ISSN 1066-5307, Mathematical Methods of Statistics, 2017, Vol. 26, No. 4, pp. 237–266. © Allerton Press, Inc., 2017.

Laguerre Deconvolution with Unknown Matrix Operator

F. Comte[1*] and G. Mabon[2]**

[1]Univ. Lyon, Univ. Claude Bernard Lyon 1
MAP5, UMR CNRS 8145, Univ. Paris Descartes, Paris, France
[2]Inst. Math. Humboldt-Univ. zu Berlin, Berlin, Germany
Received September 1, 2017

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 [17]. 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, well-suited for nonnegative 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 [23].

Keywords: convolution model, linear inverse problem, nonnegative random variables, Laguerre basis, nonparametric density estimation, random matrix, oracle inequalities, adaptive estimation.

2000 Mathematics Subject Classification: primary 62G07; secondary 62N02.

DOI: 10.3103/S1066530717040019

1. INTRODUCTION

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 independent and identically distributed (i.i.d.) variables 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 specific feature of our framework is that all random variables are 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.

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.

In the first works, g is assumed to be known, see [21] and references therein. However this assumption is 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

*E-mail: fabienne.comte@parisdescartes.fr
**E-mail: gwennaelle.mabon@hu-berlin.de

can be obtained. Neumann [20] first proposed an estimation strategy still based on Fourier inversion; for the study of convergence rates, see [20], [13] or [19]. 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 [8] and Kappus and Mabon [15] who extended it to the adaptive strategy, by Johannes and Schwarz [14] who consider a model of circular deconvolution and by Dattner et al. [9] who deal with adaptive quantile estimation via Lespki's method.

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.

Groeneboom and Wellner [11] 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 [10]. In the Uniform deconvolution problem, van Es [24] proposes a density estimator and an estimator of the c.d.f. using kernel estimators and inversion formula. The work of Mabon [17] 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.

The method relies on a a projection strategy in a specific $\mathbb{R}^+$-supported orthonormal functional basis, namely the Laguerre basis. This basis has been used for nonnegative variables in other settings: e.g. in [5] and [25] in a regression setting, or in [2] for a multiplicative censoring model.

Here, we extend the procedure proposed in [17] for known g to 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 a deterministic way when g was known. 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 in [20] and developed in [15] in the Fourier context, with the help of tools related to matrix perturbation theory (see [21]) and random matrices taken in [23]. A result of matrix perturbation theory (see Th. 8.1) is the key result to enable us to prove a lemma similar to Lemma 2.1 in [20]. Besides, Bernstein's inequality for random matrices provides useful moment inequalities. We discuss the influence of the two sample sizes n and n_0 and compare our results with the Fourier strategy outcomes, which still can be applied to nonnegative random variables.

Let us now explain the plan of the paper. In Section 2, we give notation, we define the model, the Laguerre basis and the density estimator computed on an 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 [23]). Then we discuss the resulting rates of convergence on specific subspaces of $\mathbb{L}^2(\mathbb{R}^+)$ called Laguerre–Sobolev spaces with index $s > 0$, defined in [3]. Our strategy is especially well fitted for estimating functions belonging to a collection of mixed Gamma densities. 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 the adaptive estimators through simulation experiments. Numerical results are presented and compared to the performance in 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, Assumptions and Notation

We consider the model

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \tag{2.1}$$

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

$$(Y_1', \dots, Y_{n_0}') \quad \text{and } (Y_i')_{1\le i\le n_0} \ \text{ independent of } \ (X_i, Y_i)_{1\le i\le n}, \tag{2.2}$$

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.

Now we fix some notation. 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 φ, ψ: $\mathbb{R}^+ \to \mathbb{R}$ belonging to $\mathbb{L}^2(\mathbb{R}^+)$, we denote $\|\varphi\|$ the $\mathbb{L}^2$ norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}^+} |\varphi(x)|^2\, dx$, $\langle\varphi, \psi\rangle$ the scalar product between φ and ψ defined by $\langle\varphi, \psi\rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)\, dx$.

Let d be an integer, for two vectors $\vec{u}$ and $\vec{v}$ belonging to $\mathbb{R}^d$ we denote $\|\vec{u}\|_{2,d}$ 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 $\|\mathbf{A}\|_{\text{F}} = \sqrt{\sum_{i,j} a_{ij}^2}$.

2.2. Laguerre Basis and Spaces

We define the Laguerre basis as

$$\forall k \in \mathbb{N}, \forall x \ge 0, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{x^j}{j!}. \tag{2.3}$$

The Laguerre polynomials L_k defined by Eq. (2.3) 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 $(\varphi_k)_{k\ge0}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$. We can also notice that the Laguerre basis satisfies the following inequality for any integer k

$$\sup_{x\in\mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_\infty \le \sqrt{2}. \tag{2.4}$$

Lemma 2.1. *Let D_1 be a random variable with density τ. Assume that $\tau \in \mathbb{L}^2(\mathbb{R}^+)$ and $\mathbb{E}(D_1^{-1/2}) < +\infty$. For $m \ge 1$,*

$$\sum_{k=0}^{m-1} \int_0^{+\infty} [\varphi_k(x)]^2 \tau(x)\, dx \le c^\star \sqrt{m},$$

where $c^\star$ is a constant depending on $\mathbb{E}(D_1^{-1/2})$ and $\|\tau\|$.

This result is a particular case of a Lemma proved in a work in progress by Comte and Genon-Catalot; for the sake of completeness, the proof is recalled in Section 7. The condition $\mathbb{E}(D_1^{-1/2}) < +\infty$ is rather weak and is satisfied by all classical densities. In particular, it holds if τ takes a finite value in 0. Note that if one uses (2.4), one bounds $\sum_{k=0}^{m-1} \mathbb{E}(\varphi_j^2(D_1))$ by $2m$ while with Lemma 2.1 the bound becomes $c^\star\sqrt{m}$, which is an improvement provided that $\mathbb{E}(D_1^{-1/2}) < +\infty$ and $c^\star$ is not too large.

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

$$p(x) = \sum_{k\ge0} a_k(p)\varphi_k(x), \qquad \text{where} \quad a_k(p) = \int_{\mathbb{R}^+} p(u)\varphi_k(u)\, du.$$

According to formula 22.13.14 in [1], what makes the Laguerre basis relevant in our deconvolution setting is the relation

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u)\,du = 2^{-1/2}\big(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\big), \tag{2.5}$$

where $\star$ stands for the convolution product.

Classically, to derive the rates of convergence of estimators, we need to evaluate the order of the bias term, which depends on the smoothness of the signal. To that aim, we consider Laguerre–Sobolev regularity spaces associated with the basis and defined by

$$W^s(\mathbb{R}^+, L) = \Big\{p\colon \mathbb{R}^+ \to \mathbb{R},\, p \in \mathbb{L}^2(\mathbb{R}^+),\, \sum_{k\geq 0} k^s a_k^2(p) \leq L < +\infty\Big\} \quad \text{with} \quad s \geq 0, \tag{2.6}$$

where $a_k(p) = \langle p, \varphi_k\rangle$. Bongioanni and Torrea [3] have introduced Laguerre–Sobolev space but the link with the coefficients of a function on a Laguerre basis was done in [6]. Indeed, let s be an integer, for $p\colon \mathbb{R}^+ \to \mathbb{R}$ 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 [6]. Thus, for $f \in W^s(\mathbb{R}^+, L)$ defined by (2.6) and f_m its orthogonal projection

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} a_k^2(f) = \sum_{k=m}^{\infty} a_k^2(f) k^s k^{-s} \leq L m^{-s}. \tag{2.7}$$

2.3. Projection Estimator of the Density When g is Known

Here we briefly recall the projection estimator of f when g is known established in [17]. 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 (2.1) 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 Eq. (2.5) with convention $a_{-1}(g) = 0$ implies that

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

This yields the following infinite linear triangular system $\vec{h}_\infty = \mathbf{G}_\infty \vec{f}_\infty$ with

$$\vec{h}_m = {}^t(a_0(h), \dots, a_{m-1}(h)), \qquad \vec{f}_m = {}^t(a_0(f), \dots, a_{m-1}(f))$$

and

$$[\mathbf{G}_m]_{i,j} = \begin{cases} 2^{-1/2}a_0(g) & \text{if } i = j, \\ 2^{-1/2}\big(a_{i-j}(g) - a_{i-j-1}(g)\big) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

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

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

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

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 that the projection of f on $\mathcal{S}_m$ can be estimated by

$$\hat{f}_m(x) = \sum_{k=0}^{m-1} \hat{a}_k \varphi_k(x) \qquad \text{with} \quad \hat{\vec{f}}_m = \mathbf{G}_m^{-1}\hat{\vec{h}}_m \quad \text{and} \quad \hat{a}_k(Z) = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(Z_i) \tag{2.9}$$

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

2.4. Projection Estimator of the Density When g is Unknown

Thanks to (2.2) we can easily derive an estimate of $\mathbf{G}_m$ by replacing its coefficients by their empirical versions,

$$[\widehat{\mathbf{G}}_m]_{i,j} = \begin{cases} 2^{-1/2}\hat{a}_0(Y') & \text{if } i = j, \\ 2^{-1/2}\left(\hat{a}_{i-j}(Y') - \hat{a}_{i-j-1}(Y')\right) & \text{if } j < i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.10}$$

where $\hat{a}_k(Y') = (1/n_0)\sum_{\ell=1}^{n_0} \varphi_k(Y'_\ell)$. It is clear that $\mathbb{E}[\widehat{\mathbf{G}}_m] = \mathbf{G}_m$. It is worth noting that $\widehat{\mathbf{G}}_m$ 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 $\widehat{\mathbf{G}}_m$ and $\mathbf{G}_m^{-1}$, we have to introduce a cutoff. Thus we define an inverse of $\widehat{\mathbf{G}}_m$ as follows

$$\widetilde{\mathbf{G}}_m^{-1} = \begin{cases} \widehat{\mathbf{G}}_m^{-1} & \text{if} \quad \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq \sqrt{\dfrac{n_0}{m \log m}}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

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

$$\sqrt{2}/|\hat{a}_0(Y')| = \text{spr}(\widehat{\mathbf{G}}_m^{-1}) \leq \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \tag{2.12}$$

(see Theorem 5.6.9 in [12]). Note that, for any threshold $t > 0$, $\|\mathbf{G}_m^{-1}\|_{\text{op}} \leq t$ implies $2^{-1/2}a_0(g) \geq t^{-1}$ and $\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq t$ implies $2^{-1/2}|\hat{a}_0(Y')| \geq t$.

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

$$\tilde{f}_m(x) = \sum_{k=0}^{m-1} \tilde{a}_k \varphi_k(x) \quad \text{with} \quad \tilde{\vec{f}}_m = \widetilde{\mathbf{G}}_m^{-1}\hat{\vec{h}}_m \tag{2.13}$$

with $\hat{\vec{h}}_m$ defined by (2.9), $\widetilde{\mathbf{G}}_m^{-1}$ by (2.11) and $\tilde{\vec{f}}_m = {}^t(\tilde{a}_0, \ldots, \tilde{a}_{m-1})$.

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 Eq. (2.13). Using the isomorphism between the Euclidean norm and the $\mathbb{L}^2$-norm, we show that

$$\mathbb{E}\|f_m - \tilde{f}_m\|^2 = \|f - f_m\|^2 + \mathbb{E}\|f_m - \tilde{f}_m\|^2 = \|f - f_m\|^2 + \mathbb{E}\|\vec{f}_m - \tilde{\vec{f}}_m\|_{2,m}^2 \tag{3.1}$$

$$= \|f - f_m\|^2 + \mathbb{E}\|\mathbf{G}_m^{-1}\vec{h}_m - \mathbf{G}_m^{-1}\hat{\vec{h}}_m + \mathbf{G}_m^{-1}\hat{\vec{h}}_m - \widetilde{\mathbf{G}}_m^{-1}\hat{\vec{h}}_m\|_{2,m}^2 \tag{3.2}$$

$$\leq \|f - f_m\|^2 + 2\mathbb{E}\|\mathbf{G}_m^{-1}(\vec{h}_m - \hat{\vec{h}}_m)\|_{2,m}^2 + 2\mathbb{E}\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\hat{\vec{h}}_m\|_{2,m}^2. \tag{3.3}$$

The first two terms correspond to the squared bias term and the variance term appearing in [17] 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 $\widetilde{\mathbf{G}}_m^{-1}$. For that we use some tools of random matrix theory and particularly matrix concentration inequalities (see [23]) and Paulsen dilatation trick (see the proof of Corollary 7.3).

3.1. Upper Bounds on the MISE

First we state a lemma useful to bound the $\mathbb{L}^2$ risk of $\tilde{f}_m$ along with an important corollary.

Lemma 3.1. *For $\widetilde{\mathbf{G}}_m^{-1}$ defined by Eq.* (2.11)*, $\|g\|_\infty < \infty$ and $m \log m \le n_0$, then for any integer p there exists a positive constant $\mathfrak{C}_{\mathrm{op},p}$ such that*

$$\mathbb{E}\left[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^{2p}\right] \le \mathfrak{C}_{\mathrm{op},p}\Big(\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \wedge \log m \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^4 \frac{m}{n_0}\Big)^p. \tag{3.4}$$

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

$$\mathbb{E}\left[\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2\right] \le \mathfrak{C}_{\mathrm{E}}\Big(1 \wedge \log m \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2\Big). \tag{3.5}$$

Clearly, the first bound is very general and used at several steps of the proof. It is also worth noting that Corollary 3.2 provides a better result than a rough application of Lemma 3.1 relying on $\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \le \|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2 \|h\|^2$. Relying on these key results, we can prove the main result of this subsection:

Proposition 3.3. *If f and g belong to $\mathbb{L}^2(\mathbb{R}^+)$, $\|g\|_\infty < \infty$, for $\tilde{f}_m$ defined by* (2.13) *the following result holds:*

$$\mathbb{E}\|f - \tilde{f}_m\|^2 \le \|f - f_m\|^2 + \frac{\mathfrak{C}}{n}\big(\tau_m \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\big) + 4\mathfrak{C}_{\mathrm{E}} \log m \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \tag{3.6}$$

with $\mathfrak{C} = 4 + \mathfrak{C}_{\mathrm{op},1}$. Moreover, here and in all the sequel, $\tau_m = 2m$ in the general case and $\tau_m = c^\star \sqrt{m}$ if $\mathbb{E}(1/\sqrt{Z_1}) < +\infty$ and $c^\star$ is a constant depending on $\mathbb{E}(1/\sqrt{Z_1})$.

Let us comment on the terms in the right-hand side of Eq. (3.6).

- 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 [17], where $\tau_m = 2m$).
  - The first term, $\|f - f_m\|^2$, is the squared bias term which gets smaller when m increases.
  - The second term $n^{-1}\big(\tau_m \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\big)$ has the order of the variance term when g is known, see [17], where $\tau_m = 2m$. Thanks to Lemma 3.4 in [17], we know that the spectral norm of $\mathbf{G}_m^{-1}$ grows with the dimension m, and thus that this term is increasing with m.
- The third term, of order $m \log(m) \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 / n_0$ 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 $n = n_0$. However the factor m, which cannot be reduced to $\sqrt{m}$, corresponds to the fact that the number of estimated terms in $\mathbf{G}_m$ is of order m^2 (while there are only m in $\hat{\vec{h}}_m$). This term is also increasing in m.

We illustrate hereafter that the bound in Proposition 3.3 implies upper rates of estimation.

3.2. Rates of Convergence and Examples

We have stated the bias order under regularity assumptions in (2.7). Now we have to evaluate the variance terms of Eq. (3.6) which means to assess the order of $\|\mathbf{G}_m^{-1}\|_{\rm op}^2$ and $\|\mathbf{G}_m^{-1}\|_{\rm F}^2$. First we define an integer $r \geq 1$ such that $r = 1$ if $g(0) = B_1 \neq 0$ and for $r \geq 2$,

$$\frac{d^j}{dx^j} g(x)\mid_{x=0} = 0 \text{ if } j = 0, 1, \dots, r-2 \text{ and } \frac{d^{r-1}}{dx^{r-1}} g(x)\mid_{x=0} = B_r \neq 0.$$

Comte et al. [5] give conditions on the density g giving the exact order of the Frobenius and spectral norms of $\mathbf{G}_m^{-1}$.

Lemma 3.4 (Comte at al. [5]). *Let r be defined as above. 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, $G(s) = \mathbb{E}(e^{-sY_1})$, has no zero with nonnegative real parts except for the zeros of the form $s = \infty + ib$*

are satisfied, then

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

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

We can check that, if g is a $\Gamma(q, \mu)$ density, then g satisfies (C1) and (C2) with $r = q$ and thus the variance term $\big(\tau_m \|\mathbf{G}_m^{-1}\|_{\rm op}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\rm F}^2\big)/n$ has order m^{2q}/n.

Optimizing the squared bias and the variance terms in the upper bounds stated in Propositions 3.3 implies the following result.

Proposition 3.5. *If f belongs to $W^s(\mathbb{R}^+, L)$ and g satisfies* (C1)–(C2) *for $r \geq 1$, then $\tilde{f}_{m_{opt}}$ defined by* (2.13) *with $m_{opt} \propto n^{1/s+2r} \wedge (n_0/\log n_0)^{1/s+2r+1}$ satisfies*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}\|f - \tilde{f}_{m_{opt}}\|^2 \leq C_1(s, L) n^{-s/s+2r} \vee \left(\frac{n_0}{\log n_0}\right)^{-s/s+2r+1}, \tag{3.7}$$

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

In n and n_0 have the same order, the rate is given by the term $(n_0/\log n_0)^{-s/s+2r+1}$. If n_0 is much larger than n, we can recover the rate corresponding to the known noise case: more precisely, if $n_0 \geq n^{3/2}$, then choosing $m_{opt} \propto n^{1/s+2r}$ yields $\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}\|f - \tilde{f}_{m_{opt}}\|^2 \leq C_2(s, L) n^{-s/s+2r}$, where $C_2(s, L)$ is a positive constant.

Remark. Note that if there is no noise, then the second variance term disappears and we should have $\mathbf{G}_m$ equal to $\mathbf{I}_m$, the $m \times m$ identity matrix, in the first variance term, so that $\tau_m \|\mathbf{G}_m^{-1}\|_{\rm op}^2 \wedge \|\mathbf{G}_m^{-1}\|_{\rm F}^2 = \tau_m \wedge m = O(\sqrt{m})$ if $\mathbb{E}(1/\sqrt{X_1}) < +\infty$. This order allows us to recover a classical rate of order $O(n^{-2s/(2s+1)})$ on Sobolev balls $W^s(\mathbb{R}^+, L)$.

3.3. Comparison with Fourier Rates on Some Examples

In this section we denote by $\psi^*(x) = \int e^{-iux}\psi(u)\,du$ the Fourier transform of an integrable function ψ. The Fourier estimator of f in the model defined by (2.1)–(2.2) is in fact an estimator of $f_{m,\mathrm{Fo}}(x) = (2\pi)^{-1}\int_{-\pi m}^{\pi m} e^{iux}f^*(u)\,du$, the orthogonal projection of f on the space $\mathfrak{S}_m = \{\psi \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \mathrm{support}(\psi^*) \subset [-\pi m, \pi m]\}$. It is given by

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

with

$$\widehat{h^*}(u) = \frac{1}{n}\sum_{j=1}^{n} e^{-iuZ_j}, \qquad \widehat{g^*}(u) = \frac{1}{n_0}\sum_{j=1}^{n_0} e^{-iuY'_j}, \qquad \frac{1}{\widetilde{g^*}(u)} = \frac{\mathbf{1}\{|\widehat{g^*}(u)| \geq n_0^{-1/2}\}}{\widehat{g^*}(u)}.$$

The risk bound obtained in [20] can be written as follows,

$$\mathbb{E}\|f - \hat{f}_{m,\mathrm{Fo}}\|^2 \leq \|f - f_{m,\mathrm{Fo}}\|^2 + C_1\frac{\Delta(m)}{n} + (4C_1 + 2)\frac{\Delta_f(m)}{n_0} \tag{3.8}$$

with C_1 a constant and

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

The Fourier and Laguerre estimators have a similar structure, with here a cutoff required for safe inversion of the noise characteristic function. The structure of the upper bound (3.8) is also similar to (3.6) and involves a squared bias term $\|f - f_{m,\mathrm{Fo}}\|^2$, 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 known (see [17]) that, if f is a Gamma density $\gamma(p,\theta)$, then the bias is of order $\|f - f_{m,\mathrm{Fo}}\|^2 = O(m^{-2p+1})$ in the Fourier setting while it is exponentially decreasing in the Laguerre setting, namely of order $\|f - f_m\|^2 = O(m^{2(p-1)}\exp(-\rho m))$, with $\rho = \rho(\theta) > 0$. 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 [17]).

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 [17] that both variance terms $\Delta(m)/n$ and $\|\mathbf{G}_m^{-1}\|_F^2/n$ 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 variance upper bounds have order less than m^{2a}/n.

For the variance term due to unknown noise density, it is straightforward, in the Fourier setting, that $\Delta_f(m) \leq \Delta(m)$ and thus the estimation of g does not change the Fourier risk as soon as $n_0 \geq n$. This is simpler than in the Laguerre setting.

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 the choice of small m's (namely $m = c\log(n)$ for large enough constant c) is possible and makes 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 can be used for both $\mathbb{R}$- or $\mathbb{R}^+$-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 the 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 [18]) and choose the dimension of projection spaces m as the minimizer of a penalized criterion.

First, we consider the following sets of integers:

$$\widehat{\mathcal{M}} = \Big\{1 \le m \le C\,\lfloor n/\log n\rfloor \wedge \lfloor n_0/\log n_0\rfloor,\ m\log m\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2 \le n\wedge n_0\Big\},$$
$$\mathcal{M} = \Big\{1 \le m \le C\,\lfloor n/\log n\rfloor \wedge \lfloor n_0/\log n_0\rfloor,\ 4m\log m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \le n\wedge n_0\Big\}$$

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

$$\widehat{\mathrm{pen}}_1(m) := 2\kappa_1\mathfrak{C}(\|h\|_\infty \vee 1)\frac{\log n}{n}\left(\tau_m\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2 \wedge \|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\right),$$
$$\widehat{\mathrm{pen}}_2(m) := 8\kappa_2(\|g\|_\infty \vee 1)\log n_0\frac{m}{n_0}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2,$$

where we recall that $\tau_m = 2m$ or $c^\star\sqrt{m}$ if $\mathbb{E}(Z_1^{-1/2}) < +\infty$. Then we set the random penalty as

$$\widehat{\mathrm{pen}}(m) := \widehat{\mathrm{pen}}_1(m) + \widehat{\mathrm{pen}}_2(m). \tag{4.1}$$

We also define the deterministic counterparts

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

and set the deterministic penalty as

$$\mathrm{pen}(m) := \mathrm{pen}_1(m) + \mathrm{pen}_2(m), \tag{4.2}$$

where κ_1 and κ_2 are numerical constants, see our comment in Illustration Section of Supplementary Material. 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* (2.13) *and*

$$\widehat{m} = \operatorname*{argmin}_{m\in\widehat{\mathcal{M}}}\left\{-\|\tilde{f}_m\|^2 + \widehat{\mathrm{pen}}(m)\right\}$$

with $\widehat{\mathrm{pen}}$ defined by (4.1)*, then there exists a positive numerical constant κ_1 such that*

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

where C^{ad} is a numerical constant and C depends on $\|f\|$ and $\|g\|$, pen *is defined by* (4.2)*.*

This theorem gives an oracle inequality which establishes a nonasymptotic 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 $\widehat{\mathrm{pen}}_1$ and $\widehat{\mathrm{pen}}_2$, there are four quantities which deserve some explanations: κ_1, κ_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

$$\hat{h}_{D_1}(x) = \sum_{k=0}^{D_1-1} \hat{a}_k(Z)\varphi_k(x) \quad \text{with} \quad \hat{a}_k(Z) = (1/n)\sum_{i=1}^{n} \varphi_k(Z_i), \tag{4.3}$$

$$\hat{g}_{D_2}(x) = \sum_{k=0}^{D_2-1} \hat{a}_k(Y')\varphi_k(x) \quad \text{with} \quad \hat{a}_k(Y') = (1/n_0)\sum_{i=1}^{n_0} \varphi_k(Y'_i). \tag{4.4}$$

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 $\tilde{f}_{\widetilde{m}}$ be defined by* (2.13) *and*

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

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

$$\widetilde{\text{pen}}_1(m) := 4\kappa_1 \log n \mathfrak{C}(\|\hat{h}_{D_1}\|_\infty \vee 1)\big(\tau_m \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \wedge \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2\big)/n,$$

$$\widetilde{\text{pen}}_2(m) := 16\kappa_2(\|\hat{g}_{D_2}\|_\infty \vee 1)\log n_0 m \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2/n_0,$$

where $\hat{h}_{D_1}$ and $\hat{g}_{D_2}$ are given by (4.3) *and* (4.4)*, D_1 and D_2 satisfy*

$$\log n \le D_1 \le \|h\|_\infty n/(128\sqrt{2}\log^3 n), \qquad \log n_0 \le D_2 \le \|g\|_\infty n_0/(128\sqrt{2}\log^3 n_0).$$

Then there exist positive numerical constants κ_1 and κ_2 such that

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

where C^{ad} is a positive constant.

Note that the constraints 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. NUMERICAL ILLUSTRATION

The whole implementation is conducted using Matlab software. The integrated squared error $\|f - \tilde{f}_{\widetilde{m}}\|^2$ is computed via 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) $\mathbb{E}\|f - \tilde{f}_{\widetilde{m}}\|^2$ 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 $X \sim 0.4\gamma(2, 1/2) + 0.6\gamma(16, 1/4)$ and $c = \sqrt{2.96}$ on $I_f = [0, 5]$.
- ▷ 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 noise Y with the 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 $Y \sim \mathcal{E}(\lambda)$

$$[\mathbf{G}_m]_{i,j} = \lambda/(1+\lambda)\mathbf{1}_{i=j} - 2\lambda\frac{(\lambda-1)^{i-j-1}}{(\lambda+1)^{(i-j+1)}}\mathbf{1}_{j<i}. \tag{5.1}$$

▷ For $Y \sim \gamma(2, \mu)$

$$\begin{aligned}[\mathbf{G}_m]_{i,j} &= (\mu/(1+\mu))^2\mathbf{1}_{i-1=j} - 4\mu^2/(1+\mu))^3\mathbf{1}_{i=j} \\ &\quad + 4(i-j-\mu)\mu^2\frac{(\mu-1)^{i-j-2}}{(\mu+1)^{(i-j+2)}}\mathbf{1}_{j+1<i}. \end{aligned} \tag{5.2}$$

5.2. Practical Estimation Procedure

As in [17], 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 $\tilde\kappa_0 = 0.25$ in the penalty; more precisely, the case called "direct" hereafter relies on the estimator $\hat f^{(0)}_{\hat m}$ with $\hat f^{(0)}_m = \sum_{j=0}^{m-1} \hat a^{(0)}_j \varphi_j$, $\hat a^{(0)}_k = n^{-1}\sum_{i=1}^n \varphi_k(X_i)$ and

$$\widehat m_0 = \underset{m\in\{0,1,\dots,n\}}{\operatorname{argmin}} \Big\{ -\sum_{k=0}^{m-1} (\hat a^{(0)}_k)^2 + \frac{2\tilde\kappa_0 m}{n}\Big\}.$$

We choose the general $\tau_m = 2m$ instead of its improvement to allow comparison with the results obtained in [17].

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 (5.1) and (5.2) and we apply the procedure described in [17]. We compute the estimator as given by (2.6) and select

$$\widehat m_1 = \underset{m\|\mathbf{G}_m^{-1}\|^2_{\mathrm{op}} \le n/\log(n)}{\operatorname{argmin}} \Big\{ -\|\hat f_m\|^2 + \frac{\tilde\kappa_1}{n}\big(2m\|\mathbf{G}_m^{-1}\|^2_{\mathrm{op}} \wedge \log(n)(\|g\|_\infty \vee 1)\|\mathbf{G}_m^{-1}\|^2_{\mathrm{F}}\big)\Big\}.$$

We set $\tilde\kappa_1 = 0.03$ in the penalty for known noise density, this is the value calibrated in [17], 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 $\tilde f_{\widetilde m}$ with $\tilde f_m$ given by (2.9) and $\widetilde m$ given by $\widetilde m = \operatorname{argmin}_{m\in\widehat{\mathcal M}} \big\{ -\|\tilde f_m\|^2 + \widetilde{\mathrm{pen}}(m)\big\}$, with $\widetilde{\mathrm{pen}}(m)$ defined as in Corollary 4.2 with $\tau_m = 2m$. 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 κ_1 and κ_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 $\sqrt{n}/3$.

Table 1. Results after 200 iterations of simulation 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 parentheses; third and fourth lines, mean with std in parentheses 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 $1/(2\sqrt{2})$).

			Y Exponential			Y Gamma		
f		direct	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	MISE	1.0	4.0	6.7	2.7	7.3	7.5	7.2
Gamma	(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)

5.3. Simulation Results

As in [17], 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, which we do not have in practice, and they are computed on other samples than the MISE of model selection.

Table 2. Results after 200 iterations of simulation 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 parentheses; third and fourth lines, mean with std in parentheses 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 $1/(2\sqrt{2})$).

			Y Exponential			Y Gamma		
		direct	Known noise	Noise sample	Noise sample	Known noise	Noise sample	Noise sample
f				$n_0 = 400$	$n_0 = 2000$		$n_0 = 400$	$n_0 = 2000$
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	MISE	1.6	7.2	8.4	7.0	9.0	38.2	9.1
Gamma	(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)

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 in [8], 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 the best estimation proportional to φ_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.

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 $(Z_i)_{1\le i\le n}$ from model (2.1) are available, together with an independent sample $(Y_i')_{1\le i\le n_0}$ of the nuisance process Y. All quantities related to the common density g of the $(Y_i)_{1\le i\le n_0}$ and the $(Y_i')_{1\le i\le n_0}$ 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 in [20] and developed in [15] in the Fourier context, with the help of tools related to random matrices taken in [23]; 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 nonasymptotic setting.

There remain additional questions that may be worth answering. First, in [17] 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 that context. Moreover, our framework is mainly nonasymptotic, but if we are interested in asymptotics, the question of lower bounds may be studied.

7. PROOFS

7.1. Preliminary Results

7.1.1. Proof of Lemma 2.1. The proof is a particular case of a Lemma proved in Comte and Genon-Catalot (2017). From Askey and Wainger (1965), we have for $\nu = 4k+2$, and k large enough

$$|\varphi_k(x/2)| \le C \begin{cases} (a)\quad 1 & \text{if } 0 \le x \le 1/\nu, \\ (b)\quad (x\nu)^{-1/4} & \text{if } 1/\nu \le x \le \nu/2, \\ (c)\quad \nu^{-1/4}(\nu - x)^{-1/4} & \text{if } \nu/2 \le x \le \nu - \nu^{1/3}, \\ (d)\quad \nu^{-1/3} & \text{if } \nu - \nu^{1/3} \le x \le \nu + \nu^{1/3}, \\ (e)\quad \nu^{-1/4}(x-\nu)^{-1/4} e^{-\gamma_1 \nu^{-1/2}(x-\nu)^{3/2}} & \text{if } \nu + \nu^{1/3} \le x \le 3\nu/2, \\ (f)\quad e^{-\gamma_2 x} & \text{if } x \ge 3\nu/2, \end{cases}$$

where γ_1 and γ_2 are positive and fixed constants. From these estimates, we can prove

Lemma 7.1. *Assume that a random variable R has density f_R square-integrable on $\mathbb{R}^+$ and that $\mathbb{E}(R^{-1/2}) < +\infty$. For k large enough,*

$$\int_0^{+\infty} [\varphi_k(x)]^2 f_R(x)\, dx \le \frac{c}{\sqrt{k}},$$

where $c > 0$ is a constant depending on $\mathbb{E}(R^{-1/2})$.

The result of Lemma 2.1 follows from Lemma 7.1. □

Proof of Lemma 7.1. Hereafter, we write $x \lesssim y$ when there exists a constant C such that $x \le Cy$ and recall that $\nu = 4k+2$. We have six terms to compute to find the order of

$$\int_0^{+\infty} [\varphi_k(x)]^2 f_R(x)\, dx = (1/2) \int_0^{+\infty} [\varphi_k(u/2)]^2 f_R(u/2)\, du := \sum_{\ell=1}^{6} I_\ell :$$

(a) $I_1 \lesssim \frac{1}{2}\int_0^{1/\nu} f_R(u/2)\,du \lesssim \|f_R\|\nu^{-1/2} \lesssim \|f_R\|k^{-1/2}$,

(b) $I_2 \lesssim \nu^{-1/2}\int_{1/\nu}^{\nu/2} f_R(u/2)u^{-1/2}\,du \lesssim k^{-1/2}\mathbb{E}(R^{-1/2})$,

(c) $I_3 \lesssim \nu^{-1/2}\nu^{-1/6}\int_{\nu/2}^{\nu-\nu^{1/3}} f_R(u/2)\,du = o(1/\sqrt{k})$ as $\nu - u \geq \nu^{1/3}$,

(d) $I_4 \lesssim \nu^{-2/3}\int_{\nu-\nu^{1/3}}^{\nu+\nu^{1/3}} f_R(u/2)du = o(1/\sqrt{k})$,

(e) $I_5 \lesssim \nu^{-1/2}\int_{\nu+\nu^{1/3}}^{3\nu/2}(u-\nu)^{-1/2}f_R(u/2)\,du \lesssim \nu^{-1/2}\nu^{-1/6} = o(1/\sqrt{k})$
(exp is bounded by 1, $u-\nu \geq \nu^{1/3}$),

(f) $I_6 \lesssim e^{-\gamma_2(3\nu/2)} = o(1/\sqrt{k})$.

The result of Lemma 7.1 follows from these orders. □

7.1.2. Bounds on the spectral norm.

Proposition 7.2. *Let $\widehat{\mathbf{G}}_m$ be defined by Eq.* (2.10) *and $\|g\|_\infty < \infty$, $n_0 \in \mathbb{N}\setminus\{0\}$, then for all $t > 0$*

$$\mathbb{P}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\mathrm{op}} \geq t\big] \leq 2m\exp\bigg(-\frac{n_0t^2/4}{\|g\|_\infty m + (\sqrt{2}/3)mt}\bigg).$$

Corollary 7.3. *Under the Assumptions of Proposition* 7.2, *for all $q \geq 2$, it holds that*

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

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

Proof of Proposition 7.2. To get the announced result, we apply a Bernstein matrix inequality (see Theorem 8.2). Thus we write $\widehat{\mathbf{G}}_m$ as a sum of a sequence of independent matrices

$$\widehat{\mathbf{G}}_m = \frac{1}{n_0}\sum_{i=1}^{n_0}\mathbf{K}_m(Y_i'), \quad \mathbf{K}_m(Y_i') = \begin{cases} 2^{-1/2}\varphi_0(Y_i') & \text{if } i=j,\\ 2^{-1/2}\left(\varphi_{i-j}(Y_i') - \varphi_{i-j-1}(Y_i')\right) & \text{if } j<i,\\ 0 & \text{otherwise.}\end{cases}$$

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 $L(\mathbf{K}_m) = \|\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')]\|_{\mathrm{op}}/n_0$.

First using the equivalence between the spectral and trace norms

$$\mathbf{A} \in \mathbb{R}^{m\times m}, \qquad \frac{1}{\sqrt{m}}\|\mathbf{A}\|_{\mathrm{F}} \leq \|\mathbf{A}\|_{\mathrm{op}} \leq \|\mathbf{A}\|_{\mathrm{F}} \tag{7.1}$$

we have by Eq. (7.1) that $L(\mathbf{K}_m) \leq (1/n_0)\|\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')]\|_{\mathrm{F}}$, and using Eq. (2.4)

$$\|\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')]\|_{\mathrm{F}}^2 = \sum_{1\leq i,j\leq m}|[\mathbf{K}_m(Y_1')]_{i,j} - \mathbb{E}[\mathbf{K}_m(Y_1')]_{i,j}|^2$$

$$\leq \frac{1}{2} \sum_{1 \leq i \leq m} |\varphi_0(Y_1') - \mathbb{E}[\varphi_0(Y_1')]|^2$$
$$+ \frac{1}{2} \sum_{1 \leq j < i \leq m} |\varphi_{i-j}(Y_1') - \varphi_{i-j-1}(Y_1') - \mathbb{E}[\varphi_{i-j}(Y_1') - \varphi_{i-j-1}(Y_1')]|^2$$
$$\leq \frac{1}{2} m |e^{-Y_1'} - \mathbb{E}[e^{-Y_1'}]|^2 + \frac{1}{2} \sum_{1 \leq j < i \leq m} (4\sqrt{2})^2$$
$$\leq \frac{m}{2} + 4^2 \frac{m(m-1)}{2} = \frac{16m^2 - 16m + m}{2} \leq 8m^2.$$

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

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

By definition of the operator norm we have

$$\nu(\mathbf{S}_m) = \frac{1}{n_0^2} \sup_{\|\vec{x}\|_{2,m}=1} {}^t\vec{x} \sum_{i=1}^{n_0} \mathbb{E}\Big[\,{}^t(\mathbf{K}_m(Y_i') - \mathbb{E}[\mathbf{K}_m(Y_i')])(\mathbf{K}_m(Y_i') - \mathbb{E}[\mathbf{K}_m(Y_i')])\Big]\vec{x}$$
$$= \frac{1}{n_0} \sup_{\|\vec{x}\|_{2,m}=1} {}^t\vec{x}\, \mathbb{E}\Big[\,{}^t(\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')])(\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')])\Big]\vec{x}$$
$$= \frac{1}{n_0} \sup_{\|\vec{x}\|_{2,m}=1} \mathbb{E}\big\|\big(\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')]\big)\vec{x}\big\|_{2,m}^2.$$

This implies that, for ${}^t\vec{x} = (x_0, \ldots, x_{m-1})$ and by convention $\varphi_{-1} \equiv 0$,

$$\mathbb{E}_1 := \mathbb{E}\big\|\big(\mathbf{K}_m(Y_1') - \mathbb{E}[\mathbf{K}_m(Y_1')]\big)\vec{x}\big\|_{2,m}^2$$
$$= \frac{1}{2} \sum_{i=0}^{m-1} \mathbb{E}\bigg(\sum_{j=0}^{i} \Big(\varphi_{i-j}(Y_1') - \varphi_{i-j-1}(Y_1') - \mathbb{E}[\varphi_{i-j}(Y_1') - \varphi_{i-j-1}(Y_1')]\Big) x_j \bigg)^2$$
$$= \frac{1}{2} \sum_{i=0}^{m-1} \mathrm{Var}\Big[\sum_{j=0}^{i} \big(\varphi_{i-j}(Y_1') - \varphi_{i-j-1}(Y_1')\big) x_j \Big]$$
$$\leq \frac{1}{2} \sum_{i=0}^{m-1} \mathbb{E}\Big| \sum_{j=0}^{i} \big(\varphi_{i-j}(Y_1') - \varphi_{i-j-1}(Y_1')\big) x_j \Big|^2$$
$$= \frac{1}{2} \sum_{i=0}^{m-1} \int \Big| \sum_{j=0}^{i} \big(\varphi_{i-j}(u) - \varphi_{i-j-1}(u)\big) x_j \Big|^2 g(u)\, du.$$

Therefore

$$\mathbb{E}_1 \leq \frac{\|g\|_\infty}{2} \sum_{i=0}^{m-1} \int \Big| \sum_{j=0}^{i} \big(\varphi_{i-j}(u) - \varphi_{i-j-1}(u)\big) x_j \Big|^2 du$$
$$= \frac{\|g\|_\infty}{2} \sum_{i=0}^{m-1} \Big(2 \sum_{1 \leq j,j' \leq i} \delta_{j,j'} x_j x_{j'} - \sum_{1 \leq j,j' \leq i} \delta_{j,j'+1} x_j x_{j'+1} - \sum_{1 \leq j,j' \leq i} \delta_{j,j'-1} x_j x_{j'-1} \Big)$$
$$\leq 2\|g\|_\infty m \|x\|_{2,m}^2.$$

Then we get that $\nu(\mathbf{S}_m) \le \dfrac{2\|g\|_\infty m}{n_0}$. In the end applying Theorem 8.2 implies that for all $t > 0$

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

from which we get the result of Proposition 7.2. □

Proof of Corollary 7.3. Before proving the announced result, let us explain how Theorem 8.3 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 $\lambda_{\max}(\mathcal{H}(\mathbf{A})^2) = \|\mathbf{A}\|_{\mathrm{op}}^2$ and $\lambda_{\max}(\mathcal{H}(\mathbf{A})) = \|\mathbf{A}\|_{\mathrm{op}}$.

Under the Assumptions of Proposition 7.2, we can apply Theorem A.1 in [4] (see Theorem 8.3) stated for Hermitian matrices, using the above Paulsen dilation as follows. Let $\mathbf{Y}_i$ be rectangular matrices and set $\mathbf{A} = \sum_i \mathbf{Y}_i$, then, for $q \ge 2$ and $r \ge \max(q, 2\log m)$,

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

Thus we get that

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

Now we apply this result to

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

Using the notation of the proof of Proposition 7.2, we get for $q \ge 2$, $m \ge 2$ and $r = 2\log m$

$$\begin{aligned}
\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\mathrm{op}}^q\big] &\le 2^{q-1}\big(er\nu(\mathbf{S}_m)\big)^{q/2} + 2^{q-1}\big(erL(\mathbf{K}_m)\big)^q \\
&\le 2^{q-1}\Big(er\frac{\|g\|_\infty m}{n_0}\Big)^{q/2} + 2^{q-1}\Big(er\frac{2\sqrt{2}m}{n_0}\Big)^q \\
&\le 2^{q-1}e^{q/2}\|g\|_\infty^{q/2}\Big(2\log m\frac{m}{n_0}\Big)^{q/2} + 2^{2q-1+q/2}\Big(2\log m\frac{m}{n_0}\Big)^q \\
&\le \mathfrak{C}_{\mathfrak{q}}\Big(\log m\frac{m}{n_0}\Big)^{q/2} \vee \Big(\log m\frac{m}{n_0}\Big)^q
\end{aligned}$$

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

7.2. Proofs of Results of Section 3

Proof of Lemma 3.1. First let us define the set

$$\Delta_m = \left\{ \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq \sqrt{\frac{n_0}{m \log m}} \right\} \tag{7.2}$$

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

$$\begin{aligned}\mathbb{E}\big[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\big] &= \mathbb{E}\big[\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m^c} + \|\widehat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big] \\ &= \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbb{P}[\Delta_m^c] + \mathbb{E}\big[\|\widehat{\mathbf{G}}_m^{-1}(\mathbf{G}_m - \widehat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big]. \end{aligned} \tag{7.3}$$

This proof is inspired by the proof of Lemma 2.1 in [20] 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 $\mathbb{E}\big[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\big] \lesssim \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}$. Starting from Eq. (7.3) and using the set Δ_m, we have that

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

Moreover, applying Corollary 7.3 for $q = 2p$ yields

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

• Second case: $\|\mathbf{G}_m^{-1}\|_{\text{op}} < \frac{1}{2}\sqrt{\frac{n_0}{m \log m}}$.

We prove $\mathbb{E}\big[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\big] \lesssim \Big(\log m\|\mathbf{G}_m^{-1}\|_{\text{op}}^4\dfrac{m}{n_0}\Big)^p$. Starting from (7.3) again, we get

$$\mathbb{E}\big[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\big] \leq \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbb{P}[\Delta_m^c] + \|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p}\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big]. \tag{7.4}$$

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

First let us notice that

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

Moreover applying Corollary 7.3 for $q = 2p$ and $q = 4p$ with the set Δ_m, we get

$$\begin{aligned}&\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p}\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big] \\ &\leq 2^{2p-1}\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big] + 2^{2p-1}\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p}\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big] \\ &\leq 2^{2p-1}\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big] + 2^{2p-1}\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\mathbb{E}\big[\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}}^{4p}\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}}^{2p}\mathbf{1}_{\Delta_m}\big] \\ &\leq 2^{2p-1}\mathfrak{C}_{\mathfrak{2p}}\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\Big(\frac{m \log m}{n_0}\Big)^p + 2^{2p-1}\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\Big(\frac{n_0}{m \log m}\Big)^p\mathfrak{C}_{\mathfrak{4p}}\Big(\frac{m \log m}{n_0}\Big)^{2p} \\ &\leq 2^{2p-1}(\mathfrak{C}_{\mathfrak{2p}} + \mathfrak{C}_{\mathfrak{4p}})\|\mathbf{G}_m^{-1}\|_{\text{op}}^{2p}\Big(\frac{m \log m}{n_0}\Big)^p.\end{aligned} \tag{7.5}$$

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

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

Lemma 7.4. *For Δ_m defined by Eq.* (7.2) *and* $\|\mathbf{G}_m^{-1}\|_{\mathrm{op}} < \frac{1}{2}\sqrt{\frac{n_0}{m\log m}}$, *it holds that*

$$\mathbb{P}[\Delta_m^c] = \mathbb{P}\Bigg[\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}} > \sqrt{\frac{n_0}{m\log m}}\;\Bigg] \leq 2^{2p+1}\mathfrak{C}_{2\mathfrak{p}}\Big(\frac{m\log m}{n_0}\Big)^p\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^{2p}. \tag{7.6}$$

Finally starting from Eq. (7.4) and gathering Eqs. (7.5) and (7.6), we get that

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

In conclusion, Lemma 3.1 is proved with $\mathfrak{C}_{\mathrm{op},p} = 2^{2p+1}\mathfrak{C}_{2\mathfrak{p}} + 2^{2p}\mathfrak{C}_{4\mathfrak{p}} + 1$. □

Proof of Lemma 7.4. First invoke the triangular inequality

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

which implies that

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

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

$$\mathbb{P}\Bigg[\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}} > \sqrt{\frac{n_0}{m\log m}}\;\Bigg] \leq \mathbb{P}\big[\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\mathrm{op}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big].$$

Now let us rewrite this probability as

$$\begin{aligned}&\mathbb{P}\big[\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\mathrm{op}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big] \\ &\quad= \mathbb{P}\Big[\big\{\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\mathrm{op}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big\} \cap \Big\{\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} < \frac{1}{2}\Big\}\Big] \\ &\qquad+ \mathbb{P}\Big[\big\{\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\mathrm{op}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big\} \cap \Big\{\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} \geq \frac{1}{2}\Big\}\Big] \\ &\quad\leq \mathbb{P}\Big[\big\{\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\mathrm{op}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big\} \cap \Big\{\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} < \frac{1}{2}\Big\}\Big] \\ &\qquad+ \mathbb{P}\Big[\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} \geq \frac{1}{2}\Big].\end{aligned} \tag{7.7}$$

To control the second term on the right-hand side of Eq. (7.7), we apply Markov's inequality and Corollary 7.3 for $q = 2p$

$$\begin{aligned}\mathbb{P}\Big[\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} \geq \frac{1}{2}\Big] &\leq \mathbb{P}\Big[\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\|\widehat{\mathbf{G}}_m - \mathbf{G}_m\|_{\mathrm{op}} \geq \frac{1}{2}\Big] \\ &\leq 2^{2p}\mathfrak{C}_{2\mathfrak{p}}\Big(\frac{m\log m}{n_0}\Big)^p\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^{2p}.\end{aligned} \tag{7.8}$$

Next to control the first term on the right-hand side of Eq. (7.7), we apply Theorem 8.1 (with $\mathbf{A} = \mathbf{G}_m$ and $\mathbf{B} = \widehat{\mathbf{G}}_m - \mathbf{G}_m$), which yields

$$\begin{aligned}&\mathbb{P}\Big[\big\{\|\widehat{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1}\|_{\mathrm{op}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big\} \cap \Big\{\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} < \frac{1}{2}\Big\}\Big] \\ &\quad\leq \mathbb{P}\Bigg[\Bigg\{\frac{\|\widehat{\mathbf{G}}_m - \mathbf{G}_m\|_{\mathrm{op}}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2}{1 - \|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}}} > \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\Bigg\} \cap \Big\{\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m - \mathbf{G}_m)\|_{\mathrm{op}} < \frac{1}{2}\Big\}\Bigg] \\ &\quad\leq \mathbb{P}\Big[\|\widehat{\mathbf{G}}_m - \mathbf{G}_m\|_{\mathrm{op}} > \frac{1}{2}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^{-1}\Big].\end{aligned} \tag{7.9}$$

Applying again Markov's inequality along with Corollary 7.3 we get

$$\mathbb{P}\Big[\big\{\|\widehat{\mathbf{G}}_m^{-1}-\mathbf{G}_m^{-1}\|_{\mathrm{op}}>\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}\big\}\cap\Big\{\|\mathbf{G}_m^{-1}(\widehat{\mathbf{G}}_m-\mathbf{G}_m)\|_{\mathrm{op}}<\frac{1}{2}\Big\}\Big]\le 2^{2p}\mathfrak{C}_{2\mathfrak{p}}\Big(\frac{m\log m}{n_0}\Big)^p\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^{2p}.$$

So starting from Eq. (7.7) and gathering Eqs. (7.8) and (7.9) gives

$$\mathbb{P}\bigg[\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}>\sqrt{\frac{n_0}{m\log m}}\;\bigg]\le 2^{2p+1}\mathfrak{C}_{2\mathfrak{p}}\Big(\frac{m\log m}{n_0}\Big)^p\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^{2p}.$$

□

7.2.1. Useful corollary for the Frobenius norm.

Corollary 7.5. *Under the Assumptions of Lemma* 3.1*, we have*

$$\mathbb{E}\big[\|\mathbf{G}_m^{-1}-\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\big]\le 2\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2.$$

Proof. The proof mainly follows the lines of the proof of Lemma 3.1. With Δ_m defined by Eq. (7.2), we write

$$\begin{aligned}\mathbb{E}\big[\|\mathbf{G}_m^{-1}-\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\big]&=\mathbb{E}\Big[\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\mathbf{1}_{\Delta_m^c}+\|\widehat{\mathbf{G}}_m^{-1}(\mathbf{G}_m-\widehat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\mathbf{1}_{\Delta_m}\Big]\\&=\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\mathbb{P}[\Delta_m^c]+\mathbb{E}\Big[\|\widehat{\mathbf{G}}_m^{-1}(\mathbf{G}_m-\widehat{\mathbf{G}}_m)\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\mathbf{1}_{\Delta_m}\Big].\end{aligned}\tag{7.10}$$

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

$$\|\mathbf{AB}\|_{\mathrm{F}}\le\|\mathbf{A}\|_{\mathrm{F}}\|\mathbf{B}\|_{\mathrm{op}}\qquad\text{and}\qquad\|\mathbf{AB}\|_{\mathrm{F}}\le\|\mathbf{A}\|_{\mathrm{op}}\|\mathbf{B}\|_{\mathrm{F}}.\tag{7.11}$$

Then Eqs. (7.10), (7.11), the definition of Δ_m and Lemma 3.1 for $q=2$ give

$$\begin{aligned}\mathbb{E}\big[\|\mathbf{G}_m^{-1}-\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\big]&\le\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2+\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\mathbb{E}\big[\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\|\mathbf{G}_m-\widehat{\mathbf{G}}_m\|_{\mathrm{op}}^2\mathbf{1}_{\Delta_m}\big]\\&\le\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2+\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\frac{n_0}{m\log m}\mathbb{E}\big[\|\mathbf{G}_m-\widehat{\mathbf{G}}_m\|_{\mathrm{op}}^2\big]\\&\le\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2+\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\frac{m\log m}{n_0}\frac{n_0}{m\log m}=2\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2.\end{aligned}$$

□

Proof of Corollary 3.2. The proof follows the lines of the proof of Lemma 3.1. The only difference lies in the following equation

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

with Δ_m defined by Eq. (7.2). It yields the following upper bound

$$\mathbb{E}\big[\|(\mathbf{G}_m^{-1}-\widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2\big]\le\|\vec{f}_m\|_{2,m}^2\mathbb{P}[\Delta_m^c]+\|\vec{f}_m\|_{2,m}^2\mathbb{E}\big[\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\|\mathbf{G}_m-\widehat{\mathbf{G}}_m\|_{\mathrm{op}}^2\mathbf{1}_{\Delta_m}\big].$$

And following the proof of Lemma 3.1, we get

$$\mathbb{E}\big[\|(\mathbf{G}_m^{-1}-\widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2\big]\le\|f\|^2\mathfrak{C}_{\mathrm{op}}\Big(1\wedge\log m\frac{m}{n_0}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2\Big).$$

□

Proof of Proposition 3.3. 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:

$$\begin{aligned}\|f_m-\tilde{f}_m\|^2=\|\vec{f}_m-\tilde{\vec{f}}_m\|_{2,m}^2&=\|\mathbf{G}_m^{-1}\vec{h}_m-\widetilde{\mathbf{G}}_m^{-1}\hat{\vec{h}}_m\|_{2,m}^2\\&\le 2\|\mathbf{G}_m^{-1}\vec{h}_m-\mathbf{G}_m^{-1}\hat{\vec{h}}_m\|_{2,m}^2+2\|\mathbf{G}_m^{-1}\hat{\vec{h}}_m-\widetilde{\mathbf{G}}_m^{-1}\hat{\vec{h}}_m\|_{2,m}^2.\end{aligned}\tag{7.12}$$

(i) Then according to Proposition 3.1 in [17] ($\tau_m = 2m$) and Lemma 2.1 ($\tau_m = c^\star\sqrt{m}$ under $\mathbb{E}(1/\sqrt{Z_1}) < +\infty$), we get

$$\mathbb{E}\|\mathbf{G}_m^{-1}(\vec{h}_m - \hat{\vec{h}}_m)\|_{2,m}^2 \leq \frac{\tau_m}{n}\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \frac{\|h\|_\infty}{n}\|\mathbf{G}_m^{-1}\|_{\text{F}}^2, \tag{7.13}$$

where τ_m is defined in Proposition 3.3.

(ii) Now we turn to the second term on the right-hand side of Eq. (7.12). Let us notice that

$$\begin{aligned}\|\mathbf{G}_m^{-1}\hat{\vec{h}}_m - \widetilde{\mathbf{G}}_m^{-1}\hat{\vec{h}}_m\|_{2,m}^2 &= \|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m) + (\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \\ &\leq 2\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 + 2\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2. \end{aligned} \tag{7.14}$$

(a) The first term of (7.14) can be bounded in two ways: since $(Y_1', \ldots, Y_{n_0}')$ are independent of $(Z_1, \ldots, Z_n)$, we get that

$$\mathbb{E}\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \leq \mathbb{E}\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2\mathbb{E}\|\hat{\vec{h}}_m - \vec{h}_m\|_{2,m}^2. \tag{7.15}$$

Again according to Proposition 3.1 in [17] and Lemma 2.1,

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

Applying Lemma 3.1 gives that

$$\mathbb{E}\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \leq \frac{\tau_m}{n}\mathfrak{C}_{\text{op},1}\|\mathbf{G}_m^{-1}\|_{\text{op}}^2. \tag{7.16}$$

(b) Under the assumption that $(Y_1', \ldots, Y_{n_0}')$ are independent of $(Z_1, \ldots, Z_n)$ and Proposition 3.1 in [17], we obtain

$$\mathbb{E}\big[\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2\big] \leq \mathbb{E}\big[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2\big]\frac{\|h\|_\infty}{n}.$$

And applying Corollary 7.5

$$\mathbb{E}\big[\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2\big] \leq 2\|\mathbf{G}_m^{-1}\|_{\text{F}}^2\frac{\|h\|_\infty}{n}. \tag{7.17}$$

For the second term of (7.14), we have according to Corollary 3.2

$$\mathbb{E}\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2 \leq \mathfrak{C}_{\text{E}}\log m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2\frac{m}{n_0}. \tag{7.18}$$

Finally starting from Eq. (7.12) and gathering Eqs. (7.13), (7.15), (7.16), (7.17) and (7.18) yields

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

To conclude,

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

□

Proof of Proposition 3.5. For $f \in W^s(\mathbb{R}^+, L)$ defined by (2.6), 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 3.4 we have $\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2 \asymp \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \asymp m^{2r}$. This implies that the MISE is upper bounded as follows

$$\mathbb{E}\|f - \tilde{f}_m\|^2 \leq Lm^{-s} + 2C\Big(\frac{\tau_m}{n}m^{2r} \wedge \frac{\|h\|_\infty}{n}m^{2r}\Big) + 2C\mathfrak{C}\log(m)\frac{m^{2r+1}}{n_0}. \tag{7.19}$$

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

$$Lm^{-s} + 2C(2 + \|h\|_\infty)\frac{m^{2r}}{n} \;\Rightarrow\; m_{opt_1} \propto n^{1/s+2r},$$
$$Lm^{-s} + 2C\mathfrak{C}\log(m)\frac{m^{2r+1}}{n_0} \;\Rightarrow\; m_{opt_2} \propto (n_0/\log(n_0))^{1/s+2r+1}.$$

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

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

which completes the proof of Proposition 3.5. □

7.3. Proof of Theorem 4.1

First for $m \in \mathcal{M}$, let us define the associated subspaces $\mathcal{S}_{d_1}^m \subseteq \mathbb{R}^{d_1}$:

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

These subspaces are 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

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

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 $\widetilde{\mathbf{G}}_{d_1}$ and $\widetilde{\mathbf{G}}_m$, we have

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

So we clearly have that

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

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

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

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

$$\|\tilde{f}_m - f\|^2 = \|\tilde{\vec{f}}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty}(a_j(f))^2 \quad \text{and} \quad \|f_m - f\|^2 = \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty}(a_j(f))^2. \tag{7.20}$$

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

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

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

Hence

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

which implies

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

Let us notice that

$$\nu_n(\tilde{\vec{f}}_{\widehat{m}} - \vec{f}_m) = \|\tilde{\vec{f}}_{\widehat{m}} - \vec{f}_m\|_{2,d_1}\nu_n\left(\frac{\tilde{\vec{f}}_{\widehat{m}} - \vec{f}_m}{\|\tilde{\vec{f}}_{\widehat{m}} - \vec{f}_m\|_{2,d_1}}\right)$$

and due to the relation $2ab \leq a^2/4 + 4b^2$, we have the following inequalities

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

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

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

We then have

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

which implies

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

Using Eq. (7.20) we have

$$\begin{aligned}\|\hat{f}_{\widehat{m}} - f\|^2 - \sum_{j=d_1}^{\infty}(a_j(f))^2 \leq 3\Big(\|f - f_m\|^2 - \sum_{j=d_1}^{\infty}(a_j(f))^2\Big) + 2\widehat{\text{pen}}(m) \\ + 8 \sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} \nu_n^2(\vec{t}) - 2\widehat{\text{pen}}(\widehat{m}).\end{aligned} \tag{7.21}$$

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

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

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

$$\xi^2_{1,n}(\vec{t}) = |\langle\vec{t}_{m^*}, \widetilde{\mathbf{G}}^{-1}_{d_1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}|^2, \qquad \widehat{p}_1(m,m') = 2\widehat{\text{pen}}_1(m\vee m'), \tag{7.22}$$

$$\xi^2_{2,n}(\vec{t}) = |\langle\vec{t}_{m^*}, (\widetilde{\mathbf{G}}^{-1}_{d_1} - \mathbf{G}^{-1}_{d_1})\vec{h}_{d_1}\rangle_{2,d_1}|^2, \qquad \widehat{p}_2(m,m') = 2\widehat{\text{pen}}_2(m\vee m'). \tag{7.23}$$

Let us notice that

$$\begin{aligned}&\Big[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} \nu_n^2(\vec{t}) - p(m,\widehat{m})\Big]_+ \\ &\leq \Big[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} \big|\langle\vec{t}_{m^*}, \widetilde{\mathbf{G}}^{-1}_{d_1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1}) + (\widetilde{\mathbf{G}}^{-1}_{d_1} - \mathbf{G}^{-1}_{d_1})\vec{h}_{d_1}\rangle_{2,d_1}\big|^2 - \widehat{p}_1(m,\widehat{m}) - \widehat{p}_2(m,\widehat{m})\Big]_+\end{aligned}$$

$$\leq 2\Big[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})}\xi_{1,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_1(m,\widehat{m})\Big]_+ + 2\Big[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})}\xi_{2,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_2(m,\widehat{m})\Big]_+,$$

which implies that

$$\|\tilde{f}_{\widehat{m}} - f\|^2 \leq 3\|f - f_m\|^2 + 4\widehat{\mathrm{pen}}(m) + 16\sum_{m'\in\widehat{\mathcal{M}}}\Big[\sup_{\vec{t}\in\mathcal{B}(m,m')}\xi_{1,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_1(m,m')\Big]_+$$
$$+ 16\Big[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})}\xi_{2,n}^2(\vec{t}) - \frac{1}{2}\widehat{p}_2(m,\widehat{m})\Big]_+.$$

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

Proposition 7.6. *For $m \in \mathcal{M}$, it holds that*

$$\mathbb{E}[\widehat{\mathrm{pen}}(m)] \leq C\mathrm{pen}(m) \quad \textit{with} \quad C = (2 + 2(\mathfrak{C}_{\mathrm{op}} \vee 2)).$$

Proposition 7.7. *Under the assumptions of Theorem 4.1, there exists a constant $C_1 > 0$ depending on $\|h\|_\infty$ such that for $\hat{p}_1(m,m') = 2\widehat{\mathrm{pen}}_1(m \vee m')$*

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

Proposition 7.8. *Under the assumptions of Theorem 4.1, there exists a constant $C_2 > 0$ depending on $\|h\|_\infty$ such that for $\hat{p}_2(m,m') = 2\widehat{\mathrm{pen}}_2(m \vee m')$*

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

In the end,

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

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

Proof of Proposition 7.6. Let m be in the model collection $\mathcal{M}$. By definition we have

$$\mathbb{E}[\widehat{\mathrm{pen}}(m)] = \mathbb{E}\big[\widehat{\mathrm{pen}}_1(m) + \widehat{\mathrm{pen}}_2(m)\big]$$
$$= 2\mathfrak{C}\kappa_1\log n\,\mathbb{E}\Big[\frac{\tau_m\|h\|_\infty}{n}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2 \wedge \frac{(\|h\|_\infty \vee 1)}{n}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\Big]$$
$$+ 8\kappa_2\mathfrak{C}_{\mathrm{E}}(\|g\|_\infty \vee 1)\frac{m}{n_0}\log n_0\,\mathbb{E}\big[\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\big].$$

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

$$\mathbb{E}\big[\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\big] \leq 2\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 + 2\mathbb{E}\big[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\big] \leq 2\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 + 2\mathfrak{C}_{\mathrm{op},1}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2.$$

Similarly, applying now Corollary 7.5, we get that $\mathbb{E}\big[\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\big] \leq 2\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2 + 4\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2$. Finally

$$\mathbb{E}[\widehat{\mathrm{pen}}(m)] \leq (2 + 2(\mathfrak{C}_{\mathrm{op},1} \vee 2))\mathrm{pen}(m).$$

□

Proof of Proposition 7.7. First let us notice

$$\Big\{\sup_{\vec{t}\in\mathcal{B}(m,m')}\big|\langle\vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+$$

$$= \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \mathbf{1}_{m'>m}$$

$$+ \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \mathbf{1}_{m'\le m}\mathbf{1}_{\Delta_m}$$

$$+ \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \mathbf{1}_{m'\le m}\mathbf{1}_{\Delta_m^c}$$

$$= \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \mathbf{1}_{m'>m}$$

$$+ \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \mathbf{1}_{m'\le m}\mathbf{1}_{\Delta_m}.$$

Since $\Delta_{m'} \subset \widehat{\mathcal{M}}$ and $\Delta_m \subset \widehat{\mathcal{M}}$ for $m, m' \in \widehat{\mathcal{M}}$, we have

$$\Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+$$
$$= \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widehat{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \mathbf{1}_{\Delta_{m^*}}.$$

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

$$\Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widetilde{\mathbf{G}}_{d_1}^{-1}(\widehat{\vec{h}}_{d_1} - \vec{h}_{d_1})\rangle_{2,d_1}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+$$
$$= \Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widehat{\mathbf{G}}_{m^*}^{-1}(\widehat{\vec{h}}_{m^*} - \vec{h}_{m^*})\rangle_{2,m^*}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+.$$

Now, if we define E_1

$$E_1 = \mathbb{E}\Big[\Big\{ \sup_{\vec{t}\in\mathcal{B}(m,m')} \big|\langle \vec{t}_{m^*}, \widehat{\mathbf{G}}_{m^*}^{-1}(\widehat{\vec{h}}_{m^*} - \vec{h}_{m^*})\rangle_{2,m^*}\big|^2 - \frac{1}{2}\hat{p}_1(m,m')\Big\}_+ \Big| Y'\Big], \tag{7.24}$$

then, conditionally on Y', the bound follows from the proof of Proposition 7.1 in [17] 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 $\xi^2 = \mathfrak{a}\|h\|_\infty/K_1 \log n$ with $K_1 = 1/6$ (to avoid Assumption (A2)). Note also that the proof remains valid for $2m$ replaced by τ_m. Then, as all bounds are independent of the random terms, the conditional expectation can be integrated with respect to the law of the sample $(Y_i')_{1\le i\le n_0}$ without change. □

Proof of Proposition 7.8. Let us define

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

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

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

$$E_2\mathbf{1}_{\widehat{m}\ge m} = \Big[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} \big|\langle \vec{t}_{\widehat{m}}, (\widetilde{\mathbf{G}}_{\widehat{m}}^{-1} - \mathbf{G}_{\widehat{m}}^{-1})\vec{h}_{\widehat{m}}\rangle\big|^2 - \frac{1}{2}\widehat{\mathrm{pen}}_2(\widehat{m})\Big]_+ \mathbf{1}_{\widehat{m}\ge m}$$
$$\le \Big[\|(\widehat{\mathbf{G}}_{\widehat{m}}^{-1} - \mathbf{G}_{\widehat{m}}^{-1})\vec{h}_{\widehat{m}}\|_{2,\widehat{m}}^2 - \frac{1}{2}\widehat{\mathrm{pen}}_2(\widehat{m})\Big]_+$$
$$\le \Big[\|\widehat{\mathbf{G}}_{\widehat{m}}^{-1}(\mathbf{G}_{\widehat{m}} - \widehat{\mathbf{G}}_{\widehat{m}})\mathbf{G}_{\widehat{m}}^{-1}\vec{h}_{\widehat{m}}\|_{2,\widehat{m}}^2 - \frac{1}{2}\widehat{\mathrm{pen}}_2(\widehat{m})\Big]_+$$

$$\leq \Big[\|f\|^2\|\widehat{\mathbf{G}}_{\widehat{m}}^{-1}\|_{\text{op}}^2\|\mathbf{G}_{\widehat{m}} - \widehat{\mathbf{G}}_{\widehat{m}}\|_{\text{op}}^2 - \frac{1}{2}\widehat{\text{pen}}_2(\widehat{m})\Big]_+.$$

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

$$\mathcal{M}_{\max} = \big\{m \in [\![1, n]\!], m \leq C\lfloor n/\log n\rfloor \wedge \lfloor n_0/\log n_0\rfloor\big\}. \tag{7.25}$$

We now introduce the favorable set

$$\mathcal{E}_m = \Big\{\|\mathbf{G}_m - \widehat{\mathbf{G}}_m\|_{\text{op}} \leq \sqrt{\kappa_2 4(\|g\|_\infty \vee 1)\log n_0 \frac{m}{n_0}}\ \Big\}, \quad \kappa_2 > 0, \tag{7.26}$$

and set

$$\mathcal{E} = \bigcap_{m\in\mathcal{M}_{\max}} \mathcal{E}_m. \tag{7.27}$$

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

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

On the complementary set we have that

$$\begin{aligned} \mathbb{E}\big[E_2\mathbf{1}_{\widehat{m}\geq m}\mathbf{1}_{\mathcal{E}^c}\big] &\leq \mathbb{E}\big[\|(\widetilde{\mathbf{G}}_{\widehat{m}}^{-1} - \mathbf{G}_{\widehat{m}}^{-1})\vec{h}_{\widehat{m}}\|_{2,\widehat{m}}^2\mathbf{1}_{\mathcal{E}^c}\big] \leq \mathbb{E}\Big[\sup_{m\in\mathcal{M}_{\max}}\|(\widetilde{\mathbf{G}}_m^{-1} - \mathbf{G}_m^{-1})\vec{h}_m\|_{2,m}^2\mathbf{1}_{\mathcal{E}^c}\Big] \\ &\leq \sum_{m\in\mathcal{M}_{\max}} 2\mathbb{E}\big[(\|\widetilde{\mathbf{G}}_m^{-1}\vec{h}_m\|_{2,m}^2 + \|\mathbf{G}_m^{-1}\vec{h}_m\|_{2,m}^2)\mathbf{1}_{\mathcal{E}^c}\big] \\ &\leq \sum_{m\in\mathcal{M}_{\max}} 2\mathbb{E}\big[(\|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2\|\vec{h}_m\|_{2,m}^2 + \|\vec{f}_m\|_{2,m}^2)\mathbf{1}_{\mathcal{E}^c}\big] \\ &\leq \sum_{m\in\mathcal{M}_{\max}} 2\mathbb{E}\big[(\|\vec{h}_m\|_{2,m}^2 + \|\vec{f}_m\|_{2,m}^2)n_0\mathbf{1}_{\mathcal{E}^c}\big] \leq Cn_0|\mathcal{M}_{\max}|\mathbb{P}[\mathcal{E}^c] \end{aligned}$$

and apply the following Lemma for $p = 3$.

Lemma 7.9. *For any $p \geq 1$ there exist $\kappa_2 \geq (p+2)/2$ and $C_p \geq 1$ such that $\mathbb{P}[\mathcal{E}^c] \leq \dfrac{C_p}{n_0^p}$.*

We obtain $\mathbb{E}[E_2\mathbf{1}_{\widehat{m}\geq m}\mathbf{1}_{\mathcal{E}^c}] \leq \dfrac{C_3}{n_0}$.

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

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

It implies that for $\mathcal{E}_m$ defined by (7.26)

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

According to Lemma 6.3,

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

and

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

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

$$\begin{aligned}\mathbb{E}[E_2\mathbf{1}_{\widehat{m}\leq m}\mathbf{1}_{\mathcal{E}_m^c}]&\leq\mathbb{E}\big[\|(\widetilde{\mathbf{G}}_m^{-1}-\mathbf{G}_m^{-1})\vec{h}_m\|_{2,m}^2\mathbf{1}_{\mathcal{E}_m^c}\big]\leq 2\mathbb{E}\big[(\|\widetilde{\mathbf{G}}_m^{-1}\vec{h}_m\|_{2,m}^2+\|\mathbf{G}_m^{-1}\vec{h}_m\|_{2,m}^2)\mathbf{1}_{\mathcal{E}_m^c}\big]\\&\leq 2\mathbb{E}\big[(\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\|\vec{h}_m\|_{2,m}^2+\|\vec{f}_m\|_{2,m}^2)\mathbf{1}_{\mathcal{E}_m^c}\big]\leq 2\mathbb{E}\big[(\|\vec{h}_m\|_{2,m}^2+\|\vec{f}_m\|_{2,m}^2)n_0\mathbf{1}_{\mathcal{E}_m^c}\big]\\&\leq Cn_0\mathbb{P}[\mathcal{E}_m^c].\end{aligned}$$

Moreover $\mathcal{E}\subset\mathcal{E}_m$, which implies that $\mathbb{P}[\mathcal{E}_m^c]\leq\mathbb{P}[\mathcal{E}^c]$. Then applying Lemma 7.9, we get that

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

□

Proof of Lemma 7.9. We apply Proposition 6.1 for $t=\sqrt{4\kappa_2(\|g\|_\infty\vee 1)\log n_0\dfrac{m}{n_0}}$ to obtain

$$\begin{aligned}\mathbb{P}[\mathcal{E}^c]&=\mathbb{P}\left[\exists m\in\mathcal{M}_{\max},\ \|\mathbf{G}_m-\widehat{\mathbf{G}}_m\|_{\mathrm{op}}>\sqrt{4\kappa_2(\|g\|_\infty\vee 1)\log n_0\frac{m}{n_0}}\ \right]\\&\leq\sum_{m\leq n_0}\mathbb{P}\left[\|\mathbf{G}_m-\widehat{\mathbf{G}}_m\|_{\mathrm{op}}>\sqrt{4\kappa_2(\|g\|_\infty\vee 1)\log n_0\frac{m}{n_0}}\ \right]\\&\leq 2\sum_{m\leq n_0}m\exp\left(-\frac{1}{2}\frac{4\kappa_2(\|g\|_\infty\vee 1)\log n_0 m}{\|g\|_\infty m+(2\sqrt{2}/3)m\sqrt{4\kappa_2(\|g\|_\infty\vee 1)\log n_0}\sqrt{\frac{m}{n_0}}}\right)\\&\leq 2\sum_{m\leq n_0}m\exp\left(-\frac{4\kappa_2(\|g\|_\infty\vee 1)\log n_0}{2}\left(\frac{1}{\|g\|_\infty}\wedge\frac{3}{2\sqrt{2}\sqrt{4\kappa_2(\|g\|_\infty\vee 1)\log n_0}}\sqrt{\frac{n_0}{m}}\ \right)\right)\\&\leq C\sum_{m\leq n_0}m\exp\left(-\frac{4\kappa_2(\|g\|_\infty\vee 1)\log n_0}{2\|g\|_\infty}\right)\leq C\sum_{m\leq n_0}me^{-2\kappa_2\log n_0}\leq Cn_0^2e^{-2\kappa_2\log n_0}.\end{aligned}$$

Finally we get $\mathbb{P}[\mathcal{E}^c]\leq Cn_0^2\exp(-\kappa_2 2\log n_0)=C/n_0^{2\kappa_2-2}=C/n_0^p$ with $p\geq 1$ if $\kappa_2\geq(p+2)/2$. □

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

Starting from Eq. (7.21), we get

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

We now apply Proposition 7.10 hereafter and we get the final result. □

Proposition 7.10. (i) $\mathbb{E}|\widehat{\mathrm{pen}}(m)-\widetilde{\mathrm{pen}}(m)|\lesssim\mathrm{pen}(m)+1/n_0+1/n$.

(ii) $\mathbb{E}\big(\widehat{\mathrm{pen}}(\widetilde{m})-\widetilde{\mathrm{pen}}(\widetilde{m})\big)\lesssim 1/n_0+1/n$.

Proof of Proposition 7.10. The proof relies on introducing the set such 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 $\Lambda(g) = \left\{ |\|\hat{g}_D\|_\infty - \|g\|_\infty| \leq \frac{\|g\|_\infty}{2} \right\}$.

(i) It yields

$$\mathbb{E}|\widehat{\text{pen}}_2(m) - \widetilde{\text{pen}}_2(m)|\mathbf{1}_{\Lambda(g)} = 8\kappa_2 \mathbb{E}\Big[\big|(2\|\hat{g}_D\|_\infty \vee 1) - (\|g\|_\infty \vee 1)\big| \log n_0 \frac{m}{n_0} \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \mathbf{1}_{\Lambda(g)}\Big]$$
$$\leq 8\kappa_2 \mathbb{E}\Big[4(\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0} \|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \mathbf{1}_{\Lambda(g)}\Big].$$

Moreover applying Proposition 7.6, 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

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

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

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

Now applying Lemma 5.2 in [16], it holds that for all $p > 0$ and

$$\log n_0 \leq D \leq \|g\|_\infty/(128\sqrt{2})n_0/(\log n_0)^p,$$

we get $\mathbb{P}[\Lambda^c(g)] \leq 2D/n_0^p$.

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 $\|g\|_\infty - 2\|\hat{g}_D\|_\infty \leq 0$ which implies that $(\|g\|_\infty \vee 1) - 2(\|\hat{g}_D\|_\infty \vee 1) \leq 0$, thus $(\widetilde{\text{pen}}_2(\tilde{m}) - \widehat{\text{pen}}_2(\tilde{m}))\mathbf{1}_{\Lambda(g)} \leq 0$. Moreover

$$\mathbb{E}\big[\big(\widetilde{\text{pen}}_2(\widetilde{m}) - \widehat{\text{pen}}_2(\widetilde{m})\big)\mathbf{1}_{\Lambda^c(g)}\big] \leq \mathbb{E}\big[\,\big|\widehat{\text{pen}}_2(\widetilde{m}) - \widetilde{\text{pen}}_2(\widetilde{m})\big|\mathbf{1}_{\Lambda^c(g)}\big] \leq C\mathbb{P}[\Lambda^c(g)], \tag{7.28}$$

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

8. USEFUL RESULTS

A proof of the following theorem can be found in [21].

Theorem 8.1. *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 8.2 (Bernstein matrix inequality)**.** *Consider a finite sequence* $\{\mathbf{S}_k\}$ *of independent random matrices with common dimension* $d_1 \times d_2$*. Assume that*

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

Introduce the random matrix $\mathbf{Z} = \sum_k \mathbf{S}_k$*. Let* $\nu(\mathbf{Z})$ *be the variance statistic of the sum:* $\nu(\mathbf{Z}) = \max\left\{\lambda_{\max}(\mathbb{E}[\mathbf{Z}\,{}^t\mathbf{Z}]), \lambda_{\max}(\mathbb{E}[{}^t\mathbf{Z}\mathbf{Z}])\right\}$*. 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 $t \geq 0$

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

A proof can be found in [22] or [23].

Theorem 8.3 (Matrix moment inequality, Theorem A.1 in [4])**.** *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*

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

A proof can be found in [4].

ACKNOWLEDGMENTS

Very instructive discussions with Jean Rochet are gratefully acknowledged. G. Mabon acknowledges support by CREST-ENSAE and the DFG via Research Unit 1735 *Structural Inference in Statistics*.

REFERENCES

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, (Dover, New York, 1964).
2. D. Belomestny, F. Comte, and V. Genon-Catalot, “Nonparametric Laguerre Estimation in the Multiplicative Censoring Model”, Electron. J. Statist. **10** (2), 3114–3152 (2016).
3. B. Bongioanni and J. L. Torrea, “What is a Sobolev Space for the Laguerre Function Systems?”, Studia Math. **192** (2):147–172 (2009).
4. R. Y. Chen, A. Gittens, and J. A. Tropp, “The Masked Sample Covariance Estimator: An Analysis Using Matrix Concentration Inequalities”, Inf. Inference **1** (1), 2–20 (2012).
5. F. Comte, C. A. Cuenod, M. Pensky, and Y. Rozenholc, “Laplace Deconvolution and Its Application to Dynamic Contrast Enhanced Imaging”, J. Roy. Statist. Soc., Ser. B (2015).
6. F. Comte and V. Genon-Catalot, Adaptive Laguerre Density Estimation for Mixed Poisson Models. Electron. J. Statist. **9**, 1112–1148 (2015).
7. F. Comte and V. Genon-Catalot, *Laguerre and Hermite bases for inverse problems*, Preprint hal-01449799, V2 (2017). `https://hal.archives-ouvertes.fr/hal-01449799`.
8. F. Comte and C. Lacour, “Data-Driven Density Estimation in the Presence of Additive Noise with Unknown Distribution”, J. Roy. Statist. Soc., Ser. B **73**, 601–627 (2011).
9. I. Dattner, M. Reiß, and M. Trabs, “Adaptive Quantile Estimation in Deconvolution with Unknown Error Distribution”, Bernoulli **22** (2), 143–192 (2016).
10. P. Groeneboom and G. Jongbloed, “Density Estimation in the Uniform Deconvolution Model”, Statist. Neerl. **57** (1), 136–157 (2003).
11. P. Groeneboom and J. A. Wellner, *Information Bounds and Nonparametric Maximum Likelihood Estimation* in *DMV Seminar* (Birkhäuser, Basel, 1992), vol. 19.
12. R. Horn and C. Johnson, *Matrix Analysis* (Cambridge University Press, 1990).
13. J. Johannes, “Deconvolution with Unknown Error Distribution”, Ann. Statist. **37** (5a), 2301–2323 (2009).
14. J. Johannes and M. Schwarz, “Adaptive Circular Deconvolution by Model Selection under Unknown Error Distribution”, Bernoulli **19** (5A), 1576–1611 (2013).
15. J. Kappus and G. Mabon, “Adaptive Density Estimation in Deconvolution Problems with Unknown Error Distribution”, Electron. J. Statist. **8** (2), 2879–2904 (2014).
16. G. Mabon, “Adaptive Deconvolution of Linear Functionals on the Nonnegative Real Line”, J. Statist. Plann. Inference **178**, 1–23 (2016).
17. G. Mabon, “Adaptive Deconvolution on the Nonnegative Real Line”, Scandinavian J. Statist. **44**, 707–740 (2017).
18. P. Massart, *Concentration Inequalities and Model Selection*, in *Lecture Notes in Mathematics*, Vol. 1896: *Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003* (Springer, Berlin, 2003).

19. A. Meister, *Deconvolution Problems in Nonparametric Statistics*, in *Lecture Notes in Statistics* (Springer, Berlin, 2009), Vol. 193.
20. M. H. Neumann, "On the Effect of Estimating the Error Density in Nonparametric Deconvolution", J. Nonparametric Statist. **7** (4), 307–330 (1997).
21. G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory* (Academic Press, Boston etc., 1990).
22. J. A. Tropp, "User-Friendly Tail Bounds for Sums of Random Matrices", Found. Comput. Math. **12** (4), 389–434 (2012).
23. J. A. Tropp, "An Introduction to Matrix Concentration Inequalities", Found. Trends Mach. Learn. **8** (1–2), 1–230 (2015).
24. B. van Es, "Combining Kernel Estimators in the Uniform Deconvolution Problem", Statist. Neerl. **65** (3), 275–296 (2011).
25. T. Vareschi, "Noisy Laplace Deconvolution with Error in the Operator", J. Statist. Plann. Inference **157–158**, 16–35 (2015).