:gwennaelle.mabon@hu-berlin.de).

*Date:* December 14, 2016.

Very instructive discussions with Jean Rochet are gratefully acknowledged. G. MABON acknowledges support by the CREST-ENSAE during the first part of this work then support by the DFG via Research Unit 1735 *Structural Inference in Statistics*.

In the first works,  $g$  is assumed to be known. Under this assumption, rates of convergence and their optimality for kernel estimators have been studied in [Carroll and Hall \(1988\)](#), [Stefanski \(1990\)](#), [Stefanski and Carroll \(1990\)](#), [Fan \(1991\)](#) and [Efromovich \(1997\)](#). For the study of sharp asymptotic optimality, we can cite [Butucea \(2004\)](#), [Butucea and Tsybakov \(2008a,b\)](#). For the most part, the adaptive bandwidth selection in deconvolution models has been addressed with a known error distribution, see for example [Pensky and Vidakovic \(1999\)](#) who apply wavelet technique, [Delaigle and Gijbels \(2004\)](#) for bandwidth selection, [Comte et al. \(2006\)](#) who consider adaptive model selection for projection estimators, or [Meister \(2009\)](#) and references therein.

However the assumption that the distribution of the errors is perfectly known is clearly not realistic in most fields of application. To make the problem feasible, some information on the error distribution is always required. For instance, in a physical context, a preliminary sample of the noise can be obtained. [Neumann \(1997\)](#) first proposed an estimation strategy still based on Fourier inversion; for the study of convergence rates, see [Neumann \(1997\)](#), [Johannes \(2009\)](#) or [Meister \(2009\)](#). The rigorous study of adaptive procedures in a deconvolution model with unknown errors has only recently been addressed. We are aware of the work by [Comte and Lacour \(2011\)](#) and [Kappus and Mabon \(2014\)](#) who extended it to the adaptive strategy, by [Johannes and Schwarz \(2013\)](#) who consider a model of circular deconvolution and by [Dattner et al. \(2016\)](#), who deal with adaptive quantile estimation via Lepski's method.

**1.3. The case of nonnegative variables.** In this paper, all random variables are nonnegative. Such modelization is encountered in survival analysis or reliability models. For instance,  $X$  can be the time of infection of a disease and  $Y$  the incubation time, a model used in so called back calculation problems in AIDS research. In reliability, the lifetime of interest for a component can be hidden by another one, systematically added to it. More broadly the problem of nonnegative variables appears in actuarial or insurance models. Recently, in a financial context, some papers as [Jirak et al. \(2014\)](#) or [Reiß and Selk \(2015\)](#) have addressed the problem of one-sided errors. The first authors are interested in the optimal adaptive estimation in nonparametric regression when the errors are not assumed to be centered anymore, and typically with Exponential density.

[Groeneboom and Wellner \(1992\)](#) have first introduced the problem of one-sided error in the convolution model under monotonicity of the cumulative distribution function (c.d.f.). They derive 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 \(2003\)](#). [van Es \(2011\)](#), in the Uniform deconvolution problem, proposes a density estimator and an estimator of the c.d.f. using kernel estimators and inversion formula. The work of [Mabon \(2016b\)](#) subsumes the existing ones and in this way unifies the approach to tackle the problem of density estimation for nonnegative variables in the convolution model with any known error density. Besides [Mabon \(2016b\)](#) provides a solution to estimate the survival function in this model. In [Mabon \(2016a\)](#), in the same context, the author builds a model selection estimation strategy, with respect to the pointwise risk, for the probability density function and the c.d.f among others.

Indeed, the particularity of the model allows to use a general projection strategy (see [Birgé and Massart \(1997\)](#)) in a specific  $\mathbb{R}^+$ -supported orthonormal functional basis, namely the Laguerre basis. This strategy has been used for nonnegative variables in other settings: e.g. in [Comte et al. \(2015\)](#) and [Vareschi \(2015\)](#) in a regression setting, or in [Belomestny et al. \(2016\)](#) for a multiplicative censoring model.

The present paper is the sequel of [Mabon \(2016b\)](#). She defines the density estimator through its estimated coefficients obtained as the product of the inverse of a known matrix by a random vector: a collection of density projection estimators is proposed. A penalization procedure to select the dimension of the projection space is proved to reach the adequate squared-bias/variance trade-off, in a non-asymptotic way.

Here, we extend the procedure in the case where  $g$  is no longer known: instead, all quantities related to  $g$  are estimated thanks to the independent  $(Y'_i)$ - $n_0$ -sample. This means that we estimate all coefficients of the linear system which was solved in the first step. Therefore the main difficulty is to measure the distance between the inverse of a random matrix and the inverse of its expectation.

This is what makes the problem challenging, and the solution interesting. The strategy is inspired by the one initiated by [Neumann \(1997\)](#) and developed by [Kappus and Mabon \(2014\)](#) in the Fourier context, with help of tools related to matrix perturbation theory (see [Stewart and Sun \(1990\)](#)) and random matrices taken in [Tropp \(2015\)](#). A result of matrix perturbation theory (see [Th. A.1](#)) is the key result to enable us to prove a lemma similar to Lemma 2.1 in [Neumann \(1997\)](#). Besides, Bernstein's inequality for random matrices provides useful moment inequalities. We intend to discuss the influence of the two sample sizes  $n$  and  $n_0$  and to compare our results with the Fourier strategy outcomes, which still can be applied to nonnegative random variables.

**1.4. Outline of the paper.** Let us now explain the plan of the paper. In Section 2, we give notations, we define the model, the Laguerre basis and the density estimator computed on a  $m$ -dimensional projection space. We develop in Section 3 a study of the mean integrated squared error (MISE) of the estimators, based on Bernstein's type concentration inequalities developed for random matrices (see [Tropp \(2015\)](#)). We discuss the resulting rates of convergence under different sets of assumptions. For that we introduce subspaces of  $\mathbb{L}^2(\mathbb{R}^+)$ , called Laguerre-Sobolev spaces with index  $s > 0$  which are defined in [Bongioanni and Torrea \(2009\)](#). This enables us to determine the order of the squared bias terms. This, together with variance order, provides upper bounds on the rates of convergence of the estimators of  $f$  belonging to a Laguerre-Sobolev space. We present a collection of mixed Gamma densities for which our procedure is particularly relevant, and compare our results with those of the Fourier setting. In Section 4, we define a data driven choice of the projection space by using a contrast penalization criterion and we prove an oracle inequality for the final data driven estimator. In Section 5, we study of the adaptive estimators through simulation experiments. Numerical results are then presented and compared to the performances of the direct case (direct observation of the  $X_i$ 's) and to the case of known  $g$ . The results show that our procedure works well and that the cutoff introduced for the estimation of  $g$  plays an interesting role. In the concluding Section 6 we give further possible developments or extensions of the method. All the proofs are postponed to Section 7.

## 2. ESTIMATION PROCEDURE

**2.1. Model and assumptions.** We consider the model

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \quad (1)$$

where the  $X_i$ 's are i.i.d. nonnegative variables with unknown density  $f$ . The  $Y_i$ 's are also i.i.d. nonnegative variables with unknown density  $g$ . We denote by  $h$  the density of the  $Z_i$ 's. The  $X_i$ 's and the  $Y_i$ 's are assumed to be independent. Moreover, we assume in all the following that we have at hand an auxiliary sample of the noise distribution

$$(Y'_1, \dots, Y'_{n_0}) \quad \text{and} \quad (Y'_i)_{1 \leq i \leq n_0} \quad \text{independent of} \quad (X_i, Y_i)_{1 \leq i \leq n}, \quad (2)$$

where the  $Y'_i$ 's are also i.i.d. nonnegative variables with unknown density  $g$ .

Our target is the estimation of the density  $f$  when the  $Z_i$ 's and  $Y'_i$ 's are observed.

**2.2. 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  $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$  belonging to  $\mathbb{L}^2(\mathbb{R})$ , we denote  $\|\varphi\|$  the  $\mathbb{L}^2$  norm of  $\varphi$  defined by  $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$ ,  $\langle \varphi, \psi \rangle$  the scalar product between  $\varphi$  and  $\psi$  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  $\|\vec{u}\|_{2,d}$  the Euclidean norm defined by  $\|\vec{u}\|_{2,d}^2 = {}^t\vec{u}\vec{u}$  where  ${}^t\vec{u}$  is the transpose of  $\vec{u}$ . The scalar product between  $\vec{u}$  and  $\vec{v}$  is  $\langle \vec{u}, \vec{v} \rangle_{2,d} = {}^t\vec{u}\vec{v} = {}^t\vec{v}\vec{u}$ .

We introduce the operator norm of a matrix  $\mathbf{A}$  defined by  $\|\mathbf{A}\|_{\text{op}} = \max_{\|\vec{u}\|_2=1} \|\mathbf{A}\vec{u}\|_2 = \sqrt{\lambda_{\max}({}^t\mathbf{A}\mathbf{A})}$  where  $\lambda_{\max}({}^t\mathbf{A}\mathbf{A})$  is the largest eigenvalue of  ${}^t\mathbf{A}\mathbf{A}$  in absolute value, along with the Frobenius norm defined by  $\|\mathbf{A}\|_{\text{F}} = \sqrt{\sum_{i,j} a_{ij}^2}$ .

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

$$\forall k \in \mathbb{N}, \forall x \geq 0, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}. \quad (3)$$

The Laguerre polynomials  $L_k$  defined by Equation (3) 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  $k$

$$\sup_{x \in \mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_\infty \leq \sqrt{2}. \quad (4)$$

We also introduce the space  $\mathcal{S}_m = \text{Span}\{\varphi_0, \dots, \varphi_{m-1}\}$ . 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

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u) du = 2^{-1/2} (\varphi_{k+j}(x) - \varphi_{k+j+1}(x)) \quad (5)$$

where  $\star$  stands for the convolution product.

**2.4. Projection estimator of the density function when  $g$  is known.** Here we briefly recall the projection estimator of  $f$  when  $g$  is known established in [Mabon \(2016b\)](#).

The principle of a projection method for estimation is to reduce the question of estimating  $f$  to the one of estimating  $f_m$  the projection of  $f$  on  $\mathcal{S}_m$ . We write

$$f_m(x) = \sum_{k=0}^{m-1} a_k(f)\varphi_k(x).$$

Model (1) implies that  $h = f \star g$ . If all the functions  $f, g, h$  belong to  $\mathbb{L}^2(\mathbb{R}^+)$ , then we have

$$\sum_{j \geq 0} a_j(h)\varphi_j = \sum_{k \geq 0} \sum_{\ell \geq 0} a_k(f)a_\ell(g)\varphi_k \star \varphi_\ell.$$

Thus, applying Equation (5) implies, with convention  $a_{-1}(g) = 0$ , that

$$\sum_{j \geq 0} a_j(h)\varphi_j = \sum_{k \geq 0} \sum_{\ell=0}^k 2^{-1/2} (a_{k-\ell}(g) - a_{k-\ell-1}(g)) a_\ell(f)\varphi_k.$$

This yields the following infinite linear triangular system  $\vec{h}_\infty = \mathbf{G}_\infty \vec{f}_\infty$ , with  $\vec{h}_m = {}^t(a_0(h), \dots, a_{m-1}(h))$ ,  $\vec{f}_m = {}^t(a_0(f), \dots, a_{m-1}(f))$  and

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

We can notice that  $\mathbf{G}_m$  is a lower triangular an Toeplitz matrix.

Thus for any  $m$  we can write  $\vec{h}_m = \mathbf{G}_m \vec{f}_m$ . Moreover

$$a_0(g) = \int_{\mathbb{R}^+} g(u)\varphi_0(u) du = \sqrt{2} \int_{\mathbb{R}^+} g(u)e^{-u} du = \sqrt{2}\mathbb{E}[e^{-Y}] > 0.$$

So  $\mathbf{G}_m$  is invertible and  $\mathbf{G}_m^{-1}\vec{h}_m = \vec{f}_m$ . Finally for any  $k \geq 0$   $a_k(h) = \mathbb{E}[\varphi_k(Z_1)]$  can be estimated from the observations and we obtain the projection of  $f$  on  $\mathcal{S}_m$  can be estimated by

$$\hat{f}_m(x) = \sum_{k=0}^{m-1} \hat{a}_k \varphi_k(x) \quad \text{with} \quad \hat{f}_m = \mathbf{G}_m^{-1} \hat{h}_m \quad \text{and} \quad \hat{a}_k(Z) = \frac{1}{n} \sum_{i=1}^n \varphi_k(Z_i) \quad (6)$$

with  $\hat{h}_m = {}^t(\hat{a}_0(Z), \dots, \hat{a}_{m-1}(Z))$  and  $\hat{f}_m = {}^t(\hat{a}_0, \dots, \hat{a}_{m-1})$ .

**2.5. Projection estimator of the density function when  $g$  is unknown.** Thanks to (2) we can easily derive an estimate of  $\mathbf{G}_m$  by replacing its coefficients by their empirical version,

$$[\hat{\mathbf{G}}_m]_{i,j} = \begin{cases} 2^{-1/2}\hat{a}_0(Y') & \text{if } i = j, \\ 2^{-1/2}(\hat{a}_{i-j}(Y') - \hat{a}_{i-j-1}(Y')) & \text{if } j < i, \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where  $\hat{a}_k(Y') = (1/n_0) \sum_{\ell=1}^{n_0} \varphi_k(Y'_\ell)$ . It is clear that  $\mathbb{E}[\hat{\mathbf{G}}_m] = \mathbf{G}_m$ . It is worth noting that  $\hat{\mathbf{G}}_m$  is still a lower triangular Toeplitz matrix and that, as  $\hat{a}_0(Y') = n_0^{-1} \sum_{i=0}^{n_0} \exp(-Y'_i) > 0$ , it is also invertible. However, in order to bound the distance between the inverse of  $\hat{\mathbf{G}}_m$  and  $\mathbf{G}_m^{-1}$ , we have to introduce a cutoff. Thus we define an inverse of  $\hat{\mathbf{G}}_m$  as follows

$$\tilde{\mathbf{G}}_m^{-1} = \begin{cases} \hat{\mathbf{G}}_m^{-1} & \text{if } \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq \sqrt{\frac{n_0}{m \log m}} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Under this definition of  $\tilde{\mathbf{G}}_m^{-1}$ , if we denote by  $\text{spr}(\mathbf{A})$  the spectral radius (largest eigenvalue in absolute value) of  $\mathbf{A}$ , we have

$$\frac{\sqrt{2}}{|\hat{a}_0(Y')|} = \text{spr}(\hat{\mathbf{G}}_m^{-1}) \leq \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}$$

(see Theorem 5.6.9 in [Horn and Johnson \(1990\)](#)). Note that, for any threshold  $t > 0$ ,  $\|\mathbf{G}_m^{-1}\|_{\text{op}} \leq t$  implies that  $2^{-1/2}a_0(g) \geq t^{-1}$  and  $\|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq t$  implies that  $2^{-1/2}|\hat{a}_0(Y')| \geq t$ .

We also want to emphasize that we put the constraint on the spectral norm for technical reasons. It is more convenient to deal with this norm than the trace norm thanks to some results developed in random matrix theory for largest eigenvalues of Hermitian matrices.

Finally, we estimate the projection  $f_m$  of  $f$  on the space  $\mathcal{S}_m$  as

$$\tilde{f}_m(x) = \sum_{k=0}^{m-1} \tilde{a}_k \varphi_k(x) \quad \text{with} \quad \tilde{f}_m = \tilde{\mathbf{G}}_m^{-1} \hat{h}_m \quad (9)$$

with  $\hat{h}_m$  be defined by (6),  $\tilde{\mathbf{G}}_m^{-1}$  by (8) and  $\tilde{f}_m = {}^t(\tilde{a}_0, \dots, \tilde{a}_{m-1})$

### 3. STUDY OF THE $\mathbb{L}^2$ RISK

In this section, we want to derive upper bounds on the MISE of  $\tilde{f}_m$  defined by Equation (9). Using the isomorphism between the Euclidean norm and the  $\mathbb{L}^2$ -norm, we show that

$$\begin{aligned} \mathbb{E}\|f_m - \tilde{f}_m\|^2 &= \|f - f_m\|^2 + \mathbb{E}\|f_m - \tilde{f}_m\|^2 = \|f - f_m\|^2 + \mathbb{E}\|\vec{f}_m - \tilde{f}_m\|_{2,m}^2 \\ &= \|f - f_m\|^2 + \mathbb{E}\|\mathbf{G}_m^{-1} \vec{h}_m - \mathbf{G}_m^{-1} \hat{h}_m + \mathbf{G}_m^{-1} \hat{h}_m - \tilde{\mathbf{G}}_m^{-1} \hat{h}_m\|_{2,m}^2 \\ &\leq \|f - f_m\|^2 + 2\mathbb{E}\|\mathbf{G}_m^{-1}(\vec{h}_m - \hat{h}_m)\|_{2,m}^2 + 2\mathbb{E}\|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\hat{h}_m\|_{2,m}^2. \end{aligned}$$

The first two terms correspond to the squared bias term and the variance term appearing in [Mabon \(2016b\)](#) when the density  $g$  is assumed to be known. The difficulty in this problem lies in bounding the second variance term. We need to study how large the average squared error is when we estimate  $\mathbf{G}_m^{-1}$  by  $\tilde{\mathbf{G}}_m^{-1}$ . For that we use some tools of random matrix theory and particularly matrix concentration inequalities (see [Tropp \(2015\)](#)), and Paulsen dilatation trick (see the proof of Corollary 7.2).

First we give upper bounds on the  $\mathbb{L}^2$  risk by bounding the variance term then we compute optimal rates by introducing Laguerre-Sobolev spaces and estimating the order of the bound on the variance.

**3.1. Upper bounds on the MISE.** First we state the key lemma to bound the  $\mathbb{L}^2$  risk of  $\tilde{f}_m$  along with an important corollary.

**Lemma 3.1.** *For  $\tilde{\mathbf{G}}_m^{-1}$  defined by Equation (8),  $\|g\|_\infty < \infty$  and  $m \log m \leq n_0$ , then for any integer  $p$  there exists a positive constant  $\mathfrak{C}_{\text{op},p}$  such that*

$$\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] \leq \mathfrak{C}_{\text{op},p} \left( \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^4 \frac{m}{n_0} \right)^p.$$

**Corollary 3.2.** *Under the Assumptions of Lemma 3.1, there exists a positive constant  $\mathfrak{C}_F$  such that*

$$\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_F^2 \right] \leq \mathfrak{C}_F \left( \|\mathbf{G}_m^{-1}\|_F^2 \wedge \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \|\mathbf{G}_m^{-1}\|_F^2 \frac{m}{n_0} \right).$$

**Corollary 3.3.** *Under the Assumptions of Lemma 3.1, there exists a positive constant  $\mathfrak{C}_E$  such that*

$$\mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right] \leq \mathfrak{C}_E \left( 1 \wedge \log m \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \right).$$

We can now state the main result of this subsection:

**Proposition 3.4.** *If  $f$  and  $g$  belong to  $\mathbb{L}^2(\mathbb{R}^+)$ ,  $\|g\|_\infty < \infty$ , for  $\tilde{f}_m$  defined by (9) the following result holds*

$$\mathbb{E} \|f - \tilde{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{\mathfrak{C}}{n} (2m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_F^2) + 4\mathfrak{C}_E \log m \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \quad (10)$$

with  $\mathfrak{C} = 2 + \mathfrak{C}_{\text{op},1} + \mathfrak{C}_F$ .

We can notice that, in the right hand side of Equation (10), the first two terms correspond to the upper bound on the mean integrated risk when the matrix  $\mathbf{G}_m^{-1}$  is known (see Proposition 3.1 in Mabon (2016b)). The first one is the squared bias term which gets smaller when  $m$  increases and the second one is the variance term when  $g$  is known. Thanks to Lemma 3.4 in Mabon (2016b), we know that the spectral norm of  $\mathbf{G}_m^{-1}$  grows with the dimension  $m$ . The third term is due to the estimation of the matrix  $\mathbf{G}_m^{-1}$ . This last term seems to deteriorate the rate compared to known noise case in particular if  $n = n_0$ . If we add some assumptions we can have the following Corollary

**Corollary 3.5.** *If  $m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2/n_0 < 1/4$  and  $\|f\|_{\ell_1} := \sum_j |a_j(f)| < \infty$ , then for any integer  $p \geq 2$ , we have*

$$\begin{aligned} \mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right] &\leq \frac{\mathfrak{C}'_E}{n_0} (1 \vee \|g\|_\infty)^3 (\|f\|^2 \vee \|f\|_{\ell_1}^2) (m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|\mathbf{G}_m^{-1}\|_F^2) \\ &\quad + 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{2m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n_0} \right)^p. \end{aligned}$$

It is worth mentioning that the condition  $m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2/n_0 < 1/4$  can be replaced for any  $\epsilon > 0$  by  $m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2/n_0 \leq \epsilon < 1$  and Corollary 3.5 is still valid.

Corollary 3.5 leads us to the following bound, under slightly stronger assumption than in Proposition 3.4.

**Proposition 3.6.** *If  $f$  and  $g$  belong to  $\mathbb{L}^2(\mathbb{R}^+)$ ,  $\|g\|_\infty < \infty$ ,  $\|f\|_{\ell_1} < \infty$  and  $m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2/n_0 < 1/4$ , for  $\tilde{f}_m$  defined by (9) the following result holds*

$$\begin{aligned} \mathbb{E} \|f - \tilde{f}_m\|^2 &\leq \|f - f_m\|^2 + \frac{\mathfrak{C}}{n} (2m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_F^2) \\ &\quad + \frac{2\mathfrak{C}'_E}{n_0} (1 \vee \|g\|_\infty)^3 (\|f\|^2 m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|f\|_{\ell_1}^2 \|\mathbf{G}_m^{-1}\|_F^2) + 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{2m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n_0} \right)^p. \end{aligned}$$



Note that the condition  $\|f\|_{\ell_1} < \infty$ , as the Laguerre basis is bounded, is a condition of normal convergence of the series. As  $\varphi_j(0) = \sqrt{2}$  for all  $j \geq 0$ , it is also the natural condition ensuring that  $f(0)$  is well defined.

We illustrate hereafter that this bound implies better upper rates of estimation than the one given in Proposition 3.4.

**3.2. Rates of convergence.** To derive the rates of convergence of  $\tilde{f}_m$  defined by (9), we need to evaluate the smoothness of the signal along with the order of  $\|\mathbf{G}_m^{-1}\|_{\text{op}}^2$  and  $\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$ . First of all, we assume that  $f$  belongs to a Laguerre-Sobolev space defined as

$$W^s(\mathbb{R}^+, L) = \left\{ p : \mathbb{R}^+ \rightarrow \mathbb{R}, p \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s a_k^2(p) \leq L < +\infty \right\} \quad \text{with } s \geq 0 \quad (11)$$

where  $a_k(p) = \langle p, \varphi_k \rangle$ . Bongioanni and Torrea (2009) have introduced Laguerre-Sobolev space but the link with the coefficients of a function on a Laguerre basis was done by Comte and Genon-Catalot (2015). Indeed, let  $s$  be an integer, for  $p : \mathbb{R}^+ \rightarrow \mathbb{R}$  and  $f \in \mathbb{L}^2(\mathbb{R}^+)$ , we have that  $\sum_{k \geq 0} k^s a_k^2(p) < +\infty$  is equivalent to the fact that  $p$  admits derivatives up to order  $s - 1$  with  $p^{(s-1)}$  absolutely continuous and for  $0 \leq k \leq s$ ,  $x^{k/2}(p(x)e^x)^{(k)}e^{-x} \in \mathbb{L}^2(\mathbb{R}^+)$ . For more details we refer to section 7 of Comte and Genon-Catalot (2015). Thus, for  $f \in W^s(\mathbb{R}^+, L)$  defined by (11),

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} a_k^2(f) = \sum_{k=m}^{\infty} a_k^2(f) k^s k^{-s} \leq L m^{-s}.$$

Now we have to evaluate the variance terms of Equations (10) which means assess the order of  $\|\mathbf{G}_m^{-1}\|_{\text{op}}^2$  and  $\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$ . First we define an integer  $r \geq 1$  such that

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

**Lemma 3.7** (Comte et al. (2015)). *If 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  $g$  has no zero with non negative real parts except for the zeros of the form  $\infty + ib$ .

are true, then it holds that

$$c_1 m^{2r} \leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \leq c_2 m^{2r}.$$

where  $c_1 \leq c_2$  are constants independent of  $m$ .

Optimizing the squared bias and the variance terms in the upper bounds stated in Propositions 3.4 and 3.6 imply the following results.

**Proposition 3.8.** *If  $f$  belongs to  $W^s(\mathbb{R}^+, L)$  and  $g$  satisfies (C1)-(C2) then  $\tilde{f}_{m_{\text{opt}}}$  defined by (9) and  $m_{\text{opt}} \propto n^{1/s+2r} \wedge (n_0/\log n_0)^{1/s+2r+1}$ , then*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \|f - \tilde{f}_{m_{\text{opt}}}\|^2 \leq C_1(s, L) n^{-s/s+2r} \vee \left( \frac{n_0}{\log n_0} \right)^{-s/s+2r+1}.$$

where  $C_1(s, L)$  is a positive constant.

If in addition  $n_0 \geq n^{3/2}$ , then choosing  $m_{\text{opt}} \propto n^{1/s+2r}$ , yields

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \|f - \tilde{f}_{m_{\text{opt}}}\|^2 \leq C_2(s, L) n^{-s/s+2r}.$$

where  $C_2(s, L)$  is a positive constant.



**Corollary 3.9.** *If  $f$  belong to  $W^s(\mathbb{R}^+, L)$  and  $\|f\|_{\ell_1} < \infty$ ,  $g$  satisfies (C1)-(C2) for  $\tilde{f}_{m_{opt}}$  defined by (9) and  $m_{opt} = c(n^{1/s+2r} \wedge n_0^{1/s+2r})$  for  $c$  a positive constant, the following result holds*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \|f - \tilde{f}_{m_{opt}}\|^2 \leq \mathfrak{c}(n \vee n_0)^{-s/s+2r}, \quad \mathfrak{c} \text{ a positive constant.}$$

In particular, the condition  $\|f\|_{\ell_1} < \infty$  is implied by  $f \in W^s(\mathbb{R}^+, L)$  for  $s > 1$ . Indeed,

$$\sum_{k \geq 1} |a_k(f)| \leq \sqrt{\frac{s}{s-1} \sum_{k \geq 1} k^s a_k^2(f)} \leq \sqrt{\frac{s}{s-1}} L < +\infty.$$

**3.3. Examples of rates and comparison with Fourier setting.** In this section we denote by  $\psi^*(x) = \int e^{-iux} \psi(u) du$  the Fourier transform of an integrable function  $\psi$ . The Fourier estimator of  $f$  in the model defined by (1)-(2) is in fact an estimator of  $f_{m, \text{Fo}}(x) = (2\pi)^{-1} \int_{-\pi m}^{\pi m} f^*(u) du$ , the orthogonal projection of  $f$  on the space  $\mathfrak{S}_m = \{\psi \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \text{support}(\psi^*) \subset [-\pi m, \pi m]\}$ . It is given by

$$\hat{f}_{m, \text{Fo}}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{iux} \frac{\hat{h}^*(u)}{\hat{g}^*(u)} du$$

with

$$\hat{h}^*(u) = \frac{1}{n} \sum_{j=1}^n e^{-iuZ_j}, \quad \hat{g}^*(u) = \frac{1}{n_0} \sum_{j=1}^{n_0} e^{-iuY'_j}, \quad \frac{1}{\tilde{g}^*(u)} = \frac{\mathbb{1}\{|\hat{g}^*(u)| \geq n_0^{-1/2}\}}{\hat{g}^*(u)}.$$

The risk bound obtained in Neumann (1997) can be written as follows,

$$\mathbb{E} \|f - \hat{f}_{m, \text{Fo}}\|^2 \leq \|f - f_{m, \text{Fo}}\|^2 + C_1 \frac{\Delta(m)}{n} + (4C_1 + 2) \frac{\Delta_f(m)}{n_0}$$

with  $C_1$  a constant and

$$\Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|g^*(u)|^2} du, \quad \Delta_f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|g^*(u)|^2} du.$$

The structure of the estimator is similar to the structure of ours, with a cutoff required for safe inversion of the noise characteristic function. The structure of the upper bound is also similar and involves a squared bias term  $\|f - f_{m, \text{Fo}}\|^2$ , a variance term corresponding to known  $g$ ,  $\Delta(m)/n$  and the price for estimating  $g$ ,  $\Delta_f(m)/n_0$ .

There are also several differences. The bias term does not refer to the same regularity. It is clearly explained in Mabon (2016b) that, if  $f$  is a Gamma density  $\gamma(p, \theta)$ , then the bias is of order  $m^{-2p+1}$  in the Fourier setting while it is exponentially decreasing in the Laguerre setting (namely of order  $m^{2(p-1)} \exp(-\rho m)$ , for  $\rho = \rho(\theta) > 0$ ). Thus, most reasonably, our method, dedicated to  $\mathbb{R}^+$ -supported function estimation, performs at best for Gamma and all types of mixed Gamma densities (see Section 2.3.3 in Mabon (2016b)).

The first variance term is simpler in the Fourier setting than in the Laguerre setting in the sense that there is no choice between two quantities, and the characteristic function of the noise is better known than the trace and operator norms of  $\mathbf{G}_m^{-1}$ . However, for  $g$  following a Gamma or a beta distribution, it is checked in Mabon (2016b) that these variance terms have the same order with respect to  $m$  in Laguerre and Fourier settings: if  $g$  follows a  $\gamma(q, \mu)$  density, both upper bounds have order less than  $m^{2q}/n$ ; if  $g$  follows a  $\beta(a, b)$  density with  $b > a \geq 1$ , both upper bounds have order less than  $m^{2a}/n$ .

For the second variance term, it is straightforward that  $\Delta_f(m) \leq \Delta(m)$  and thus the estimation of  $g$  does not change the Fourier risk as soon as  $n_0 \geq n$ . A similar result is obtained for Laguerre estimator, in Proposition 3.6, with a condition on  $\|f\|_{\ell_1}$  similar to than the property  $|f^*(u)| \leq 1$ .

As a consequence, the Laguerre estimator has smaller upper bounds on the rates than the Fourier methods when the function  $f$  under estimation belongs to a class of mixed Gamma densities: the exponential decrease of the Laguerre bias, implies that choice of small  $m$ 's (namely  $m = c \log(n)$  for large enough constant  $c$ ) are possible, and make also the variance small. In this case, the rates

are of order  $(\log n)^\alpha/n$  with  $\alpha > 0$ . However, the Fourier method remains more general and has to be used for  $\mathbb{R}$ -supported functions.

Now, as we are about to deal with model selection, we can mention that in the Laguerre method, the quantity  $m$  to be selected is a dimension and is therefore searched among a set of integers, while in the Fourier method, fractional  $m$ 's are often considered and it is a real difficulty to determine which set of values is wise to be visited in the selection procedure.

#### 4. MODEL SELECTION AND ADAPTATION

The aim of this section is to select an integer  $m$  that enables us to compute an estimator of the unknown density  $f$  with the  $\mathbb{L}^2$  risk as close as possible to the oracle risk  $\inf_m \mathbb{E} \|f - \hat{f}_m\|^2$ . We follow the model selection paradigm (see [Birgé and Massart \(1997\)](#), [Birgé \(1999\)](#), [Massart \(2003\)](#)) and choose the dimension of projection spaces  $m$  as the minimizer of a penalized criterion.

First, we consider the following sets of integers:

$$\begin{aligned}\widehat{\mathcal{M}} &= \left\{ 1 \leq m \leq C \lfloor n/\log n \rfloor \wedge \lfloor n_0/\log n_0 \rfloor, m \log m \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \leq n \wedge n_0 \right\}, \\ \mathcal{M} &= \left\{ 1 \leq m \leq C \lfloor n/\log n \rfloor \wedge \lfloor n_0/\log n_0 \rfloor, 4m \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \leq n \wedge n_0 \right\},\end{aligned}$$

with  $C$  a positive constant. Next, we define the two parts of the random penalty

$$\begin{aligned}\widehat{\text{pen}}_1(m) &:= 2\kappa_1 \mathfrak{C}(\|h\|_\infty \vee 1) \frac{\log n}{n} \left( 2m \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \wedge \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2 \right) \\ \widehat{\text{pen}}_2(m) &:= 8\kappa_2 (\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0} \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2.\end{aligned}$$

Then we set the random penalty as

$$\widehat{\text{pen}}(m) := \widehat{\text{pen}}_1(m) + \widehat{\text{pen}}_2(m). \quad (12)$$

We also define the deterministic counterparts

$$\begin{aligned}\text{pen}_1(m) &:= 2\kappa_1 \mathfrak{C}(\|h\|_\infty \vee 1) \frac{\log n}{n} \left( 2m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|\mathbf{G}_m^{-1}\|_{\text{F}}^2 \right) \\ \text{pen}_2(m) &:= 8\kappa_2 (\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2\end{aligned}$$

and set the deterministic penalty as

$$\text{pen}(m) := \text{pen}_1(m) + \text{pen}_2(m) \quad (13)$$

where  $\kappa_1$  and  $\kappa_2$  are numerical constants, see our comment in Section 5. Then we can prove the following result.

**Theorem 4.1.** *Assume that  $f$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$  with  $\|g\|_\infty < \infty$ . Let  $\hat{f}_{\widehat{m}}$  be defined by (9) and*

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

*with  $\widehat{\text{pen}}$  defined by (12), then there exists a positive numerical constant  $\kappa_1$  such that*

$$\mathbb{E} \|f - \tilde{f}_{\widehat{m}}\|^2 \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \|f - f_m\|^2 + \text{pen}(m) \right\} + \frac{C}{n \wedge n_0},$$

*where  $C^{ad}$  is a numerical constant and  $C$  depends on  $\|f\|$  and  $\|g\|$ ,  $\text{pen}$  is defined by (13).*

This Theorem gives an oracle inequality which establishes a non asymptotic oracle bound. It shows that the squared bias variance trade-off is automatically made up to a loss of logarithmic factor and a multiplicative constant. Theorem 4.1 is derived under mild assumptions.

Some comments for practical use are in order. Indeed in the penalty terms  $\widehat{\text{pen}}_1$  and  $\widehat{\text{pen}}_2$ , there are four quantities which deserve some explanations:  $\kappa_1$ ,  $\kappa_2$ ,  $\|g\|_\infty$  and  $\|h\|_\infty$ . It follows from the proof that  $\kappa_1 = 196$  and  $\kappa_2 = 5/2$  would suit. But in practice, values obtained from the theory are generally too large and constants are calibrated by simulations. Once chosen, they remain fixed for all simulation experiments. There are still two unknown terms in the penalty,  $\|g\|_\infty$  and  $\|h\|_\infty$ , that must be estimated. We have to check that we can derive an oracle inequality when those terms are estimated, which is done in the following Corollary.

Beforehand let us define projection estimators of  $h$  and  $g$

$$\hat{h}_{D_1}(x) = \sum_{k=0}^{D_1-1} \hat{a}_k(Z) \varphi_k(x) \quad \text{with} \quad \hat{a}_k(Z) = (1/n) \sum_{i=1}^n \varphi_k(Z_i), \quad (14)$$

$$\hat{g}_{D_2}(x) = \sum_{k=0}^{D_2-1} \hat{a}_k(Y') \varphi_k(x) \quad \text{with} \quad \hat{a}_k(Y') = (1/n_0) \sum_{i=1}^{n_0} \varphi_k(Y'_i). \quad (15)$$

We can see that  $\hat{h}_{D_1}$  and  $\hat{g}_{D_2}$  are respectively unbiased estimators of  $h_{D_1}(x) = \sum_{k=0}^{D_1-1} a_k(h) \varphi_k(x)$  and  $g_{D_2}(x) = \sum_{k=0}^{D_2-1} a_k(g) \varphi_k(x)$ .

**Corollary 4.2.** *Assume that  $f$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$  with  $\|g\|_\infty < \infty$ . Let  $\tilde{f}_{\tilde{m}}$  be defined by (9) and*

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

with  $\widetilde{\text{pen}}$  defined by  $\widetilde{\text{pen}}(m) := \widetilde{\text{pen}}_1(m) + \widetilde{\text{pen}}_2(m)$  with

$$\begin{aligned} \widetilde{\text{pen}}_1(m) &:= 4\kappa_1 \log n \mathfrak{C}(\|\hat{h}_{D_1}\|_\infty \vee 1) \left( 2m \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \wedge \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2 \right) / n \\ \widetilde{\text{pen}}_2(m) &:= 16\kappa_2 (\|\hat{g}_{D_2}\|_\infty \vee 1) \log n_0 m \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 / n_0, \end{aligned} \quad (16)$$

where  $\hat{h}_{D_1}$  and  $\hat{g}_{D_2}$  are given by (14) and (15),  $D_1$  and  $D_2$  satisfy  $\log n \leq D_1 \leq \|h\|_\infty n / (128\sqrt{2} \log^3 n)$  and  $\log n_0 \leq D_2 \leq \|g\|_\infty n_0 / (128\sqrt{2} \log^3 n_0)$ . Then there exist positive numerical constants  $\kappa_1$  and  $\kappa_2$  such that

$$\mathbb{E} \|f - \tilde{f}_{\tilde{m}}\|^2 \leq C^{ad} \inf_{m \in \tilde{\mathcal{M}}} \left\{ \|f - f_m\|^2 + \text{pen}(m) \right\} + \frac{C}{n \wedge n_0},$$

Note that the constraint on  $D_1$  and  $D_2$  are fulfilled respectively for  $n$  and  $n_0$  large enough as soon as  $D_1 \simeq \sqrt{n}$  and  $D_2 \simeq \sqrt{n_0}$  for instance. In this sense Corollary 4.2 has rather an asymptotic flavor.

## 5. ILLUSTRATION

The whole implementation is conducted using Matlab software. The integrated squared error  $\|f - \tilde{f}_{\tilde{m}}\|^2$  is computed via a standard approximation and discretization (over 100 points) of the integral on an interval of  $\mathbb{R}^+$  denoted by  $I_f$ . Then the mean integrated squared error (MISE)  $\mathbb{E} \|f - \tilde{f}_{\tilde{m}}\|^2$  is computed as the empirical mean of the approximated ISE over 200 simulation samples.

**5.1. Simulation setting.** We consider the following six densities with unit variance.

- ▷ An exponential density  $\mathcal{E}(1)$  with parameter 1, on  $I_f = [0, 5]$
- ▷ A Gamma density  $X = 2\gamma(4, 1/4)$ , on  $I_f = [0, 10]$
- ▷ A mixed Gamma  $X/c$  with  $X \sim 0.4\gamma(2, 1/2) + 0.6\gamma(16, 1/4)$ , with  $c = \sqrt{2.96}$ ,
- ▷ A Weibull density,  $X/c$  with  $f(x) = kx^{k-1}e^{-x^k} \mathbf{1}_{\mathbb{R}^+}(x)$  with  $c = \sqrt{\Gamma(7/3) - \Gamma(5/3)^2}$  on  $I_f = [0, 5]$ ,
- ▷ A Rayleigh density  $X \sim f$  with  $f(x) = (x/\sigma^2) \exp(-x^2/(2\sigma^2))$  with  $\sigma^2 = 2/(4 - \pi)$  on  $I_f = [0, 5]$ ,
- ▷ A beta density  $X/c$  with  $X \sim \beta(4, 5)$ ,  $c = \sqrt{2}/9$  on  $I_f = [0, 1/c]$ .

We also consider two types of noises  $Y$  with same variance, namely an exponential density  $\mathcal{E}(\lambda)$  with  $\lambda = 2$ , and a gamma density  $\gamma(2, 1/\lambda')$  with  $\lambda' = 2\sqrt{2}$ . In both cases, the variance is equal to  $1/4$ .

In the case where the noise density is assumed to be known, we can compute analytically the matrix  $\mathbf{G}_m$  and use the exact formulae:

▷ For  $Y \sim \mathcal{E}(\lambda)$

$$[\mathbf{G}_m]_{i,j} = \lambda/(1+\lambda)\mathbb{1}_{i=j} - 2\lambda \frac{(\lambda-1)^{i-j-1}}{(\lambda+1)^{(i-j+1)}} \mathbb{1}_{j<i} \quad (17)$$

▷ For  $Y \sim \gamma(2, \mu)$

$$[\mathbf{G}_m]_{i,j} = (\mu/(1+\mu))^2 \mathbb{1}_{i-1=j} - 4\mu^2/(1+\mu)^3 \mathbb{1}_{i=j} + 4(i-j-\mu)\mu^2 \frac{(\mu-1)^{i-j-2}}{(\mu+1)^{(i-j+2)}} \mathbb{1}_{j+1<i} \quad (18)$$

**5.2. Practical estimation procedure.** As in [Mabon \(2016b\)](#), to illustrate the loss implied by the noise, we apply the density estimation method on the true  $X_i$ 's, for comparison, with a specific  $\tau_0 = 0.25$  in the penalty; more precisely, the case called "direct" hereafter relies on the estimator  $\hat{f}_{\hat{m}}^{(0)}$  with  $\hat{f}_m^{(0)} = \sum_{j=0}^{m-1} \hat{a}_j^{(0)} \varphi_j$ ,  $\hat{a}_k^{(0)} = n^{-1} \sum_{i=1}^n \varphi_k(X_i)$  and

$$\hat{m}_0 = \arg \min_{m \in \{0,1,\dots,n\}} \left\{ - \sum_{k=0}^{m-1} (\hat{a}_k^{(0)})^2 + \frac{2\tau_0 m}{n} \right\}.$$

To study if the estimation of  $\mathbf{G}_m$  implies a loss, we implement the "known noise" case. We compute  $\mathbf{G}_m$  as given by (17) and (18) and we apply the procedure described in [Mabon \(2016b\)](#). We compute the estimator as given by (6) and select

$$\hat{m}_1 = \arg \min_{m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \leq n/\log(n)} \left\{ -\|\hat{f}_m\|^2 + \frac{\tau_1}{n} (2m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \log(n)(\|g\|_{\infty} \vee 1) \|\mathbf{G}_m^{-1}\|_{\text{F}}^2) \right\}.$$

We set  $\tau_1 = 0.03$  in the penalty for known noise density, this is the value calibrated in [Mabon \(2016b\)](#), and  $\|g\|_{\infty}$  is known in this setting.

For the case of estimated  $\mathbf{G}_m$  which is specifically studied in the present work, we compute  $\tilde{f}_{\tilde{m}}$  with  $\tilde{f}_m$  given by (9) and  $\tilde{m}$  given by  $\tilde{m} = \arg \min_{m \in \tilde{\mathcal{M}}} \left\{ -\|\tilde{f}_m\|^2 + \widetilde{\text{pen}}(m) \right\}$ , with  $\widetilde{\text{pen}}(m)$  defined by (16). The constant calibrations were done with intensive preliminary simulations, including other densities than the ones mentioned above (to avoid overfitting): the selected values are  $\kappa_1 = 0.01$  and  $\kappa_2 = 0.01/4$ . It can be noted that the values of  $\kappa_1$  and  $\kappa_2$  are much smaller than what comes in theory. The infinite norms  $\|h\|_{\infty}$  and  $\|g\|_{\infty}$  are estimated by taking the maximum of a projection estimator in the Laguerre basis of the density of  $Z$  (resp. of  $Y'$ ) with dimension taken as the integral part of  $\sqrt{n}/3$ .

**5.3. Simulation results.** As in [Mabon \(2016b\)](#), we consider two sample sizes  $n = 200$  and  $n = 2000$ . For each distribution, we present in [Tables 1 and 2](#) the MISE computed over 200 repetitions, together with the standard deviation, both being multiplied by 100 for small sample size 200 ([Table 1](#)) and by 1000 for larger sample size ([Table 2](#)). For simplicity, the dimension is selected in all cases among 30 values. We also provide "oracles", with mean values and standard deviations also multiplied by the same factor as the MISE: we compute over 200 repetitions the MISE which would be obtained if we were choosing the best proposal in our family of thirty estimators. These oracles use the knowledge of the true, that we do not have in practice, and they are computed on other samples than the MISE of model selection.

We can see by comparing [Tables 1 and 2](#) (recall that the multiplying factor is 100 for the first table and 1000 for the second), that the results are improved when  $n$  increases. Estimating the matrix  $\mathbf{G}_m$  does not seem to really increase the error when we compare with the case where it is known; it even sometimes happens that the estimation of  $\mathbf{G}_m$  improves the MISE. In deconvolution setting, the same remark had been made by [Comte and Lacour \(2011\)](#), it seems that the cutoff in the estimation procedure is often safe. For fixed  $n$  and estimated  $\mathbf{G}_m$ , increasing  $n_0$  systematically improves the results, except in the case where  $f$  is exponential with parameter 1. But this case corresponds to a best estimation proportional to  $\varphi_0$ , a simplicity which seems to be difficult for the estimation algorithm. We can also see that the mixed Gamma distribution has the highest errors and is clearly more difficult to estimate:  $n = 200$  seems too small to get a good account of the bimodality. We can also see that increasing the degree of the inverse problem when going from

Exponential to Gamma distribution for  $Y$  always increases the errors, even if the signal-to-noise ratio is unchanged.

$f$		direct	Y Exponential			Y Gamma		
			Known noise	Noise sample $n_0 = 50$	Noise sample $n_0 = 200$	Known noise	Noise sample $n_0 = 50$	Noise sample $n_0 = 200$
Exp(1)	MISE	0.5	8.2	2.1	3.3	4.2	1.9	2.2
	(std)	(0.9)	(33)	(3.1)	(6.4)	(23)	(3.3)	(4.1)
	Oracles	0.10	0.13	0.25	0.15	0.13	0.29	0.16
	(std)	(0.1)	(0.2)	(0.3)	(0.2)	(0.2)	(0.5)	(0.2)
Gamma	MISE	0.37	1.0	1.6	0.8	2.2	1.2	1.7
	(std)	(0.4)	(0.7)	(0.7)	(0.6)	(0.3)	(0.3)	(0.7)
	Oracles	0.2	0.3	0.5	0.4	0.4	1.5	0.4
	(std)	(0.2)	(0.3)	(0.4)	(0.4)	(0.4)	(0.7)	(0.3)
Mixed Gamma	MISE	1.0	4.0	6.7	2.7	7.3	7.5	7.2
	(std)	(0.4)	(2.6)	(1.9)	(2.1)	(0.8)	(1.1)	(0.8)
	Oracles	0.7	1.6	5.1	2.0	2.4	7.0	6.1
	(std)	(0.4)	(1.1)	(1.8)	(1.3)	(1.5)	(1.0)	(1.0)
Weibull	MISE	0.4	0.8	1.1	0.9	1.0	1.1	0.9
	(std)	(0.4)	(0.8)	(1.1)	(1.1)	(0.9)	(0.7)	(0.8)
	Oracles	0.3	0.4	0.6	0.5	0.5	0.8	0.5
	(std)	(0.2)	(0.4)	(0.6)	(0.5)	(0.5)	(0.9)	(0.5)
Rayleigh	MISE	0.4	0.8	1.0	0.6	1.1	1.1	1.0
	(std)	(0.4)	(0.4)	(0.3)	(0.5)	(0.2)	(0.2)	(0.3)
	Oracles	0.2	0.3	0.4	0.4	0.3	0.4	0.3
	(std)	(1.2)	(1.5)	(1.6)	(0.3)	(0.3)	(0.3)	(0.3)
Beta	MISE	0.3	1.4	1.7	0.8	1.7	1.8	1.7
	(std)	(0.2)	(0.6)	(0.3)	(0.6)	(0.1)	(0.2)	(0.1)
	Oracles	0.2	0.3	0.5	0.3	0.4	1.7	0.6
	(std)	(0.2)	(0.2)	(0.3)	(0.2)	(0.3)	(0.2)	(0.3)

TABLE 1. Results after 200 iterations of simulations of the six considered densities, for sample sizes  $n = 200$  and  $n_0 = 50$ ,  $n_0 = 200$ . For each density : first two lines, MISE  $\times 100$  with (std  $\times 100$ ) in parenthesis; third and fourth lines, mean with std in parenthesis of oracles. First column, direct observations of the  $X_i$ 's. Columns 2, 3 and 4, noise is  $\mathcal{E}(\lambda)$  with  $\lambda = 2$  (mean  $1/2$ ). Columns 5, 6 and 7, noise is  $\gamma(2, \lambda')$  with  $\lambda' = 2\sqrt{2}$  (mean  $1/(2\sqrt{2})$ ).

## 6. CONCLUDING REMARKS

In this work, we have defined a projection estimator of the density  $f$  of unobserved i.i.d. random variables  $X_i$ ,  $i = 1, \dots, n$ , when data  $(Z_i)_{1 \leq i \leq n}$  from model (1) are available, together with an independent sample  $(Y'_i)_{1 \leq i \leq n_0}$  of the nuisance process  $Y$ . All quantities related to the common density  $g$  of the  $(Y_i)_{1 \leq i \leq n_0}$  and the  $(Y'_i)_{1 \leq i \leq n_0}$  are estimated thanks to the independent  $(Y'_i)$ - $n_0$ -sample. This means that we estimate a matrix whose inverse is involved in the definition of the coefficients of the estimator. Therefore the main difficulty is to measure the distance between the inverse of a random matrix and the inverse of its expectation. Our strategy is inspired by the one initiated by Neumann (1997) and developed by Kappus and Mabon (2014) in the Fourier context, with help of tools related to random matrices taken in Tropp (2015); it relies on the use of a relevant cutoff for the inversion of the estimated matrix. We obtain risk bounds generalizing the case where  $g$  is known and showing that, if both sample sizes  $n$  and  $n_0$  have the same order, it is possible that no loss in the order of the upper bound occurs. We also provide a model selection

procedure for which a risk bound states that the bias-variance compromise is adequately performed, in a non-asymptotic setting.

There remains additional questions that may be worth answering.

First, in [Mabon \(2016b\)](#), the problem of survival function estimation for known  $g$  is also studied: the question is left open here, to determine if the strategy developed in the present work could be extended to this context.

Moreover, our framework is mainly nonasymptotic, but if we are interested in asymptotics, the question of lower bounds may be studied; we may wonder if the tedious strategy proposed in [Belomestny et al. \(2016\)](#) can be extended to the present context, for the terms corresponding to known  $g$ ; the additional terms due to estimating  $\mathbf{G}_m$  would have to be studied also.

$f$		direct	Y Exponential			Y Gamma		
			Known noise	Noise sample $n_0 = 400$	Noise sample $n_0 = 2000$	Known noise	Noise sample $n_0 = 400$	Noise sample $n_0 = 2000$
Exp(1)	MISE	0.6	3.8	2.3	3.4	1.2	1.8	2.1
	(std)	(1.2)	(14.2)	(8.1)	(8.8)	(3.8)	(3.8)	(5.2)
	Oracles	0.10	0.14	0.36	0.17	0.15	0.30	0.17
	(std)	(0.1)	(0.2)	(0.6)	(0.2)	(0.2)	(0.4)	(0.2)
Gamma	MISE	0.6	0.8	1.6	0.8	3.4	4.6	2.3
	(std)	(0.3)	(0.3)	(1.6)	(0.4)	(1.4)	(2.1)	(1.7)
	Oracles	0.3	0.6	0.7	0.6	0.7	1.1	0.8
	(std)	(0.3)	(0.4)	(0.4)	(0.4)	(0.5)	(0.9)	(0.6)
Mixed Gamma	MISE	1.6	7.2	8.4	7.0	9.0	38.2	9.1
	(std)	(0.8)	(1.6)	(1.7)	(1.6)	(3.7)	(20.8)	(3.8)
	Oracles	1.0	2.9	4.8	3.5	4.8	24.5	7.6
	(std)	(0.6)	(1.9)	(2.0)	(1.9)	(2.4)	(8.0)	(2.6)
Weibull	MISE	0.9	1.2	1.2	1.3	1.1	1.5	1.1
	(std)	(0.4)	(0.9)	(0.8)	(0.6)	(5.0)	(1.3)	(0.6)
	Oracles	0.7	1.0	1.2	1.0	1.1	1.3	1.5
	(std)	(0.3)	(0.5)	(0.8)	(0.5)	(0.6)	(0.8)	(1.1)
Rayleigh	MISE	0.5	0.9	0.9	0.3	1.1	1.5	1.1
	(std)	(0.3)	(0.4)	(0.8)	(0.4)	(0.6)	(1.3)	(0.6)
	Oracles	0.3	0.5	0.6	0.5	0.6	0.8	0.6
	(std)	(0.2)	(0.3)	(0.4)	(0.3)	(0.4)	(0.5)	(0.4)
Beta	MISE	0.5	1.9	3.0	1.9	3.0	10.0	3.0
	(std)	(0.2)	(0.2)	(0.5)	(0.3)	(0.4)	(6.6)	(0.4)
	Oracles	0.3	0.5	0.5	0.5	0.5	2.1	0.6
	(std)	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	(0.4)	(0.3)

TABLE 2. Results after 200 iterations of simulations of the six considered densities, for sample sizes  $n = 2000$  and  $n_0 = 400$ ,  $n_0 = 2000$ . For each density : first two lines, MISE  $\times 1000$  with (std  $\times 1000$ ) in parenthesis; third and fourth lines, mean with std in parenthesis of oracles. First column, direct observations of the  $X_i$ 's. Columns 2, 3 and 4, noise is  $\mathcal{E}(\lambda)$  with  $\lambda = 2$  (mean  $1/2$ ). Columns 5, 6 and 7, noise is  $\gamma(2, \lambda')$  with  $\lambda' = 2\sqrt{2}$  (mean  $1/(2\sqrt{2})$ ).

## 7. PROOFS

## 7.1. Preliminary results.

## 7.1.1. Bounds on the spectral norm.

**Proposition 7.1.** For  $\widehat{\mathbf{G}}_m$  defined by Equation (7) and  $\|g\|_\infty < \infty$ ,  $n_0 \in \mathbb{N} \setminus \{0\}$ , then for all  $t > 0$

$$\mathbb{P} \left[ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}} \geq t \right] \leq 2m \exp \left( - \frac{n_0 t^2 / 4}{\|g\|_\infty m + (\sqrt{2}/3) m t} \right).$$

**Corollary 7.2.** Under the Assumptions of Proposition 7.1, for all  $q \geq 2$ , it holds that

$$\mathbb{E} \left[ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^q \right] \leq \mathfrak{C}_q (\log m)^{q/2} \frac{m^{q/2}}{n_0^{q/2}} \vee (\log m)^q \frac{m^q}{n_0^q}$$

with  $\mathfrak{C}_q = 2^{q-1} e^{q/2} \|g\|_\infty^{q/2} (q+2)^{q/2} + 2^{2q-1+q/2} (q+2)^{q/2}$ .

*Proof of Proposition 7.1.* To get the announced result, we apply a Bernstein matrix inequality (see Theorem A.2). Thus we write  $\widehat{\mathbf{G}}_m$  as a sum of a sequence of independent matrices

$$\widehat{\mathbf{G}}_m = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbf{K}_m(Y'_i)$$

with

$$\mathbf{K}_m(Y'_i) = \begin{cases} 2^{-1/2} \varphi_0(Y'_i) & \text{if } i = j, \\ 2^{-1/2} (\varphi_{i-j}(Y'_i) - \varphi_{i-j-1}(Y'_i)) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases}$$

We put

$$\mathbf{S}_m = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbf{K}_m(Y'_i) - \mathbb{E} [\mathbf{K}_m(Y'_i)].$$

- Bound on  $L(\mathbf{K}_m) = \|\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]\|_{\text{op}} / n_0$ .

First using the equivalence between the spectral and trace norms

$$\mathbf{A} \in \mathbb{R}^{m \times m}, \quad \frac{1}{\sqrt{m}} \|\mathbf{A}\|_{\text{F}} \leq \|\mathbf{A}\|_{\text{op}} \leq \|\mathbf{A}\|_{\text{F}} \quad (19)$$

we have by Equation (19) that

$$L(\mathbf{K}_m) \leq \frac{1}{n_0} \|\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]\|_{\text{F}},$$

and using Equation (4)

$$\begin{aligned} \|\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]\|_{\text{F}}^2 &= \sum_{1 \leq i, j \leq m} |[\mathbf{K}_m(Y'_1)]_{i,j} - \mathbb{E} [\mathbf{K}_m(Y'_1)]_{i,j}|^2 \\ &\leq \frac{1}{2} \sum_{1 \leq i \leq m} |\varphi_0(Y'_1) - \mathbb{E} [\varphi_0(Y'_1)]|^2 + \frac{1}{2} \sum_{1 \leq j < i \leq m} |\varphi_{i-j}(Y'_1) - \varphi_{i-j-1}(Y'_1) - \mathbb{E} [\varphi_{i-j}(Y'_1) - \varphi_{i-j-1}(Y'_1)]|^2 \\ &\leq \frac{1}{2} m |e^{-Y'_1} - \mathbb{E} [e^{-Y'_1}]|^2 + \frac{1}{2} \sum_{1 \leq j < i \leq m} (4\sqrt{2})^2 \leq \frac{m}{2} + 4^2 \frac{m(m-1)}{2} = \frac{16m^2 - 16m + m}{2} \leq 8m^2. \end{aligned}$$

So we get that  $L(\mathbf{K}_m) \leq \frac{2\sqrt{2}m}{n_0}$ .

- Bound on  $\nu(\mathbf{S}_m) = \|\sum_{i=1}^{n_0} \mathbb{E} [{}^t(\mathbf{K}_m(Y'_i) - \mathbb{E} [\mathbf{K}_m(Y'_i)])(\mathbf{K}_m(Y'_i) - \mathbb{E} [\mathbf{K}_m(Y'_i)])]\|_{\text{op}} / n_0^2$ .



By definition of the operator norm we have

$$\begin{aligned}\nu(\mathbf{S}_m) &= \frac{1}{n_0^2} \sup_{\|\vec{x}\|_{2,m}=1} {}^t \vec{x} \sum_{i=1}^{n_0} \mathbb{E} \left[ {}^t (\mathbf{K}_m(Y'_i) - \mathbb{E} [\mathbf{K}_m(Y'_i)]) (\mathbf{K}_m(Y'_i) - \mathbb{E} [\mathbf{K}_m(Y'_i)]) \vec{x} \right] \\ &= \frac{1}{n_0} \sup_{\|\vec{x}\|_{2,m}=1} {}^t \vec{x} \mathbb{E} \left[ {}^t (\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]) (\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]) \vec{x} \right] \\ &= \frac{1}{n_0} \sup_{\|\vec{x}\|_{2,m}=1} \mathbb{E} \left\| (\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]) \vec{x} \right\|_{2,m}^2\end{aligned}$$

It yields that, for  ${}^t \vec{x} = (x_0, \dots, x_{m-1})$ , and convention  $\varphi_{-1} \equiv 0$ ,

$$\begin{aligned}\mathbb{E} \left\| (\mathbf{K}_m(Y'_1) - \mathbb{E} [\mathbf{K}_m(Y'_1)]) \vec{x} \right\|_{2,m}^2 &= \frac{1}{2} \sum_{i=0}^{m-1} \mathbb{E} \left( \sum_{j=0}^i (\varphi_{i-j}(Y'_1) - \varphi_{i-j-1}(Y'_1) - \mathbb{E}[\varphi_{i-j}(Y'_1) - \varphi_{i-j-1}(Y'_1)]) x_j \right)^2 \\ &= \frac{1}{2} \sum_{i=0}^{m-1} \mathbb{V}\text{ar} \left[ \sum_{j=0}^i (\varphi_{i-j}(Y'_1) - \varphi_{i-j-1}(Y'_1)) x_j \right] \\ &\leq \frac{1}{2} \sum_{i=0}^{m-1} \mathbb{E} \left| \sum_{j=0}^i (\varphi_{i-j}(Y'_1) - \varphi_{i-j-1}(Y'_1)) x_j \right|^2 = \frac{1}{2} \sum_{i=0}^{m-1} \int \left| \sum_{j=0}^i (\varphi_{i-j}(u) - \varphi_{i-j-1}(u)) x_j \right|^2 g(u) \, du \\ &\leq \frac{\|g\|_\infty}{2} \sum_{i=0}^{m-1} \int \left| \sum_{j=0}^i (\varphi_{i-j}(u) - \varphi_{i-j-1}(u)) x_j \right|^2 \, du \\ &= \frac{\|g\|_\infty}{2} \sum_{i=0}^{m-1} \left( 2 \sum_{1 \leq j, j' \leq i} \delta_{j, j'} x_j x_{j'} - \sum_{1 \leq j, j' \leq i} \delta_{j, j'+1} x_j x_{j'+1} - \sum_{1 \leq j, j' \leq i} \delta_{j, j'-1} x_j x_{j'-1} \right) \\ &\leq 2\|g\|_\infty m \|\vec{x}\|_{2,m}^2.\end{aligned}$$

Then we get that  $\nu(\mathbf{S}_m) \leq \frac{2\|g\|_\infty m}{n_0}$ .

In the end applying Theorem A.2 yields that for all  $t > 0$

$$\mathbb{P} \left[ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}} \geq t \right] \leq 2m \exp \left( -\frac{t^2/2}{2\|g\|_\infty m/n_0 + (2\sqrt{2}/3)mt/n_0} \right).$$

from which we get the result of Proposition 7.1.  $\square$

*Proof of Corollary 7.2.* Before proving the announced result, let us explain how Theorem A.3 for Hermitian matrices can be extended to non-Hermitian matrices. This is due to the so-called *Paulsen dilation* which corresponds to the following isomorphism trick for a rectangular matrix  $\mathbf{A}$

$$\mathbf{A} \mapsto \mathcal{H}(\mathbf{A}) = \begin{pmatrix} 0 & \mathbf{A} \\ \mathbf{A}^\dagger & 0 \end{pmatrix}$$

where  $\mathbf{A}^\dagger$  denotes the conjugate transpose of  $\mathbf{A}$ . Obviously  $\mathcal{H}(\mathbf{A})$  is an Hermitian matrix. We can also notice that

$$\mathcal{H}(\mathbf{A})^2 = \begin{pmatrix} \mathbf{A}\mathbf{A}^\dagger & 0 \\ 0 & \mathbf{A}^\dagger\mathbf{A} \end{pmatrix}$$

So we get  $\lambda_{\max}(\mathcal{H}(\mathbf{A})^2) = \|\mathbf{A}\|_{\text{op}}^2$  and  $\lambda_{\max}(\mathcal{H}(\mathbf{A})) = \|\mathbf{A}\|_{\text{op}}$ .

Under the Assumptions of Proposition 7.1, we can apply Theorem A.1 in Chen et al. (2012) (see Theorem A.3) stated for Hermitian matrices, using the above Paulsen dilation as follows. Let  $\mathbf{Y}_i$  be rectangular matrices and set  $\mathbf{A} = \sum_i \mathbf{Y}_i$  it yields that, for  $q \geq 2$  and  $r \geq \max(q, 2 \log m)$ ,

$$\mathcal{H}(\mathbf{A}) = \begin{pmatrix} 0 & \sum_i \mathbf{Y}_i \\ \sum_i \mathbf{Y}_i^\dagger & 0 \end{pmatrix} = \sum_i \begin{pmatrix} 0 & \mathbf{Y}_i \\ \mathbf{Y}_i^\dagger & 0 \end{pmatrix} = \sum_i \mathcal{H}(\mathbf{Y}_i).$$

Thus we get that

$$\begin{aligned} [\mathbb{E} \|\mathbf{A}\|_{\text{op}}^q]^{1/q} &= \left[ \mathbb{E} \lambda_{\max} \left( \mathcal{H} \left( \sum_i \mathbf{Y}_i \right) \right)^{q/2} \right]^{1/q} \leq \sqrt{er} \lambda_{\max}^{1/2} \left( \sum_i \mathbb{E} \mathcal{H}(\mathbf{Y}_i)^2 \right) + 2er \left[ \mathbb{E} \max_i \lambda_{\max} (\mathcal{H}(\mathbf{Y}_i))^q \right]^{1/q} \\ &\leq \sqrt{er \max(\lambda_{\max}(\mathbb{E} \mathbf{A} \mathbf{A}^\dagger), \lambda_{\max}(\mathbb{E} \mathbf{A}^\dagger \mathbf{A}))} + 2er \left[ \mathbb{E} \max_i \|\mathbf{Y}_i\|_{\text{op}}^q \right]^{1/q}. \end{aligned}$$

Now we apply this result for

$$\mathbf{A} = \mathbf{G}_m - \widehat{\mathbf{G}}_m = \mathbf{S}_m = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbf{K}_m(Y_i') - \mathbb{E} [\mathbf{K}_m(Y_i')].$$

Using the notations of the proof of Proposition 7.1, we get for  $q \geq 2$ ,  $m \geq 2$  and  $r = 2 \log m$

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^q \right] &\leq 2^{q-1} (er \nu(\mathbf{S}_m))^{q/2} + 2^{q-1} (er L(\mathbf{K}_m))^q \\ &\leq 2^{q-1} \left( er \frac{\|g\|_{\infty} m}{n_0} \right)^{q/2} + 2^{q-1} \left( er \frac{2\sqrt{2}m}{n_0} \right)^q \\ &\leq 2^{q-1} e^{q/2} \|g\|_{\infty}^{q/2} \left( 2 \log m \frac{m}{n_0} \right)^{q/2} + 2^{2q-1+q/2} \left( 2 \log m \frac{m}{n_0} \right)^q \\ &\leq \mathfrak{C}_q \left( \log m \frac{m}{n_0} \right)^{q/2} \vee \left( \log m \frac{m}{n_0} \right)^q, \end{aligned}$$

with  $\mathfrak{C}_q = 2^{q-1} e^{q/2} \|g\|_{\infty}^{q/2} (q+2)^{q/2} + 2^{2q-1+q/2} (q+2)^{q/2}$ .  $\square$

## 7.2. Proofs of results of Section 3.

7.2.1. *Proof of Lemma 3.1.* First let us define the set

$$\Delta_m = \left\{ \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq \sqrt{\frac{n_0}{m \log m}} \right\} \quad (20)$$

and notice that

$$\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1} = \mathbf{1}_{\Delta_m^c} \mathbf{G}_m^{-1} + \mathbf{1}_{\Delta_m} (\mathbf{G}_m^{-1} - \widehat{\mathbf{G}}_m^{-1}) = \mathbf{1}_{\Delta_m^c} \mathbf{G}_m^{-1} - \mathbf{1}_{\Delta_m} \widehat{\mathbf{G}}_m^{-1} (\mathbf{G}_m - \widehat{\mathbf{G}}_m) \mathbf{G}_m^{-1}.$$

Then we can write that

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] &= \mathbb{E} \left[ \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m^c} + \|\widehat{\mathbf{G}}_m^{-1} (\mathbf{G}_m - \widehat{\mathbf{G}}_m) \mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &= \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[ \|\widehat{\mathbf{G}}_m^{-1} (\mathbf{G}_m - \widehat{\mathbf{G}}_m) \mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right]. \end{aligned} \quad (21)$$

This proof is inspired of the proof of Lemma 2.1 in Neumann (1997), in the sense that we divide the proof in two cases according to the comparison of  $\|\mathbf{G}_m^{-1}\|_{\text{op}}$  with the threshold.

- First case:  $\|\mathbf{G}_m^{-1}\|_{\text{op}} > \frac{1}{2} \sqrt{\frac{n_0}{m \log m}}$ . Let us prove that  $\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] \lesssim \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}$ .

Starting from Equation (21) and using the set  $\Delta_m$ , we have that

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] &\leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} + \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[ \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &\leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} + \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \left( \frac{n_0}{m \log m} \right)^p \mathbb{E} \left[ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p} \right]. \end{aligned}$$

Besides applying Corollary 7.2 for  $q = 2p$  yields

$$\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] \leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} + \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \left( \frac{n_0}{m \log m} \right)^p \mathfrak{C}_{2p} \left( \frac{m \log m}{n_0} \right)^p \leq (1 + \mathfrak{C}_{2p}) \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}.$$

• Second case:  $\|\mathbf{G}_m^{-1}\|_{\text{op}} < \frac{1}{2} \sqrt{\frac{n_0}{m \log m}}$ . Let us prove that  $\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] \lesssim \left( \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^4 \frac{m}{n_0} \right)^p$ .

Starting from (21) again, we get

$$\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] \leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{P}[\Delta_m^c] + \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right]. \quad (22)$$

i) Upper bound on  $\mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right]$ .

First let us notice that

$$\|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \leq 2^{2p-1} \|\hat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}}^{2p} + 2^{2p-1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}.$$

Moreover applying Corollary 7.2 for  $q = 2p$  and  $q = 4p$  with the set  $\Delta_m$ , we get

$$\begin{aligned} & \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ & \leq 2^{2p-1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] + 2^{2p-1} \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ & \leq 2^{2p-1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] + 2^{2p-1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^{4p} \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ & \leq 2^{2p-1} \mathfrak{C}_{2p} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \left( \frac{m \log m}{n_0} \right)^p + 2^{2p-1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \left( \frac{n_0}{m \log m} \right)^p \mathfrak{C}_{4p} \left( \frac{m \log m}{n_0} \right)^{2p} \\ & \leq 2^{2p-1} (\mathfrak{C}_{2p} + \mathfrak{C}_{4p}) \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p} \left( \frac{m \log m}{n_0} \right)^p. \end{aligned} \quad (23)$$

ii) Upper bound on  $\mathbb{P}[\Delta_m^c] = \mathbb{P} \left[ \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{n_0}{m \log m}} \right]$ .

The upper bound is given by the following Lemma proved afterwards.

**Lemma 7.3.** *For  $\Delta_m$  defined by Equation (20) and  $\|\mathbf{G}_m^{-1}\|_{\text{op}} < \frac{1}{2} \sqrt{\frac{n_0}{m \log m}}$ , it holds that*

$$\mathbb{P}[\Delta_m^c] = \mathbb{P} \left[ \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{n_0}{m \log m}} \right] \leq 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{m \log m}{n_0} \right)^p \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}. \quad (24)$$

Finally starting from Equation (22) and gathering Equations (23) with (24), we get that

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] & \leq 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{m \log m}{n_0} \right)^p \|\mathbf{G}_m^{-1}\|_{\text{op}}^{4p} + 2^{2p-1} (\mathfrak{C}_{2p} + \mathfrak{C}_{4p}) \|\mathbf{G}_m^{-1}\|_{\text{op}}^{4p} \left( \frac{m \log m}{n_0} \right)^p \\ & \leq (2^{2p+1} \mathfrak{C}_{2p} + 2^{2p} \mathfrak{C}_{4p}) \left( \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^4 \frac{m}{n_0} \right)^p. \end{aligned}$$

In conclusion, we just proved the following upper bound

$$\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p} \right] \leq \mathfrak{C}_{\text{op},p} \left( \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^4 \frac{m}{n_0} \right)^p$$

with  $\mathfrak{C}_{\text{op},p} = 2^{2p+1} \mathfrak{C}_{2p} + 2^{2p} \mathfrak{C}_{4p} + 1$ . □

*Proof of Lemma 7.3.* First invoke the triangular inequality

$$\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} + \|\mathbf{G}_m^{-1}\|_{\text{op}}$$

which implies that

$$\mathbb{P} \left[ \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{n_0}{m \log m}} \right] \leq \mathbb{P} \left[ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \sqrt{\frac{n_0}{m \log m}} - \|\mathbf{G}_m^{-1}\|_{\text{op}} \right].$$

Moreover we assume that  $\|\mathbf{G}_m^{-1}\|_{\text{op}} < \frac{1}{2} \sqrt{\frac{n_0}{m \log m}}$ , so

$$\mathbb{P} \left[ \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{n_0}{m \log m}} \right] \leq \mathbb{P} \left[ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right].$$

Now let us rewrite this probability, as

$$\begin{aligned} & \mathbb{P} \left[ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right] \\ &= \mathbb{P} \left[ \left\{ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right\} \cap \left\{ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ & \quad + \mathbb{P} \left[ \left\{ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right\} \cap \left\{ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} \geq \frac{1}{2} \right\} \right] \\ &\leq \mathbb{P} \left[ \left\{ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right\} \cap \left\{ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ & \quad + \mathbb{P} \left[ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} \geq \frac{1}{2} \right]. \end{aligned} \tag{25}$$

To control the second term of the right hand side of Equation (25), we apply Markov inequality and Corollary 7.2 for  $q = 2p$

$$\mathbb{P} \left[ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} \geq \frac{1}{2} \right] \leq \mathbb{P} \left[ \|\mathbf{G}_m^{-1}\|_{\text{op}} \|\widehat{\mathbf{G}}_m - \mathbf{G}_m\|_{\text{op}} \geq \frac{1}{2} \right] \leq 2^{2p} \mathfrak{C}_{2p} \left( \frac{m \log m}{n_0} \right)^p \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}. \tag{26}$$

Next to control the first term on the right hand side of Equation (25), we apply Theorem A.1 (with  $\mathbf{A} = \mathbf{G}_m$  and  $\mathbf{B} = \widehat{\mathbf{G}}_m - \mathbf{G}_m$ ), it yields

$$\begin{aligned} & \mathbb{P} \left[ \left\{ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right\} \cap \left\{ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ &\leq \mathbb{P} \left[ \left\{ \frac{\|\widehat{\mathbf{G}}_m - \mathbf{G}_m\|_{\text{op}} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{1 - \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right\} \cap \left\{ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ &\leq \mathbb{P} \left[ \|\widehat{\mathbf{G}}_m - \mathbf{G}_m\|_{\text{op}} > \frac{1}{2} \|\mathbf{G}_m^{-1}\|_{\text{op}}^{-1} \right], \end{aligned} \tag{27}$$

again applying Markov inequality along with Corollary 7.2 gets

$$\mathbb{P} \left[ \left\{ \|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}} > \|\mathbf{G}_m^{-1}\|_{\text{op}} \right\} \cap \left\{ \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \leq 2^{2p} \mathfrak{C}_{2p} \left( \frac{m \log m}{n_0} \right)^p \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}$$

So starting from Equation (25) and gathering Equations (26) and (27) gives

$$\mathbb{P} \left[ \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{n_0}{m \log m}} \right] \leq 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{m \log m}{n_0} \right)^p \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}$$

□

7.2.2. *Useful corollary for the Frobenius norm.*

**Corollary 7.4.** *Under the Assumptions of Lemma 3.1, we have*

$$\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\mathbb{F}}^2 \right] \leq 2\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$$

*Proof of Corollary 7.4.* The proof mainly follows the lines of the proof of Lemma 3.1.  $\Delta_m$  is defined by Equation (20), then we write

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\mathbb{F}}^2 \right] &= \mathbb{E} \left[ \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \mathbf{1}_{\Delta_m^c} + \|\hat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \hat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \mathbf{1}_{\Delta_m} \right] \\ &= \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[ \|\hat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \hat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \mathbf{1}_{\Delta_m} \right]. \end{aligned} \quad (28)$$

The definition of  $\Delta_m$  along with the equivalence between norms (see Eq. (19)) implies that on these set

$$\|\hat{\mathbf{G}}_m^{-1}\|_{\mathbb{F}}^2 \leq \frac{n_0}{\log m}. \quad (29)$$

Let us recall that for two matrices  $\mathbf{A}$  and  $\mathbf{B}$

$$\|\mathbf{AB}\|_{\mathbb{F}} \leq \|\mathbf{A}\|_{\mathbb{F}} \|\mathbf{B}\|_{\text{op}} \quad \text{and} \quad \|\mathbf{AB}\|_{\mathbb{F}} \leq \|\mathbf{A}\|_{\text{op}} \|\mathbf{B}\|_{\mathbb{F}}. \quad (30)$$

Starting from Equation (28) plus Equation (29) and Lemma 3.1 for  $q = 2$  along with Equation (30) gives

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\mathbb{F}}^2 \right] &\leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 + \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \mathbb{E} \left[ \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^2 \mathbf{1}_{\Delta_m} \right] \\ &\leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 + \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \frac{n_0}{m \log m} \mathbb{E} \left[ \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^2 \right] \\ &\leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 + \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \frac{m \log m}{n_0} \frac{n_0}{m \log m} = 2\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \end{aligned}$$

□

7.2.3. *Proof of Corollary 3.3.* The proof follows the lines of the proof of Lemma 3.1. The only difference lies in the following equation

$$\begin{aligned} \mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right] &= \|\mathbf{G}_m^{-1}\vec{h}_m\|_{2,m}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[ \|\hat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \hat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\vec{h}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \right] \\ &= \|\vec{f}_m\|_{2,m}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[ \|\hat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \hat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\vec{h}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \right], \end{aligned}$$

with  $\Delta_m$  defined by Equation (20). It yields the following upper bound

$$\mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right] \leq \|\vec{f}_m\|_{2,m}^2 \mathbb{P}[\Delta_m^c] + \|\vec{f}_m\|_{2,m}^2 \mathbb{E} \left[ \|\hat{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \|\mathbf{G}_m - \hat{\mathbf{G}}_m\|_{\text{op}}^2 \mathbf{1}_{\Delta_m} \right].$$

And following the proof of Lemma 3.1, we get

$$\mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right] \leq \|f\|^2 \mathbf{c}_{\text{op}} \left( 1 \wedge \log m \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \right).$$

□

## 7.2.4. Proof of Corollary 3.5.

$$\begin{aligned}
\mathbb{E} \left[ \left\| (\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}) \vec{h}_m \right\|_{2,m}^2 \right] &= \left\| \mathbf{G}_m^{-1} \vec{h}_m \right\|_{2,m}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[ \left\| (\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}) \vec{h}_m \right\|_{2,m}^2 \mathbf{1}_{\Delta_m} \right] \\
&\leq \|f\|^2 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{2m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n_0} \right)^p \\
&\quad + 2\mathbb{E} \left[ \left\| \mathbf{G}_m^{-1} (\mathbf{G}_m - \hat{\mathbf{G}}_m) \vec{f}_m \right\|_{2,m}^2 \mathbf{1}_{\Delta_m} \right] \\
&\quad + 2\mathbb{E} \left[ \left\| (\hat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) (\mathbf{G}_m - \hat{\mathbf{G}}_m) \vec{f}_m \right\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m} \right] \\
&\quad + 2\mathbb{E} \left[ \left\| (\hat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) (\mathbf{G}_m - \hat{\mathbf{G}}_m) \vec{f}_m \right\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m^c} \right]
\end{aligned} \tag{31}$$

where the last inequality follows from Lemma 7.3, and we define

$$E_m = \left\{ \left\| \mathbf{G}_m - \hat{\mathbf{G}}_m \right\|_{\text{op}}^2 < \mathfrak{a}^2 \frac{m \log(m)}{n_0} \right\},$$

along with

$$T_2 := \mathbb{E} \left[ \left\| \mathbf{G}_m^{-1} (\mathbf{G}_m - \hat{\mathbf{G}}_m) \vec{f}_m \right\|_{2,m}^2 \mathbf{1}_{\Delta_m} \right]$$

Bound on  $T_2$ :

First let us notice that, with convention  $\varphi_{-1} \equiv 0$ ,

$$\begin{aligned}
T_2 &\leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \mathbb{E} \left[ \left\| (\mathbf{G}_m - \hat{\mathbf{G}}_m) \vec{f}_m \right\|_{2,m}^2 \right] \leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \mathbb{E} \left[ \sum_{k=1}^m \left( \sum_{j=1}^m [\mathbf{G}_m - \hat{\mathbf{G}}_m]_{k,j} a_{j-1}(f) \right)^2 \right] \\
&= \frac{1}{2n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \sum_{k=0}^{m-1} \text{Var} \left( \sum_{j=0}^k (\varphi_{k-j}(Y'_1) - \varphi_{k-j-1}(Y'_1)) a_j(f) \right).
\end{aligned}$$

It yields that

$$\begin{aligned}
T_2 &\leq \frac{1}{2n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \sum_{k=0}^{m-1} \mathbb{E} \left( \sum_{j=0}^k (\varphi_{k-j}(Y'_1) - \varphi_{k-j-1}(Y'_1)) a_j(f) \right)^2 \\
&\leq \frac{1}{2n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \sum_{k=0}^{m-1} \int_{\mathbb{R}^+} \left( \sum_{j=0}^k (\varphi_{k-j}(y) - \varphi_{k-j-1}(y)) a_j(f) \right)^2 g(y) dy \\
&\leq \frac{\|g\|_{\infty}}{2n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \sum_{k=0}^{m-1} \int_{\mathbb{R}^+} \left( \sum_{j=1}^k (\varphi_{k-j}(y) - \varphi_{k-j-1}(y)) a_{j-1}(f) \right)^2 dy \\
&\leq 2 \frac{\|g\|_{\infty}}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \sum_{k=0}^{m-1} \sum_{j=0}^k a_j^2(f) \leq 2 \|f\|^2 \|g\|_{\infty} \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2.
\end{aligned}$$

Secondly we propose an alternative bound on  $T_2$ .

$$\begin{aligned}
T_2 &= \mathbb{E} \|\mathbf{G}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \vec{f}_m\|_{2,m}^2 = \mathbb{E} \sum_{k=1}^m \left[ \mathbf{G}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \vec{f}_m \right]_k^2 \\
&= \mathbb{E} \sum_{k=1}^m \left[ \sum_{1 \leq j, l \leq m} [\mathbf{G}_m^{-1}]_{k,l} [\mathbf{G}_m - \widehat{\mathbf{G}}_m]_{l,j} a_{j-1}(f) \right]^2 \\
&= \sum_{k=1}^m \frac{1}{2n_0} \mathbb{V} \text{ar} \left[ \sum_{1 \leq j < l \leq m} [\mathbf{G}_m^{-1}]_{k,l} (\varphi_{l-j}(Y_1') - \varphi_{l-j-1}(Y_1')) a_j(f) \right] \\
&\leq \frac{1}{2n_0} \sum_{k=1}^m \mathbb{E} \left| \sum_{1 \leq j < l \leq m} [\mathbf{G}_m^{-1}]_{k,l} (\varphi_{l-j}(Y_1') - \varphi_{l-j-1}(Y_1')) a_j(f) \right|^2
\end{aligned}$$

Following the same reasoning as previously, we get

$$T_2 \leq 2 \frac{\|g\|_\infty}{n_0} \sum_{k=1}^m \sum_{\substack{1 \leq j < l \leq m \\ 1 \leq j' < l' \leq m}} a_{j-1}(f) a_{j'-1}(f) [\mathbf{G}_m^{-1}]_{k,l} [\mathbf{G}_m^{-1}]_{k,l'} \delta_{l-j, l'-j'}.$$

Moreover, we can notice that

$$\sum_{1 \leq j < l \leq m} a_{j-1}(f) [\mathbf{G}_m^{-1}]_{k,l} \delta_{l-j, l'-j'} = \sum_{j=1}^m \sum_{l=j+1}^m a_{j-1}(f) [\mathbf{G}_m^{-1}]_{k,l} \delta_{l-j, l'-j'} = \sum_{j=1}^m \sum_{l'=1}^{m-j} a_{l+j}(f) [\mathbf{G}_m^{-1}]_{k, l'+j} \delta_{l, l'-j'},$$

it yields

$$\begin{aligned}
T_2 &\leq 2 \frac{\|g\|_\infty}{n_0} \sum_{k=1}^m \sum_{j, l, j'} |a_{j-1}(f)| |a_{j'-1}(f)| |[\mathbf{G}_m^{-1}]_{k,l}| |[\mathbf{G}_m^{-1}]_{k, l-j-j'}| \\
&\leq 2 \frac{\|g\|_\infty}{n_0} \|\mathbf{G}_m^{-1}\|_F^2 \left( \sum_{j=1}^m |a_{j-1}(f)| \right)^2 \leq 2 \|f\|_{\ell_1}^2 \|g\|_\infty \frac{\|\mathbf{G}_m^{-1}\|_F^2}{n_0}.
\end{aligned}$$

So we get

$$T_2 \leq \frac{2}{n_0} \|g\|_\infty (\|f\|^2 m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|f\|_{\ell_1}^2 \|\mathbf{G}_m^{-1}\|_F^2). \quad (32)$$

Next using the definition of  $\Delta_m$  and  $E_m$

$$\mathbb{E}[\|(\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1})(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \vec{f}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m}] = \mathbb{E}[\|\widehat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \mathbf{G}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \vec{f}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m}]$$

$$\leq \mathbb{E}[\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \mathbf{1}_{\Delta_m} \|(\mathbf{G}_m - \widehat{\mathbf{G}}_m)\|_{\text{op}}^2 \mathbf{1}_{E_m} \|\mathbf{G}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \vec{f}_m\|_{2,m}^2] \leq \mathbf{a}^2 T_2. \quad (33)$$

For the last term, applying Lemma 3.1 and Corollary 7.2

$$\begin{aligned}
&\mathbb{E}[\|(\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1})(\mathbf{G}_m - \widehat{\mathbf{G}}_m) \vec{f}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m^c}] \\
&\leq \|f\|^2 \mathbb{E}[\|\widetilde{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}}^2 \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m^c}] \\
&\leq \|f\|^2 \mathbb{E}^{1/4}[\|\widetilde{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}}^8] \mathbb{E}^{1/4}[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^8] \sqrt{\mathbb{P}[E_m^c]} \\
&\leq \|f\|^2 \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \mathfrak{c}_4 \frac{m \log(m)}{n_0} \sqrt{\mathbb{P}(E_m^c)}
\end{aligned}$$

Now, we apply Proposition 7.1 with  $t = \mathbf{a} \sqrt{\frac{m \log(m)}{n_0}}$ , and taking into account that, given our definition of  $\mathcal{M}$ ,  $m \log(m)/n_0 \leq 1/4$ ,

$$\|g\|_\infty + (2\sqrt{2}/3) \mathbf{a} \sqrt{\frac{m \log(m)}{n_0}} \leq (1 + \mathbf{a})(1 \vee \|g\|_\infty) \leq 2\mathbf{a}(1 \vee \|g\|_\infty)$$



if  $\mathbf{a} \geq 1$ . We get

$$\mathbb{P}(E_m^c) \leq 2m \exp\left(-\frac{n_0}{m} \frac{\mathbf{a}}{2(1 \vee \|g\|_\infty)} \frac{m \log(m)}{n_0}\right) = 2m^{1-\mathbf{a}/(2(1 \vee \|g\|_\infty))}.$$

Choose  $\mathbf{a} = \mathbf{a}(q)$  such that

$$1 - \frac{\mathbf{a}(q)}{2(1 \vee \|g\|_\infty)} = -2q$$

that is

$$\mathbf{a}(q) = 2(2q + 1)(1 \vee \|g\|_\infty)$$

yields  $\mathbb{P}(E_m^c) \leq 2m^{-2q}$ . Thus, we obtain

$$\mathbb{E}[\|(\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1})(\mathbf{G}_m - \widehat{\mathbf{G}}_m)\vec{f}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m^c}] \leq 2\mathfrak{C}_4 \|f\|^2 \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \frac{m \log(m)}{n_0} m^{-2q}$$

and we use that  $\log(m)/m \leq 1$  for  $m \geq 2$  (take  $q = 1/2$ ) and  $m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 / m^2 \leq \|\mathbf{G}_m^{-1}\|_{\text{F}}^2$  (take  $q = 1$ ). Therefore, for  $\mathbf{a} = \mathbf{a}(1) = 6(1 \vee \|g\|_\infty)$ , we obtain

$$\mathbb{E}[\|(\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1})(\mathbf{G}_m - \widehat{\mathbf{G}}_m)\vec{f}_m\|_{2,m}^2 \mathbf{1}_{\Delta_m} \mathbf{1}_{E_m^c}] \leq \frac{2\mathfrak{C}_4}{n_0} \|f\|^2 (m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|\mathbf{G}_m^{-1}\|_{\text{F}}^2).$$

Plugging (32), (33) and (34) into (31) gives

$$\begin{aligned} \mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right] &\leq \frac{\mathfrak{C}'_{\text{E}}}{n_0} (1 \vee \|g\|_\infty)^3 (\|f\|^2 \vee \|f\|_{\ell_1}^2) (m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|\mathbf{G}_m^{-1}\|_{\text{F}}^2) \\ &\quad + 2^{2p+1} \mathfrak{C}_{2p} \left( \frac{2m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n_0} \right)^p. \end{aligned}$$

with  $\mathfrak{C}'_{\text{E}} = 2(196 + 2\mathfrak{C}_4)$ . □

### 7.2.5. Proofs of Propositions 3.4 and 3.6.

*Proof of Proposition 3.4.* By Pythagoras theorem, we have

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

Let us rewrite the second term of the above equality:

$$\begin{aligned} \|f_m - \tilde{f}_m\|^2 &= \|\vec{f}_m - \tilde{\vec{f}}_m\|_{2,m}^2 = \|\mathbf{G}_m^{-1} \vec{h}_m - \widetilde{\mathbf{G}}_m^{-1} \hat{\vec{h}}_m\|_{2,m}^2 \\ &= \|\mathbf{G}_m^{-1} \vec{h}_m - \mathbf{G}_m^{-1} \hat{\vec{h}}_m + \mathbf{G}_m^{-1} \hat{\vec{h}}_m - \widetilde{\mathbf{G}}_m^{-1} \hat{\vec{h}}_m\|_{2,m}^2 \\ &\leq 2\|\mathbf{G}_m^{-1} \vec{h}_m - \mathbf{G}_m^{-1} \hat{\vec{h}}_m\|_{2,m}^2 + 2\|\mathbf{G}_m^{-1} \hat{\vec{h}}_m - \widetilde{\mathbf{G}}_m^{-1} \hat{\vec{h}}_m\|_{2,m}^2. \end{aligned} \tag{35}$$

i) First let us bound the first term on the right hand side of Equation (35).

a) First we apply (4) and get

$$\begin{aligned} \mathbb{E} \|\mathbf{G}_m^{-1}(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 &\leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \mathbb{E} \|\hat{\vec{h}}_m - \vec{h}_m\|_{2,m}^2 \leq \mathbb{E} \left[ \sum_{j=1}^m \left( \frac{1}{n} \sum_{i=1}^n \varphi_j(Z_i) - \mathbb{E}[\varphi_j(Z_1)] \right)^2 \right] \\ &\leq \frac{\|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n} \sum_{j=1}^m \mathbb{E}[\varphi_j^2(Z_1)] \leq \frac{2m}{n} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2. \end{aligned}$$

**b)** Secondly we also have

$$\begin{aligned} \mathbb{E}\|\mathbf{G}_m^{-1}(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 &= \mathbb{E}\sum_{k=1}^m \left( \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} (a_{j-1}(h) - \hat{a}_{j-1}(Z)) \right)^2 \\ &= \sum_{k=1}^m \text{Var} \left[ \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \hat{a}_{j-1}(Z) \right], \end{aligned}$$

and since  $\hat{a}_{j-1}(h) - a_{j-1}(Z) = (1/n) \sum_{i=1}^n (\varphi_{j-1}(Z_i) - \mathbb{E}[\varphi_{j-1}(Z_i)])$ , it yields that

$$\begin{aligned} \mathbb{E}\|\mathbf{G}_m^{-1}(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 &= \frac{1}{n} \sum_{k=1}^m \sum_{j=1}^m \text{Var} \left[ [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(Z_1) \right] \\ &\leq \frac{1}{n} \sum_{k=1}^m \mathbb{E} \left[ \left( \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(Z_1) \right)^2 \right] \leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \int_{\mathbb{R}^+} \left( \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(u) \right)^2 du \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \sum_{1 \leq j, j' \leq m} [\mathbf{G}_m^{-1}]_{k,j} [\mathbf{G}_m^{-1}]_{k,j'} \int_{\mathbb{R}^+} \varphi_{j-1}(u) \varphi_{j'-1}(u) du \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j}^2 = \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_F^2. \end{aligned}$$

In conclusion, the first term on the right-hand-side of Equation (35) is upper bounded as follows

$$\mathbb{E}\|\mathbf{G}_m^{-1}(\vec{h}_m - \hat{\vec{h}}_m)\|_{2,m}^2 \leq \frac{2m}{n} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_F^2. \quad (36)$$

**ii)** Now we turn to the second term on the right-hand-side of Equation (35). Let us notice that

$$\begin{aligned} \|\mathbf{G}_m^{-1} \hat{\vec{h}}_m - \tilde{\mathbf{G}}_m^{-1} \hat{\vec{h}}_m\|_{2,m}^2 &= \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m) + (\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \\ &\leq 2\|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 + 2\|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2. \end{aligned} \quad (37)$$

**a)** The first term of (37) can be bounded in two ways: since  $(Y'_1, \dots, Y'_{n_0})$  is independent of  $(Z_1, \dots, Z_n)$ , we get that

$$\mathbb{E}\|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \leq \mathbb{E}\|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \mathbb{E}\|\hat{\vec{h}}_m - \vec{h}_m\|_{2,m}^2 \quad (38)$$

and by using (4) we get

$$\mathbb{E}\|\hat{\vec{h}}_m - \vec{h}_m\|_{2,m}^2 = \mathbb{E} \left[ \sum_{j=1}^m \left( \frac{1}{n} \sum_{i=1}^n \varphi_j(Z_i) - \mathbb{E}[\varphi_j(Z_1)] \right)^2 \right] \leq \frac{1}{n} \sum_{j=1}^m \mathbb{E}[\varphi_j^2(Z_1)] \leq \frac{2m}{n}. \quad (39)$$

Applying Lemma 3.1 gives that

$$\mathbb{E}\|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \leq \frac{2m}{n} \mathfrak{C}_{\text{op},1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2$$

**b)** Secondly, repeating the same scheme as in **i)b)** under the assumption that  $(Y'_1, \dots, Y'_{n_0})$  is independent of  $(Z_1, \dots, Z_n)$ , we obtain

$$\mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \right] \leq \mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_F^2 \right] \frac{\|h\|_\infty}{n}.$$

And applying Corollary 7.4

$$\mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})(\hat{h}_m - \vec{h}_m)\|_{2,m}^2 \right] \leq \mathfrak{C}_F \|\mathbf{G}_m^{-1}\|_F^2 \frac{\|h\|_\infty}{n}. \quad (40)$$

For the second term of (37), we have according to Corollary 3.3

$$\mathbb{E} \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \leq \mathfrak{C}_E \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \frac{m}{n_0}. \quad (41)$$

Finally starting from Equation (35) and gathering Equations (36), (38), (39), (40) and (41) yields

$$\mathbb{E} \|f_m - \tilde{f}_m\|^2 \leq (2 + \mathfrak{C}_{\text{op},1} + \mathfrak{C}_F) \left( \frac{2m}{n} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_F^2 \right) + 4\mathfrak{C}_E \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \frac{m}{n_0}.$$

To conclude

$$\mathbb{E} \|f_m - \tilde{f}_m\|^2 \leq \|f - f_m\|^2 + \mathfrak{C} \left( \frac{2m}{n} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_F^2 \right) + 4\mathfrak{C}_E \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \frac{m}{n_0}. \quad \square$$

*Proof of Proposition 3.6.* The proof follows the lines of the proof of Proposition 3.4. The difference lies in the bounding of  $\mathbb{E} \left[ \|(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \right]$ . For  $\|f\|_{\ell_1} < \infty$ , we can apply Corollary 3.5 which yields the announced bound on  $\mathbb{E} \|f - \tilde{f}_m\|^2$ .  $\square$

*Proofs of Proposition 3.8 and Corollary 3.9.* For  $f \in W^s(\mathbb{R}^+, L)$  defined by (11), we have

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} a_k^2(f) = \sum_{k=m}^{\infty} a_k^2(f) k^s k^{-s} \leq Lm^{-s},$$

and according to Lemma 3.7

$$\|\mathbf{G}_m^{-1}\|_F^2 \asymp \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \asymp m^{2r}.$$

It yields that the MISE is upper bounded as follows

$$\mathbb{E} \|f - \tilde{f}_m\|^2 \leq Lm^{-s} + 2C \left( \frac{2m}{n} m^{2r} \wedge \frac{\|h\|_\infty}{n} m^{2r} \right) + 2C \mathfrak{C} \frac{m^{2r+1}}{n_0}$$

Now we have to counterbalance the bias and the variance terms as follows

$$\begin{aligned} Lm^{-s} + 2C(2 + \|h\|_\infty) \frac{m^{2r}}{n} &\Rightarrow m_{\text{opt}_1} \propto n^{1/s+2r} \\ Lm^{-s} + 2C\mathfrak{C} \log(m) \frac{m^{2r+1}}{n_0} &\Rightarrow m_{\text{opt}_2} \propto (n_0 / \log(n_0))^{1/s+2r+1} \end{aligned}$$

For  $m_{\text{opt}} \propto n^{1/s+2r} \wedge (n_0 / \log(n_0))^{1/s+2r+1}$  we get

$$\mathbb{E} \|f - \tilde{f}_{m_{\text{opt}}}\|^2 \lesssim n^{-s/s+2r} \vee \left( \frac{n_0}{\log n_0} \right)^{-s/s+2r+1}.$$

which ends the proof of Proposition 3.8.

For Corollary 3.9, we start from Proposition 3.6. Assume that  $\sum_j j^s |a_j(f)|^2 < +\infty$ , then  $\sum_j |a_j(f)| \leq \sqrt{\sum_j j^s |a_j(f)|^2 \sum_j j^{-s}}$  is finite is  $s > 1$ . Assume also that  $\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \asymp m^{-2r}$  and  $\|\mathbf{G}_m^{-1}\|_F^2 \asymp m^{-2r}$ , then for  $m_{\text{opt},1} = n^{1/(s+2r)}$ , we get

$$\inf_m \left( \|f - f_m\|^2 + \frac{C}{n} (m \|\mathbf{G}_m^{-1}\|_{\text{op}} \wedge \|\mathbf{G}_m^{-1}\|_F^2) \right) \asymp n^{-s/(s+2r)},$$

and for  $m_{\text{opt},2} = n_0^{1/(s+2r)}$ , we

$$\inf_m \left( \|f - f_m\|^2 + \frac{C}{n_0} (m \|\mathbf{G}_m^{-1}\|_{\text{op}} \wedge \|\mathbf{G}_m^{-1}\|_F^2) \right) \asymp n_0^{-s/(s+2r)}.$$

The last term in the bound of the Corollary becomes

$$\left( \frac{2m \log(m) \|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n_0} \right)^p \asymp \left( \log(n_0) n_0^{-\frac{s-1}{s+2r}} \right)^p.$$

This term is  $o(n_0^{-s/(s+2r)})$  if

$$-p \frac{s-1}{s+2r} < -\frac{s}{s+2r} \Leftrightarrow s > \frac{p}{p-1} = 1 + \frac{1}{p-1}.$$

If one takes  $p = 2$ , we get  $s > 2$ . But for any  $s > 1$ , there exists  $\epsilon > 0$  such that  $s = 1 + \epsilon$  and we just have to choose  $p = 2\epsilon^{-1} + 1$ . Thus for  $s > 1$ , the rate is

$$(n \vee n_0)^{-\frac{s}{s+2r}}.$$

□

**7.3. Proof of Theorem 4.1.** First for  $m \in \mathcal{M}$ , let us define the associated subspaces  $\mathcal{S}_{d_1}^m \subseteq \mathbb{R}^{d_1}$

$$\mathcal{S}_{d_1}^m = \left\{ \vec{t}_m \in \mathbb{R}^{d_1} / \vec{t}_m = {}^t(a_0(t), a_1(t), \dots, a_{m-1}(t), 0, \dots, 0) \right\}.$$

This space is defined to give nested models. When we increase the dimension from  $m$  to  $m+1$  we only compute one more coefficient. Then for any  $\vec{t} \in \mathbb{R}^{d_1}$ , we define the following contrast for the density estimation

$$\gamma_n(\vec{t}) = \|\vec{t}\|_{2,d_1}^2 - 2\langle \vec{t}, \tilde{\mathbf{G}}_{d_1}^{-1} \hat{h}_{d_1} \rangle_{2,d_1}.$$

Let us notice that for  $\vec{t}_m \in \mathcal{S}_{d_1}^m$ , thanks to the null coordinates of  $\vec{t}_m$  and the lower triangular form of  $\tilde{\mathbf{G}}_{d_1}$  and  $\tilde{\mathbf{G}}_m$ , we have

$$\langle \vec{t}_m, \tilde{\mathbf{G}}_{d_1}^{-1} \hat{h}_{d_1} \rangle_{2,d_1} = \langle \vec{t}_m, \tilde{\mathbf{G}}_m^{-1} \hat{h}_m \rangle_{2,m} = \langle \vec{t}_m, \tilde{f}_m \rangle_{2,m}.$$

So we clearly have that

$$\tilde{f}_m = \underset{\vec{t}_m \in \mathcal{S}_{d_1}^m}{\text{argmin}} \gamma_n(\vec{t}_m).$$

Now let  $m, m' \in \mathcal{M}$ ,  $\vec{t}_m \in \mathcal{S}_{d_1}^m$  and  $\vec{s}_{m'} \in \mathcal{S}_{d_1}^{m'}$ . Notice that

$$\gamma_n(\vec{t}_m) - \gamma_n(\vec{s}_{m'}) = \|\vec{t}_m - \vec{f}\|_{2,d_1}^2 - \|\vec{s}_{m'} - \vec{f}\|_{2,d_1}^2 - 2\langle \vec{t}_m - \vec{s}_{m'}, \tilde{\mathbf{G}}_{d_1}^{-1} (\hat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}$$

and due to orthonormality of Laguerre basis, for any  $m$  we have the following relations between the  $\mathbb{L}^2$  norm and the Euclidean norms,

$$\|\tilde{f}_m - f\|^2 = \|\tilde{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty} (a_j(f))^2 \quad \text{and} \quad \|f_m - f\|^2 = \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty} (a_j(f))^2 \quad (45)$$

We set  $\nu_n(\vec{t}) = \langle \vec{t}, \tilde{\mathbf{G}}_{d_1}^{-1} (\hat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}$  for  $\vec{t} \in \mathbb{R}^{d_1}$ .

According to the definition of  $\hat{m} \in \hat{\mathcal{M}}$ , for any  $m$  in the model collection  $\mathcal{M}$ , we have the following inequality

$$\gamma_n(\tilde{f}_{\hat{m}}) + \widehat{\text{pen}}(\hat{m}) \leq \gamma_n(\vec{f}_m) + \widehat{\text{pen}}(m).$$

It yields that

$$\|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 - \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 - 2\nu_n(\tilde{f}_{\hat{m}} - \vec{f}_m) \leq \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m})$$

which implies

$$\|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\nu_n(\tilde{f}_{\hat{m}} - \vec{f}_m) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}).$$

Let us notice that  $\nu_n(\tilde{f}_{\hat{m}} - \vec{f}_m) = \|\tilde{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1} \nu_n \left( \frac{\tilde{f}_{\hat{m}} - \vec{f}_m}{\|\tilde{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}} \right)$  and due to the relation  $2ab \leq a^2/4 + 4b^2$ , we have the following inequalities

$$\begin{aligned} \|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 &\leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\|\tilde{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1} \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n(\vec{t}) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}) \\ &\leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \frac{1}{4}\|\tilde{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}^2 + 4 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}) \end{aligned}$$

where  $\mathcal{B}(m, \hat{m}) = \{\vec{t}_{m \vee \hat{m}} \in \mathcal{S}_{d_1}^{m \vee \hat{m}}, \|\vec{t}_{m \vee \hat{m}}\|_{2,d_1} = 1\}$ . Now notice that

$$\|\tilde{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}^2 \leq 2\|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 + 2\|\vec{f}_m - \vec{f}\|_{2,d_1}^2$$

we then have

$$\|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \frac{1}{2}\|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 + \frac{1}{2}\|\vec{f} - \vec{f}_m\|_{2,d_1}^2 + 4 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m})$$

which implies

$$\|\tilde{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq 3\|\vec{f} - \vec{f}_m\|_{2,d_1}^2 + 2\widehat{\text{pen}}(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\widehat{\text{pen}}(\hat{m}).$$

Using Equation (45), we have

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 - \sum_{j=d_1}^{\infty} (a_j(f))^2 &\leq 3 \left( \|f - f_m\|^2 - \sum_{j=d_1}^{\infty} (a_j(f))^2 \right) + 2\widehat{\text{pen}}(m) \\ &\quad + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\widehat{\text{pen}}(\hat{m}). \end{aligned} \quad (46)$$

Now let  $\hat{p}$  be a function such that for any  $m, m'$ , we have :  $4\hat{p}(m, m') \leq \widehat{\text{pen}}(m) + \widehat{\text{pen}}(m')$ .

$$\|\tilde{f}_{\hat{m}} - f\|^2 \leq 3\|f - f_m\|^2 + 4\widehat{\text{pen}}(m) + 8 \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - \hat{p}(m, \hat{m}) \right]_+$$

Let us define  $m^* = m \vee \hat{m}$  and

$$\begin{aligned} \xi_{1,n}^2(\vec{t}) &= |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\hat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}|^2 & \hat{p}_1(m, m') &= 2\widehat{\text{pen}}_1(m \vee m') \\ \xi_{2,n}^2(\vec{t}) &= |\langle \vec{t}_{m^*}, (\tilde{\mathbf{G}}_{d_1}^{-1} - \mathbf{G}_{d_1}^{-1})\vec{h}_{d_1} \rangle_{2,d_1}|^2 & \hat{p}_2(m, m') &= 2\widehat{\text{pen}}_2(m \vee m') \end{aligned}$$

Let us notice that

$$\begin{aligned} &\left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - \hat{p}(m, \hat{m}) \right]_+ \\ &\leq \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\hat{h}_{d_1} - \vec{h}_{d_1}) + (\tilde{\mathbf{G}}_{d_1}^{-1} - \mathbf{G}_{d_1}^{-1})\vec{h}_{d_1} \rangle_{2,d_1}|^2 - \hat{p}_1(m, \hat{m}) - \hat{p}_2(m, \hat{m}) \right]_+ \\ &\leq 2 \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \xi_{1,n}^2(\vec{t}) - \frac{1}{2}\hat{p}_1(m, \hat{m}) \right]_+ + 2 \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \xi_{2,n}^2(\vec{t}) - \frac{1}{2}\hat{p}_2(m, \hat{m}) \right]_+, \end{aligned}$$

it yields that

$$\begin{aligned} \|\tilde{f}_{\hat{m}} - f\|^2 &\leq 3\|f - f_m\|^2 + 4\widehat{\text{pen}}(m) + 16 \sum_{m' \in \widehat{\mathcal{M}}} \left[ \sup_{\vec{t} \in \mathcal{B}(m, m')} \xi_{1,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_1(m, m') \right]_+ \\ &\quad + 16 \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \xi_{2,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_2(m, \hat{m}) \right]_+ \end{aligned}$$

We now use the three following results which ensure the validity of Theorem 4.1.

**Proposition 7.5.** *For  $m \in \mathcal{M}$ , it holds that*

$$\mathbb{E} [\widehat{\text{pen}}(m)] \leq C \text{pen}(m),$$

with  $C = (6 + 2\mathfrak{C}_{\text{op}} + \mathfrak{C}_{\text{F}})$ .

**Proposition 7.6.** *Under the assumptions of Theorem 4.1, there exists a constant  $C_1 > 0$  depending on  $\|h\|_\infty$  such that for  $\widehat{p}_1(m, m') = 2\widehat{\text{pen}}_1(m \vee m')$*

$$\mathbb{E} \left[ \sum_{m' \in \widehat{\mathcal{M}}} \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} \xi_{1,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_1(m, m') \right\} \right]_+ \leq \frac{C_1}{n}.$$

**Proposition 7.7.** *Under the assumptions of Theorem 4.1, there exists a constant  $C_2 > 0$  depending on  $\|h\|_\infty$  such that for  $\widehat{p}_2(m, m') = 2\widehat{\text{pen}}_2(m \vee m')$*

$$\mathbb{E} \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \xi_{2,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_2(m, \hat{m}) \right]_+ \leq C_2 \left( \frac{1}{n_0} + \text{pen}_2(m) \right).$$

In the end:

$$\mathbb{E}\|f - \hat{f}_{\hat{m}}\|^2 \leq 4C \inf_{m \in \mathcal{M}_n} \{ \|f - f_m\|^2 + \text{pen}(m) \} + \frac{C_1}{n} + \frac{C_2}{n_0},$$

as soon as  $\kappa_1 \geq 196$  and  $\kappa_2 \geq 5/2$ . □

7.3.1. *Proof of Proposition 7.5.* Let  $m$  be in the model collection  $\mathcal{M}$ . By definition we have

$$\begin{aligned} \mathbb{E} [\widehat{\text{pen}}(m)] &= \mathbb{E} [\widehat{\text{pen}}_1(m) + \widehat{\text{pen}}_2(m)] \\ &= 2\mathfrak{C}_{\kappa_1} \log n \mathbb{E} \left[ \frac{2m\|h\|_\infty}{n} \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \wedge \frac{(\|h\|_\infty \vee 1)}{n} \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2 \right] \\ &\quad + 8\kappa_2 \mathfrak{C}_{\text{E}} (\|g\|_\infty \vee 1) \frac{m}{n_0} \log n_0 \mathbb{E} \left[ \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \right]. \end{aligned}$$

Applying Lemma 3.1 for  $p = 1$ , we get that

$$\mathbb{E} \left[ \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \right] \leq 2\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 + 2\mathbb{E} \left[ \|\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \right] \leq 2\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 + 2\mathfrak{C}_{\text{op},1} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2.$$

Besides applying Corollary 7.4, we get that

$$\mathbb{E} \left[ \|\tilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2 \right] \leq 2\|\mathbf{G}_m^{-1}\|_{\text{F}}^2 + 2\mathfrak{C}_{\text{F}} \|\mathbf{G}_m^{-1}\|_{\text{F}}^2.$$

Finally

$$\mathbb{E} [\widehat{\text{pen}}(m)] \leq (4 + 2\mathfrak{C}_{\text{op},1} + 2\mathfrak{C}_{\text{F}}) \text{pen}(m). \quad \square$$

7.3.2. *Proof of Proposition 7.6.* First let us notice

$$\begin{aligned}
& \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \\
&= \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \mathbb{1}_{m' > m} \\
&+ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \mathbb{1}_{m' \leq m} \mathbb{1}_{\Delta_m} \\
&+ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \mathbb{1}_{m' \leq m} \mathbb{1}_{\Delta_m^c} \\
&= \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \mathbb{1}_{m' > m} \\
&+ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \mathbb{1}_{m' \leq m} \mathbb{1}_{\Delta_m}
\end{aligned}$$

Since  $\Delta_{m'} \subset \widehat{\mathcal{M}}$  and  $\Delta_m \subset \widehat{\mathcal{M}}$  for  $m, m' \in \widehat{\mathcal{M}}$ , it yields that

$$\begin{aligned}
& \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \\
&= \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \widehat{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \mathbb{1}_{\Delta_{m^*}}.
\end{aligned}$$

Since  $\Delta_{m^*} \subset \widehat{\mathcal{M}}$  for  $m' \in \widehat{\mathcal{M}}$ , it yields that

$$\begin{aligned}
& \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \tilde{\mathbf{G}}_{d_1}^{-1}(\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2, d_1}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \\
&= \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \widehat{\mathbf{G}}_{m^*}^{-1}(\widehat{h}_{m^*} - \vec{h}_{m^*}) \rangle_{2, m^*}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+.
\end{aligned}$$

Now, if we define  $E_1$

$$E_1 = \mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\langle \vec{t}_{m^*}, \widehat{\mathbf{G}}_{m^*}^{-1}(\widehat{h}_{m^*} - \vec{h}_{m^*}) \rangle_{2, m^*}|^2 - \frac{1}{2} \hat{p}_1(m, m') \right\}_+ \middle| Y' \right],$$

then, conditionally on  $Y'$ , the bound follows from the proof of Proposition 7.1 in Mabon (2016b) with  $\mathbf{G}_{m^*}$  replaced by  $\widehat{\mathbf{G}}_{m^*}$ ,  $\mathcal{M}$  by  $\widehat{\mathcal{M}}$  and  $\xi^2 = 1/2$  in the first case **i**) increased as  $\xi^2 = \mathbf{a} \|h\|_\infty / K_1 \log n$  with  $K_1 = 1/6$ . Then, as all bounds are independent of the random terms, the conditional expectation can be integrated with respect to the law of the sample  $(Y'_i)_{1 \leq i \leq n_0}$  without change.  $\square$

7.3.3. *Proof of Proposition 7.7.* Let us define

$$E_2 := \left[ \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} \xi_{2, n}^2(\vec{t}) - \frac{1}{2} \widehat{p}_2(m, \widehat{m}) \right]_+$$



with  $\frac{1}{2}\widehat{p}_2(m, \widehat{m}) = \widehat{\text{pen}}_2(m \vee \widehat{m})$ .

First case :  $\widehat{m} \geq m$ . Since  $\widehat{m} \in \widehat{\mathcal{M}}$ ,  $\widetilde{\mathbf{G}}_{\widehat{m}}^{-1} = \widehat{\mathbf{G}}_{\widehat{m}}^{-1}$ , it yields that

$$\begin{aligned} E_2 \mathbf{1}_{\widehat{m} \geq m} &= \left[ \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} |\langle \vec{t}_{\widehat{m}}, (\widetilde{\mathbf{G}}_{\widehat{m}}^{-1} - \mathbf{G}_{\widehat{m}}^{-1}) \vec{h}_{\widehat{m}} \rangle|^2 - \frac{1}{2} \widehat{\text{pen}}_2(\widehat{m}) \right]_+ \mathbf{1}_{\widehat{m} \geq m} \\ &\leq \left[ \|(\widehat{\mathbf{G}}_{\widehat{m}}^{-1} - \mathbf{G}_{\widehat{m}}^{-1}) \vec{h}_{\widehat{m}}\|_{2, \widehat{m}}^2 - \frac{1}{2} \widehat{\text{pen}}_2(\widehat{m}) \right]_+ \leq \left[ \|\widehat{\mathbf{G}}_{\widehat{m}}^{-1} (\mathbf{G}_{\widehat{m}} - \widehat{\mathbf{G}}_{\widehat{m}}) \mathbf{G}_{\widehat{m}}^{-1} \vec{h}_{\widehat{m}}\|_{2, \widehat{m}}^2 - \frac{1}{2} \widehat{\text{pen}}_2(\widehat{m}) \right]_+ \\ &\leq \left[ \|f\|^2 \|\widehat{\mathbf{G}}_{\widehat{m}}^{-1}\|_{\text{op}}^2 \|\mathbf{G}_{\widehat{m}} - \widehat{\mathbf{G}}_{\widehat{m}}\|_{\text{op}}^2 - \frac{1}{2} \widehat{\text{pen}}_2(\widehat{m}) \right]_+ \end{aligned}$$

Let us define the set  $\mathcal{M}_{\max}$  such that

$$\mathcal{M}_{\max} = \{m \in \llbracket 1, n \rrbracket, m \leq C \lfloor n / \log n \rfloor \wedge \lfloor n_0 / \log n_0 \rfloor\}.$$

We now introduce the favorable set

$$\mathcal{E}_m = \left\{ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}} \leq \sqrt{\kappa_2 4 (\|g\|_{\infty} \vee 1) \log n_0 \frac{m}{n_0}} \right\}, \quad \kappa_2 > 0. \quad (50)$$

and set

$$\mathcal{E} = \bigcap_{m \in \mathcal{M}_{\max}} \mathcal{E}_m.$$

Thus we can notice that for  $\widehat{m} \in \widehat{\mathcal{M}} \subset \mathcal{M}_{\max}$ . It yields that

$$\begin{aligned} E_2 \mathbf{1}_{\mathcal{E}} \mathbf{1}_{\widehat{m} \geq m} &\leq \left[ \|f\|^2 \|\widehat{\mathbf{G}}_{\widehat{m}}^{-1}\|_{\text{op}}^2 \kappa_2 4 (\|g\|_{\infty} \vee 1) \log n_0 \frac{\widehat{m}}{n_0} - \frac{1}{2} \widehat{\text{pen}}_2(\widehat{m}) \right]_+ \mathbf{1}_{\mathcal{E}} \\ &= \left[ \|\widehat{\mathbf{G}}_{\widehat{m}}^{-1}\|_{\text{op}}^2 \kappa_2 4 (\|g\|_{\infty} \vee 1) \log n_0 \frac{\widehat{m}}{n_0} - \frac{1}{2} \widehat{\text{pen}}_2(\widehat{m}) \right]_+ \mathbf{1}_{\mathcal{E}} = 0. \end{aligned}$$

On the complementary set we have that

$$\begin{aligned} \mathbb{E} [E_2 \mathbf{1}_{\widehat{m} \geq m} \mathbf{1}_{\mathcal{E}^c}] &\leq \mathbb{E} \left[ \|(\widetilde{\mathbf{G}}_{\widehat{m}}^{-1} - \mathbf{G}_{\widehat{m}}^{-1}) \vec{h}_{\widehat{m}}\|_{2, \widehat{m}}^2 \mathbf{1}_{\mathcal{E}^c} \right] \leq \mathbb{E} \left[ \sup_{m \in \mathcal{M}_{\max}} \|(\widetilde{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) \vec{h}_m\|_{2, m}^2 \mathbf{1}_{\mathcal{E}^c} \right] \\ &\leq \sum_{m \in \mathcal{M}_{\max}} 2\mathbb{E} \left[ (\|\widetilde{\mathbf{G}}_m^{-1} \vec{h}_m\|_{2, m}^2 + \|\mathbf{G}_m^{-1} \vec{h}_m\|_{2, m}^2) \mathbf{1}_{\mathcal{E}^c} \right] \\ &\leq \sum_{m \in \mathcal{M}_{\max}} 2\mathbb{E} \left[ (\|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \|\vec{h}_m\|_{2, m}^2 + \|\vec{f}_m\|_{2, m}^2) \mathbf{1}_{\mathcal{E}^c} \right] \\ &\leq \sum_{m \in \mathcal{M}_{\max}} 2\mathbb{E} \left[ (\|\vec{h}_m\|_{2, m}^2 + \|\vec{f}_m\|_{2, m}^2) n_0 \mathbf{1}_{\mathcal{E}^c} \right] \leq C n_0 |\mathcal{M}_{\max}| \mathbb{P} [\mathcal{E}^c] \end{aligned}$$

and applying the following Lemma for  $p = 3$

**Lemma 7.8.** *For all  $p \geq 1$ , there exists  $\kappa_2 \geq (p+2)/2$  and  $C_p \geq 1$  such that  $\mathbb{P} [\mathcal{E}^c] \leq \frac{C_p}{n_0^p}$ ,*

gives that  $\mathbb{E} [E_2 \mathbf{1}_{\mathcal{E}^c}] \leq \frac{C_3}{n_0}$ .

Second case:  $\widehat{m} \leq m$ . We have that

$$\begin{aligned} E_2 \mathbf{1}_{\widehat{m} \leq m} &= \left[ \sup_{\vec{t} \in \mathcal{B}(m, m)} |\langle \vec{t}_m, (\widetilde{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) \vec{h}_m \rangle|^2 - \widehat{\text{pen}}_2(m) \right]_+ (\mathbf{1}_{\Delta_m} + \mathbf{1}_{\Delta_m^c}) \\ &= \left[ \sup_{\vec{t} \in \mathcal{B}(m, m)} |\langle \vec{t}_m, (\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) \vec{h}_m \rangle|^2 - \widehat{\text{pen}}_2(m) \right]_+ \mathbf{1}_{\Delta_m} + \sup_{\vec{t} \in \mathcal{B}(m, m)} |\langle \vec{t}_m, \vec{f}_m \rangle|^2 \mathbf{1}_{\Delta_m^c}. \end{aligned}$$

It implies that for  $\mathcal{E}_m$  defined by (50)

$$\mathbb{E}[E_2 \mathbf{1}_{\widehat{m} \leq m}] \leq \mathbb{E} \left[ \left[ \sup_{\vec{t} \in \mathcal{B}(m, m)} |\langle \vec{t}_m (\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) \vec{h}_m \rangle|^2 - \widehat{\text{pen}}_2(m) \right]_+ \mathbf{1}_{\Delta_m} \mathbf{1}_{\mathcal{E}_m} \right] + \|f\|^2 \mathbb{P}[\Delta_m^c].$$

According to Lemma 7.3,

$$\|f\|^2 \mathbb{P}[\Delta_m^c] \leq \|f\|^2 8\mathfrak{C}_2 \log m \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \frac{m}{n_0} \leq \|f\|^2 8\mathfrak{C}_2 \log n_0 \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \frac{m}{n_0} \lesssim \text{pen}_2(m)$$

and

$$\mathbb{E} \left[ \left[ \sup_{\vec{t} \in \mathcal{B}(m, m)} |\langle \vec{t}_m (\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) \vec{h}_m \rangle|^2 - \widehat{\text{pen}}_2(m) \right]_+ \mathbf{1}_{\mathcal{E}_m} \right] \leq \frac{C_3}{n_0}$$

On  $\mathcal{E}_m^c$ , we have

$$\begin{aligned} \mathbb{E} [E_2 \mathbf{1}_{\widehat{m} \leq m} \mathbf{1}_{\mathcal{E}_m^c}] &\leq \mathbb{E} \left[ \|(\widetilde{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}) \vec{h}_m\|_{2, m}^2 \mathbf{1}_{\mathcal{E}_m^c} \right] \leq 2\mathbb{E} \left[ (\|\widetilde{\mathbf{G}}_m^{-1} \vec{h}_m\|_{2, m}^2 + \|\mathbf{G}_m^{-1} \vec{h}_m\|_{2, m}^2) \mathbf{1}_{\mathcal{E}_m^c} \right] \\ &\leq 2\mathbb{E} \left[ (\|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \|\vec{h}_m\|_{2, m}^2 + \|\vec{f}_m\|_{2, m}^2) \mathbf{1}_{\mathcal{E}_m^c} \right] \leq 2\mathbb{E} \left[ (\|\vec{h}_m\|_{2, m}^2 + \|\vec{f}_m\|_{2, m}^2) n_0 \mathbf{1}_{\mathcal{E}_m^c} \right] \\ &\leq C n_0 \mathbb{P}[\mathcal{E}_m^c] \end{aligned}$$

Besides  $\mathcal{E} \subset \mathcal{E}_m$  which implies that  $\mathbb{P}[\mathcal{E}_m^c] \leq \mathbb{P}[\mathcal{E}^c]$ . Then applying Lemma 7.8, we get that

$$\mathbb{E} [E_2 \mathbf{1}_{\widehat{m} \leq m} \mathbf{1}_{\mathcal{E}_m^c}] \leq \frac{C_2}{n_0}$$

□

*Proof of Lemma 7.8.* We apply Lemma 7.1 for  $t = \sqrt{4\kappa_2(\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0}}$

$$\begin{aligned} \mathbb{P}[\mathcal{E}^c] &= \mathbb{P} \left[ \exists m \in \mathcal{M}_{\max}, \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}} > \sqrt{4\kappa_2(\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0}} \right] \\ &\leq \sum_{m \leq n_0} \mathbb{P} \left[ \|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}} > \sqrt{4\kappa_2(\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0}} \right] \\ &\leq 2 \sum_{m \leq n_0} m \exp \left( -\frac{1}{2} \frac{4\kappa_2(\|g\|_\infty \vee 1) \log n_0 m}{\|g\|_\infty m + (2\sqrt{2}/3)m \sqrt{4\kappa_2(\|g\|_\infty \vee 1) \log n_0} \sqrt{\frac{m}{n_0}}} \right) \\ &\leq 2 \sum_{m \leq n_0} m \exp \left( -\frac{4\kappa_2(\|g\|_\infty \vee 1) \log n_0}{2} \left( \frac{1}{\|g\|_\infty} \wedge \frac{3}{2\sqrt{2} \sqrt{4\kappa_2(\|g\|_\infty \vee 1) \log n_0} \sqrt{\frac{n_0}{m}}} \right) \right) \\ &\leq C \sum_{m \leq n_0} m \exp \left( -\frac{4\kappa_2(\|g\|_\infty \vee 1) \log n_0}{2\|g\|_\infty} \right) \\ &\leq C \sum_{m \leq n_0} m \exp(-2\kappa_2 \log n_0) \leq C n_0^2 \exp(-2\kappa_2 \log n_0) \end{aligned}$$

Finally we get

$$\mathbb{P}[\mathcal{E}^c] \leq C n_0^2 \exp(-\kappa_2 2 \log n_0) = \frac{C}{n_0^{2\kappa_2 - 2}} = \frac{C}{n_0^p}$$

with  $p \geq 1$  if  $\kappa_2 \geq (p+2)/2$ .

□

**7.4. Proof of Corollary 4.2.** The beginning of the proof follows exactly the same lines as in Theorem 4.1 except that  $\widehat{\text{pen}}$  and  $\widehat{m}$  are respectively replaced by  $\widetilde{\text{pen}}$  and  $\widetilde{m}$ .

Starting from Equation (46), we get

$$\begin{aligned} \|\widehat{f}_{\widetilde{m}} - f\|^2 &\leq 3\|f - f_m\|^2 + 2\widetilde{\text{pen}}(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \widetilde{m})} \nu_n^2(\vec{t}) - 2\widetilde{\text{pen}}(\widehat{m}) \\ &\leq 3\|f - f_m\|^2 + 2(\widetilde{\text{pen}}(m) - \widehat{\text{pen}}(m)) + 2\widehat{\text{pen}}(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \widetilde{m})} \nu_n^2(\vec{t}) - 2\widehat{\text{pen}}(\widetilde{m}) \\ &\quad + 2(\widehat{\text{pen}}(\widetilde{m}) - \widetilde{\text{pen}}(\widetilde{m})) \end{aligned}$$

We now apply Proposition 7.9 hereafter and we get the final result.  $\square$

**Proposition 7.9.**

$$\begin{aligned} \text{(i)} \quad \mathbb{E}|\widehat{\text{pen}}(m) - \widetilde{\text{pen}}(m)| &\lesssim \text{pen}(m) + \frac{1}{n_0} + \frac{1}{n} \\ \text{(ii)} \quad \mathbb{E}(\widehat{\text{pen}}(\widetilde{m}) - \widetilde{\text{pen}}(\widetilde{m})) &\lesssim \frac{1}{n_0} + \frac{1}{n}. \end{aligned}$$

*Proof of Proposition 7.9.* The proof relies on the introduction of the set so that the estimators of the sup-norms of  $h$  and  $g$  are under control around their true values. As it works exactly the same for both functions, we only detail the proof for  $g$ .

Let us define the set

$$\Lambda(g) = \left\{ \left| \|\widehat{g}_D\|_\infty - \|g\|_\infty \right| \leq \frac{\|g\|_\infty}{2} \right\}.$$

(i) It yields that

$$\begin{aligned} \mathbb{E}|\widehat{\text{pen}}_2(m) - \widetilde{\text{pen}}_2(m)| \mathbf{1}_{\Lambda(g)} &= 8\kappa_2 \mathbb{E} \left[ \left| (2\|\widehat{g}_D\|_\infty \vee 1) - (\|g\|_\infty \vee 1) \right| \log n_0 \frac{m}{n_0} \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \mathbf{1}_{\Lambda(g)} \right] \\ &\leq 8\kappa_2 \mathbb{E} \left[ 4(\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0} \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \mathbf{1}_{\Lambda(g)} \right] \end{aligned}$$

Moreover applying Proposition 7.5, we get that

$$\mathbb{E}|\widehat{\text{pen}}_2(m) - \widetilde{\text{pen}}_2(m)| \mathbf{1}_{\Lambda(g)} \leq C \text{pen}_2(m)$$

On the set  $\Lambda^c(g)$  with the definition of  $\mathcal{M}$ , we have

$$\begin{aligned} \mathbb{E}|\widehat{\text{pen}}_2(m) - \widetilde{\text{pen}}_2(m)| \mathbf{1}_{\Lambda^c(g)} &= 8\kappa_2 \mathbb{E} \left[ \left| (2\|\widehat{g}_D\|_\infty \vee 1) - (\|g\|_\infty \vee 1) \right| \log n_0 \frac{m}{n_0} \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \right] \\ &\leq 8\kappa_2 \mathbb{E} \left[ \left| (2\|\widehat{g}_D\|_\infty \vee 1) - (\|g\|_\infty \vee 1) \right| \right] \\ &\leq 8\kappa_2 \mathbb{E} \left[ (2\|\widehat{g}_D\|_\infty \vee 1) \mathbf{1}_{\Lambda^c(g)} \right] + \mathbb{E} \left[ (\|g\|_\infty \vee 1) \mathbf{1}_{\Lambda^c(g)} \right] \end{aligned}$$

Yet  $\|\widehat{g}_D\|_\infty \leq \|\sum_k \varphi_k\|_\infty \leq 2D \leq 2n_0$ , then

$$\mathbb{E}|\widehat{\text{pen}}_2(m) - \widetilde{\text{pen}}_2(m)| \mathbf{1}_{\Lambda^c(g)} \leq C \mathbb{P}[\Lambda^c(g)]$$

Now applying Lemma 5.2 in Mabon (2016a), it holds that for all  $p > 0$  and  $\log n_0 \leq D \leq \|g\|_\infty / (128\sqrt{2})n_0 / (\log n_0)^p$ ,

$$\mathbb{P}[\Lambda^c(g)] \leq \frac{2D}{n_0^p}$$

The proof follows exactly the same lines for controlling  $\mathbb{E}|\widehat{\text{pen}}_1(m) - \widetilde{\text{pen}}_1(m)|$  by defining  $\Lambda(h)$  and replacing  $n_0$  by  $n$ .

(ii) On  $\Lambda(g)$ , we have  $\|g\|_\infty - 2\|\widehat{g}_D\|_\infty \leq 0$  which yields that  $(\|g\|_\infty \vee 1) - 2(\|\widehat{g}_D\|_\infty \vee 1) \leq 0$ , thus  $(\widehat{\text{pen}}_2(\widetilde{m}) - \widetilde{\text{pen}}_2(\widetilde{m})) \mathbf{1}_{\Lambda(g)} \leq 0$ . Moreover

$$\mathbb{E}[(\widehat{\text{pen}}_2(\widetilde{m}) - \widetilde{\text{pen}}_2(\widetilde{m})) \mathbf{1}_{\Lambda^c(g)}] \leq \mathbb{E}[|\widehat{\text{pen}}_2(\widetilde{m}) - \widetilde{\text{pen}}_2(\widetilde{m})| \mathbf{1}_{\Lambda^c(g)}] \leq C \mathbb{P}[\Lambda^c(g)],$$

as above since  $\tilde{m} \in \widehat{\mathcal{M}}$ . This gives the result for  $\widehat{\text{pen}}_2$ . The same reasoning holds for  $\widehat{\text{pen}}_1(\tilde{m}) - \widetilde{\text{pen}}_1(\tilde{m})$   $\square$

## APPENDIX A. APPENDIX

A proof of the following theorem can be found in [Stewart and Sun \(1990\)](#).

**Theorem A.1.** *Let  $\mathbf{A}$ ,  $\mathbf{B}$  be  $(m \times m)$  matrices. If  $\mathbf{A}$  is invertible and  $\|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}} < 1$ , then  $\tilde{\mathbf{A}} := \mathbf{A} + \mathbf{B}$  is invertible and it holds*

$$\|\tilde{\mathbf{A}}^{-1} - \mathbf{A}^{-1}\|_{\text{op}} \leq \frac{\|\mathbf{B}\|_{\text{op}} \|\mathbf{A}^{-1}\|_{\text{op}}^2}{1 - \|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}}}$$

**Theorem A.2** (Bernstein Matrix inequality). *Consider a finite sequence  $\{\mathbf{S}_k\}$  of independent, random matrices with common dimension  $d_1 \times d_2$ . Assume that*

$$\mathbb{E}\mathbf{S}_k = 0 \quad \text{and} \quad \|\mathbf{S}_k\|_{\text{op}} \leq L \quad \text{for each index } k.$$

Introduce the random matrix

$$\mathbf{Z} = \sum_k \mathbf{S}_k.$$

Let  $\nu(\mathbf{Z})$  be the variance statistic of the sum:  $\nu(\mathbf{Z}) = \max\{\lambda_{\max}(\mathbb{E}[\mathbf{Z}^t\mathbf{Z}]), \lambda_{\max}(\mathbb{E}[{}^t\mathbf{Z}\mathbf{Z}])\}$ . Then

$$\mathbb{E}\|\mathbf{Z}\|_{\text{op}} \leq \sqrt{2\nu(\mathbf{Z}) \log(d_1 + d_2)} + \frac{1}{3}L \log(d_1 + d_2).$$

Furthermore, for all  $t \geq 0$

$$\mathbb{P}[\|\mathbf{Z}\|_{\text{op}} \geq t] \leq (d_1 + d_2) \exp\left(-\frac{t^2/2}{\nu(\mathbf{Z}) + Lt/3}\right).$$

A proof can be found in [Tropp \(2012\)](#) or [Tropp \(2015\)](#).

**Theorem A.3** (Matrix moment inequality, Theorem A.1 in [Chen et al. \(2012\)](#)). *Suppose that  $q \geq 2$  and fix  $r \geq \max(q, 2 \log p)$ . Consider a finite sequence  $\{\mathbf{Y}_i\}$  of independent, symmetric, random, self-adjoint matrices with dimension  $p \times p$ . Then*

$$\left[ \mathbb{E} \lambda_{\max} \left( \sum_i \mathbf{Y}_i \right)^q \right]^{1/q} \leq \sqrt{er \lambda_{\max} \left( \sum_i \mathbb{E} \mathbf{Y}_i^2 \right)} + 2er \left[ \mathbb{E} \max_i \lambda_{\max}^q(\mathbf{Y}_i) \right]^{1/q}.$$

A proof can be found in [Chen et al. \(2012\)](#).

## 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.
- Belomestny, D., Comte, F., and Genon-Catalot, V. (2016). Nonparametric laguerre estimation in the multiplicative censoring model. *Electron. J. Statist.*, 10(2):3114–3152.
- Birgé, L. (1999). An alternative point of view on Lepski's method. In *State of the art in probability and statistics: Festschrift for Willem R. van Zwet*, pages 113–133. IMS Lecture Notes-Monograph Series.
- Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer.
- 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. (2004). Deconvolution of supersmooth densities with smooth noise. *Can. J. Stat.*, 32(2):181–192.
- 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.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Am. Stat. Assoc.*, 83(404):1184–1186.
- Chen, R. Y., Gittens, A., and Tropp, J. A. (2012). The masked sample covariance estimator: an analysis using matrix concentration inequalities. *Inf. Inference*, 1(1):2–20.
- Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2015). Laplace deconvolution and its application to dynamic contrast enhanced imaging. *J. R. Stat. Soc., Ser. B.*
- Comte, F. and Genon-Catalot, V. (2015). Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.*, 9:1112–1148.
- Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc., Ser. B*, 73:601–627.
- Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.*, 3(34):431–452.
- Dattner, I., Reiß, M., and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli*, 22(2):143–192.
- Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Stat. Math.*, 56(1):19–47.
- Efromovich, S. (1997). Density estimation for the case of supersmooth measurement errors. *J. Am. Stat. Assoc.*, 92:526–535.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.
- Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Stat. Neerl.*, 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.
- Jirak, M., Meister, A., and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.*, 42(5):1970–2002.
- Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.*, 37(5a):2301–2323.
- Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli*, 19(5A):1576–1611.
- Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.*, 8(2):2879–2904.
- Mabon, G. (2016a). Adaptive deconvolution of linear functionals on the nonnegative real line. *J. Stat. Plann. Inference*, 178:1–23.
- Mabon, G. (2016b). Adaptive deconvolution on the nonnegative real line. *preprint MAP5 2014-33*.
- Massart, P. (2003). *Concentration Inequalities and Model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003.
- Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.
- Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametric Stat.*, 7(4):307–330.
- Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053.
- Reiß, M. and Selk, L. (2015). Estimating nonparametric functionals efficiently under one-sided errors. Preprint, arXiv:1407.4229, To appear in *Bernoulli*.
- Stefanski, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Stat. Probab. Lett.*, 9:229–235.
- Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21:169–184.

- Stewart, G. W. and Sun, J.-G. (1990). *Matrix perturbation theory*. Boston etc.: Academic Press, Inc.
- Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434.
- Tropp, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230.
- van Es, B. (2011). Combining kernel estimators in the uniform deconvolution problem. *Stat. Neerl.*, 65(3):275–296.
- Vareschi, T. (2015). Noisy Laplace deconvolution with error in the operator. *J. Stat. Plann. Inference*, 157–158:16–35.