Noisy Laplace deconvolution with error in the operator

Thomas Vareschi

Université Denis Diderot Paris 7, Bâtiment Sophie-Germain, rue Alice-Domon et Léonie-Duquet, 75013 Paris, France

E-mail: thomas.vareschi@univ-paris-diderot.fr

Abstract. We adress the problem of Laplace deconvolution with random noise in a regression framework. The time set is not considered to be fixed, but grows with the number of observation points. Moreover, the convolution kernel is unknown, and accessible only through experimental noise. We make use of a recent procedure of estimation based on a Galerkin projection of the operator on Laguerre functions ([9]), and couple it with a threshold performed both on the operator and the observed signal. We establish the minimax optimality of our procedure under the squared loss error, when the smoothness of the signal is measured in a Laguerre-Sobolev sense and the kernel satisfies fair blurring assumptions. It is important to stress that the resulting process is adaptive with regard both to the target function's smoothness and to the kernel's blurring properties. We end this paper with a numerical study emphazising the good practical performances of the procedure on concrete examples.

Keywords: Laplace convolution; blind deconvolution; nonparametric adaptive estimation; linear inverse problems; error in the operator.
Mathematical Subject Classification: 62G05, 62G99, 65J20, 65J22.

1. Introduction

Laplace deconvolution is motivated by a wide set of practical applications, ranging from population dynamics or physics to computational tomography or fluorescence spectroscopy (Linz [22, Chap. 2], Ameloot et al. [4], Comte et al. [9]). In the corresponding setting we observe $\boldsymbol{q}$, the result of the action of a kernel $\boldsymbol{g}$ on the function of interest $\boldsymbol{f}$, according to the following equation

$$\boldsymbol{q}(t) = \int_0^t \boldsymbol{g}(t-\tau)\boldsymbol{f}(\tau)d\tau,\ t \geqslant 0 \tag{1}$$

Equation (1) is also refered to as Volterra integral equation. One of its main features is its causal property, since $\boldsymbol{q}(t)$ is affected only by the values of $\boldsymbol{f}$ and $\boldsymbol{g}$ at times anterior to t. Of course, only finite samples of $\boldsymbol{q}(t)$ are accessible in practice. Moreover, the

presence of additional noise justifies the empirical modelization of (1) by the classical regression model, inspired by Abramovich et al. [3]

$$\boldsymbol{y}(t_i) = \int_0^{t_i} \boldsymbol{g}(t_i - \tau)\boldsymbol{f}(\tau)d\tau + \sigma\eta_i,\ i = 1, ..., n \tag{2}$$

where $0 \leqslant t_1 \leqslant ... \leqslant t_n \leqslant T_n$ are the points of observation, $(\eta_i)_{i=1,...,n}$ are independent standard gaussian variables, and σ is a fixed factor accounting for the precision of the observations. T_n is supposed to grow with the number of observations n.

As pointed out in Abramovich et al. [3] and Comte et al. [9], in spite of its apparent similarity with the Fourier deconvolution problem, the theoritical features of equation (1), as well as the practical problems raised during its resolution are deeply different. More precisely, setting artificially $\boldsymbol{g}(t) = \boldsymbol{f}(t) = 0$ for $t < 0$ amounts to solving the classical Fourier deconvolution problem

$$\boldsymbol{y}(t_i) = \int_0^{T_n} \boldsymbol{g}(t_i - \tau)\boldsymbol{f}(\tau)d\tau + \sigma\eta_i,\ i = 1, ..., n \tag{3}$$

A first notable objection is that the framework of classical Fourier deconvolution assumes periodicity of the function f and the kernel g on $[0, T]$, a meaningless notion when applied to a varying time set $[0, T_n]$. Even more problematic is the fact that this modelization totally ignores the causal feature of Laplace convolution, creating unwanted interferences between different time sets. To finish, the manipulation consisting in artificially expanding $\boldsymbol{q}$ and $\boldsymbol{g}$ for $t < 0$ creates artifacts on the estimated function at times $t < 0$ as well.

Another approach is to treat equation (2) as a general ill-posed problem and apply a Tikhonov regularization (Golubev [15]). However the direct implementation of this method also destroys the causal nature of equation (1), and tends to oversmooth the solution (Cinzori and Lamm [7]). Subsequent adaptations which remedy these shortcomings are present in Lamm [20] and Cinzori and Lamm [7]. However in these works the time set is considered to be fixed.

A more suitable theoritical tool in solving (1) is the use of Laplace transform, which allows to derive a closed form of the solution. However, its direct implementation is compromised by numerical problems, since the generic expression of the inverse Laplace transform is not easily computable in general. This motivates the widespread use of inversion tables, unfortunately irrelevant when the image function is not known exactly but approximated via a numerical scheme.

In this paper, following Comte et al. [9], we will exploit the properties of Laguerre functions, which can be used either to compute the inverse Laplace transform (Abate et al. [1], Lien et al. [21]), or to solve directly equation (1) (Keilson and Nunn [19]). More precisely, a Galerkin method applied to (2) shows that, even if their role is not entirely symmetric to the role played by harmonics in the framework of Fourier deconvolution, they allow a sparse analysis of equation (1).

All the previous mentionned works only concerned the case of a deterministic noise

at best. The presence of random noise requires an additionnal treatment, and calls for specific statistical tools. In the setting of random noise, Dey et al. [11] considered a kernel of the form e^{-at} and used a regularized inversion of the inverse Laplace transform. More recently, Abramovich et al. [3] conceived an optimal procedure in the minimax sense on Hölder spaces $\boldsymbol{H}^s(\mathbb{R}_+)$. This procedure used an exact expression of the solution involving the derivatives of $\boldsymbol{q}$, which were then estimated via Lepskii's method. However a shortcoming of the procedure is its strong dependence on the kernel $\boldsymbol{g}$, in the sense that a small error in $\boldsymbol{g}$ can translate into a wide difference in the result. In other words there seems to be a trade off between the closed form of the solution, and the unstability with regard to the kernel. Moreover, the fact that $\boldsymbol{g}$ is seldom observed directly in practice, but is usually subject to experimental noise should prompt us to privilege stability over exactitude.

In that spirit, Comte et al. [9] took advantage of the algebraic properties of Laguerre functions in the context of (2). With an adequate penalty term, they proposed an estimator which mimicks the oracle risk to within logarithmic terms. This modelization has the non negligible advantage of practical simplicity and efficiency, since solving equation (1) amounts to the inversion of a lower triangular Toeplitz matrix.

Even if this latter procedure proves to be more stable with regard to $\boldsymbol{g}$ experimentally, no systematic study has been conducted on the subject yet. In this paper we attempt to fill in this gap: we suppose that the observation of $\boldsymbol{g}$ is contaminated by a gaussian white noise, and show how Laguerre functions allow to handle this issue. We place ourselves under the minimax point of view and suppose that $\boldsymbol{f}$ belongs to a Laguerre-Sobolev space and that $\boldsymbol{g}$ satisfies standard blurring assumptions. We apply recent techniques for the treatment of noisy operators in the context of inverse problems (Hoffmann and Reiß [17],Delattre et al. [10]), which consist in a preliminary processing of the operator $\boldsymbol{K}$ coupled with a classical thresholding procedure applied to $\boldsymbol{y}$.

2. Discretization of Laplace deconvolution

2.1. Laguerre functions

Suppose that the target function $\boldsymbol{f}$ and the kernel $\boldsymbol{K}$ both lie in $L^2(\mathbb{R}_+)$. Define the Laguerre polynomials (see Gradshteyn and Ryzhik [16])

$$L_\ell(t) = \sum_{j=0}^{\ell} (-1)^j \binom{\ell}{j} \frac{t^j}{j!} \tag{4}$$

and, following Comte et al. [9], the ensuing Laguerre functions, depending on the parameter $a > 0$,

$$\boldsymbol{\varphi}_\ell(t) = \sqrt{2a} e^{-at} L_\ell(2at),\ \ell \in \mathbb{N} \tag{5}$$

The parameter a is a tuning parameter used to fit experimental curves. The Laguerre functions constitute a Hilbert basis of $L^2(\mathbb{R}_+)$. Any function $f \in L^2(\mathbb{R}_+)$ satisfies

$$f = \sum_{\ell \geqslant 0} \check{f}_\ell \boldsymbol{\varphi}_\ell(t), \;\; \check{f}_\ell \triangleq \int_0^\infty f(\tau)\boldsymbol{\varphi}_\ell(\tau)d\tau \tag{6}$$

The following proposition illustrates the conveniency of Laguerre functions in the framework of equation (1).

Proposition 2.1 (Gradshteyn and Ryzhik [16], Formula 7.411.4)**.**

$$\forall a > 0, \forall t \geqslant 0, \;\; \int_0^t \boldsymbol{\varphi}_k(x)\boldsymbol{\varphi}_\ell(t-x)dx = (2a)^{-1/2}\big(\boldsymbol{\varphi}_{\ell+m}(t) - \boldsymbol{\varphi}_{\ell+m+1}(t)\big) \tag{7}$$

From now on, except if explicitly mentionned, we will suppose $a = \frac{1}{2}$.

2.2. Galerkin method

Proposition 2.1 prompted Comte et al. [9] to apply a Galerkin scheme to equation (1). Galerkin schemes rely on the choice of a set of functions which discretize the inverse problem at stake in a convenient way. They were beneficially applied in the context of inverse problems (Cohen et al. [8]), and blind deconvolution (Efromovich and Koltchinskii [14], Hoffmann and Reiß [17] and Delattre et al. [10]). To this end we will remind briefly the underlying methodology of a Galerkin scheme and show how it conveniently applies to equation (1).

Let $f \in L^2(\mathbb{R}_+)$ and K an operator of $L^2(\mathbb{R}_+)$, and suppose we want to recover f from the observation $q = Kf$. Note $\boldsymbol{V}_\ell$ the finite dimensional space spanned by the orthogonal set of Laguerre functions $\{\boldsymbol{\varphi}_k\}_{k\leqslant\ell}$. The Galerkin approximation f^ℓ of f on $\boldsymbol{V}_\ell$ is the solution of the equation

$$\begin{aligned} &\langle Kf^\ell, v\rangle = \langle g, v\rangle, \;\; \forall v \in \boldsymbol{V}_\ell \\ \Leftrightarrow &\sum_{k\leqslant\ell} \langle K\boldsymbol{\varphi}_k, \boldsymbol{\varphi}_k'\rangle\langle f^\ell, \boldsymbol{\varphi}_k\rangle = \langle g, \boldsymbol{\varphi}_k\rangle, \;\; \forall k' \leqslant \ell \end{aligned} \tag{8}$$

We shall note K^ℓ the Galerkin matrix $(K^\ell)_{i,j} = \langle K\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i\rangle$, $i, j \leqslant \ell$. Note hence $\boldsymbol{K}$ the operator of $L^2(\mathbb{R}_+)$ mapping f onto $t \mapsto \int_0^t f(t-\tau)g(\tau)d\tau$. We can reformulate (1) as

$$\boldsymbol{q}^\ell = \boldsymbol{K}^\ell \boldsymbol{f}^\ell \tag{9}$$

Moreover, Proposition 2.1 implies:

Proposition 2.2 (Comte et al. [9], Lemma 1)**.** *The Galerkin matrix $\boldsymbol{K}^\ell$ is lower triangular, Toeplitz. More precisely, note $\dot{\boldsymbol{g}}$ the function with Laguerre coefficients*

$$\check{\dot{\boldsymbol{g}}}_\ell = \check{\boldsymbol{g}}_0 \mathbf{1}_{\{\ell=0\}} + \big(\check{\boldsymbol{g}}_\ell - \check{\boldsymbol{g}}_{\ell-1}\big)\mathbf{1}_{\{\ell\geqslant 1\}}, \quad \ell \in \mathbb{N}$$

Then

$$\boldsymbol{K}^{\ell} = \begin{pmatrix} \check{\boldsymbol{g}}_0 & 0 & \dots & 0 \\ \check{\boldsymbol{g}}_1 & \check{\boldsymbol{g}}_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \check{\boldsymbol{g}}_\ell & \dots & \check{\boldsymbol{g}}_1 & \check{\boldsymbol{g}}_0 \end{pmatrix}$$

In the sequel, for any function $f \in L^2(\mathbb{R}_+)$, we will note $T(f)$ the infinite Toeplitz matrix such that $T(f)_{i,1} = f_{i+1}$ for all $i \geqslant 0$, and $T_\ell(f)$ the extracted matrix defined by $T_\ell(f)_{i,1} = T(f)_{i,1}$, $i \leqslant \ell + 1$. In particular,

$$\boldsymbol{K}^{\ell} = T_\ell(\dot{\boldsymbol{g}})$$

The resolution of the linear system (8) now shows great practical conveniency, provided that $\boldsymbol{K}^{\ell}$ is invertible. This is equivalent to $\check{\boldsymbol{g}}_0 \neq 0$, an assumption we will make in the sequel.

2.3. Application to the regression model with irregular design

It remains to incorporate two supplementary features of equation (2) in the inversion of (9). First, the presence of the random noise $\boldsymbol{\eta}$ and secondly, the possible irregularity of the design points. This construction is due to Comte et al. [9]. Due to the fact that the observation points t_i are imposed by the problem, the estimation of the Laguerre coefficients $\check{\boldsymbol{q}}_\ell$ of the function $\boldsymbol{q}$ suffers from two potential drawbacks. First, the infinite support of the Laguerre polynomials as well as the function $\boldsymbol{q}$ which should not be too problematic, provided that T_n is large enough and that the functions decrease sufficiently to infinity. More problematic is the fact that the observation points t_i are sometimes subject to experimental constraints, which affect their repartition on $\mathbb{R}_+$. The consistency of the estimation of $\check{\boldsymbol{q}}_\ell$ is hereby deteriorated.
We will hence suppose that the following conditions are fulfilled:

- There exists an integer n_0 such that $\frac{n}{T_n} > \sigma$ for all $n \geqslant n_0$.
- $\lim_{n\to\infty} T_n = \infty$, and $\lim_{n\to\infty} \dfrac{T_n}{n} = 0$

To take into account the irregularity of the design, we follow Comte et al. [9] and define $P_n : [0; T_n] \to [0; T_n]$ a regular non decreasing function such that

$$P_n(0) = 0,\ P_n(T_n) = T_n,\ P_n(t_i) = \frac{i}{n} T_n \text{ for } i \leqslant n \tag{10}$$

Note $\boldsymbol{\Phi}_\ell$ the $(\ell+1) \times n$ matrix with entries $(\boldsymbol{\Phi}_\ell)_{k,i} = \boldsymbol{\varphi}_k(t_i)$. For any function $h \in L^2(\mathbb{R})$, we have

$$\boldsymbol{P}_\ell h(t_i) = \sum_{k \leqslant \ell} \boldsymbol{\varphi}_k(t_i) \check{h}_k = \boldsymbol{\Phi}_\ell h^\ell$$
$$\Leftrightarrow h^\ell = \left({}^{\boldsymbol{t}}\boldsymbol{\Phi}_\ell \boldsymbol{\Phi}_\ell\right)^{-1} {}^{\boldsymbol{t}}\boldsymbol{\Phi}_\ell \boldsymbol{P}_\ell h(t_i)$$

where $\boldsymbol{P}_\ell$ is the orthogonal projector onto $\boldsymbol{V}_\ell$. We deduce that

$$\boldsymbol{y}^\ell = \left({}^t\boldsymbol{\Phi}_\ell\boldsymbol{\Phi}_\ell\right)^{-1}{}^t\boldsymbol{\Phi}_\ell\boldsymbol{P}_\ell\big[\boldsymbol{q}+\sigma\boldsymbol{\eta}\big](t_i) = \boldsymbol{K}^\ell\boldsymbol{f}^\ell + \sigma\left({}^t\boldsymbol{\Phi}_\ell\boldsymbol{\Phi}_\ell\right)^{-1}{}^t\boldsymbol{\Phi}_\ell\boldsymbol{\eta}_n \tag{11}$$

where $\boldsymbol{\eta}_\ell \sim \mathcal{N}(0,\boldsymbol{I}_n)$. Let us take a closer look to the matrix $\left({}^t\boldsymbol{\Phi}_\ell\boldsymbol{\Phi}_\ell\right)$. Its general term is

$$\begin{aligned}({}^t\boldsymbol{\Phi}_\ell\boldsymbol{\Phi}_\ell)_{\ell,k} &= \sum_{i=1}^{n}\boldsymbol{\varphi}_k(P^{-1}(\frac{i}{n}T_n))\boldsymbol{\varphi}_\ell(P^{-1}(\frac{i}{n}T_n))\\ &\sim\frac{n}{T_n}\int_0^{T_n}\boldsymbol{\varphi}_k(P^{-1}(\tau))\boldsymbol{\varphi}_\ell(P^{-1}(\tau))d\tau\\ &= \frac{n}{T_n}\int_0^{T_n}\boldsymbol{\varphi}_k(\tau)\boldsymbol{\varphi}_\ell(\tau)P'(\tau)d\tau\end{aligned}$$

for n, T_n large enough. If the points t_i are equispaced, taking $P(\tau)=\tau$ in (10) entails that $T_n n^{-1}\left({}^t\boldsymbol{\Phi}_\ell\boldsymbol{\Phi}_\ell\right)$ is close to the identity provided that T_n is large enough. As in Comte et al. [9], we hence reformulate (11) as the sequential model

$$\boldsymbol{y}^\ell = \boldsymbol{K}^\ell\boldsymbol{f}^\ell + \sigma\sqrt{\frac{T_n}{n}}\boldsymbol{\xi}_\ell$$

where $\boldsymbol{\xi}_\ell \sim \mathcal{N}(0,\boldsymbol{\Omega}_\ell)$ and $\boldsymbol{\Omega}_\ell = nT_n^{-1}\left({}^t\boldsymbol{\Phi}_\ell\boldsymbol{\Phi}_\ell\right)^{-1}$. In general, $\boldsymbol{\Omega}_\ell$ somehow quantifies the distance to the uniform design case. To ensure that the design is not too ill conditionned, we will suppose that the following assumption is fulfilled.

Assumption 2.3. *Let $L \in \mathbb{N}$. There exists $C \geqslant 0$, such that for all $\ell \leqslant L$, for all $\lambda \in Sp(\boldsymbol{\Omega}_\ell)$, $\lambda \leqslant C$*

This assumption is dependent on the integer L, which plays the role of a maximal resolution level, and will be adapted to the case of interest later. The inversion of (11) now requires controls of the variable $(\boldsymbol{K}^\ell)^{-1}\boldsymbol{\xi}_\ell$. Under suitable properties of $\boldsymbol{f}$ and $\boldsymbol{g}$, we shall be able to apply a classical inverse/thresholding procedure, and derive rates of convergence over specific regularity spaces. These properties are the subject of Part 3.

2.4. Error in the operator

We already mentionned the fact that the resolution of (1) is usually unstable with respect to $\boldsymbol{g}$ (Abramovich et al. [3]). Furthermore, in practice, inference on the kernel $\boldsymbol{g}$ is possible only through experimental noise, and requires a preliminary step of estimation giving way to imprecision. This additionnal error might significantly contaminate the result of any procedure of estimation if not properly treated. Let us see how Laguerre functions $\boldsymbol{\varphi}_\ell$ allow to handle this issue: in section 2.2, we established that the discretization of (13) with Laguerre functions involved a Toeplitz matrix with entries constituted of the Laguerre coefficients of $\dot{\boldsymbol{g}}$. We can thus consider $\dot{\boldsymbol{g}}$ as the finite impulse response of the operator $\boldsymbol{K}$ when applied to the system $(\boldsymbol{\varphi}_\ell)_{\ell\geqslant 0}$. To take into account

the imprecision in the observations of $\dot{\boldsymbol{g}}$, we adopt the framework of blind deconvolution and suppose that $\dot{\boldsymbol{g}}$ is not known exactly, but that we have acces the noisy version

$$\dot{\boldsymbol{g}}_\delta = \dot{\boldsymbol{g}} + \delta \boldsymbol{b} \tag{12}$$

where $\boldsymbol{b}$ is a gaussian white noise on $L^2(\mathbb{R}_+)$. The generic problem of blind deconvolution is motivated by numerous scientific fields, including for example electronic microscopy or astrophysics, where the corresponding kernel is seldom known nor directly observed. It was adequatly discussed in Efromovich and Koltchinskii [14] and Hoffmann and Reiß [17].

Taking into account the observations (12), the projection $\dot{\boldsymbol{g}}^\ell$ is changed to $\dot{\boldsymbol{g}}^\ell_\delta = \dot{\boldsymbol{g}}^\ell + \delta \boldsymbol{b}^\ell$ where $\boldsymbol{b}^\ell$ is a gaussian vector with covariance $\boldsymbol{I}_\ell$. The new model, adjusted from (11) becomes

$$\begin{cases} \boldsymbol{y}^\ell & = \boldsymbol{K}^\ell \boldsymbol{f}^\ell + \sigma\sqrt{\frac{T_n}{n}}\boldsymbol{\xi}_\ell \\ \boldsymbol{K}^\ell_\delta & = \boldsymbol{K}^\ell + \delta \boldsymbol{B}^\ell \end{cases} \tag{13}$$

where $\boldsymbol{B}^\ell = T_\ell(\boldsymbol{b})$ is a random Toeplitz matrix. In the sequel, for the sake of clarity, we note $\varepsilon = \sigma\sqrt{\frac{T_n}{n}}$.

Remark 2.4. *We could as well suppose that we observe $\boldsymbol{g}_\delta = \boldsymbol{g} + \delta\boldsymbol{b}$, yet it is more convenient to work with $\dot{\boldsymbol{g}}$ (the entries of the noisy Toeplitz matrix $\boldsymbol{B}$ are directly i.i.d standard gaussian variables). In the former case, the rest of the paper however adapts with no change in the algorithms, since inequality (27) is satisfied as well. A modification of the proof of Theorem 4.6 should also provide the lower bound for the second procedure.*

3. Features of the target function and the kernel

3.1. Sobolev spaces associated to Laguerre functions

We proceed to the description of regularity spaces associated with the resolution of (13). The following material is classical, we refer to Bongioanni and Torrea [5] or Rathnakumar [26] for example.

Since $f \mapsto \sqrt{2a}f(2a.)$ is an isometry of $L^2(\mathbb{R}_+)$, the structures defined for different values of a are equivalent. Hence we shall only concentrate on the mainstream case where $a = 1/2$. Define the operator $\mathfrak{L}$ on $L^2(\mathbb{R}_+, dx)$ by

$$\mathfrak{L} = -\Big[x\frac{d^2}{dx^2} + \frac{d}{dx} - \frac{x}{4}\Big] \tag{14}$$

The functions $\boldsymbol{\varphi}_\ell$ are the eigenfunctions of $\mathfrak{L}$ associated with eigenvalues $(\ell + \frac{1}{2})$. We hence define the Sobolev space $\mathcal{W}^s$ associated with Laguerre functions as

$$\mathcal{W}^s = \{f \in L^2(\mathbb{R}_+, dx) \text{ s.t. } \mathfrak{L}^s f \in L^2(\mathbb{R}_+, dx)\}$$

For a function $f \in L^2(\mathbb{R}_+, dx)$, we have the straightforward equivalence

$$f \in \mathcal{W}^s \Leftrightarrow \sum_{\ell \geqslant 0} \left|\left(\ell + \frac{1}{2}\right)^s \langle f, \boldsymbol{\varphi}_\ell \rangle\right|^2 < \infty$$

and the associated norm

$$\|f\|_{\mathcal{W}^s} = \sum_{\ell \geqslant 0} \left|\left(\ell + \frac{1}{2}\right)^s \langle f, \boldsymbol{\varphi}_\ell \rangle\right|^2$$

For $M \geqslant 0$, we shall note $\mathcal{W}^s(M)$ the Sobolev ball of radius M. Finally, we remind that, as $\|\boldsymbol{\varphi}_\ell\|_\infty \leqslant 1$ for all $\ell \geqslant 0$, we have $s > 1/2 \Rightarrow \mathcal{W}^s \subset \mathcal{C}^0(\mathbb{R}_+)$. From now on, we will hence suppose that there exists $s > 1/2$ such that $\boldsymbol{f} \in \mathcal{W}^s$.

3.2. Banded Toeplitz matrices

Before entering into details about the kernel features, we introduce basic material on Toeplitz matrices. Most of it is inspired by Böttcher and Grudsky [6] and Comte et al. [9].

Let $a = (a_\ell) \in \ell^1(\mathbb{Z})$ be a sequence of real numbers. We remind from section 2.2 that we note $T(a)$ the infinite Toeplitz matrix defined by

$$T(a) = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots & \dots \\ a_1 & a_0 & a_{-1} & \dots & \dots \\ a_2 & a_1 & a_0 & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots \end{pmatrix}$$

and $T_\ell(a) \in M_\ell(\mathbb{R})$ the truncated Toeplitz matrix defined as

$$(T_\ell(a))_{i,j} = (T(a))_{i,j},\ i, j \leqslant \ell + 1$$

The Toeplitz matrices $T(a)$ and $T_\ell(a)$ are naturally linked to the two respective Laurent series

$$a(z) = \sum_{k=-\infty}^{\infty} a_k z^k \text{ and } a_\ell(z) = \sum_{k=-\ell}^{\ell} a_k z^k$$

We will indifferently refer to the vector a or the corresponding Laurent serie. The spectral norm of $T(a)$ is related to the behaviour of $a(z)$, as illustrated in the following proposition.

Proposition 3.1. *Let $a \in \ell^1(\mathbb{Z})$. Let $\mathcal{C}$ stand for the complex unit circle. We have*

$$\|T(a)\|_{op} = \|a(z)\|_{circ}$$

where $\|a(z)\|_{circ} \triangleq \sup_{z \in \mathcal{C}} \left| \sum_{\ell=-\infty}^{\infty} a_\ell z^\ell \right|$. A simple corollary is the following inequality

$$\|T(a)\|_{op} \leqslant \sum_{\ell=-\infty}^{\infty} |a_\ell|$$

In particular, Proposition 3.1 applies to the case of truncated Toeplitz matrices $T_\ell(a)$. Moreover, if a has no zero on the complex unit circle, we have

$$\limsup_{\ell\to\infty}\|T_\ell(a)\|_{\text{op}} < \infty \text{ and } \lim_{\ell\to\infty}\|T_\ell(a)\|_{\text{op}} = \|T(a)\|_{\text{op}} \tag{15}$$

Now suppose that a and a' both generate lower triangular Toeplitz matrices (i.e. $a_k = a'_k = 0$ if $k < 0$). Then the following equalities hold for all $\ell \geqslant 0$:

$$T_\ell(a)T_\ell(a') = T_\ell(a')T_\ell(a) = T_\ell(aa') \text{ and } T_\ell(a)^{-1} = T_\ell(1/a) \tag{16}$$

In other words, the matrix multiplication (resp. inversion) is equivalent to a power serie multiplication (resp. inversion).

3.3. Degree of ill posedness

We now need to precise the properties of $\boldsymbol{K}$ as a blurring operator of $L^2(\mathbb{R}_+)$. Usually the operator $\boldsymbol{K}$ is not compact, and the problem (1) is ill-posed. This results in practical unstabilities when trying to invert equation (11) from discrete observations. The quantification of the ill-posedness of the problem is specified by the introduction of a constant, called degree of ill-posedness (DIP) of the problem (see Nussbaum and Pereverzev [25], Mathe and Pereverzev [23] for a generic review). We adapt this concept to our framework, and make the following assumption.

Assumption 3.2 (Degree of ill-posedness of $\boldsymbol{g}$)**.** *There exists $\nu \geqslant 0$, $Q \geqslant 0$ such that, for all $\ell \geqslant 0$,*

$$\|(\boldsymbol{K}^\ell)^{-1}\|_{op} \leqslant Q(\ell \vee 1)^\nu$$

ν is called degree of ill-posedness of $\boldsymbol{g}$ (or equivalently of $\boldsymbol{K}$). We note $\mathcal{K}_\nu(Q)$ the set of functions which satisfy this assumption.

We shall see examples of kernels satisfying this assumption further. For the moment, we concentrate on the treatment of observations (13) in the context we just described.

3.4. Algorithms and rates of convergence

The main challenge which remains to be treated now is to articulate the two critical steps of inversion and regularization, via adapted procedures. For example, let us give a brief overview of the methodology in Comte et al. [9]: Let $\ell \in \mathbb{N}$, and let Λ be the following contrast function, defined on $\mathbb{R}^\ell$ by

$$\Lambda : t \mapsto \|t\|^2 - 2\langle t, (\boldsymbol{K}^\ell)^{-1}\boldsymbol{y}^\ell\rangle$$

Note $\|.\|_{\text{op}}$ the spectral norm and $\|.\|_{\text{HS}}$ the Hilbert Schmidt norm. A model selection is performed on the maximal level L, by introducing the following penalizing factor ($B > 0$ is an arbitrary constant):

$$\text{pen}(\ell) = 4\sigma^2 T_n n^{-1}\Big((1+B)\|\sqrt{\boldsymbol{Q}^\ell}\|^2_{\text{HS}} + (1+B)^{-1}(\nu+1)\|\sqrt{\boldsymbol{Q}^\ell}\|^2_{\text{op}}\log\ell\Big)$$

where $\boldsymbol{Q}^\ell = (\boldsymbol{K}^\ell)^{-1}\boldsymbol{\Omega}_\ell{}^t(\boldsymbol{K}^\ell)^{-1}$ and $\sqrt{\boldsymbol{Q}^\ell}$ is a lower triangular matrix satisfying $\sqrt{\boldsymbol{Q}^\ell}{}^t\sqrt{\boldsymbol{Q}^\ell} = \boldsymbol{Q}^\ell$. The maximal level $\tilde{L}$ is hence chosen as

$$\tilde{L} = \underset{\ell \leqslant \ell(n)}{\operatorname{argmin}}\{\Lambda^2\big((\boldsymbol{K}^\ell)^{-1}\boldsymbol{y}^\ell\big) + \operatorname{pen}(\ell)\}$$

where $\ell(n)$ is a large enough resolution level, possibly depending on n, and the ensuing estimator of $\boldsymbol{f}$ is

$$(\boldsymbol{K}^{\tilde{L}})^{-1}\boldsymbol{y}^{\tilde{L}}$$

We follow here a different path: we suppose that the target function belongs to a Sobolev-Laguerre space, and perform thresholding techniques in a minimax framework. Furthermore, our results are asymptotic with regard to ε, δ. Would $\boldsymbol{g}$ be known, the estimation of $\boldsymbol{f}$ from observations (13) amounts to solving a standard inverse problem with signal noise. To this end, a prolific litterature is at disposal (a selected list is Donoho [12], Abramovich and Silverman [2], Cohen et al. [8]). In order to take into account the presence of noise in the operator, we shall hence apply a preliminar regularizing thresholding procedure to the noisy operator $\boldsymbol{K}_\delta$ in order to ensure the stability of the further inversion step. To that end, define the maximal level as

$$L^{\mathbf{I}} = \lambda\Big(\varepsilon\sqrt{|\log\varepsilon|} \vee \delta|\log\delta|\Big)^{\frac{-1}{\nu+1}} \tag{17}$$

with λ a positive constant. Define also the two thresholding levels

$$O_{\ell,\delta} = \kappa\big((\ell \vee 1)\log(\ell \vee 2)\big)^{1/2}\delta\sqrt{|\log\delta|} \tag{18}$$

$$S^{\mathbf{I}}_{\ell,n} = (\ell \vee 1)^\nu\Big(\tau_{sig}\varepsilon\sqrt{|\log\varepsilon|} \vee \tau_{\mathrm{op}}\delta|\log\delta|\Big) \tag{19}$$

For $\ell \geqslant 0$, note $\boldsymbol{\zeta}_\ell = \big\langle(\boldsymbol{K}^\ell_\delta)^{-1}\mathbf{1}_{\left\{\|(\boldsymbol{K}^\ell_\delta)^{-1}\|_{\mathrm{op}} < O^{-1}_{\delta,l}\right\}}\boldsymbol{y}^\ell, \boldsymbol{\varphi}_\ell\big\rangle$. The estimator $\widetilde{\boldsymbol{f}}^{\mathbf{I}}$ of $\boldsymbol{f}$ is defined by

$$\widetilde{\boldsymbol{f}}^{\mathbf{I}} = \sum_{\ell \leqslant L^{\mathbf{I}}} \boldsymbol{\zeta}_\ell \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S_\ell\}}\boldsymbol{\varphi}_\ell$$

We call this procedure **Algorithm I**. The preliminary threshold performed on $(\boldsymbol{K}^\ell_\delta)^{-1}$ ensures its proximity with $(\boldsymbol{K}^\ell)^{-1}$ with high probability (see Lemma 6.2). We now study the squared loss performance of the procedure.

Theorem 3.3. *Let $M \geqslant 0$, $s > 1/2$. Let $\nu \geqslant 0$, $Q \geqslant 0$. Suppose that Assumption 2.3 holds for $L = L^{\boldsymbol{I}}$. Then for sufficiently large thresholding constants κ, τ_{sig} and τ_{op},*

$$\sup_{\substack{\boldsymbol{f}\in\mathcal{W}^s(M)\\ \boldsymbol{g}\in\mathcal{K}_\nu(Q)}} \mathbb{E}\|\widetilde{\boldsymbol{f}}^{\boldsymbol{I}} - \boldsymbol{f}\| \lesssim \Big(\delta|\log\delta|\Big)^{\frac{2s}{2(s+\nu)+1}} \vee \Big(\varepsilon\sqrt{|\log\varepsilon|}\Big)^{\frac{2s}{2(s+\nu)+1}}$$

where $\lesssim$ means inequality up to a constant depending only on $\lambda, \kappa, \tau_{sig}, \tau_{op}, s, M, \nu, Q$.

The rates in Theorem 3.3 reveal two components, accounting respectively for the imprecision in the observation of the operator and the signal. The latter is fairly classic in non parametric statistics (Nussbaum and Pereverzev [25], Johnstone et al. [18]) where it is also optimal, while the former is standard (and optimal too) in blind deconvolution on Hilbert spaces (Efromovich and Koltchinskii [14], Hoffmann and Reiß [17]). Thus, we do not study the optimality of these rates in this paper, but rather concentrate on a more specific framework related to the problem of interest.

4. Adaptation to the standard framework of Laplace deconvolution

We now discuss the adapation of our algorithm in the mainstream framework of Laplace deconvolution, as exposed in Abramovich et al. [3] or Comte et al. [9]. As we shall see, this more restrictive framework allows to treat observations (13) more efficiently. To this end, we first define a more restrictive version of the degree of ill-posedness.

Assumption 4.1 (Second kind degree of ill-posedness)**.** *Note* $\boldsymbol{\gamma}_k = \langle (1/\dot{\boldsymbol{g}}), \boldsymbol{\varphi}_k \rangle$*, so that* $(1/\dot{\boldsymbol{g}})(z) = \sum_{k\geqslant 0} \boldsymbol{\gamma}_k z^k$*. There exists* $\nu > 0$*, there exists* $Q_2, Q_1 > 0$*, such that for all* $\ell \geqslant 0$*,*

$$\sum_{k=0}^{\ell} \boldsymbol{\gamma}_k^2 \leqslant Q_2(\ell \vee 1)^{2\nu-1} \tag{20}$$

$$\sum_{k=0}^{\ell}\sum_{n=0}^{k} \boldsymbol{\gamma}_n^2 \geqslant Q_1(\ell \vee 1)^{2\nu} \tag{21}$$

For $Q = (Q_1, Q_2)$, we note $\mathcal{G}_\nu(Q)$ the set of functions $\boldsymbol{g} \in L^2(\mathbb{R}_+)$ such that Assumption 4.1 holds. Note that the validity of this assumption automatically entails $Q_1 \leqslant \big(1 + \frac{2^{2\nu}}{2\nu}\big)Q_2$. Note also that the left term in (21) is the Hilbert-Schmidt norm of $(\boldsymbol{K}^\ell)^{-1}$. Thus, Assumption 4.1 is more restrictive that Assumption 3.2. However, it is satisfied by a natural class of functions $\boldsymbol{g}$:

Proposition 4.2 (Comte et al. [9], Lemma 3/ Lemma 5)**.** *Suppose that there exists* $C, \nu > 1/2$*,* $\mu \in \mathcal{C}$ *and* $w(z) = \prod_{i=1}^N (z - \mu_i)$*,* $|\mu_i| > 1$ *a polynomial function with no pole inside of the complex unit disc, such that*

$$\dot{\boldsymbol{g}}(z) = Cw(z)(\mu - z)^\nu \tag{22}$$

Then Assumption 4.1 is satisfied. Furthermore, if $w \equiv 1$ *and* $\nu \geqslant 0$*, then* $|\boldsymbol{\gamma}_\ell| \sim \frac{\ell^{\nu-1}}{\Gamma(\nu)}$.

For completeness, we give a proof of Proposition 4.2 in section 6. We now turn to the standard framework of Laplace deconvolution, as exposed in Abramovich et al. [3] and Comte et al. [9]. To this end, we define the following assumptions concerning the kernel $\boldsymbol{g}$.

Assumption 4.3.

(A1) There exists an integer $r \geqslant 1$ such that

$$\frac{d^j \boldsymbol{g}}{dt^j}\Big|_{t=0} = \begin{cases} 0 & \text{if } j = 0, 1, ..., r-2 \\ B_r \neq 0 & \text{if } j = r-1 \end{cases}$$

(A2) $\boldsymbol{g} \in L^1\big([0,+\infty)\big)$ is r times differentiable and $\boldsymbol{g}^{(r)} \in L^1\big([0,+\infty)\big)$.

(A3) The Laplace transform of $\boldsymbol{g}$ has no zeros with non negative real parts except for the zeros of the form $\infty + ib$.

The consequences of these assumptions are well formulated in the terms of the preceding framework:

Proposition 4.4 (Comte et al. [9], Lemma 3)**.** *Suppose that Assumptions (A1), (A2) and (A3) hold. Then the hypotheses of Proposition 4.2 are satisfied with $\mu = 1$, $\nu = r$.*

Hence, Assumption 4.1 is verified with $\nu = r$ and **Algorithm I** applies. However, Assumption 4.1 provides additional information on the behaviour of $(1/\dot{\boldsymbol{g}})$. We adapt **Algorithm I** to this new framework, by operating the following changes:

- Set the maximal level to

$$L^{\mathbf{II}} = \lambda\Big(\varepsilon\sqrt{|\log\varepsilon|} \vee \delta|\log\delta|\Big)^{-1}$$

- Set the signal thresholding level to

$$S^{\mathbf{II}}_{\ell,n} = \begin{cases} \|(\boldsymbol{K}^\ell_\delta)^{-1}\|_{\mathrm{HS}}(\ell \vee 1)^{-1/2}\Big(\tau_{sig}\varepsilon\sqrt{|\log\varepsilon|} \vee \tau_{\mathrm{op}}\delta|\log\delta|\Big) & \text{if } \|(\boldsymbol{K}^\ell_\delta)^{-1}\|_{\mathrm{op}} < O^{-1}_{\ell,\delta} \\ +\infty & \text{if } \|(\boldsymbol{K}^\ell_\delta)^{-1}\|_{\mathrm{op}} \geqslant O^{-1}_{\ell,\delta} \end{cases} \tag{23}$$

where $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}({}^tAA)}$ is the Hilbert-Schmidt norm. We call the modified procedure **Algorithm II** and note $\widetilde{\boldsymbol{f}}^{\mathbf{II}}$ the corresponding estimated function. A notable gain of this new algorithm is its independence with regard to the parameter ν. Indeed, Assumption 4.1 allows us to use $\|(\boldsymbol{K}^\ell_\delta)^{-1}\|_{\mathrm{HS}}$ in (19) as a substitute of ℓ^ν, and to overesimate the 'true' maximal level $L^{\mathbf{I}}$. Its performances are exposed in the next theorem:

Theorem 4.5. *Let $M \geqslant 0$. Let $\nu > 0$, $Q_2, Q_1 > 0$ and $s > 1/2$. Suppose that Assumption 2.3 holds with $L = L^{\boldsymbol{II}}$. Then for sufficiently large thresholding constants κ, τ_{sig} and τ_{op},*

$$\sup_{\substack{\boldsymbol{f}\in\mathcal{W}^s(M) \\ \boldsymbol{K}\in\mathcal{G}_\nu(Q)}} \mathbb{E}\|\widetilde{\boldsymbol{f}}^{\boldsymbol{II}} - \boldsymbol{f}\| \lesssim \Big(\delta|\log\delta|\Big)^{\frac{s}{s+\nu}} \vee \Big(\varepsilon\sqrt{|\log\varepsilon|}\Big)^{\frac{s}{s+\nu}}$$

where $\lesssim$ means inequality up to a constant depending only on $\lambda, \kappa, \tau_{sig}, \tau_{op}, s, M, \nu, Q_1, Q_2$.

Thus, in addition to the adaptivity over the parameter ν, the strengthening of Assumption 3.2 via (20) and (21) allows to improve on the rates of Theorem 3.3 with regard both to the operator and signal noise. Our next result shows that the rate achieved in Theorem 4.5 is indeed optimal, up to logarithmic terms. The lower bound will not decrease for increasing noise levels δ and ε, whence it suffices to provide separately the cases $\delta = 0$ and $\varepsilon = 0$.

Theorem 4.6. *Let $s > 1/2$, let $M \geqslant 0$ $\nu > 1/2$ and $Q_2 \geqslant c_\nu Q_1 > 0$. Here c_ν is a constant depending only on ν which will will not seek to precise. We have*

$$\inf_{\tilde{f}} \sup_{\substack{f\in\mathcal{W}^s(M)\\ g\in\mathcal{G}_\nu(Q)}} \mathbb{E}\|\tilde{\boldsymbol{f}} - \boldsymbol{f}\| \gtrsim \delta^{\frac{s}{s+\nu}} |\log\delta|^{-1} \vee \varepsilon^{\frac{s}{s+\nu}} |\log\varepsilon|^{-1}$$

where the infimum is taken among all estimators $\tilde{f}$ of f based on observations (13).

Combining Theorem 3.3 together with Theorem 4.6, we conclude that our algorithm is minimax over $\mathcal{W}^s(M)$ to within logarithmic terms in ε and δ, uniformly with regard to the blurring kernel $\boldsymbol{g} \in \mathcal{G}_\nu(Q)$.

5. Practical performances

In this section we study the practical performances of the two procedures developped above. Note that three potential sources of errors may contaminate the quality of the observations in (13) : the signal precision $\sigma\sqrt{\frac{T_n}{n}}$, the operator precision δ and the design quality $\|\boldsymbol{\Omega}_\ell\|_{\text{op}}$. We shall hence emphasize their influence in the estimation of $\boldsymbol{f}$, as well as their respective interactions.

Our first aim is to study the interaction between the effect of signal and kernel noise in the two procedures of reconstruction. To this end, we will isolate them from the effect of the design, and suppose that the latter is ideally conditionned by setting $\boldsymbol{\Omega}_\ell = \boldsymbol{I}_\ell$. Let us start by a few precisions concerning the tuning parameters of **Algorithm I** and **II**. The setting up of these procedures requires the preliminary definition of $\lambda, \kappa, \tau_{sig}$ and τ_{op}.

Tuning parameters: for the definition of the maximal level of resolution, we set $\lambda = 1$ for both algorithms. The concrete choice of adequate thresholding constants κ and τ is a complex issue. Our practical choices will be based on the following remark, inspired by Donoho and Johnstone [13]: in the case of direct estimation on real line, the universal threshold which is both efficient and simple to implement, takes the form $2\sqrt{|\log\varepsilon|}$. A consistent interpretation is to consider that this threshold should kill any pure noise signal. We will adapt this reasoning to the case of interest.

Choice of κ : we use as a benchmark the case where $\boldsymbol{g} \equiv 0$. Given δ large enough, we define κ as the smallest value κ_δ such that , for all $\ell \leqslant 10$, $\mathbf{1}_{\left\{\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{\text{op}} < O_{\delta,l}^{-1}\right\}} = 0$. The results are reported in Table 1 and give $\kappa = 0.3$.

Choice of τ_{sig} and τ_{op}: It is clear that the role of τ_{sig} and τ_{op} is to control the influence of the signal (resp. the operator) error. To choose τ_{sig} (resp. τ_{op}), we therefore set $\varepsilon_{sig} > \delta_{sig} > 0$ (resp. $\delta_{op} > \varepsilon_{op} > 0$) large enough. We resort to the case $\boldsymbol{f} \equiv 0$ as a benchmark: we have $\langle \boldsymbol{f}, \boldsymbol{\varphi}_\ell\rangle = 0$ for $\ell \geqslant 1$, consequently the observations $\langle \boldsymbol{g}_{\varepsilon_{sig}}, \boldsymbol{\varphi}_\ell\rangle$, $\ell \geqslant 0$ are pure noise. We hence simulate $\boldsymbol{K}_{\delta_{sig}}$ and, integrating the precedently computed value of κ, apply the procedure for increasing values of τ_{sig} (resp. τ_{op}) until all the computed coefficients $\langle \tilde{\boldsymbol{f}}^i, \boldsymbol{\varphi}_\ell\rangle$ (i =**I**,**II**) are killed for $\ell \leqslant 10$. The results are reported in Table 2.

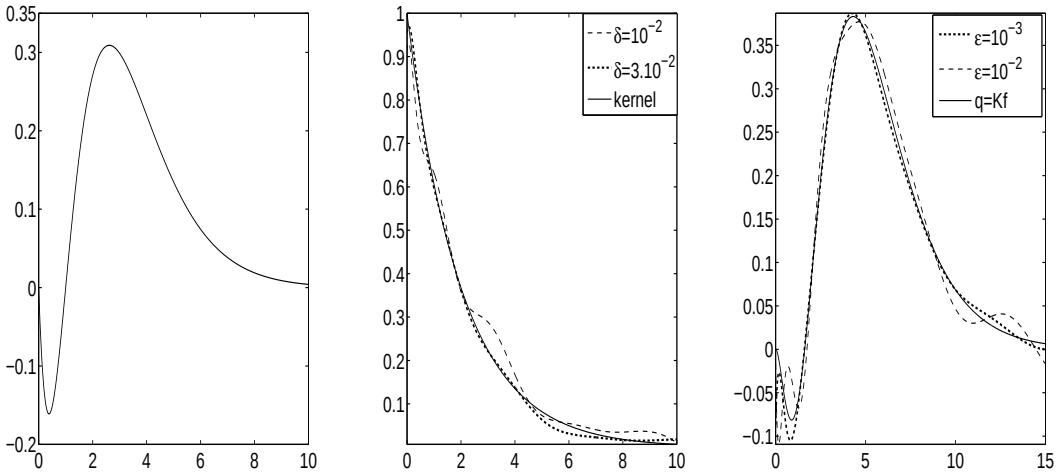
κ	0.1	0.2	0.3
N	3	1	0

Table 1. Choosing of κ. N is the average number, computed on a basis of 10 realizations, of levels $\ell \leqslant 10$ such that $\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{\mathrm{op}} < O_{\delta,l}^{-1}(\kappa)$. We have $\delta = 10^{-2}$.

τ_{sig}	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$N_{\mathbf{I}}$	1	1	0	0	0	0	0	0
$N_{\mathbf{II}}$	7	5	4	3	2	1	1	0

τ_{op}	0.1
$N_{\mathbf{I}}$	0
$N_{\mathbf{II}}$	0

Table 2. Choosing of τ. For $(\delta_{sig}, \varepsilon_{sig}) = (\varepsilon_{\mathrm{op}}, \delta_{\mathrm{op}}) = (10^{-2}, 10^{-1})$ and each value of τ, we computed 10 times the described procedure and reported N_i the average number of remaining Laguerre coefficients for **Algorithm** i.



(a) **Target function** $\boldsymbol{f}_1$ (b) **Kernel** $\boldsymbol{g}$ (c) $\boldsymbol{q} = \boldsymbol{K}\boldsymbol{f}$

Figure 1. Datas and noisy observations of $\boldsymbol{g}$ and $\boldsymbol{q}$

We now apply the two procedures to the case where $\boldsymbol{f}_1(t) = (t^2 - t)\exp(-t)$ and $\boldsymbol{g} = \boldsymbol{\varphi}_0$ (a graphical representation of these two functions is presented in Figure 1).

We have

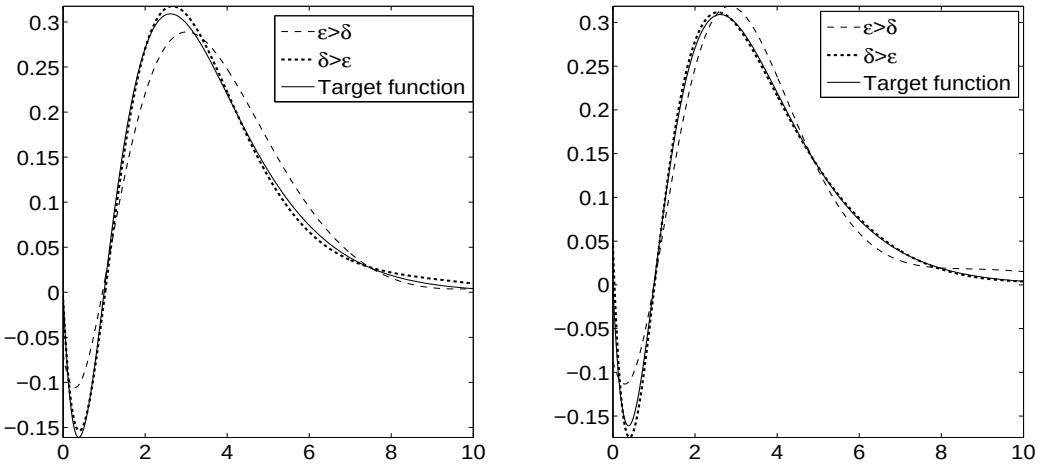
$$\big(1/\dot{\boldsymbol{g}}\big)(z) = (1-z)^{-1} = \sum_{\ell \geqslant 0} z^\ell$$

hence Assumptions 3.2 and 4.1 are both satisfied taking $\nu = 1$. For several values of ε and δ, we report the corresponding squared loss, computed on a basis of 500 realisations with the use of Parseval's identity, in Table 3. The corresponding results are presented in Figure 2 for one particular realization of $\boldsymbol{\xi}, \boldsymbol{b}$. The results indicate that the transition on the two types of errors occur when δ is higher than ε, translating a prevailing effect of the signal noise ε over the operator error δ in practice. As Theorems 3.3 and 4.5 suggest, the second Algorithm overperforms the first in (almost) every case.

Discussion on the design irregularity: to control the squared risk of the two

	Algorithm I				**Algorithm II**			
δ \ ε	0	10^{-3}	10^{-2}	3.10^{-2}	0	10^{-3}	10^{-2}	3.10^{-2}
0	0	0.020	0.141	0.348	0	0.012	0.109	0.312
10^{-3}	0.004	0.020	0.141	0.352	0.005	0.012	0.108	0.301
10^{-2}	0.047	0.054	0.143	0.344	0.053	0.039	0.116	0.318
3.10^{-2}	0.170	0.169	0.190	0.348	0.118	0.109	0.145	0.324

Table 3. Normalized mean squared error of the two procedures applied to the functions $\boldsymbol{f}_1$ and $\boldsymbol{g}$. The computations were performed using a monte carlo method on 500 realizations.



(a) **Algorithm I** (b) **Algorithm II**

Figure 2. Estimation of $\boldsymbol{f}_1$ for predominant signal noise $(\varepsilon, \delta) = (10^{-2}, 10^{-3})$ and predominant operator noise $(\delta, \varepsilon) = (10^{-3}, 10^{-2})$.

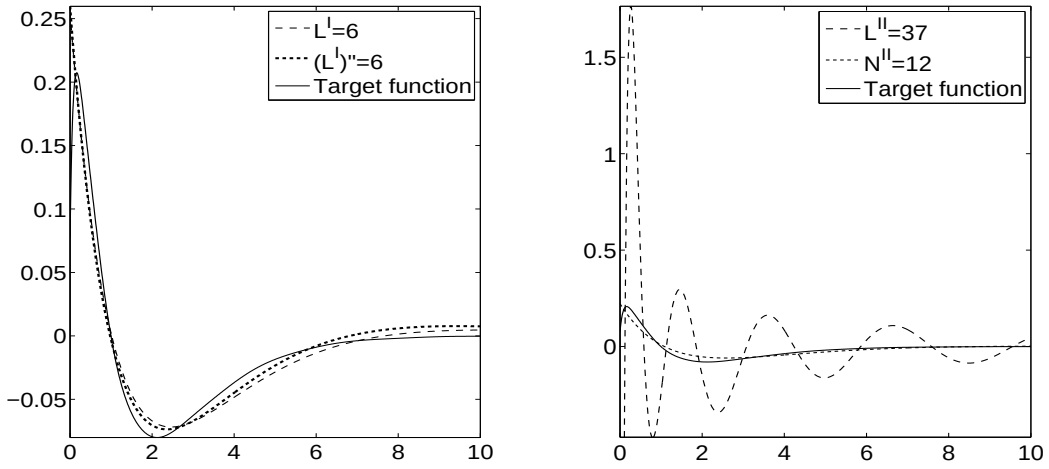
procedures, one needs condition 2.3 to be fulfilled. If not, the eigenvalues of the matrix $\boldsymbol{\Omega}_L$ become potentially too large, and observations (11) are not conveniently treatable. In this case, it is preferable to lower the maximal level down to a point where $\|\boldsymbol{\Omega}_L\|_{op}$ remains under control. To this end, we change the maximal level of the two respective procedures to

$$N^i = L^i \wedge \max\{\ell \geqslant 0 \text{ s.t. } \|\boldsymbol{\Omega}_\ell\|_{op} \leqslant \alpha\}, \quad i = \mathbf{I}, \mathbf{II}$$

where α is an arbitrary thresholding constant, set to 1.5 in the sequel. We now fix $\sigma = \delta = 10^{-2}$ and chose the design points t_i as $t_i = 100i/n$ for $n = 200, 250, 750$ and 1000. Taking the same kernel $\boldsymbol{g} = \boldsymbol{\varphi}_0$, and setting $\boldsymbol{f}_2(t) = (t^{1/2} - t)\exp(-t)$, we compare the performances of the new choice N^i to the previous one L^i, by computing the respective mean squared losses on a basis of 500 observations and report the result in Table 4. The results show a minor effect of the design ill-posedness on **Algorithm I**,

	$n =$	200	250	500	750
Algorithm I	$(L^{\mathbf{I}}, N^{\mathbf{I}})$	(6,6)	(6,6)	(6,6)	(6,6)
	MSE, $L^{\mathbf{I}}$	0.273	0.270	0.264	0.258
	MSE, $N^{\mathbf{I}}$	0.275	0.272	0.264	0.257
Algorithm II	$(L^{\mathbf{II}}, N^{\mathbf{II}})$	(37,12)	(37,15)	(37,27)	(37,27)
	MSE, $L^{\mathbf{II}}$	1.336	0.559	0.289	0.253
	MSE, $N^{\mathbf{II}}$	0.294	0.291	0.284	0.256

Table 4. Normalized mean squared error of the two procedures when the design is constituted of 200 equispaced points on the interval $[0; 100]$. We compare the performances of the two maximal resolution levels L^i and N^i for the parameters $\sigma = \delta = 10^{-2}$, $\boldsymbol{g} = \boldsymbol{\varphi}_0$ and $\boldsymbol{f}_2(t) = (t^{1/2} - t)\exp(-t)$.



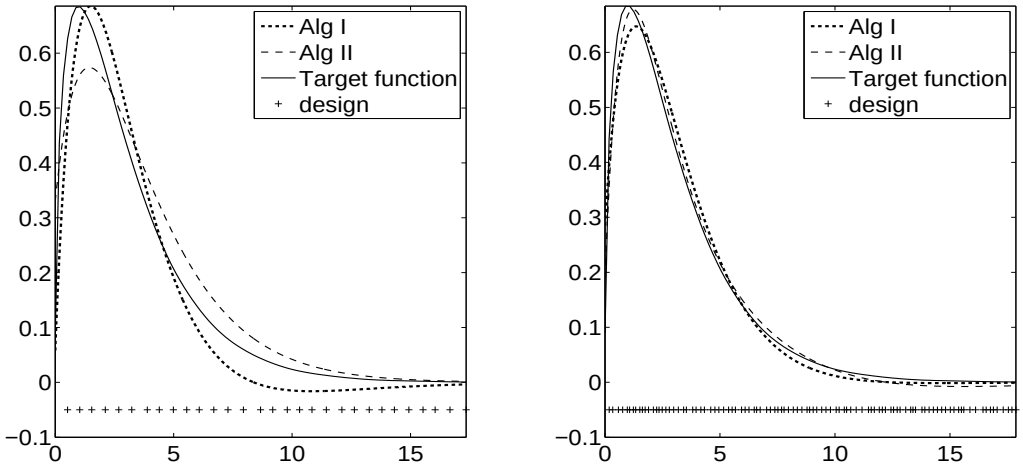
(a) **Algorithm I** (b) **Algorithm II**

Figure 3. Result of the two different maximal levels L^i and N^i to estimate $\boldsymbol{f}_2$, for a particular realizaton of $\boldsymbol{b}$ and $\boldsymbol{\xi}$. The design is constituted of 200 equidistant points of observations in $[0; 100]$. The noise levels are $\sigma = \delta = 10^{-2}$.

since L^i is usually already smaller than N^i. However, the gain is notable for **Algorithm II** when $n \leqslant 250$. To illustrate this point, we plot in Figure 3 the corresponding results when $n = 200$.

Back to the regression model

We now turn back to the original model (2) to apply the two procedures. It is well known that this model is asymptotically equivalent to (11), in the sense that a fine



(a) step=5.10^{-1}; n=30 (MSE**I**,MSE**II**)=(0.166,0.177)

(b) step=10^{-1}; n=100 (MSE**I**,MSE**II**)=(0.118,0.133)

Figure 4. Adaptation of the procedures to the regression framework. Here, MSE denotes the normalized mean squared error for each algorithm (computed with 500 realizations). The target function is $\boldsymbol{f}_3$ and the noise levels are $\sigma = \delta = 5.10^{-2}$.

enough design will provide an estimation of the Laguerre coefficients with a negligible error when $n \to \infty$. We work with $\boldsymbol{f}_3(z) = (1-z)^{1/2}$, $\boldsymbol{g} = \boldsymbol{\varphi}_0$, $\delta = 10^{-2}$, and suppose that the design is constituted of the points $t_i = \sum_{j=1}^{i} (\text{step} + |X_j|)$ where $(X_j)_{j \leqslant n}$ is an i.i.d sequence of $\mathcal{N}(0, 10^{-2})$ variables. We observe the noisy values $\boldsymbol{y}(t_i) = \boldsymbol{q}(t_i) + \sigma\eta_i$ where $\boldsymbol{q}(z) = (1-z)^{3/2}$, and compute the Laguerre coefficients $\check{\boldsymbol{q}}_\ell$ via the approximation

$$\check{\boldsymbol{q}}_\ell \sim \sum_{i=1}^{n-1} \frac{\boldsymbol{q}(t_i)\boldsymbol{\varphi}_\ell(t_i) + \boldsymbol{q}(t_{i+1})\boldsymbol{\varphi}_\ell(t_{i+1})}{2}(t_{i+1} - t_i)$$

We apply the two procedures and present the results on Figure 4.

6. Proofs

In the sequel, for the sake of clarity, we suppose that $\tau_{sig} = \tau_{\text{op}} \stackrel{\Delta}{=} \tau$.

6.1. Proof of Proposition 4.2

Proof. We can restrict ourselves to the case where $\mu = 1$. Proposition 16 applied to equality (22) entails

$$\forall \ell \geqslant 0,\ T_\ell\big((1/\dot{\boldsymbol{g}})\big) = C^{-1} T_\ell\big(w^{-1}\big) T_\ell\big((1-z)^{-\nu}\big)$$
$$T_\ell\big((1-z)^{-\nu}\big) = C T_\ell\big(w\big) T_\ell\big((1/\dot{\boldsymbol{g}})\big)$$

As a consequence,

$$\sum_{k=0}^{\ell} \boldsymbol{\gamma}_k^2 = \|(1/\dot{\boldsymbol{g}})^\ell\|^2 \leqslant C^{-1}\|T_\ell(w^{-1})\|_{\mathrm{op}}\|(1-z)_\ell^{-\nu}\|^2 \tag{24}$$

$$\text{and} \quad \|(\boldsymbol{K}^\ell)^{-1}\|_{\mathrm{HS}} \geqslant C^{-1}\|T_\ell\big(w\big)\|_{\mathrm{op}}^{-1}\|T_\ell\big((1-z)^{-\nu}\big)\|_{\mathrm{HS}} \tag{25}$$

Since w is assumed to have no zeros on $\mathcal{C}$, we deduce from Proposition 3.1 that both $\|w^{-1}\|_{circ}$ and $\|w\|_{circ}$ are finite, and from (15) that

$$\|T_\ell\big(w^{-1}\big)\|_{\mathrm{op}} \asymp 1 \text{ and } \|T_\ell\big(w\big)\|_{\mathrm{op}} \asymp 1$$

It remains to treat the binomial serie $(1-z)^{-\nu}$. This serie can be expanded as

$$(1-z)^{-\nu} = \sum_{\ell\geqslant 0}(-1)^\ell \binom{-\nu}{\ell} z^\ell$$

, where $\binom{-\nu}{\ell} \triangleq \frac{\Gamma(-\nu+1)}{\Gamma(\ell+1)\Gamma(-\nu-\ell+1)}$ is the generalized binomial coefficient. Furthermore, we have

$$\binom{-\nu}{\ell} \underset{\ell\to\infty}{\sim} \frac{(-1)^\ell}{\Gamma(\nu)\ell^{-\nu+1}} \tag{26}$$

which is a direct consequence of Euler's definition of the Gamma function $\Gamma(z) = \lim_{k\to\infty} \frac{k!k^z}{\Pi_{i=0}^k(z+i)}$. Since $\nu > 1/2$, the serie $\sum_k \binom{-\nu}{k}^2$ is hence divergent, and there exists $\widetilde{Q}_2, \widetilde{Q}_1 > 0$ such that, for all $\ell \geqslant 0$,

$$\sum_{k=0}^{\ell}\binom{-\nu}{k}^2 \leqslant \widetilde{Q}_1(\ell\vee 1)^{2\nu-1} \quad \text{and} \quad \sum_{k=0}^{\ell}\sum_{n=0}^{k}\binom{-\nu}{n}^2 \geqslant \widetilde{Q}_2(\ell\vee 1)^{2\nu}$$

The proof is complete thanks to (24) and (25). □

6.2. Proofs of theorems 3.3 and 4.5

6.2.1. Preliminary lemmas We begin with the following lemmas. Lemma 6.1 is a concentration inequality on the variable $\|\boldsymbol{B}^\ell\|_{\mathrm{op}}$, which results from a concentration inequality on subgaussian processes. Lemma 6.2 states that $\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{\mathrm{op}}$ behaves as $\|(\boldsymbol{K}^\ell)^{-1}\|_{\mathrm{op}}$ on a set with large probability. Finally, Lemma 6.3 establishes deviations bounds on the variables $\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell$ which will be useful throughout the proofs of Theorem 3.3 and Theorem 4.5.

Lemma 6.1. *There exists β_0, c_0 independent from $\ell \geqslant 0$, such that, for all $\ell \geqslant 0$, for all $t \geqslant \beta_0$,*

$$\mathbb{P}\Big(\frac{1}{\sqrt{(\ell\vee 1)\log(\ell\vee 2)}}\|\boldsymbol{B}^\ell\|_{op} > t\Big) \leqslant \exp(-c_0 t^2)$$

This readily entails the following moments control, available for all $\ell \geqslant 0$, $p \geqslant 1$

$$\mathbb{E}\|\boldsymbol{B}^\ell\|_{op}^p \lesssim \big(\ell\log\ell\big)^{p/2} \vee 1$$

Proof. The proof is a slight modification of Meckes [24, Theorem 1], to which we refer for a complete study. Lemma 6.1 is trivially satisfied if $\ell = 0, 1$, hence we will suppose that $\ell \geqslant 2$. From Proposition 3.1, we derive that

$$\mathbb{E}\|\boldsymbol{B}^{\ell}\|_{\text{op}} \leqslant \mathbb{E}\|T\big((\boldsymbol{b})_k\big)\|_{\text{op}} = \mathbb{E}\sup_{x\in[0,1]}|Y_x|, \qquad Y_x \stackrel{\Delta}{=} \sum_{k=0}^{\ell} \boldsymbol{b}_\ell e^{2i\pi kx}$$

We claim the two following facts:

- Let $a_0, ..., a_\ell \in \mathbb{R}$. There exists $c \geqslant 0$ such that ofor all $t > 0$,

$$\mathbb{P}\big(\big|\sum_{k=0}^{\ell} a_k \boldsymbol{b}_k\big| > t\big) \leqslant \exp\big(\frac{-ct^2}{\sum_{k=0}^{\ell} a_k^2}\big) \tag{27}$$

- $d(x, y) \stackrel{\Delta}{=} \sqrt{\mathbb{E}|Y_x - Y_y|^2} \leqslant 4\ell^{3/2}|x - y| \wedge 2\sqrt{\ell}$

The first point is readily verified since $(\boldsymbol{b}_k)_{k\leqslant\ell}$ is a standard Gaussian vector, while the second point directly results from the bound

$$\big|e^{2i\pi kx} - e^{2i\pi ky}\big| \leqslant 2 \wedge 2\pi k|x - y| \text{ for all } x, y \in [0, 1],\ k \geqslant 0$$

A direct application of Dudley's entropy bound (Talagrand [27, Proposition 2.1]) now entails

$$\mathbb{E}\sup_{x\in[0,1]}|Y_x| \lesssim (\ell \log \ell)^{1/2}$$

(see Meckes [24] for the rest of the proof). The deviation bound is now a consequence of Talagrand [27, Lemma 5.3]. Indeed, for all $x \in [0, 1]$,

$$E|Y_x|^2 = \mathbb{E}\big|\sum_{k=0}^{\ell} \boldsymbol{b}_k e^{2i\pi kx}\big|^2 \lesssim \ell$$

which ends the proof. □

Lemma 6.2. *Let $\ell \geqslant 0$, $a_\ell = \rho O_{\ell,\delta}$ for some $0 < \rho < \frac{1}{2}$. Note $\boldsymbol{\gamma}_\delta(z) = \sum_{\ell\geqslant 0} \boldsymbol{\gamma}_{k,\delta} z^k$ the power series associated to $(\boldsymbol{K}_\delta^\ell)^{-1}$. On $\boldsymbol{A}_\ell \stackrel{\Delta}{=} \{\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{op} \leqslant O_{\ell,\delta}^{-1}\}$ and $\boldsymbol{B}_\ell \stackrel{\Delta}{=} \{\|\delta \boldsymbol{B}^{\boldsymbol{\ell}}\|_{op} \leqslant a_\ell\}$, the following inequalities hold*

$$\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{op} \leqslant \frac{\rho}{1-\rho}\|(\boldsymbol{K}^\ell)^{-1}\|_{op} \text{ and } \|(\boldsymbol{K}^\ell)^{-1}\|_{op} \leqslant (1-\rho)^{-1}\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{op} \tag{28}$$

$$\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{HS} \leqslant \frac{\rho}{1-\rho}\|(\boldsymbol{K}^\ell)^{-1}\|_{HS} \text{ and } \|(\boldsymbol{K}^\ell)^{-1}\|_{HS} \leqslant (1-\rho)^{-1}\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{HS} \tag{29}$$

$$\sum_{k=0}^{\ell} \boldsymbol{\gamma}_{k,\delta}^2 \leqslant \frac{\rho}{1-\rho}\sum_{k=0}^{\ell} \boldsymbol{\gamma}_k^2 \text{ and } \sum_{k=0}^{\ell} \boldsymbol{\gamma}_k^2 \leqslant (1-\rho)^{-1}\sum_{k=0}^{\ell} \boldsymbol{\gamma}_{k,\delta}^2 \tag{30}$$

Proof. First, we have

$$\big(\boldsymbol{K}^\ell\big)^{-1} = \big(\boldsymbol{K}^\ell_\delta - \delta \boldsymbol{B}^\ell\big)^{-1} = \big(\boldsymbol{I} - \delta(\boldsymbol{K}^\ell_\delta)^{-1}\boldsymbol{B}^\ell\big)^{-1}\big(\boldsymbol{K}^\ell_\delta\big)^{-1}$$

On $\boldsymbol{A}_\ell \cap \boldsymbol{B}_\ell$, since a_ℓ satisfies $O^{-1}_{\ell,\delta}\, a_\ell = \rho < \frac{1}{2}$, by a usual Neumann series argument, we have

$$\begin{aligned}
\|\big(\boldsymbol{K}^\ell\big)^{-1}\|_{\mathrm{op}} &= \Big\|\Big[\sum_{k\geqslant 0}\big(-\delta(\boldsymbol{K}^\ell_\delta)^{-1}\boldsymbol{B}^\ell\big)^k\Big]\big(\boldsymbol{K}^\ell_\delta\big)^{-1}\Big\|_{\mathrm{op}} \\
&\leqslant \Big[\sum_{k\geqslant 0}\delta^k\|\big(\boldsymbol{K}^\ell_\delta\big)^{-1}\|^k_{\mathrm{op}}\|\boldsymbol{B}^\ell\|^k_{\mathrm{op}}\Big]\|\big(\boldsymbol{K}^\ell_\delta\big)^{-1}\|_{\mathrm{op}} \\
&\leqslant \Big[\sum_{k\geqslant 0}\rho^k\Big]\|\big(\boldsymbol{K}^\ell_\delta\big)^{-1}\|_{\mathrm{op}} \\
&\leqslant (1-\rho)^{-1}\|\big(\boldsymbol{K}^\ell_\delta\big)^{-1}\|_{\mathrm{op}} \qquad (31)
\end{aligned}$$

Secondly, we have

$$\big(\boldsymbol{K}^\ell_\delta\big)^{-1} = \big(\boldsymbol{K}^\ell + \delta \boldsymbol{B}^\ell\big)^{-1} = \big(\boldsymbol{I} + \delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\big)^{-1}\big(\boldsymbol{K}^\ell\big)^{-1}$$

Moreover, thanks to (31), on $\boldsymbol{A}_\ell \cap \boldsymbol{B}_\ell$, we have

$$\|\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\|_{\mathrm{op}} \leqslant (1-\rho)^{-1}O^{-1}_{\ell,\delta}\, a_\ell \leqslant \frac{\rho}{1-\rho} < 1 \qquad (32)$$

So that we can now similarly derive

$$\|\big(\boldsymbol{K}^\ell_\delta\big)^{-1}\|_{\mathrm{op}} \leqslant \frac{\rho}{1-\rho}\|\big(\boldsymbol{K}^\ell\big)^{-1}\|_{\mathrm{op}}$$

This prooves (28). The proofs of (29) and (30) follow the same lines, since $\|AB\|_{\mathrm{HS}} \leqslant \|A\|_{\mathrm{op}}\|B\|_{\mathrm{HS}}$, and $\|Ab\| \leqslant \|A\|_{\mathrm{op}}\|b\|$. □

6.2.2. Proof of theorem 3.3

Lemma 6.3. *Under Assumption 3.2, we have, for all $\ell \geqslant 0$,*

$$\mathbb{E}\Big[\Big|\langle(\boldsymbol{K}^\ell_\delta)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\big(-\delta\boldsymbol{B}^\ell\boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}_{L^{\boldsymbol{I}}}\big), \boldsymbol{\varphi}_\ell\rangle\Big|^q\Big] \lesssim (\ell\vee 1)^{q\nu}(\varepsilon\vee\delta)^q \qquad (33)$$

$$\mathbb{P}\Big(\Big|\langle(\boldsymbol{K}^\ell_\delta)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\big(-\delta\boldsymbol{B}^\ell\boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}_{L^{\boldsymbol{I}}}\big), \boldsymbol{\varphi}_\ell\rangle\Big| > S^{\boldsymbol{I}}_{\ell,\varepsilon}\Big) \lesssim \varepsilon^{\tau^2}\vee\delta^\tau \qquad (34)$$

Proof. In order to prove Inequalities (33) and (34), it suffices to study the tails of the random variables $\Big|\langle(\boldsymbol{K}^\ell_\delta)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\big(-\delta\boldsymbol{B}^\ell\boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}_{L^{\mathrm{I}}}\big), \boldsymbol{\varphi}_\ell\rangle\Big|$. For convenience we will only treat the case where $\ell \geqslant 2$, otherwise the result follows by identical arguments. To this end, we study each term apart. On $\boldsymbol{A}_\ell \cap \boldsymbol{B}_\ell$, Lemma 6.2 and Assumption 3.2 entail

$$\|(\boldsymbol{K}^\ell_\delta)^{-1}\|_{\mathrm{op}} \leqslant \frac{Q\rho}{1-\rho}\ell^\nu$$

Thus, combining Assumption 2.3 with the latter inequality, a brief conditionning argument readily yields

$$\mathbb{P}\Big(\Big|\langle(\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\varepsilon\boldsymbol{\xi}_{L^{\mathbf{I}}},\boldsymbol{\varphi}_\ell\rangle\Big|>t\Big)\lesssim\exp\big(-\frac{t^2}{\varepsilon^2\ell^{2\nu}}\big)$$

Let us study the second term. On $\boldsymbol{A}_\ell\cap\boldsymbol{B}_\ell$, we have

$$\delta(\boldsymbol{K}_\delta^\ell)^{-1}\boldsymbol{B}^\ell=\sum_{k\geqslant1}\big(\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\big)^k=\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell+\sum_{k\geqslant2}\big(\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\big)^k$$

Hence,

$$\delta(\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\boldsymbol{B}^\ell\boldsymbol{f}^\ell=r_1+r_2 \tag{35}$$

where

$$\begin{cases}r_1&=\langle\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\boldsymbol{f}^\ell,\boldsymbol{\varphi}_\ell\rangle\\ r_2&=\langle\big(\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\big)^2\big(\boldsymbol{I}+\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\big)^{-1}\boldsymbol{f}^\ell,\boldsymbol{\varphi}_\ell\rangle\end{cases} \tag{36}$$

Let's now bound separately r_1 and r_2. We first apply equality (16) to get

$$\begin{aligned}\langle(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\boldsymbol{f}^\ell,\boldsymbol{\varphi}_\ell\rangle&=\langle(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell,{}^{\boldsymbol{t}}\boldsymbol{B}^\ell\boldsymbol{\varphi}_\ell\rangle\\&=\langle(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell,(\boldsymbol{b}^\ell)'\rangle\end{aligned}$$

where $(\boldsymbol{b}^\ell)'_k=(\boldsymbol{b}^\ell)_{\ell-k}$. The result is a centred gaussian variable with variance

$$\|\delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell\|^2\leqslant\delta^2Q^2M^2\ell^{2\nu}$$

which hence satisfies

$$\mathbb{P}(|r_1|>t)\lesssim\exp\big(\frac{-t^2}{\delta^2\ell^{2\nu}}\big)$$

Let us study the term r_2. Since the maximal level L verifies $L\leqslant\lambda(\delta|\log\delta|)^{-\frac{1}{\nu+1}}$, we have, for all $\ell\leqslant L$, $\delta\ell^{\nu+1}\log\ell\lesssim1$. We deduce that

$$\begin{aligned}\mathbb{P}(|r_2|>t)&\leqslant\mathbb{P}(\delta^2\ell^{2\nu}\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2>t)\\&\leqslant\mathbb{P}(\frac{1}{\ell\log\ell}\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2>\delta^{-2}\ell^{-2\nu-1}(\log\ell)^{-1}t)\\&\lesssim\mathbb{P}(\frac{1}{\ell\log\ell}\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2>\delta^{-1}\ell^{-\nu}t)\\&\lesssim\exp(-t\big(\delta\ell^\nu\big)^{-1})\mathbf{1}_{\{t>\beta_0\delta\ell^\nu\}}+\mathbf{1}_{\{t\leqslant\beta_0\delta\ell^\nu\}}\end{aligned}$$

inequality (34) directly follows, and inequality (33) is now a direct application of the well known formula

$$\mathbb{E}[X^2]=\int_{t>0}2t\mathbb{P}(|X|>t)dt$$

□

Proof of theorem 3.3. We apply Parseval's formula to derive

$$\mathbb{E}\|\tilde{\boldsymbol{f}}^{\mathbf{I}} - \boldsymbol{f}\|_2^2 = \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\langle \tilde{\boldsymbol{f}} - \boldsymbol{f}, \boldsymbol{\varphi}_\ell\rangle^2 + \sum_{\ell > L^{\mathbf{I}}} \check{\boldsymbol{f}}_\ell^2$$

The second term is easily handled. Remark first that, since $s > 1/2$, we have $\frac{2s}{\nu+1} > \frac{2s}{s+\nu+1/2}$ and we can write

$$\begin{aligned}\sum_{\ell > L^{\mathbf{I}}} \check{\boldsymbol{f}}_\ell^2 &\leqslant \big(L^{\mathbf{I}}\big)^{-2s} \\ &\leqslant \big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{\frac{2s}{\nu+1}} \vee \big(\delta|\log\delta|\big)^{\frac{2s}{\nu+1}} \\ &\leqslant \big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{\frac{4s}{2(s+\nu)+1}} \vee \big(\delta|\log\delta|\big)^{\frac{4s}{2(s+\nu)+1}}\end{aligned}$$

In order to lighten the notations, we will only consider the indexes $\ell \geqslant 2$ in the first term. This is of course not problematic, since an identical reasoning allows to bound the two remaining summands by the desired rates of convergence. We hence write the following decomposition

$$\begin{aligned}\sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\langle \tilde{\boldsymbol{f}} - \boldsymbol{f}, \boldsymbol{\varphi}_\ell\rangle^2 &= \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell)^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\}} + \mathbb{E}\sum_{\ell \leqslant L^{\mathbf{I}}} \check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| \leqslant S^{\mathbf{I}}_{\ell,\varepsilon}\}} \\ &\lesssim I + II + III + IV\end{aligned}$$

where

$$\begin{aligned}I &= \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\big(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell\big)^2 \mathbf{1}_{\boldsymbol{A}_\ell} \mathbf{1}_{\boldsymbol{B}_\ell} \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\}} \\ II &= \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| \leqslant S^{\mathbf{I}}_{\ell,\varepsilon}\}} \mathbf{1}_{\boldsymbol{A}_\ell} \mathbf{1}_{\boldsymbol{B}_\ell} \\ III &= \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\big(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell\big)^2 \mathbf{1}_{\boldsymbol{A}_\ell} \mathbf{1}_{\boldsymbol{B}_\ell^c} \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\}} + \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| \leqslant S^{\mathbf{I}}_{\ell,\varepsilon}\}} \mathbf{1}_{\boldsymbol{B}_\ell^c} \\ IV &= \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\boldsymbol{A}_\ell^c} \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| \leqslant S^{\mathbf{I}}_{\ell,\varepsilon}\}}\end{aligned}$$

- ***Term I and II.*** On $\boldsymbol{A}_\ell$, we have

$$\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell = \big\langle (\boldsymbol{K}_\delta^\ell)^{-1}\big(-\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}\big), \boldsymbol{\varphi}_\ell\big\rangle \tag{37}$$

Hence we can decompose further I as

$$\begin{aligned}I &\lesssim \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\langle (\boldsymbol{K}_\delta^\ell)^{-1} \mathbf{1}_{\boldsymbol{A}_\ell} \mathbf{1}_{\boldsymbol{B}_\ell}\big(-\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}\big), \boldsymbol{\varphi}_\ell\rangle^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\}} \\ &\lesssim \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\langle (\boldsymbol{K}_\delta^\ell)^{-1} \mathbf{1}_{\boldsymbol{A}_\ell} \mathbf{1}_{\boldsymbol{B}_\ell}\big(-\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}\big), \boldsymbol{\varphi}_\ell\rangle^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\}} \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell| \geqslant S^{\mathbf{I}}_{\ell,\varepsilon}/2\}} \\ &\quad + \sum_{\ell \leqslant L^{\mathbf{I}}} \mathbb{E}\langle (\boldsymbol{K}_\delta^\ell)^{-1} \mathbf{1}_{\boldsymbol{A}_\ell} \mathbf{1}_{\boldsymbol{B}_\ell}\big(-\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}\big), \boldsymbol{\varphi}_\ell\rangle^2 \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\}} \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell| < S^{\mathbf{I}}_{\ell,\varepsilon}/2\}} \\ &\triangleq V + VI\end{aligned}$$

Let us first treat the term VI. From Lemmas 6.2 and 6.3 and Cauchy-Schwarz inequality, we derive

$$\begin{aligned} VI &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\Big[\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\big(\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}\big), \boldsymbol{\varphi}_\ell\rangle^4\Big]^{1/2} \\ &\qquad\qquad .\mathbb{P}(|\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon})^{1/2} \\ &\lesssim \sum_{\ell\leqslant L^{\mathbf{I}}} \big[(\delta\vee\varepsilon)^4 \ell^{4\nu}\big]^{1/2}\big(\delta^{\tau/2}\vee\varepsilon^{\tau^2/2}\big) \end{aligned}$$

which is less than the desired bound for τ large enough. As for term V, we split it in two and write

$$\begin{aligned} V &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\big(\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}\big), \boldsymbol{\varphi}_\ell\rangle^2 \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\geqslant S^{\mathbf{I}}_{\ell,\varepsilon}/2\}} \\ &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\delta \boldsymbol{B}^\ell \boldsymbol{f}^\ell, \boldsymbol{\varphi}_\ell\rangle^2 \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\geqslant \ell^\nu\delta|\log\delta|/2\}} \\ &\quad + \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\varepsilon\boldsymbol{\xi}, \boldsymbol{\varphi}_\ell\rangle^2 \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\geqslant \ell^\nu\varepsilon\sqrt{|\log\varepsilon|}/2\}} \\ &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \delta^2\ell^{2\nu}\big(\check{\boldsymbol{f}}_\ell^2\big(\ell^\nu\,\delta|\log\delta|\big)^{-2}\wedge 1\big) \\ &\qquad\qquad \vee \sum_{\ell\leqslant L^{\mathbf{I}}} \varepsilon^2\ell^{2\nu}\big(\check{\boldsymbol{f}}_\ell^2\big(\ell^\nu\varepsilon\sqrt{|\log\varepsilon|}\big)^{-2}\wedge 1\big) \end{aligned}$$

Note $\ell_\delta = \big(\delta|\log\delta|\big)^{\frac{-2}{2(s+\nu)+1}}$ and write

$$\begin{aligned} \sum_{\ell\leqslant L^{\mathbf{I}}} \delta^2\ell^{2\nu}\big(\check{\boldsymbol{f}}_\ell^2\big(\ell^\nu\delta|\log\delta|\big)^{-2}\wedge 1\big) &\leqslant \sum_{\ell\leqslant \ell_\delta}\delta^2\ell^{2\nu} + \sum_{\ell>\ell_\delta}\check{\boldsymbol{f}}_\ell^2|\log\delta|^{-2} \\ &\lesssim \big(\delta|\log\delta|\big)^{\frac{4s}{2(s+\nu)+1}} \end{aligned}$$

The ε-term is treated similarly by taking $\ell_\varepsilon = \big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{\frac{-2}{2s+2\nu+1}}$ and leads to the desired convergence rate. As for the term II, a similar reasonning leads to

$$\begin{aligned} II &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2\mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|\leqslant S^{\mathbf{I}}_{\ell,\varepsilon}\}}\mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\leqslant 2S^{\mathbf{I}}_{\ell,\varepsilon}\}} + \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2\mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|\leqslant S^{\mathbf{I}}_{\ell,\varepsilon}\}}\mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|> 2S^{\mathbf{I}}_{\ell,\varepsilon}\}} \\ &\triangleq VII + VIII \end{aligned}$$

The term $VIII$ is handled as the term VI. Indeed,

$$VIII \leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \check{\boldsymbol{f}}_\ell^2\mathbb{P}\big(|\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell| > S^{\mathbf{I}}_{\ell,\varepsilon}\big) \leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \check{\boldsymbol{f}}_\ell^2(\varepsilon^{\tau^2}\vee\delta^\tau)$$

which is less than the desired rate for τ large enough. Finally, we have

$$\begin{aligned} VII &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\leqslant 2S^{\mathbf{I}}_{\ell,\varepsilon}\}} \\ &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\leqslant 2\tau\ell^\nu\varepsilon\sqrt{|\log\varepsilon|}\}} + \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \mathbf{1}_{\{|\check{\boldsymbol{f}}_\ell|\leqslant 2\tau\ell^\nu\delta|\log\delta|\}} \\ &\lesssim \sum_{\ell\leqslant\ell_\varepsilon} \ell^{2\nu}\varepsilon^2|\log\varepsilon| + \sum_{\ell>\ell_\varepsilon} \check{\boldsymbol{f}}_\ell^2 + \sum_{\ell\leqslant\ell_\delta} \ell^{2\nu}\delta^2|\log\delta| + \sum_{\ell>\ell_\delta} \check{\boldsymbol{f}}_\ell^2 \\ &\lesssim \big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{\frac{4s}{2(s+\nu)+1}} \vee \big(\delta|\log\delta|^2\big)^{\frac{4s}{2(s+\nu)+1}} \end{aligned}$$

□

- ***Term III***. We have

$$\begin{aligned} III &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\Big(\big(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell\big)^2 \mathbf{1}_{\boldsymbol{A}_\ell} + \check{\boldsymbol{f}}_\ell^2\Big)\mathbf{1}_{\boldsymbol{B}_\ell^c} \\ &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\Big[\big(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell\big)^4 \mathbf{1}_{\boldsymbol{A}_\ell}\Big]^{1/2} \mathbb{P}\big(\boldsymbol{B}_\ell^c\big)^{1/2} + \sum_{\ell\leqslant L^{\mathbf{I}}} \check{\boldsymbol{f}}_\ell^2 \mathbb{P}(\boldsymbol{B}_\ell^c) \end{aligned}$$

Moreover, Lemma 6.1 entails

$$\mathbb{P}\big(\boldsymbol{B}_\ell^c\big) \leqslant \delta^{\kappa^2\rho^2}$$

for all $\ell \geqslant 1$. It is hence clear that for κ large enough, the term III is less than the announced rate. □

- ***Term IV***. We claim that

$$\mathbf{1}_{\{\boldsymbol{A}_\ell^c\}} \leqslant \mathbf{1}_{\{\|(\boldsymbol{K}^\ell)^{-1}\|_{\mathrm{op}}\geqslant O_{\ell,\delta}^{-1}/2\}} + \mathbf{1}_{\{\|\delta\boldsymbol{B}^\ell\|_{\mathrm{op}}\geqslant O_{\ell,\delta}\}}$$

for all $\ell \geqslant 0$ (see Delattre et al. [10], Lemma 5.3). Hence,

$$\begin{aligned} IV &\leqslant \sum_{\ell\leqslant L^{\mathbf{I}}} \mathbb{E}\check{\boldsymbol{f}}_\ell^2 \big(\mathbf{1}_{\{\|(\boldsymbol{K}^\ell)^{-1}\|_{\mathrm{op}}\geqslant O_{\ell,\delta}^{-1}/2\}} + \mathbf{1}_{\{\|\delta\boldsymbol{B}^\ell\|_{\mathrm{op}}\geqslant O_{\ell,\delta}\}}\big) \\ &\triangleq VIII + IX \end{aligned}$$

Since $\|(\boldsymbol{K}^\ell)^{-1}\|_{\mathrm{op}} \leqslant Q_2\ell^\nu$, we have $\{\|(\boldsymbol{K}^\ell)^{-1}\|_{\mathrm{op}} \geqslant O_{\ell,\delta}^{-1}/2\} \subset \{\ell^{\nu+1/2}\sqrt{\log\ell} \geqslant c\big(\delta|\log\delta|\big)^{-1}\}$ where c is a constant depending only on Q_2 and κ. Hence

$$\begin{aligned} VIII &\leqslant \sum_{\ell\geqslant c\left(\delta|\log\delta|^{3/2}\right)^{-\frac{2}{2\nu+1}}} \check{\boldsymbol{f}}_\ell^2 \\ &\lesssim \big(\delta|\log\delta|\big)^{\frac{4s}{2\nu+2s+1}} \end{aligned}$$

As for IX, a quick application of 6.1 entails

$$\mathbb{P}(\|\delta\boldsymbol{B}^\ell\|_{\mathrm{op}} \geqslant O_{\ell,\delta}) \leqslant \delta^{\kappa^2}$$

,so that

$$IX \leqslant \sum_{\ell \leqslant L^{\mathrm{I}}} \check{\boldsymbol{f}}_\ell^2 \delta^{\kappa^2} \lesssim \delta^{\kappa^2}$$

which is less than the announced rate for κ large enough. □

It remains to put together the bounds of the four terms above to get the desired rates of convergence in theorem 3.3. □

6.2.3. Proof of theorem 4.5

Lemma 6.4. *Note, for* $\ell \geqslant 0$, $\overline{S^{II}_{\ell,\varepsilon}} \triangleq (\ell \vee 1)^{\nu-1/2}\big(\tau_{sig}\varepsilon\sqrt{|\log\varepsilon|} \vee \tau_{op}\delta|\log\delta|\big)$. *Under Assumption 4.1, we have, for all* $\ell \geqslant 0$, *for all* $q \geqslant 0$,

$$\mathbb{E}\Big[\big|\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}(-\delta\boldsymbol{B}^\ell\boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}_{L^{II}}), \boldsymbol{\varphi}_\ell\rangle\big|^q\Big] \lesssim (\ell \vee 1)^{q(\nu-1/2)}(\varepsilon \vee \delta)^q \tag{38}$$

$$\mathbb{P}\Big(\big|\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}(-\delta\boldsymbol{B}^\ell\boldsymbol{f}^\ell + \varepsilon\boldsymbol{\xi}_{L^{II}}), \boldsymbol{\varphi}_\ell\rangle\big| > \overline{S^{II}_{\ell,\varepsilon}}\Big) \lesssim \varepsilon^{\tau^2} \vee \delta^\tau \tag{39}$$

Proof. The proof is very similar to Lemma 6.3, whence we will just mention the notable changes compared to it. Once more, we shall only treat the case $\ell \geqslant 2$. First, we have

$$\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\varepsilon\boldsymbol{\xi}, \boldsymbol{\varphi}_\ell\rangle = \langle \varepsilon\boldsymbol{\xi}_{L^{\mathrm{II}}}, {}^t(\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\boldsymbol{\varphi}_\ell\rangle = \varepsilon\langle \boldsymbol{\xi}_{L^{\mathrm{II}}}, ((1/\dot{\boldsymbol{g}}_\delta)^\ell)'\rangle$$

so that a brief contitionning argument, combined with (30) and Assumption 2.3 entails

$$\mathbb{P}(|\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\varepsilon\boldsymbol{\xi}_{L^{\mathrm{II}}}, \boldsymbol{\varphi}_\ell\rangle| > t) \lesssim \exp(\frac{-t^2}{\varepsilon^2\ell^{2\nu-1}})$$

In order to treat the term $\mathbb{P}(|\langle (\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\delta\boldsymbol{B}^\ell\boldsymbol{f}^\ell, \boldsymbol{\varphi}_\ell\rangle| > t)$, we first establish a useful result for the sequel: if $\boldsymbol{g}$ satisfies (20), then

$$\|(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell\| = \|T_\ell(\boldsymbol{f})(1/\dot{\boldsymbol{g}})^\ell\| \leqslant \|T_\ell(\boldsymbol{f})\|_{\mathrm{op}}\|(1/\dot{\boldsymbol{g}})^\ell\|$$

Furthermore, thanks to Proposition 3.1 we have

$$\|T_\ell(\boldsymbol{f})\|_{\mathrm{op}} \leqslant \sum_{\ell\geqslant 0}|\check{\boldsymbol{f}}_\ell| \leqslant \sum_{\ell\geqslant 0}\ell^{2s}|\check{\boldsymbol{f}}_\ell|^2\sum_{\ell\geqslant 0}\ell^{-2s} \lesssim 1$$

since $\boldsymbol{f} \in \mathcal{W}^s(M)$ and $s > 1/2$. We derive that

$$\|(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell\|^2 \lesssim \ell^{2\nu-1} \tag{40}$$

Let us now bound the term of interest. Once more, we decompose it as $r_1 + r_2$ where r_1 and r_2 are defined in (36). We now apply Proposition 16 and (32) and derive

$$\begin{aligned}\mathbb{P}\Big(|r_1| > t\Big) &= \mathbb{P}\big(|\langle \delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{B}^\ell\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\boldsymbol{f}^\ell, \boldsymbol{\varphi}_\ell\rangle| > t\big)\\ &\leqslant \mathbb{P}\big(|\langle \delta(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell, {}^t\boldsymbol{B}^\ell\boldsymbol{\varphi}_\ell\rangle| > t\big)\end{aligned}$$

The latter is a gaussian random variable with variance $\delta^2\|(\boldsymbol{K}^\ell)^{-1}\boldsymbol{f}^\ell\|^2 \lesssim \delta^2\ell^{2\nu-1}$ where we used (40). Turning to the term r_2, we apply Proposition 16 to derive

$$\begin{aligned}\mathbb{P}\Big(|r_2|>t\Big) &= \mathbb{P}\big(\big|\langle\delta\big(\boldsymbol{B}^\ell\big)^2(\boldsymbol{K}_\delta^\ell)^{-1}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\boldsymbol{f}^{\ell,t}(\boldsymbol{K}^\ell)^{-1}\boldsymbol{\varphi}_\ell\rangle\big|>t\big)\\ &\lesssim \mathbb{P}\big(\delta\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2\|(\boldsymbol{K}_\delta^\ell)^{-1}\boldsymbol{f}^\ell\|\|^t(\boldsymbol{K}^\ell)^{-1}\boldsymbol{\varphi}_\ell\|\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}>t\big)\end{aligned}$$

We now apply (30) and (40) to get

$$\|(\boldsymbol{K}_\delta^\ell)^{-1}\boldsymbol{f}^\ell\|\|^t(\boldsymbol{K}^\ell)^{-1}\boldsymbol{\varphi}_\ell\| \lesssim \ell^{2\nu-1}$$

Hence,

$$\begin{aligned}\mathbb{P}\Big(|r_2|>t\Big) &\lesssim \mathbb{P}\big(\delta^2\ell^{2\nu-1}\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}>t\big)\\ &\lesssim \mathbb{P}\big(\frac{1}{\ell\log\ell}\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}>t(\delta^2\ell^{2\nu}\log\ell)^{-1}\big)\end{aligned}$$

Let us take a look back to Lemma 6.2. On $\boldsymbol{A}_\ell\cap\boldsymbol{B}_\ell$ we have prooved that

$$\|(\boldsymbol{K}_\delta^\ell)^{-1}\|_{\mathrm{HS}} \geqslant (1-\rho)Q_1\ell^\nu$$

so that $\boldsymbol{A}_\ell\cap\boldsymbol{B}_\ell \subset \Big\{\delta\ell^{\nu+1/2}\log\ell \lesssim 1\Big\}$. We deduce

$$\begin{aligned}\mathbb{P}(|r_2|>t) &\lesssim \mathbb{P}\Big(\frac{1}{\ell\log\ell}\|\boldsymbol{B}^\ell\|_{\mathrm{op}}^2>t(\delta\ell^{\nu-1/2})^{-1}\Big)\\ &\lesssim \exp\big(\frac{-t}{\delta\ell^{\nu-1/2}}\big)\mathbf{1}_{\left\{t>\beta_0^2\delta\ell^{\nu-1/2}\right\}}+\mathbf{1}_{\left\{t\leqslant\beta_0^2\delta\ell^{\nu-1/2}\right\}}\end{aligned}$$

The end of the proof is identical to Lemma 6.3.

□

Proof of Theorem 4.5. The proof is very similar to Theorem 3.3, whence we will just emphasize the notable changes compared to it. First, we apply Parseval's formula to derive

$$\mathbb{E}\|\tilde{\boldsymbol{f}}^{\mathbf{II}}-\boldsymbol{f}\|_2^2 = \sum_{\ell\leqslant L^{\mathbf{II}}}\mathbb{E}\langle\tilde{\boldsymbol{f}}^{\mathbf{II}}-\boldsymbol{f},\boldsymbol{\varphi}_\ell\rangle^2+\sum_{\ell>L^{\mathbf{II}}}\check{\boldsymbol{f}}_\ell^2$$

The second term is easily handled, since

$$\begin{aligned}\sum_{\ell>L^{\mathbf{II}}}\check{\boldsymbol{f}}_\ell^2 &\leqslant (L^{\mathbf{II}})^{-2s}\\ &\leqslant \big(\varepsilon\sqrt{|\log\varepsilon|}\vee\delta|\log\delta|\big)^{2s}\\ &\leqslant \big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{\frac{2s}{s+\nu}}\vee\big(\delta|\log\delta|\big)^{\frac{2s}{s+\nu}}\end{aligned}$$

To bound the first sum, we write the following decomposition

$$\begin{aligned}\sum_{\ell\leqslant L^{\mathbf{II}}}\mathbb{E}\|\tilde{\boldsymbol{f}}^{\mathbf{II}}-\boldsymbol{f}\|^2 &= \sum_{\ell\leqslant L^{\mathbf{II}}}\mathbb{E}\big(\boldsymbol{\zeta}_\ell-\check{\boldsymbol{f}}_\ell\big)^2\mathbf{1}_{\left\{|\boldsymbol{\zeta}_\ell|>S_{\ell,\varepsilon}^{\mathbf{II}}\right\}}+\mathbb{E}\sum_{\ell\leqslant L}\check{\boldsymbol{f}}_\ell^2\mathbf{1}_{\left\{|\boldsymbol{\zeta}_\ell|\leqslant S_{\ell,\varepsilon}^{\mathbf{II}}\right\}}\\ &\lesssim I+II+III+IV\end{aligned}$$

where

$$
\begin{aligned}
I &= \sum_{\ell\leqslant L^{\mathbf{II}}} \mathbb{E}\big(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell\big)^2 \mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell}\mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|>S^{\mathbf{II}}_{\ell,\varepsilon}\}} \\
II &= \sum_{\ell\leqslant L^{\mathbf{II}}} \mathbb{E}\check{\boldsymbol{f}}^2_\ell \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|\leqslant S^{\mathbf{II}}_{\ell,\varepsilon}\}}\mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}_\ell} \\
III &= \sum_{\ell\leqslant L^{\mathbf{II}}} \mathbb{E}\big(\boldsymbol{\zeta}_\ell - \check{\boldsymbol{f}}_\ell\big)^2 \mathbf{1}_{\boldsymbol{A}_\ell}\mathbf{1}_{\boldsymbol{B}^c_\ell}\mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|>S^{\mathbf{II}}_{\ell,\varepsilon}\}} + \sum_{\ell\leqslant L^{\mathbf{II}}} \mathbb{E}\check{\boldsymbol{f}}^2_\ell \mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|\leqslant S^{\mathbf{II}}_{\ell,\varepsilon}\}}\mathbf{1}_{\boldsymbol{B}^c_\ell} \\
IV &= \sum_{\ell\leqslant L^{\mathbf{II}}} \mathbb{E}\check{\boldsymbol{f}}^2_\ell \mathbf{1}_{\boldsymbol{A}^c_\ell}\mathbf{1}_{\{|\boldsymbol{\zeta}_\ell|\leqslant S^{\mathbf{II}}_{\ell,\varepsilon}\}}
\end{aligned}
$$

Thanks to Lemma 6.2 and the definition of $S^{\mathbf{II}}_{\ell,\varepsilon}$, we have

$$
\frac{Q_1}{1-\rho}\overline{S^{\mathbf{II}}_{\ell,\varepsilon}} \leqslant S^{\mathbf{II}}_{\ell,\varepsilon} \leqslant \frac{Q_2\rho}{1-\rho}\overline{S^{\mathbf{II}}_{\ell,\varepsilon}}
$$

on $\boldsymbol{A}_\ell \cap \boldsymbol{B}_\ell$. Thus, the Terms I and II can be treated identically to the preceding proof and yield the desired rates of convergence. The terms III and IV are treated exactly as in the preceding proof. □

6.3. Proof of theorem 4.6

Proof. The lower bound will not decrease for increasing noise levels δ and ε, whence it suffices to provide the case $\delta = 0$ and the case $\varepsilon = 0$ separately. In the sequel, c_i will denote a positive constant to be adjusted later and L will play the role of a maximal level. Also, we will note $\boldsymbol{K}_\nu$ (resp. $\dot{\boldsymbol{g}}_\nu$) the operator (resp. the function) associated with the Laurent serie $(1-z)^\nu_L$. The function $\dot{\boldsymbol{g}}_{-1/2}$ will play an essential role in the sequel. Unfortunately it is not square integrable. We thus begin with a preliminary lemma, which states that a minor modification corrects this defect.

Lemma 6.5. *Let $\boldsymbol{h}$ be the function associated to the Laurent serie* $\displaystyle\sum_{\ell\geqslant 0}\frac{(-1)^\ell}{\log(\ell\vee 2)}\binom{-1/2}{\ell}z^\ell$. *Then $\boldsymbol{h}$ is square integrable. Furthermore, for all $\nu \geqslant 0$, for all $\ell \leqslant L$,*

$$
\frac{(\ell\vee 1)^{\nu-1/2}}{\log(\ell\vee 2)} \lesssim |\langle \boldsymbol{K}_{-\nu}\boldsymbol{h}, \boldsymbol{\varphi}_\ell\rangle| \lesssim (\ell\vee 1)^{\nu-1/2}
$$

Proof of Lemma 6.5. $\boldsymbol{h}$ is trivially squared integrable thanks to (26) and Parseval's formula. Now, we have

$$
\langle \boldsymbol{K}_{-\nu}\boldsymbol{h}, \boldsymbol{\varphi}_\ell\rangle = (-1)^\ell \sum_{k=0}^{\ell}\binom{-\nu}{k}\binom{-1/2}{\ell-k}\log^{-1}\big((\ell-k)\vee 2\big)
$$

Moreover, since the product $\binom{-\nu}{k}\binom{-1/2}{\ell-k}$ has a constant sign for all $k\leqslant\ell$, we derive

$$
\log^{-1}(\ell\vee 2)\Big|\sum_{k=0}^{\ell}\binom{-\nu}{k}\binom{-1/2}{\ell-k}\Big| \leqslant |\langle \boldsymbol{K}_{-\nu}\boldsymbol{h}, \boldsymbol{\varphi}_\ell\rangle| \leqslant \Big|\sum_{k=0}^{\ell}\binom{-\nu}{k}\binom{-1/2}{\ell-k}\Big|
$$

but $(-1)^\ell \sum_{k=0}^{\ell} \binom{-\nu}{k}\binom{-1/2}{\ell-k}$ is precisely the ℓ^{th} coefficient of the power serie $(1-z)^{-\nu}(1-z)^{-1/2} = (1-z)^{-\nu-1/2}$ which satisfies, thanks to (26),

$$\left|(-1)^\ell \sum_{k=0}^{\ell} \binom{-\nu}{k}\binom{-1/2}{\ell-k}\right| \sim \frac{\ell^{\nu-1/2}}{\Gamma(\nu+1/2)}$$

This entails the result. □

• *Case* $\delta = 0$. For more clarity, we will suppose that $\boldsymbol{\xi}$ is a white noise (the proof readily adapts otherwise). Let hence $\boldsymbol{K}^0 = c_1\boldsymbol{K}_\nu$. Then $\boldsymbol{K}^0 \in \mathcal{G}_\nu(Q)$ for an appropriate constant c_1, thanks to Proposition 4.2. Following the arguments of Willer [28], it suffices to find $\boldsymbol{f}_0$, $\boldsymbol{f}_1$ such that

i) $\boldsymbol{f}_0, \boldsymbol{f}_1 \in \mathcal{W}^s(M)$
ii) $\|\boldsymbol{f}_0 - \boldsymbol{f}_1\|^2 \gtrsim \varepsilon^{\frac{2s}{s+\nu}} |\log\varepsilon|^{-2}$
iii) $K(\mathbb{P}_1, \mathbb{P}_2) \lesssim 1$ where $\mathbb{P}_i$ is the law of $\boldsymbol{y}$ under the hypothesis $\boldsymbol{f}_i$, and K is the Kullback-Leibler divergence.

Let $L = c_2\varepsilon^{\frac{-1}{s+\nu}}$. Set $\boldsymbol{f}_0 = 0$ and define $\boldsymbol{f}_1 = c_3\boldsymbol{K}_{-\nu}\boldsymbol{h}$.
Point i): $\boldsymbol{f}_0$ trivially belongs to the considered set. Moreover, Lemma 6.5 entails

$$\|\boldsymbol{f}_1\|^2_{\mathcal{W}^s} \lesssim \varepsilon^2 \sum_{\ell=0}^{L} \ell^{2s}\ell^{2\nu-1} \lesssim 1$$

Point ii): again, thanks to Lemma 6.5, we have

$$\|\boldsymbol{f}_0 - \boldsymbol{f}_1\|^2 \gtrsim \varepsilon^2 \sum_{\ell\leqslant L} \frac{\ell^{2\nu-1}}{\log^2(\ell\vee 2)} \gtrsim \varepsilon^2 L^{2\nu}(\log L)^{-2} \gtrsim \varepsilon^{\frac{2s}{s+\nu}}|\log\varepsilon|^{-2}$$

Point iii): the expression of the Kullback-Leibler divergence in this case is

$$K(\mathbb{P}_1, \mathbb{P}_2) = \frac{1}{2}\|\varepsilon^{-1}\boldsymbol{K}^0(\boldsymbol{f}_0 - \boldsymbol{f}_1)\|^2 = \frac{c_3}{2}\|\boldsymbol{h}\|^2 \lesssim 1$$

thanks to Lemma 6.5. The choice of appropriate constants c_i clearly yields the result and the proof is complete. □

• *Case* $\varepsilon = 0$. Let $L = c_1\delta^{\frac{-1}{s+\nu}}$. Following the lines of Hoffmann and Reiß [17], we set $\boldsymbol{f}_0 = c_2\varphi_0$, $\boldsymbol{K}^0 = c_3\boldsymbol{K}_\nu$ and we only consider couples $(\boldsymbol{K}, \boldsymbol{f})$ such that $\boldsymbol{K}\boldsymbol{f} = \boldsymbol{q}_0$ for a fixed $\boldsymbol{q}_0 = \boldsymbol{K}^0\boldsymbol{f}_0$. It is clear that, for well chosen c_2 and c_3, we have $\boldsymbol{f}_0 \in \mathcal{W}^s(M)$ and $\boldsymbol{K}^0 \in \mathcal{G}_\nu(Q)$. We thus define $\boldsymbol{H}$ the operator associated to the kernel $\boldsymbol{h}$ and introduce $\boldsymbol{K}^\delta = \boldsymbol{K}_\nu + c_4\delta\boldsymbol{H}$ a perturbation of $\boldsymbol{K}_\nu$. We shall refer to $\dot{\boldsymbol{g}}^\delta$ for the corresponding kernel. Remark that we have

$$\boldsymbol{f}_1 - \boldsymbol{f}_0 = c_4\delta(\boldsymbol{K}^\delta)^{-1}\boldsymbol{H}\boldsymbol{f}_0 = c_4\delta(\boldsymbol{K}^\delta)^{-1}\boldsymbol{h} \tag{41}$$

Furthermore, for c_4 small enough, we have thanks to Lemma 6.5 and Proposition 4.2,

$$\|\boldsymbol{K}^{\delta} - \boldsymbol{K}^{0}\|_{\text{op}} \lesssim \delta L^{1/2} \lesssim \delta^{\frac{2s+2\nu-1}{2(s+\nu)}} < \frac{1}{2}$$

since $s > 1/2$. Hence, the same Neumann serie arguments as in Lemma 6.2 entail that $\boldsymbol{K}^{\delta}$ belongs to $\mathcal{G}_{\nu}(Q)$. We now need to check that i), ii) and iii) are satisfied, replacing ε with δ.
Point i) : (41) and the preceding remark entail

$$\|\boldsymbol{f}_1 - \boldsymbol{f}_0\|_{\mathcal{W}^s}^2 = c_4^2\delta^2 \sum_{\ell \leqslant L} \ell^{2s} \langle (\boldsymbol{K}^{\delta})^{-1}\boldsymbol{h}, \boldsymbol{\varphi}_{\ell}\rangle \lesssim \delta^2 \sum_{\ell \leqslant L} \ell^{2s+2\nu-1} \lesssim 1$$

Point ii) : we precise (41) and write

$$\boldsymbol{f}_1 - \boldsymbol{f}_0 = c_4\delta\boldsymbol{K}^{-1}\boldsymbol{h} + c_4^2\delta^2(\boldsymbol{K}^{\delta})^{-1}\boldsymbol{K}^{-1}\boldsymbol{H}\boldsymbol{h}$$

Moreover, Lemma 6.5 and the preceding remark entail

$$\|\delta\boldsymbol{K}^{-1}\boldsymbol{H}\boldsymbol{f}_0\|^2 = \|\delta\boldsymbol{K}^{-1}\boldsymbol{h}\|^2 \gtrsim \delta^2 \sum_{\ell \leqslant L} \frac{\ell^{2\nu-1}}{\log(\ell \vee 2)} \gtrsim \delta^2 L^{2\nu}(\log L)^{-2} \gtrsim \delta^{\frac{2s}{s+\nu}} |\log\delta|^{-2}$$

$$\|\delta^2(\boldsymbol{K}^{\delta})^{-1}\boldsymbol{K}^{-1}\boldsymbol{H}^2\boldsymbol{f}_0\|^2 \lesssim \delta^4 \|(\boldsymbol{K}^{\delta})^{-1}\|_{\text{HS}}\|\boldsymbol{K}^{-1}\boldsymbol{H}\|_{\text{HS}}\|\boldsymbol{h}\| \lesssim \delta^4 L^{4\nu+1} \lesssim \delta^{\frac{2s-1}{\nu+s}}\delta^2 L^{2\nu}$$

Since $s > 1/2$, the second term is negligible with respect to the first. This proves the point ii).
Point iii) : Since we work with couples $(\boldsymbol{K}, \boldsymbol{f})$ such that $\boldsymbol{K}\boldsymbol{f}$ is fixed, we have

$$K(\mathbb{P}_0, \mathbb{P}_1) = \frac{1}{2}\delta^{-2}\|\dot{\boldsymbol{g}}^{\delta} - \dot{\boldsymbol{g}}_{\nu}\|^2 = \frac{c_4}{2}\|\boldsymbol{h}\|^2 \lesssim 1$$

thanks to Lemma 6.5 and the proof is complete. □

It remains to piece together the two cases $\delta = 0$ and $\varepsilon = 0$ to get the desired result. □

Acknowledgements The author would like to thank his advisor Dominique Picard for her help and her support.

References

[1] J. Abate, G.L. Choudhury, and W. Whitt. On the laguerre method for numerically inverting laplace transforms. *INFORMS Journal on Computing*, 8:413–427, 1996.

[2] F. Abramovich and B. W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85:115129, 1998.

[3] F. Abramovich, M. Pensky, and Y. Rozenholc. Laplace deconvolution with noisy observations. *ArXiv:1107.2766v2*, 2012.

[4] M. Ameloot, J.M. Beechem, and L. Brand. Simultaneous analysis of multiple fluorescence decay curves by laplace transforms. deconvolution with reference or excitation profiles. *Biophys Chem.*, 23(3-4):155–71, 1986.

[5] B. Bongioanni and J. L. Torrea. What is a Sobolev space for the Laguerre function systems? *Studia Mathematica*, 192:147–172, 2009. doi: 10.4064/sm192-2-4.

[6] A. Böttcher and S.M. Grudsky. *Spectral Properties of Banded Toeplitz Matrices*. Siam edition, 2005.

[7] A.C. Cinzori and P.K. Lamm. Future polynomial regularization of ill-posed volterra equations. *SIAM J. Num. Anal.*, 37:949–979, 2000.

[8] A. Cohen, M. Hoffmann, and M. Reiß. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Numer. Anal.*, 42:1479–1501, 2004.

[9] F. Comte, C. Guenod, M. Pensky, and Y. Rozenholc. Laplace deconvolution and its application to dynamic contrast enhanced imaging. *hal-00715943*, 2012.

[10] S. Delattre, M. Hoffmann, D. Picard, and T. Vareschi. Blockwise svd with error in the operator and application to blind deconvolution. *Electronic Journal of Statistics*, 6:2274–2308, 2012.

[11] A.K. Dey, C.F. Martin, and F.H. Ruymgaart. Input recovery from noisy output data, using regularized inversion of laplace transform. *IEEE Trans. Inform. Theory*, 44:1125–1130, 1998.

[12] D. Donoho. Nonlinear solution of linear inverse problems by wavelet vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2:101–126, 1995.

[13] D. Donoho and I Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[14] S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transf. Inf. Theory*, 47:2876–2894, 2001.

[15] Y. Golubev. On universal oracle inequalities related to high-dimensional linear models. *AOS*, 38-(5):2751–2780, 2010.

[16] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. Academic press, new york edition, 1980.

[17] M. Hoffmann and M. Reiß. Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.*, 36:310–336, 2008.

[18] I. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66: 547–573, 2004.

[19] J. Keilson and W.R. Nunn. Laguerre transformation as a tool for the numerical solution of integral equations of convolution type. *Appl. Math. and Comput.*, 5(4): 313–359, 1979.

[20] P.K. Lamm. Variable-smoothing local regularization methods for first-kind integral equations. *Inverse problems*, 19:195–216, 2003.

[21] T.N. Lien, D.D Trong, and A.P.N. Dinh. Laguerre polynomials and the inverse laplace transform using discrete data. *J. Math. Anal. Appl.*, 337:1302–1314, 2008.

[22] P. Linz. *Analytical and Numerical Methods for Volterra Equations*. Siam edition, 1985.

[23] P. Mathe and S.V. Pereverzev. Geometry of linear ill-posed problems in variable hilbert scales. *Inverse problems*, 19:789–803, 2003.

[24] M.W. Meckes. On the spectral norm of a random toeplitz matrix. *arXiv:math/0703134v2*, 2007.

[25] M. Nussbaum and S.V. Pereverzev. The degree of ill-posedness in stochastic and deterministic models. *Preprint No. 509, Weierstrass Institute (WIAS), Berlin*, 1999.

[26] P.K. Rathnakumar. A localization theorem for laguerre expansions. *Proceedings of the Indian Academy of Sciences - Mathematical Sciences*, 105(3):303–314, 1995.

[27] M. Talagrand. Majorizing measures: the generic chaining. *Ann. Probab.*, 24(3): 1049–1103, 1996.

[28] Thomas Willer. Optimal bound for inverse problems with jacobi-type eigenfunctions. *Statist. Sinica*, 19(2):785–800, 2009.