Electronic Journal of Statistics
Vol. 6 (2012) 2274–2308
ISSN: 1935-7524
DOI: 10.1214/12-EJS745

Blockwise SVD with error in the operator and application to blind deconvolution

S. Delattre

Université Denis Diderot and CNRS-UMR 7099
175 rue du Chevaleret
75013 Paris, France

M. Hoffmann

Université Paris-Dauphine and CNRS-UMR 7534
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16, France

D. Picard and T. Vareschi

Université Denis Diderot and CNRS-UMR 7099
175 rue du Chevaleret
75013 Paris, France

Abstract: We consider linear inverse problems in a nonparametric statistical framework. Both the signal and the operator are unknown and subject to error measurements. We establish minimax rates of convergence under squared error loss when the operator admits a blockwise singular value decomposition (blockwise SVD) and the smoothness of the signal is measured in a Sobolev sense. We construct a nonlinear procedure adapting simultaneously to the unknown smoothness of both the signal and the operator and achieving the optimal rate of convergence to within logarithmic terms. When the noise level in the operator is dominant, by taking full advantage of the blockwise SVD property, we demonstrate that the block SVD procedure outperforms classical methods based on Galerkin projection [14] or nonlinear wavelet thresholding [18]. We subsequently apply our abstract framework to the specific case of blind deconvolution on the torus and on the sphere.

AMS 2000 subject classifications: Primary 62G05, 62G99; secondary 65J20, 65J22.
Keywords and phrases: Blind deconvolution, blockwise SVD, circular and spherical deconvolution, nonparametric adaptive estimation, linear inverse problems, error in the operator.

Received April 2012.

1. Introduction

1.1. Motivation

Consider the following idealised statistical problem: estimate a function f (a signal, an image) from data

$$y_n = Kf + n^{-1/2}\dot{W}, \tag{1.1}$$

where

$$K : \mathbb{H} \to \mathbb{G}$$

is a linear operator between two Hilbert spaces $\mathbb{H}$ and $\mathbb{G}$. The observation of the unknown $f \in \mathbb{H}$ is challenged by the action of the linear degradation K as well as contaminated by an experimental Gaussian white noise $\dot{W}$ on $\mathbb{G}$ with vanishing noise level $n^{-1/2}$ as $n \to \infty$. Alternatively, in a density estimation setting, we observe a random sample $(Z_1, \dots, Z_n)$ drawn from a probability distribution[1] with density Kf. In each case, we do not know the operator K exactly, but we have access to

$$K_\delta = K + \delta \dot{B}, \tag{1.2}$$

where $\dot{B}$ is a Gaussian white noise on $\mathbb{H} \times \mathbb{G}$ thanks to preliminary experiments or calibration through trial functions. This setting has been discussed in details in [14, 18]. In this paper, we are interested in operators K admitting a singular value decomposition (SVD) or a blockwise SVD. In essence, we know the typical eigenfunctions of K but not the eigenvalues. We cover two specific examples of interest: spherical and circular deconvolution.

Spherical deconvolution. Used for the analysis of data distributed on the celestial sphere, see Section 4.1 below. One observes a random sample $(Z_1, \dots, Z_n)$ with

$$Z_i = \varepsilon_i X_i, \quad i = 1, \dots, n$$

where the ε_i are random elements in $\mathcal{SO}(3)$, the group of 3×3 rotation matrices, and the X_i are independent and identically distributed on the sphere $\mathbb{S}^2$, with common density f with respect to the uniform probability distribution μ on $\mathbb{S}^2$. In this setting, if the ε_i have common density g with respect to the Haar measure du on $\mathcal{SO}(3)$, we have

$$Kf(x) = g \star f(x) = \int_{\mathcal{SO}(3)} g(u) f(u^{-1}x) du, \quad x \in \mathbb{S}^2.$$

We are interested in the case where the exact form g is unknown. However, K is block-diagonal in the spherical harmonic basis. □

Circular deconvolution. Used for restoring signal or images, see Section 4.2 below. We take $\mathbb{H} = \mathbb{G} = L^2(\mathbb{T})$ the space of square integrable functions on the torus $\mathbb{T} = [0, 1]$ (or $[0, 1]^d$) appended with periodic boundary conditions. We have

$$Kf(x) = g \star f(x) = \int_{\mathbb{T}} g(u) f(x - u) du, \quad x \in \mathbb{T}.$$

The degradation process $K = g \star \bullet$ is characterised by the impulse response function g which we do not know exactly. However, K is diagonal in the Fourier basis. □

[1] In that setting, Kf must therefore also be a probability density .

Although the problem of estimating f is fairly classical and well understood when K is known (a selected literature is [36, 8, 12, 1, 17, 33, 32] and the references therein), only moderate attention has been paid in the case of an unknown K despite its relevance in practice. When the eigenfunctions of K are known solely, we have the results of Cavalier and Hengartner [5], Cavalier and Raimondo [6] but they are confined to the case where the error in the operator is negligible $\delta \ll n^{-1/2}$. In a general setting with error in the operator, Efromovitch and Kolchinskii [14] and later Hoffmann and Reiß [18] studied the recovery of f when the eigenfunctions and eigenvalues of K are unknown. In both contributions, a marginal attention is paid to the case of sparse or diagonal operators, but it is shown in both papers that unusual rates of convergence can be obtained when $n^{-1/2} \ll \delta$. In a univariate setting, Neumann [28], Johannes [20] and Comte and Lacour [9] consider the case of deconvolution with an error density, known only through an auxiliary set of m learning data. This formally corresponds to having $\delta = m^{-1/2}$ in our setting. Minimax rates and adaptive estimators are derived in both regimes $m \ll n$ and $n \ll m$. We address in the paper the following program:

i) Construction of a feasible procedure $\widehat{f}_{n,\delta}$ estimating f from data (1.1) and (1.2) that achieves optimal rates of convergence (up to inessential logarithmic terms). We require $\widehat{f}_{n,\delta}$ to be adaptive with respect to smoothness constraints on f and K.

ii) Identification of best achievable accuracy for f under smoothness constraints on f and K so that the interplay between $n^{-1/2}$ and δ can be explicitly related in the asymptotic $\delta \to 0$ and $n \to \infty$; this includes the comparison with earlier results of [28, 14, 18] in the context of blockwise SVD.

iii) Application to spherical deconvolution on $\mathbb{S}^2$ or circular deconvolution on the torus; this includes the discussion of our findings in terms of the existing literature on the topic [7, 29, 26] and some practical aspects of numerical implementation.

1.2. Main results and organisation of the paper

In Section 2, we present an abstract framework that allows for operators K to admit a so-called blockwise SVD. This property is simply turned into the existence of pairs of increasing finite dimensional spaces $(H_\ell,\ G_\ell)$ that are stable under the action of K. The blockwise SVD property is further appended with a smoothness condition quantified by the arithmetic decay of the operator norm of K and its inverse on H_ℓ (resp. G_ℓ) (the so-called *ordinary smooth* assumption, see *e.g.* [36]). By means of a reconstruction formula, we obtain in Section 2.2 an estimator $\widehat{f}_{n,\delta}$ of f by first inverting K_δ on H_ℓ with a thresholding tuned with δ and then filter the resulting signal by a block thresholding tuned with $n^{-1/2}$. As for i) and ii), we establish in Theorems 1 and 3.4 of Section 3 the minimax rates of convergence for Sobolev constraints on f under squared error loss and we demonstrate that $\widehat{f}_{n,\delta}$ is optimal and adaptive to within logarithmic terms. The explicit interplay between δ and $n^{-1/2}$ is revealed and discussed in the case of a sparse operator when $n^{-1/2} \ll \delta$, completing earlier findings in [14, 18] and to some extent [28] in the univariate case for density deconvolution. In particular, we exhibit a certain parametric regime when the smoothness of the signal dominates the smoothing properties of the operator. Concerning iii), the method is applied to the case of spherical and circular deconvolution in Section 4 where harmonic Fourier analysis enables to provide explicit blockwise SVD for the convolution operator. We illustrate the numerical feasability of $\widehat{f}_{n,\delta}$ and the phenomena that appear in the case $n^{-1/2} \ll \delta$. Section 5 is devoted to the proofs.

We choose to state and prove our results in the white Gaussian model generated by the observation of y_n and K_δ defined by (1.1) and (1.2). The extension to the case of density estimation, when y_n is replaced by the observation of a random sample of size n drawn from the distribution Kf, like for instance in [28, 9, 20] readily carries over in the context of circular deconvolution, as proved in the discussion Section 3.2. The extension to a general blockwise setting is a bit more involved, due to the intrinsic heteroscedasticity that appears when enforcing a formal analogy with the Gaussian setting (1.1), see Section 3.2. Finally an anonymous referee is greatly acknowledged for pointing out the similarities of the present work with the independent recent paper of Johannes and Schwarz [22], see also [21] and the discussion Section 3.2.

2. Estimation by blockwise SVD

2.1. The blockwise SVD property

Let $\mathcal{G}$ denote a family of linear operators

$$K : \mathbb{H} \to \mathbb{G}$$

between two Hilbert spaces $\mathbb{H}$ and $\mathbb{G}$ that shall represent our parameter set of unknown K.

A fundamental property (Assumption 1 below) is that an explicit singular value decomposition (SVD) or blockwise SVD is known for all $K \in \mathcal{G}$ simultaneously. More specifically, we suppose that there exist two explicitly known bases $(e_\lambda,\ \lambda \in \Lambda)$ of $\mathbb{H}$ and $(g_\lambda,\ \lambda \in \Lambda)$ of $\mathbb{G}$, as well as a partition of $\Lambda = \cup_{\ell \geq 1}\Lambda_\ell$ with $\Lambda_\ell \cap \Lambda_{\ell'} = \emptyset$ if $\ell \neq \ell'$, and a constant $d \geq 1$ such that:

$$\ell^{d-1} \lesssim |\Lambda_\ell| \lesssim \ell^{d-1}, \tag{2.1}$$

where $\lesssim$ means inequality up to a multiplicative constant that does not depend on ℓ. Here $|\Lambda_\ell|$ is a short-hand notation for the cardinality of the set Λ_ℓ.

In our examples as well as in the rates of convergence that we will exhibit later, d plays the role of a dimension. In particular, $d = 1$ includes the standard SVD case of diagonal operators, whereas $d > 1$ creates increasing blocks with ℓ and deserves the name of 'blockwise SVD'. However, there is no need in the paper to assume that d is an integer. Set

$$H_\ell = \mathrm{Span}\{e_\lambda, \lambda \in \Lambda_\ell\} \text{ and } G_\ell = \mathrm{Span}\{g_\lambda, \lambda \in \Lambda_\ell\}.$$

The Galerkin projection of K onto (H_ℓ, G_ℓ) is denoted by K_ℓ and defined by $K_\ell = P_\ell K_{|H_\ell}$, where P_ℓ is the orthogonal projector onto G_ℓ. We denote by $\|K_\ell\|_{H_\ell \to G_\ell} = \sup_{v \in H_\ell, \|v\|_{\mathbb{H}}=1} \|K_\ell v\|_{\mathbb{G}}$ the operator norm of K_ℓ.

Assumption 1 (Blockwise SVD)**.** *For every $K \in \mathcal{G}$, and every $\ell \geq 1$, we have*

$$K_{|H_\ell} = K_\ell.$$

Moreover, K_ℓ is invertible and there exists $\nu \geq 0$ such that

$$Q_1(K) = \sup_{\ell \geq 1} \ell^{-\nu} \|(K_\ell)^{-1}\|_{G_\ell \to H_\ell} < \infty$$

and

$$Q_2(K) = \sup_{\ell \geq 1} \ell^{\nu} \|K_\ell\|_{H_\ell \to G_\ell} < \infty.$$

For every $f \in \mathbb{H}$, we have a decomposition associated to $(e_\lambda, \lambda \in \Lambda)$

$$f = \sum_{\ell \geq 1} \sum_{\lambda \in \Lambda_\ell} \langle f, e_\lambda \rangle \, e_\lambda$$

where $\langle \bullet, \bullet \rangle$ denotes the inner product (either in $\mathbb{H}$ or $\mathbb{G}$) and a scale of Sobolev spaces defined by

$$\mathcal{W}^s = \Big\{ f \in \mathbb{H}, \;\; \|f\|^2_{\mathcal{W}^s} = \sum_{\ell \geq 1} \ell^{2s} \sum_{\lambda \in \Lambda_\ell} \langle f, e_\lambda \rangle^2 < \infty \Big\}, \;\; s \in \mathbb{R}. \tag{2.2}$$

For every $g \in \mathbb{G}$, we have a decomposition $g = \sum_{\ell \geq 1} \sum_{\lambda \in \Lambda_\ell} \langle g, g_\lambda \rangle \, g_\lambda$ associated to $(g_\lambda, \lambda \in \Lambda)$, likewise. For $\nu \geq 0$, Assumption 1 implies that

$$K : \mathcal{W}^{-\nu/2} \to \widetilde{\mathcal{W}}^{\nu/2}$$

is continuous, where $\widetilde{\mathcal{W}}^s = \big\{ g \in \mathbb{G}, \;\; \|g\|^2_{\widetilde{\mathcal{W}}^s} = \sum_{\ell \geq 1} \ell^{2s} \sum_{\lambda \in \Lambda_\ell} \langle g, g_\lambda \rangle^2 < \infty \big\}$. In particular, when $\nu > 0$, the operator K is (mildly) ill-posed with degree ν, see for instance [30].

2.2. Blockwise SVD reconstruction with noisy data

Under Assumption 1 we have the inversion formula

$$f = \sum_{\ell \geq 0} (K_\ell)^{-1} \sum_{\lambda \in \Lambda_\ell} \langle Kf, g_\lambda \rangle \, e_\lambda \quad \text{for every } f \in \mathbb{H}\,. \tag{2.3}$$

However, since the observation of both K_ℓ and $\langle Kf, g_\lambda \rangle$ is blurred by noise, formula (2.3) is ineffective for the reconstruction process, unless some appropriate regularisation is operated. By the observed blurred version K_δ of K in (1.2), we obtain a family of estimators of $(K_\ell)^{-1}$ from data (1.2) by considering the operator

$$\mathbf{1}_{\left\{\|(K_{\delta,\ell})^{-1}\|_{G_\ell\to H_\ell}\leq\kappa\right\}}(K_{\delta,\ell})^{-1}, \tag{2.4}$$

where $\kappa > 0$ is a cutoff level, possibly depending on ℓ. Likewise, the coefficient $\langle Kf, g_\lambda\rangle$ can be estimated by

$$z_{n,\lambda} = \langle y_n, g_\lambda\rangle. \tag{2.5}$$

Mimicking (2.3) with the estimates (2.5) and (2.4), we finally obtain a (family of) estimator(s) of f by setting

$$\widehat{f}_{n,\delta} = \sum_{0\leq\ell\leq L}\big(K_{\delta,\ell}\big)^{-1}\Big(\sum_{\lambda\in\Lambda_\ell} z_{n,\lambda}e_\lambda\mathbf{1}_{\left\{\sum_{\lambda\in\Lambda_\ell} z_{n,\lambda}^2\geq\tau_\ell^2\right\}}\Big)\mathbf{1}_{\mathcal{E}_{\delta,\ell}(\kappa_\ell)}$$

where

$$\mathcal{E}_{\delta,\ell}(\kappa_\ell) = \big\{\|(K_{\delta,\ell})^{-1}\|_{G_\ell\to H_\ell}\leq\kappa_\ell\big\}.$$

The procedure is specified by the maximal frequency level L and the threshold levels

$$\kappa_\ell = \Big(\lambda_0|\Lambda_\ell|^{-1/2}\big(\delta^2|\log\delta|\big)^{-1/2}\Big)\bigwedge n^{1/2} \tag{2.6}$$

and

$$\tau_\ell = \mu_0|\Lambda_\ell|^{1/2}\big(n^{-1}\log n\big)^{1/2}, \tag{2.7}$$

for some prefactors $\lambda_0, \mu_0 > 0$. The threshold rule we introduce in both the signal (with level τ_ℓ) and the operator (with level κ_ℓ) is inspired by classical block thresholding [24, 4, 3] and will enable to adapt with respect to the smoothness properties of both the signal f and the operator K, see below.

3. Main results

3.1. Minimax rates of convergence

We assess the performance of the estimator $\widehat{f}_{n,\delta}$ defined in Section 2.2 over the Sobolev spaces linked to the basis $(e_\lambda, \lambda\in\Lambda)$ defined in (2.2). Define the Sobolev balls $\mathcal{W}^s(M) = \{f\in\mathcal{W}^s,\ \|f\|_{\mathcal{W}^s}\leq M\}$ for $M>0$ and let

$$\mathcal{G}^\nu(Q) = \big\{K\in\mathcal{G},\ \ Q_i(K)\leq Q_i, i=1,2\big\}. \tag{3.1}$$

for $Q=(Q_1,Q_2)$ with $Q_1>0$, $Q_1Q_2\geq 1$, where the mapping constants $Q_i(K)$ are defined in Assumption 1.

Theorem 1 (Upper bounds)**.** *Let $\mathcal{G}$ be a class of operators satisfying Assumption 1. Assume we observe (y_n, K_δ) given by (1.1) and (1.2), with $n\geq 1$ and $\delta\leq\delta_0<1$. Specify $\widehat{f}_{n,\delta}$ with*

$$L = \lfloor(\delta^2)^{-1/(2\nu+d-1)}\rfloor\bigwedge\lfloor n^{1/(2\nu+d)}\rfloor \tag{3.2}$$

and κ_ℓ, τ_ℓ as in (2.6) and (2.7). For sufficiently small λ_0 and sufficiently large μ_0, for every $s, M > 0$, $Q = (Q_1, Q_2)$ with $Q_1 > 0$ and such that $Q_1 Q_2 \geq 1$, we have

$$\sup_{f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)} \mathbb{E}\Big[\big\|\widehat{f}_{n,\delta} - f\big\|_{\mathbb{H}}^2\Big]$$
$$\lesssim \big(\delta^2 |\log \delta|\big)^{1 \bigwedge 2s/(2\nu+d-1)} \bigvee \big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} \tag{3.3}$$

where $\lesssim$ means inequality up to a multiplicative constant that depends on d, s, ν, M, Q and μ_0, λ_0 only.

The bounds for μ_0 and λ_0 are explicitly computable. In the model generated by y_n in (1.1) and K_δ in (1.2), they depend on the dimension d and on the absolute constants c_0 and c_1 of the concentration lemmas 5.3 and 5.6 below. However, they are in practice much too conservative, as is well known in the signal detection case [13] or the classical inverse problem case [1], see the numerical implementation Section 4. Our next result shows that the rate achieved by $\widehat{f}_{n,\delta}$ is indeed optimal, up to logarithmic terms.

Theorem 2 (Lower bounds)**.** *Let $\mathcal{G}$ be a class of operators satisfying Assumption 1. Assume we observe (y_n, K_δ) given by (1.1) and (1.2). For every $s, M > 0$, $Q = (Q_1, Q_2)$ with $Q_1 > 0$ and $Q_2 > 1/Q_1$, for sufficiently small δ and large n, we have*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)} \mathbb{E}\Big[\big\|\widehat{f} - f\big\|_{\mathbb{H}}^2\Big] \gtrsim \big(\delta^2\big)^{1 \bigwedge 2s/(2\nu+d-1)} \bigvee n^{-2s/(2(s+\nu)+d)} \tag{3.4}$$

where $\gtrsim$ means inequality up to a positive multiplicative constant that depends on d, s, ν, M and Q only.

Combining (3.3) together with (3.4) and the results of [30], we conclude that $\widehat{f}_{n,\delta}$ is minimax over $\mathcal{W}^s(M)$ to within logarithmic terms in n and δ, and that this result is uniform over the nuisance parameter $K \in \mathcal{G}^\nu(Q)$.

3.2. Discussion

The case of diagonal operators and relation to other works

It is interesting to notice that the condition $Q_2 > 1/Q_1$ in Theorem 2 excludes the case where K is diagonal. In this particular case, considered especially in the deconvolution example of Section 4.2 below, a closer inspection of the proof of the upper bound shows that the rate

$$n^{-s/(2(s+\nu)+d)} \bigvee \delta^{1 \bigwedge s/\nu}$$

can be obtained (up to some extra logarithmic factors) as in the case where $d = 1$, which improves on the rate

$$n^{-s/(2(s+\nu)+d)} \bigvee \delta^{1 \bigwedge 2s/(2\nu+d-1)}$$

provided by Theorem 1. This sheds some light on the role of the number d. It is in fact twofolds: it acts as a 'dimension' in the term $n^{-s/(2(s+\nu)+d)}$; in the term involving error in the operator δ, it reflects the distance to the diagonal case expanding from $\delta^{1 \bigwedge s/\nu}$ in the diagonal case, to $\delta^{1 \bigwedge 2s/(2\nu+d-1)}$ in the case $Q_2 > 1/Q_1$. It is very plausible, though beyond the scope of this paper, to express conditions on K leading to rates of the form $2s/(2\nu+\alpha)$, with α continuously varying from 0 to $d-1$. Note that in the case $d=1$, we recover the minimax rate of density deconvolution with unknown error as proved by Neumann [28], see also [9, 20, 22].

While the present work was under completion, Johannes and Schwarz released a preprint [22] that shares interesting similarities with our approach in the restricted case of diagonal operators. Indeed, Johannes and Schwarz study the adaptive estimation by model selection of f under the same observation scheme (1.1) and (1.2). While we confine ourselves to the loss in the $\mathbb{H}$-metric $\|f\|_{\mathbb{H}} = \sum_{\ell \geq 0} \langle f, e_\lambda \rangle^2$, Johannes and Schwarz consider weighted metrics of the form $\|f\|_\omega = \sum_{\ell \geq 0} \sum_{\lambda \in \Lambda_\ell} \langle f, e_\lambda \rangle^2 \omega_\lambda$, where the $\omega_\lambda \geq 0$ are given weights. and include severly ill-posed operators when the behaviour of behaviour K_ℓ and $(K_\ell)^{-1}$ can be exponential with ℓ in operator norm, leading to logarithmic rates of convergence. Such extensions could presumably be obtained in our case, at an additional technical cost. The framework of Johannes and Schwarz is however restricted to the special case of diagonal operators: this imposes $d=1$ and moreover $|\Lambda_\ell| = 1$. It excludes in particular the interesting example of spherical deconvolution.

Relation to other works in the case of sparse operators

For an unknown signal f with smoothness $s > 0$ and unknown operator with degree of ill-posedness $\nu \geq 0$, the optimal rates of convergence are

$$n^{-\alpha(s,\nu)/2} \bigvee \delta^{\beta(s,\nu)}, \tag{3.5}$$

up to inessential logarithmic terms. The exponents $\alpha(s,\nu)$ and $\beta(s,\nu)$ are linked respectively to the error in the signal y_n and the error in the operator K_δ. Efromovitch and Kolchinskii [14] established that under fairly general conditions on the operator K, the optimal exponents are given by

$$\alpha(s,\nu) = \beta(s,\nu) = \frac{2s}{2(s+\nu)+d}.$$

They noted however that if certain sparsity properties on K are moreover assumed (and that we shall not describe here, for instance if K is diagonal in an appropriate basis) then the exponent $\beta(s,\nu) = \frac{2s}{2(s+\nu)+d}$ is no longer optimal, while $\alpha(s,\nu)$ remains unchanged. In the related context of operators acting on Besov spaces $B^s_{p,p}([0,1]^d)$ of functions with smoothness s measured in L^p-norm, Hoffmann and Reiß [18] introduce an *ad hoc* hypothesis on the sparsity of the unknown operator (that we shall not describe here either), expressed in terms of the wavelet discretization of K. They subsequently obtain new rates of convergence for a certain nonlinear wavelet procedure, and these rates overperform (3.5) as expected from the results by [14]. In particular, if one considers the estimation of $f \in B^s_{2,2}$, in the extreme case where the operator K is diagonal in a wavelet basis, the procedure in [18] achieves the rate

$$n^{-\alpha(s,\nu)/2} \bigvee (\delta^2)^{1 \bigwedge (s-d/2)/\nu} \tag{3.6}$$

up to extra logarithmic terms. We may compare our results with the rate (3.6). In our setting, if we pick $(e_\lambda, \lambda \in \Lambda)$ as the Fourier basis described in Section 4.2, then we have $\mathcal{W}^s = B^s_{2,2}([0,1]^d)$. Assuming K to be diagonal in the basis $(e_\lambda, \lambda \in \Lambda)$ which is the exact counterpart of the approach of Hoffmann and Reiß with K being diagonal in a wavelet basis, then by Theorem 1, our estimator $\widehat{f}_{n,\delta}$ (nearly) achieves the rate

$$n^{-\alpha(s,\nu)/2} \bigvee (\delta^2)^{1 \bigwedge 2s/(2\nu+d-1)}$$

which already outperforms the rate (3.6) whenever the error in the signal y_n is dominated by the error in the operator and s is small compared to ν, as follows from the elementary inequality

$$2s/(2\nu+d-1) > (s-d/2)/\nu \;\text{ for }\; 2\nu+d-1 \geq 2s.$$

The superiority of the blockwise SVD in this setting is explained by the fact that the wavelet procedure in [18] is agnostic to the diagonal structure of K in the wavelet basis, in contrast to $\widehat{f}_{n,\delta}$ that takes full advantage of the block structure of K. As already explained in the preceding section, one could actually improve further this result in the specific case of K being diagonal in $(e_\lambda, \lambda \in \Lambda)$ and show that $\widehat{f}_{n,\delta}$ (nearly) achieves the rate $n^{-\alpha(s,\nu)/2} \bigvee (\delta^2)^{1 \bigwedge s/\nu}$, thus deleting the 'dimensional effect' of d for the error in the operator.

Adaptation over the scales $\{\mathcal{W}^s, s > 0\}$ and $\{\mathcal{G}^\nu, \nu \geq 0\}$

The estimator $\widehat{f}_{n,\delta}$ is adaptive over the family $\{W^s(M), s > 0, M > 0\}$ (in the sense that $\widehat{f}_{n,\delta}$ does not require the knowledge of s nor M). However, the knowledge of the degree of ill-posedness ν of K is required through the choice of the maximal frequency L in (3.2). This restriction can actually be relaxed further in dimension $d \geq 2$. Indeed, setting formally $\nu = 0$ in (3.2), one readily checks that $\widehat{f}_{n,\delta}$ becomes adaptive over $\{\mathcal{W}^s(M), s > 0, M > 0\}$ and $\{\mathcal{G}^\nu(Q), \nu \geq 0, Q = (Q_1, Q_2), Q_1 Q_2 \geq 1\}$ simultaneously. In dimension $d = 1$ however, setting $\nu = 0$ in (3.2) is forbidden, but an alternative adaptivity result can be obtained by taking $L = \lfloor (\delta^2)^{-1/s_0} \rfloor \wedge n$ for some $s_0 > 0$, in which case $\widehat{f}_{n,\delta}$ is fully adaptive over the scale $\{\mathcal{G}^\nu(Q), \nu \geq 0, Q = (Q_1, Q_2), Q_1 Q_2 \geq 1\}$ and the restricted family $\{W^s(M), s \geq s_0, M > 0\}$.

Extension to density estimation

Suppose that instead of y_n we observe a random sample $Z_1, \dots, Z_n$ drawn from Kf assumed to be a probability density w.r.t. some measure μ over a domain $\mathcal{D} \subset \mathbb{R}^d$. We take $\mathbb{G} = L^2(\mathcal{D}, \mu)$. By analogy to (2.5), we have an estimator of $\langle Kf, g_\lambda \rangle$ replacing $z_n = \langle y_n, g_\lambda \rangle$ in (2.5) with

$$n^{-1} \sum_{i=1}^{n} g_\lambda(Z_i).$$

Writing

$$n^{-1} \sum_{i=1}^{n} g_\lambda(Z_i) = \langle Kf, g_\lambda \rangle + n^{-1/2} \eta_{n,\lambda},$$

with $\eta_{n,\lambda} = n^{1/2} \big(n^{-1} \sum_{i=1}^{n} g_\lambda(Z_i) - \langle Kf, g_\lambda \rangle \big)$, an inspection of the proof of Theorem 1 reveals that an extension to the density estimation setting carries over as soon as the vector $(\eta_{n,\lambda}, \lambda \in \Lambda_\ell)$ satisfies a concentration inequality, namely

$$\exists \beta_1 > 0, c_1 > 0,\ \forall \beta \geq \beta_1,\ \ \mathbb{P}\Big(|\Lambda_\ell|^{-1} \sum_{\lambda \in \Lambda_\ell} \eta_{n,\lambda}^2 \geqslant \beta^2 \Big) \leq \exp\big(-c_1 \beta^2 |\Lambda_\ell| \big), \quad (3.7)$$

as follows from (5.6) in Lemma 3. In order to obtain (3.7), we may apply a concentration inequality by Bousquet [2] as developed for instance in Massart [27], Eq (5.51) p. 171. The precise control of this extension requires further properties on the basis $(g_\lambda, \lambda \in \Lambda)$ and on the density Kf via the behaviour of $\sum_{\lambda \in \Lambda_\ell} \mathrm{Var}\big(g_\lambda(Z_1)\big)$, see Eq. (5.52) p. 171 in [27]. There is however one instance when (3.7) is easily obtained: suppose that the basis $(g_\lambda, \lambda \in \Lambda)$ associated with the class $\mathcal{G}$ satisfies

$$\max_{\ell \geq 1, |\lambda| = \ell} \sup_{x \in \mathcal{D}} |g_\lambda(x)| < \infty \ \text{ and } \ \max_{\ell \geq 1} |\Lambda_\ell| < \infty. \quad (3.8)$$

Note that the second part of (3.8) simply amounts to assume that $d = 1$ in (2.1). In that case, we obtain the following extension of Theorem 1.

Corollary 1. *Let $\mathcal{G}$ be a class of operators satisfying Assumption 1. Assume we observe a drawn $(Z_1, \dots, Z_n)$ with density Kf w.r.t. a measure μ on $\mathcal{D} \subset \mathbb{R}^d$ and K_δ given by (1.2), with $n \geq 1$ and $\delta \leq \delta_0 < 1$. Assume that $\mathbb{G} = L^2(\mathcal{D}, \mu)$ is equipped with a basis that satisfies (3.8). Put $z_n = n^{-1} \sum_{i=1}^{n} g_\lambda(Z_i)$ in the definition of $\widehat{f}_{n,\delta}$, specified further by (3.2) and κ_ℓ, τ_ℓ as in (2.6) and (2.7). For sufficiently small λ_0 and sufficiently large μ_0, for every $s, M > 0$, $Q = (Q_1, Q_2)$ with $Q_1 > 0$ and such that $Q_1 Q_2 \geq 1$, we have (3.3).*

The proof is given in Section 5.3. The assumptions of Corollary 1 are satisfied in the case of circular deconvolution, see Section 4.2, where $(g_\lambda, \lambda \in \Lambda)$ is the trigonometrical basis and $\mathcal{D} = \mathbb{T}$ is the torus. This framework is in accordance with Neumann [28], Comte et Lacour [9], Johannes [20] and Johannes and Schwarz [22].

4. Application to blind deconvolution

4.1. Spherical deconvolution

Scientific context. A common challenge in astrophysics is the analysis of complex data sets consisting of a number of objects or events such as galaxies of a particular type or ultra high energy cosmic rays (UHECR) and that are genuinely distributed over the celestial sphere. Such objects or events are distributed according to a probability density distribution f on the sphere, depending itself on the physics that governs the production of these objects or events. For instance, UHECR are particles of unknown nature arriving at the earth from apparently random directions of the sky. They could originate from long-lived relic particles from the Big Bang. Alternatively, they could be generated by the acceleration of standard particles, such as protons, in extremely violent astrophysical phenomena. They could also originate from Active Galactic Nuclei (AGN), or from neutron stars surrounded by extremely high magnetic fields. As a consequence, in some hypotheses, the underlying probability distribution for observed UHECRs would be a finite sum of point-like sources. In other hypotheses, the distribution could be uniform, or smooth and correlated with the local distribution of matter in the universe. The distribution could also be a superposition of the above. Identifying between these hypotheses is of primordial importance for understanding the origin and mechanism of production of UHECRs. The observations, denoted by X_i, are often perturbated by an experimental noise, say ε_i, that lead to the deconvolution problem described in Section 1.1. Following van Rooj and Ruymgart [37], Healy *et al.* [16], Kim and Koo [26], Kim, Koo and Luo [25] and Kerkyacharian *et al.* [29], we assume the following model: we observe an n-sample $(Z_1, \dots, Z_n)$ with

$$Z_i = \varepsilon_i X_i, \quad i = 1, \dots, n$$

where the X_i are distributed on the sphere $\mathbb{S}^2$, with common density f with respect to the uniform probability distribution $\mu(d\omega)$ on $\mathbb{S}^2$ and independent of the ε_i that have a common density g with respect to the Haar probability measure dr on the group $\mathcal{SO}(3)$ of 3×3 rotation matrices. One proves in [16, 26] that the density of the Z_i is

$$Kf(\omega) = g \star f(\omega) := \int_{\mathcal{SO}(3)} g(r) f(r^{-1}x)dr, \quad \omega \in \mathbb{S}^2 \tag{4.1}$$

and we are interested in the case where the exact form g of the convolution operator $K = g \star \bullet$ is unknown, due for instance to insufficient knowledge of the device that is used to measure the observations. However, we assume approximate knowledge of g through K_δ as defined in (1.2). □

Checking the blockwise SVD Assumption 1. We closely follow the exposition of [16, 26, 29] for an overview of Fourier theory on $\mathbb{S}^2$ and $\mathcal{SO}(3)$ in order to establish rigorously the connection to Theorem 1 and 3.4. Define

$$u(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } a(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

where $\varphi \in [0, 2\pi)$, $\theta \in [0, \pi)$. Every rotation $r \in \mathcal{SO}(3)$ has representation $r = u(\varphi)a(\theta)u(\psi)$ for some $\varphi, \psi \in [0, 2\pi), \theta \in [0, \pi)$. Define the rotational harmonics

$$D^l_{mn}(r) = D^l_{mn}(\varphi, \theta, \psi) = e^{-i(m\varphi + n\psi)} P^l_{mn}\big(\cos(\theta)\big)$$

for $l \in \mathbb{N}, -l \leqslant m, n \leqslant l$ where P^l_{mn} are the second type Legendre functions described in details in [38]. The D^l_{mn} are the eigenfunctions of the Laplace-Beltrami operator on $\mathcal{SO}(3)$ hence the family $(\sqrt{2l+1} D^l_{mn})$ forms a complete orthonormal basis of $L^2(dr)$ on $\mathcal{SO}(3)$, where dr is the Haar probability measure. Every $h \in L^2(dr)$ has a rotational Fourier transform

$$\mathcal{F}(h)^l_{mn} = \int_{\mathcal{SO}(3)} h(u) D^l_{mn}(u) du, \;\; l \in \mathbb{N}, -l \leqslant m, n \leqslant l,$$

and for every $h \in L^2(dr)$ we have a reconstruction formula

$$\begin{aligned} h &= \sum_{l \in \mathbb{N}} \sum_{-l \le m, n \le l} \mathcal{F}(h)^l_{mn} \overline{D^l_{mn}} \\ &= \sum_{l \in \mathbb{N}} \sum_{-l \le m, n \le l} \mathcal{F}(h)^l_{mn} D^l_{mn}(\bullet^{-1}) \end{aligned}$$

An analogous analysis is available on $\mathbb{S}^2$. Any point $\omega \in \mathbb{S}^2$ is determined by its spherical coordinates $\omega = \big(\sin(\theta)\cos(\varphi), \sin(\theta)\sin(\varphi), \cos(\theta)\big)$ for some $\theta \in [0, \pi), \varphi \in [0, 2\pi)$. Define

$$Y^l_m(\omega) = Y^m_l(\theta, \varphi) = (-1)^m \sqrt{\tfrac{2l+1}{4\pi} \tfrac{(l-m)!}{(l+m)!}} P^l_m\big(\cos(\theta)\big) e^{im\varphi} \tag{4.2}$$

for $l \in \mathbb{N}, -l \leqslant m \leqslant l$ where P^l_m are the Legendre functions. We have $Y^l_{-m} = (-1)^m Y^l_m$ and the (Y^l_m) constitute an orthonormal basis of $L^2(\mu)$ on $\mathbb{S}^2$, generally referred to as the spherical harmonic basis. Any $f \in L^2(\mu)$ has a spherical Fourier transform

$$\mathcal{F}(f)^l_m = \int_{\mathbb{S}^2} f(\omega) \overline{Y^l_m(\omega)} \mu(d\omega)$$

and a reconstruction formula

$$f = \sum_{\ell \in \mathbb{N}} \sum_{-l \le m \le l} \mathcal{F}(f)^l_m Y^l_m.$$

If $g \in L^2\big(\mathcal{SO}(3)\big)$ the spherical convolution operator $Kf = g \star f$ defined in (4.1) satisfies

$$\mathcal{F}(g \star f)^l_m = \sum_{n=-l}^{l} \mathcal{F}(g)^l_{mn} \mathcal{F}(f)^l_n \tag{4.3}$$

and we retrieve the blockwise SVD formalism of Section 2.1 in dimension $d = 2$ by setting $\mathbb{H} = \mathbb{G} = L^2(\mathbb{S}^2, \mu)$, where μ the probability Haar measure on $\mathbb{S}^2$ and

$$e_\lambda = g_\lambda = Y_m^\ell \ \text{ with } \ \lambda = (m, \ell), \ \ \Lambda_\ell = \{(m, \ell), \ \ -\ell \leqslant m \leqslant \ell\}.$$

We have $|\Lambda_\ell| = 2\ell + 1$ and by (4.3), K_ℓ is the finite dimension operator stable on $\mathrm{Span}\{e_\lambda, \lambda \in \Lambda_\ell\}$ with matrix having entries

$$(K_\ell)_{mn} = \mathcal{F}(g)^\ell_{mn}.$$

Hence the first part of Assumption 1 is satisfied. Notice also that in this case K_ℓ is generally not diagonal. Assumption 1 is satisfied as we assume that g is *ordinary smooth* in the terminology of Kim and Koo [26]. Our Assumption 1 exactly matches the constraint (3.6) in their paper with examples given by the Laplace distribution on the sphere ($\nu = 2$) or the Rosenthal distribution ($\nu > 0$ arbitrary). □

Numerical implementation. Following Kerkyacharian, Pham Ngoc and Picard [29] in their Example 2, we take $f(\omega) = C \exp(-4\|\omega - \omega_1\|^2)$ with $\omega_1 = (0, 1, 0)$ and $C = 1/0.7854$. We have $\|f\|_{L^2(\mu)} = 0.7469$.

g is the density of a Laplace distribution on $\mathcal{SO}(3)$, defined through $\mathcal{F}(g)^\ell_{mn} = \delta_{mn}\big(1 + \ell(\ell+1)\big)^{-1}$. Hence, the matrices $(K_\ell)_{mn}$ are homotheties whose ratios behave as ℓ^{-2}. We have $\nu = 2$.

We plot in Figures 1 a 1000-sample of X_i with density f on the sphere, and the action by ε_i on the X_i, where the ε_i are distributed according to g in Figure 2. Note that for the estimation of g, we have access to a noisy version of g with noise level δ only.

We display below the (renormalised) empirical squared error of $\widehat{f}_{10^8,\delta}$ (oracle choice $\lambda_0 = 1, \mu_0 = 1$) for 1000 Monte-Carlo for several values of δ. The noise level δ is to to be compared with the noise level $n^{-1/2} = 10^{-4}$. The latter is chosen non-negative, in order to show the interaction between the two types of error, and sufficiently small to emphasize the influence of δ on the process of estimation.

Noise level δ	0	10^{-3}	$3\,10^{-3}$	$5\,10^{-3}$	10^{-2}
Mean error	0.0466	0.0542	0.1732	0.2784	0.4335
Standard dev.	0.0011	0.0022	0.0126	0.0355	0.0466

Finally, on a specific sample of $n = 10^8$ data, we plot the target density f (Figure 3) and its reconstruction for $n = 10^8$ data with $\delta = 0$ (Figure 4) and $\delta = 3\,10^{-3}$ (Figure 5). At a visual level, we oversimplify the representation by plotting f and its reconstruction with a view from above the sphere through the Oz axis. We see that the contour in Figure 5 is not well recovered in the regions where f is small (on the right side of the graph in Figure 5). The choice of λ_0, μ_0 remains unchanged. □

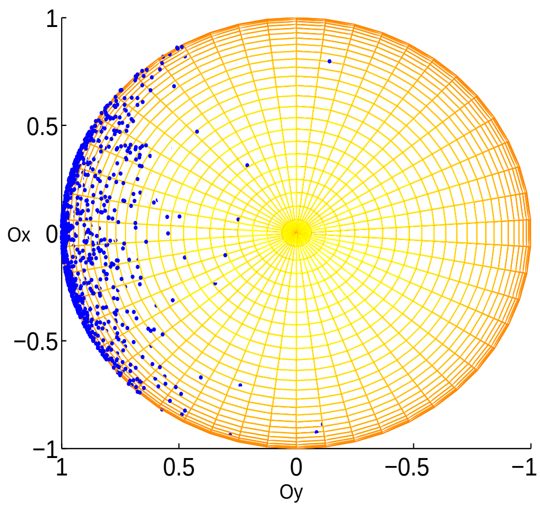

FIG 1. ***Data from*** $\boldsymbol{f}$***.*** *Plot of* $n = 1000$ *data with common distribution* f *on the sphere* $\mathbb{S}$ *(planar representation).*

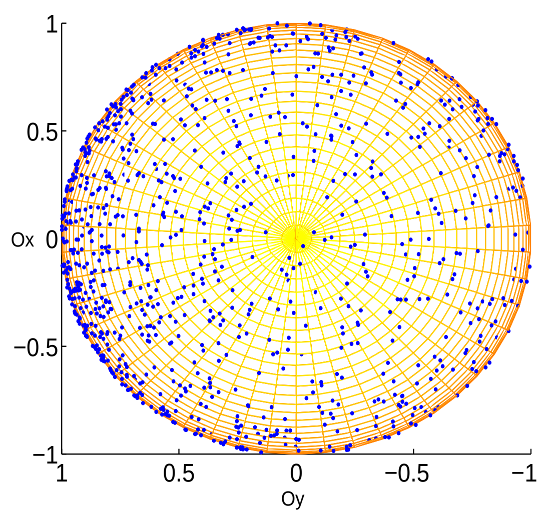

FIG 2. ***Data from*** $\boldsymbol{f \star g}$***.*** *Plot of* $n = 1000$ *data* $\varepsilon_i X_i$ *on the sphere* $\mathbb{S}$ *with common distribution* $Kf = f \star g$*. The* X_i *are the data pictured in Figure 1 and the* ε_i *are sampled according to* g *(planar representation).*

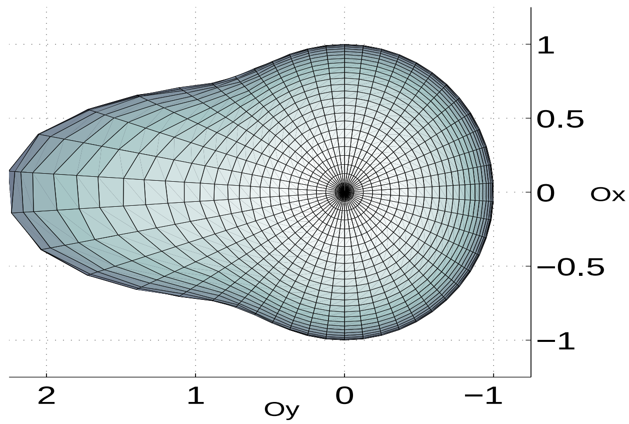

Fig 3. ***Target density*** f. *The representation is simplified through a view from above the sphere through the* Oz*-axis.*

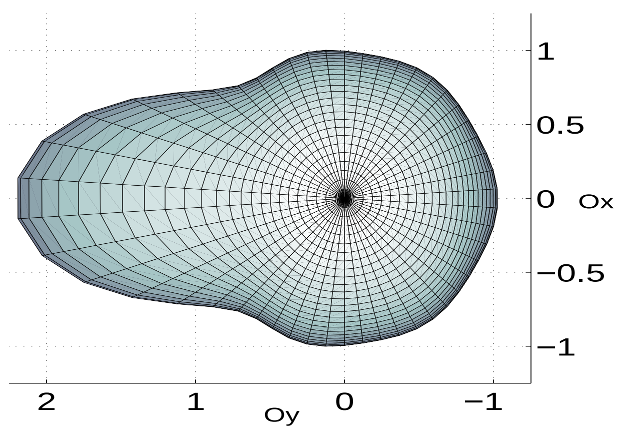

Fig 4. ***Reconstruction for*** $n = 10^8$ ***and*** $\delta = 0$.

4.2. Circular deconvolution

Scientific context. In many engineering problems, the observation of a signal f or image is distorted by the action of a linear operator K. We assume for simplicity that f lives on the torus $\mathbb{T} = [0,1]$ (or $[0,1]^d$) appended with periodic boundary conditions. In many instances, the restoration of f from the noisy observation of Kf is challenged by the additional uncertainty about the operator K. This is the case for instance in electronic microscopy [31] for the restoration of fluorescence Confocal Laser Scanning Microscope (CLSM) images. In other words, the quality of the image suffers from two physical limitations: error mea-

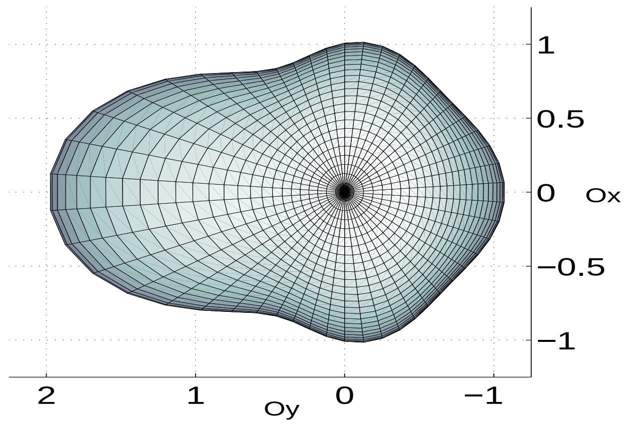

FIG 5. ***Reconstruction for*** $n = 10^8$ ***and*** $\delta = 3\,10^{-3}$. *The reconstruction is polluted simultaneously by the limited number of observations* n *and the noise level* δ *in the blurring* g.

surements or limited accuracy, and the fact that the exact PSF (the incoherent point spread function) that accounts for the blurring of f (mathematically the action of K) is not precisely known. This is a classical issue that goes back to [35, 15]. An idealised additive Gaussian model for the noise contamination yields the observation (1.1) with

$$Kf(x) = g \star f(x) := \int_{\mathbb{T}^d} g(u)f(x-u)\mu(du), \;\; x \in \mathbb{T}^d,$$

and μ is the uniform probability measure on $\mathbb{T}^d$. The degradation $K = g \star \bullet$ is characterised by the impulse response function g. In most cases of interest, we do not know the exact form of g. In a condensed idealised statistical setup, we have access to

$$g_\delta = g + \delta \dot{W}', \tag{4.4}$$

where $\dot{W}'$ is another Gaussian white noise defined on $L^2(\mu) = L^2(\mathbb{T}^d, \mu)$ and independent of $\dot{W}$. Experimental approaches that justify representation (4.4) are described in [10, 23, 34]. □

Checking Assumption 1. We obviously have $\mathbb{H} = \mathbb{G} = L^2(\mu)$ and the bases (e_λ) and (g_λ) will coincide with the d-dimensional extension of the circular trigonometric basis $(e^{2i\pi kx}, k \in \mathbb{Z})$ if we set:

$$e_\lambda(x_1, \ldots, x_d) = \prod_{j=1}^{d} e^{2i\pi k_j x_j}, \;\; (x_1, \ldots, x_d) \in \mathbb{T}^d,$$

where we put

$$\lambda = (k_1, \ldots, k_d), \;\; \ell = |\lambda| = 1 + \sum_{j=1}^{d} |k_j|, \;\; \text{and} \;\; \ell \geq 1.$$

Any $f \in L^2(\mu)$ has a Fourier transform $\mathcal{F}(f)_\lambda = \int_{\mathbb{T}^d} f(x)\overline{e_\lambda(x)}\mu(dx)$ and moreover, if $g \in L^2(\mu)$, we have

$$\mathcal{F}(f \star g)_\lambda = \mathcal{F}_\lambda(f)\mathcal{F}_\lambda(g).$$

Therefore, K is diagonal in the basis $(e_\lambda, \lambda \in \Lambda)$. With $\Lambda_\ell = \{\lambda, |\lambda| = \ell\}$, we have $|\Lambda_\ell| = \binom{\ell-1+d}{d-1} \sim \ell^{d-1}$. Moreover $K_\ell = \mathrm{Diag}\big(\mathcal{F}_\lambda(g), \lambda \in \Lambda_\ell\big)$ and the first part of Assumption 1. Assuming that g satisfies $c|\lambda|^{-\nu} \le \big|\mathcal{F}(g)_\lambda\big| \le c'|\lambda|^{-\nu}$ for some $\nu \ge 0$ and constants $c, c' > 0$, we readily obtain the second part of Assumption 1. Note also that since K is diagonal in the basis $(e_\lambda, \lambda \in \Lambda)$ observing g_δ in the representation (4.4) is equivalent to observing K_δ in (1.2). □

Numerical implementation. We numerically implement $\widehat{f}_{n,\delta}$ in dimension $d = 1$ in the case where there is no noise in the signal (formally $n^{-1/2} = 0$) in order to illustrate the parametric effect that dominates in the optimal rate of convergence in Theorems 1 and 3.4 that becomes $\big(\delta^2\big)^{s/\nu\wedge 1}$ in that case. We take as target function $f : \mathbb{T} \to \mathbb{R}$ belonging to $\mathcal{W}^{5-\alpha}$ for all $\alpha > 1/2$ and defined by its Fourier coefficients

$$\mathcal{F}(f)_\lambda = |\lambda|^{-5}, \ \ \lambda \in \{-1000, \dots, 1000\}.$$

We pick a family of blurring functions g_ν defined in the same manner by the formula

$$\mathcal{F}(g_\nu)_\lambda = |\lambda|^{-\nu}, \ \ \lambda \in \{-1000, \dots, 1000\}, \ \ \nu \in \{1, 4, 5, 6, 8\}.$$

We show in Figure 6 in a log-log plot the mean-squared error of $\widehat{f}_{\infty,\delta}$ for the oracle choice $\mu_0 = 0, \lambda_0 = 1$ over 1000 Monte-Carlo simulations for $\nu \in \{1, 4, 5, 6, 8\}$

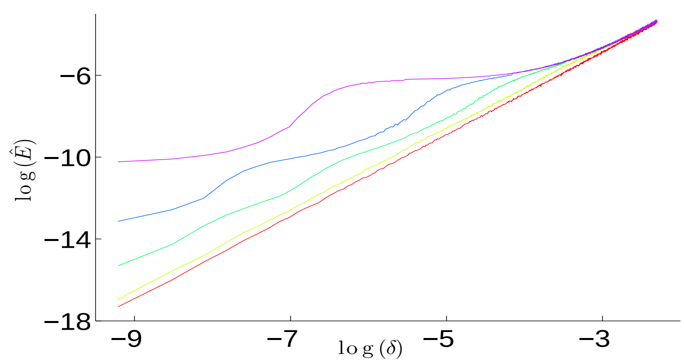

FIG 6. ***Estimation of the rate exponent when*** $n^{-1/2} \ll \delta$. *Empirical squared-error* $\hat{E}$ *versus* δ *in log-log scale. Top-to-bottom:* $\nu = 8, 6, 5, 4, 1$. *The target function has smoothness* $s = 5-\alpha$ *for all* $\alpha > 1/2$. *For* $\nu < 4.5$, *the slope of the curve is constant and close to* 2, *confirming the parametric rate predicted by the theory when the smoothess of the signal dominates the degree of ill-posedness of the operator. The empirical errors were computed using* 1000 *Monte-Carlo simulations.*

and $\delta \in [10^{-4}, 10^{-1}]$. For small values of δ the predicted slope of the curve gives a rough estimate of the rate of convergence. We visually see that for the critical case $\nu \le s = 5 - \alpha$ with $\alpha > 1/2$ and below, the slope is close to 2 confirming the parametric rate that is obtained whenever $\nu \le s$. □

5. Proofs

5.1. Preliminary estimates

Preparation. Recall that $H_\ell = \text{Span}\{e_\lambda, \lambda \in \Lambda_\ell\}$, $G_\ell = \text{Span}\{g_\lambda, \lambda \in \Lambda_\ell\}$ and that P_ℓ (resp. Q_ℓ) denotes the orthogonal projector onto G_ℓ (resp. H_ℓ). We repeatedly rely on the following consequence of Assumption 1

Lemma 1. *We have*

$$P_\ell K = K_\ell Q_\ell. \tag{5.1}$$

Proof. For $h \in \mathbb{H}$, we have

$$P_\ell K h = P_\ell K Q_\ell h + P_\ell K(\text{Id} - Q_\ell)h.$$

The result is then a straighforward consequence of $P_\ell K(\text{Id} - Q_\ell)h = 0$. Indeed, by definition, $(\text{Id} - Q_\ell)h \in H_\ell^\perp$ hence $(\text{Id} - Q_\ell)h = \sum_{\ell' \neq \ell} Q_{\ell'}(\text{Id} - Q_\ell)h$. Since $Q_{\ell'}(\text{Id} - Q_\ell)h \in H_{\ell'}$, we have $KQ_{\ell'}(\text{Id} - Q_\ell)h \in G_{\ell'}$ by Assumption 1. It follows that

$$K(\text{Id} - Q_\ell)h = \sum_{\ell' \neq \ell} KQ_{\ell'}(\text{Id} - Q_\ell)h \in \overline{\text{Span}\{G_{\ell'}, \ell' \neq \ell\}}^{\mathbb{G}}$$

Therefore $K(\text{Id} - Q_\ell)h \in G_\ell^\perp$ and $P_\ell K(\text{Id} - Q_\ell)h = 0$ follows. □

In turn, we have a convenient description of the observation K_δ defined in (1.2) and y_n defined in (1.1) and in terms of a sequence space model that we shall now describe. □

Notation. If $h \in \mathbb{G}$, we denote by $\boldsymbol{h}_\ell$ the (column) vector of coordinates of $P_\ell h$ in the basis $(g_\lambda, \lambda \in \Lambda_\ell)$. If $T : \mathbb{H} \to \mathbb{G}$ is a linear operator, we write $\boldsymbol{T}_\ell$ for the matrix of the Galerkin projection $T_\ell = P_\ell T_{|H_\ell}$ of T. □

Sequence model for error in the operator. The observation of K_δ in (1.2) leads to the representation $K_{\delta,\ell} = K_\ell + \delta \dot{B}_\ell$, or equivalently, in matrix notation

$$\boldsymbol{K}_{\delta,\ell} = \boldsymbol{K}_\ell + \delta \dot{\boldsymbol{B}}_\ell, \ \ \ell \ge 1, \tag{5.2}$$

where $\dot{\boldsymbol{B}}_\ell$ is a $|\Lambda_\ell| \times |\Lambda_\ell|$ matrix with entries that are independent centred Gaussian random variables, with unit variance. The following estimate is a classical concentration property of random matrices. For $\ell \le L$, $\|\bullet\|_{\text{op}}$ denotes the operator norm for $|\Lambda_\ell| \times |\Lambda_\ell|$ matrices (we shall skip the dependence upon ℓ in the notation).

Lemma 2 ([11], Theorem II.4)**.** *There are positive constants β_0, c_0 such that*

$$\text{For all } \beta \geq \beta_0, \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{B}}_\ell\|_{op} \geq \beta\big) \leq \exp\big(-c_0\beta^2|\Lambda_\ell|^2\big). \tag{5.3}$$

An immediate consequence of Lemma 2 is the following moment bound:

$$\text{For every } p > 0, \; \mathbb{E}\big[\|\dot{\boldsymbol{B}}_\ell\|_{\text{op}}^p\big] \lesssim |\Lambda_\ell|^{p/2}. \tag{5.4}$$

□

Sequence model for error in the signal. From (1.1), we observe the Gaussian measure y_n, or equivalently, thanks to (5.1)

$$P_\ell y_n = P_\ell K f + n^{-1/2} P_\ell \dot{W} = K_\ell Q_\ell f + n^{-1/2}\boldsymbol{\eta}_\ell, \;\; \ell \geq 1$$

or, using the notation introduced in (2.5), in matrix notation

$$\boldsymbol{z}_{n,\ell} = \boldsymbol{K}_\ell \boldsymbol{f}_\ell + n^{-1/2}\boldsymbol{\eta}_\ell, \;\; \ell \geq 1 \tag{5.5}$$

where we used (5.1), with $\boldsymbol{\eta}_\ell$ denoting a vector of $|\Lambda_\ell|$ independent centred Gaussian random variables with unit variance.

The following result is a direct consequence of the fact that $\|\boldsymbol{\eta}_\ell\|^2$ has a χ-square distribution with $|\Lambda_\ell|$ degrees of freedom. The proof is standard

Lemma 3. *There are positive constant β_1, c_1 such that*

$$\text{For all } \beta \geq \beta_1, \; \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\boldsymbol{\eta}_\ell\| \geq \beta\big) \leq \exp\big(-c_1\beta^2|\Lambda_\ell|\big), \tag{5.6}$$

□

5.2. *Proof of Theorem 1*

We have

$$\|\widehat{f}_n - f\|_{\mathbb{H}}^2 = \sum_{\ell \geq 1} \|\widehat{\boldsymbol{f}}_{n,\ell} - \boldsymbol{f}_\ell\|^2 = \sum_{\ell=1}^{L} \|\widehat{\boldsymbol{f}}_{n,\ell} - \boldsymbol{f}_\ell\|^2 + \sum_{\ell > L} \|\boldsymbol{f}_\ell\|^2$$

where $\|\bullet\|$ denotes the Euclidean norm on $\mathbb{R}^{|\Lambda_\ell|}$ (we shall omit any reference to ℓ when no confusion is possible). Concerning the bias term, we have

$$\sum_{\ell > L} \|\boldsymbol{f}_\ell\|^2 \leq \|f\|_{\mathcal{W}^s}^2 L^{-2s} \tag{5.7}$$

and this term has the right order by definition of L in (3.2). Concerning the stochastic term, thanks to our preliminary analysis, we may write

$$\widehat{\boldsymbol{f}}_{n,\ell} = (\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{z}_{n,\ell}\mathbf{1}_{\{\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{\text{op}} \leq \kappa_\ell\}}\mathbf{1}_{\{\|\boldsymbol{z}_{n,\ell}\| \geq \tau_\ell\}},$$

We set

$$\mathcal{A}_\ell = \{\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{\text{op}} \leq \kappa_\ell\} \text{ and } \mathcal{B}_\ell = \{\|\boldsymbol{z}_{n,\ell}\| \geq \tau_\ell\}.$$

We thus obtain the decomposion of the variance term as

$$\sum_{\ell=1}^{L} \|\widehat{\boldsymbol{f}}_{n,\ell} - \boldsymbol{f}_\ell\|^2 \leq I + II + III,$$

with

$$I = \sum_{\ell=1}^{L} \|(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{z}_{n,\ell} - \boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell}$$

$$II = \sum_{\ell=1}^{L} \|\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell^c},$$

$$III = \sum_{\ell=1}^{L} \|\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{B}_\ell^c}.$$

We shall successively bound each term I, II and III.

- *The term I, preliminary decomposition.* Writing

$$\boldsymbol{z}_{n,\ell} = \big(\boldsymbol{K}_{\delta,\ell} - \delta\dot{\boldsymbol{B}}_\ell\big)\boldsymbol{f}_\ell + n^{-1/2}\boldsymbol{\eta}_\ell,$$

we obtain

$$(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{z}_{n,\ell} - \boldsymbol{f}_\ell = -\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell\boldsymbol{f}_\ell + n^{-1/2}(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{\eta}_\ell.$$

We introduce further the event $\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \leq a_\ell\}$ with $a_\ell = \frac{\rho}{\kappa_\ell}$ for some $0 < \rho < \frac{1}{2}$ and the condition $\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\| \geq \frac{\tau_\ell}{2}\}$. We thus have

$$I \lesssim IV + V + VI + VII,$$

with

$$IV = \sum_{\ell=1}^{L} \|\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell\cap\mathcal{B}_\ell}\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\right\}}\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|\geq\frac{\tau_\ell}{2}\right\}},$$

$$V = \sum_{\ell=1}^{L} \|n^{-1/2}(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{\eta}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell\cap\mathcal{B}_\ell}\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\right\}}\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|\geq\frac{\tau_\ell}{2}\right\}},$$

$$VI = \sum_{\ell=1}^{L} \|\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell\cap\mathcal{B}_\ell}\Big(\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}> a_\ell\right\}} + \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|<\frac{\tau_\ell}{2}\right\}}\Big),$$

$$VII = \sum_{\ell=1}^{L} \|n^{-1/2}(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{\eta}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell\cap\mathcal{B}_\ell}\Big(\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}> a_\ell\right\}} + \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|<\frac{\tau_\ell}{2}\right\}}\Big).$$

We shall next successively bound each term IV, V, VI and VII. □

• *The term IV.* First, we have

$$\begin{aligned}(\boldsymbol{K}_\ell)^{-1} &= (\boldsymbol{K}_{\delta,\ell} - \delta\dot{\boldsymbol{B}}_\ell)^{-1}\\ &= (\boldsymbol{I} - \delta\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{B}})^{-1}(\boldsymbol{K}_{\delta,\ell})^{-1}.\end{aligned}$$

On $\mathcal{A}_\ell = \{\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{\mathrm{op}} \leq \kappa_\ell\}$ and $\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \leq a_\ell\}$, since a_ℓ satisfies $\kappa_\ell\, a_\ell = \rho < \frac{1}{2}$, by a usual Neumann series argument,

$$\begin{aligned}\|\big(\boldsymbol{I} - \delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}\big)^{-1}\|_{\mathrm{op}} &= \|\sum_{i\geq 0}(-\boldsymbol{K}_{\delta,\ell})^i(\delta\dot{\boldsymbol{B}})^i\|_{\mathrm{op}}\\ &\leq \sum_{i\geq 0}\|\boldsymbol{K}_{\delta,\ell}\|_{\mathrm{op}}^i\|\delta\dot{\boldsymbol{B}}\|_{\mathrm{op}}^i\\ &\leq \sum_{i\geq 0}\rho^i = (1-\rho)^{-1}.\end{aligned}$$

Therefore, on $\mathcal{A}_\ell$ and $\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \leq a_\ell\}$, we have

$$\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}} \leq (1-\rho)^{-1}\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{\mathrm{op}} \leq (1-\rho)^{-1}\kappa_\ell. \tag{5.8}$$

Second, we now write

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

hence, on $\mathcal{A}_\ell$ and $\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \leq a_\ell\}$, we have by (5.8)

$$\|(\boldsymbol{K}_\ell)^{-1}\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \leq (1-\rho)^{-1}\kappa_\ell a_\ell \leq \frac{\rho}{1-\rho} < 1$$

since $\rho < \frac{1}{2}$ by assumption. The same Neumann series argument now entails

$$\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{\mathrm{op}} \leq \frac{\rho}{1-\rho}\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}}. \tag{5.9}$$

We are ready to bound the term IV itself. We have

$$\begin{aligned}&\|\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell\boldsymbol{f}_\ell\|^2\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\}}\\ &\leq \|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{\mathrm{op}}^2\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}^2\|\boldsymbol{f}_\ell\|^2\,\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\}}\\ &\lesssim \|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}}^2\kappa_\ell^{-2}\|\boldsymbol{f}_\ell\|^2\,\mathbf{1}_{\{\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}}\leq(1-\rho)^{-1}\kappa_\ell\}},\end{aligned}$$

where we successively used (5.8) and (5.9). It follows that

$$\begin{aligned}\mathbb{E}\big[IV\big] &\lesssim \sum_{\ell=1}^{L}\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}}^2\kappa_\ell^{-2}\|\boldsymbol{f}_\ell\|^2\,\mathbf{1}_{\{\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}}\leq(1-\rho)^{-1}\kappa_\ell\}}\\ &\lesssim \sum_{\ell=1}^{L}\ell^{2\nu}\kappa_\ell^{-2}\|\boldsymbol{f}_\ell\|^2\end{aligned}$$

where we used Assumption 1. The bound is uniform in $K \in \mathcal{G}^\nu(Q)$. By definition of κ_ℓ and using that $|\Lambda_\ell|$ is of order ℓ^{d-1}, we derive

$$\mathbb{E}[IV] \lesssim \left((\delta^2|\log\delta|)\bigvee n^{-1}\right)\sum_{\ell=1}^{L}\ell^{2\nu+d-1}\|\boldsymbol{f}_\ell\|^2.$$

If $2\nu + d - 1 \leq 2s$, we have

$$\sum_{\ell=1}^{L}\ell^{2\nu+d-1}\|\boldsymbol{f}_\ell\|^2 \leq \|f\|^2_{\mathcal{W}^s},$$

therefore

$$\begin{aligned}\mathbb{E}\big[IV\big] &\lesssim \delta^2|\log\delta| + L^{-2s}\\ &\lesssim \left(\delta^2|\log\delta|\right)^{1\bigwedge 2s/(2\nu+d-1)}\bigvee n^{-2s/(2\nu+d)}\end{aligned} \tag{5.10}$$

by definition of L in (3.2), and this result is uniform in $f \in \mathcal{W}^s(M)$. If $2\nu+d-1 \geq 2s$, we have

$$\begin{aligned}\sum_{\ell=1}^{L}\ell^{2\nu+d-1}\|\boldsymbol{f}_\ell\|^2 &\leq L^{2(\nu-s)+d-1}\sum_{\ell=1}^{L}\ell^{2s}\|\boldsymbol{f}_\ell\|^2\\ &\leq L^{2(\nu-s)+d-1}\|f\|^2_{\mathcal{W}^s}.\end{aligned}$$

By definition of L again we derive

$$\begin{aligned}\mathbb{E}\big[IV\big] &\lesssim L^{-2s}L^{2\nu+d-1}(n^{-1}\bigvee\delta^2|\log\delta|)\\ &\lesssim L^{-2s}(n^{-1/(2\nu+d)}\bigvee 1) \leqslant L^{-2s}\end{aligned} \tag{5.11}$$

and this bound is uniform in $f \in \mathcal{W}^s(M)$. Putting together (5.10) and (5.11), we finally obtain

$$\mathbb{E}\big[IV\big] \lesssim \left(\delta^2|\log\delta|\right)^{1\bigwedge 2s/(2\nu+d-1)}\bigvee n^{-2s/(1\nu+d)} \tag{5.12}$$

uniformly in $f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)$. □

- *The term V.* We have

$$\begin{aligned}&\|n^{-1/2}(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{\eta}_\ell\|^2\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\right\}}\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|\geq\frac{\tau_\ell}{2}\right\}}\\ &\leq n^{-1}\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|^2_{\mathrm{op}}\|\boldsymbol{\eta}_\ell\|^2\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\right\}}\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|\geq\frac{\tau_\ell}{2}\right\}}\\ &\lesssim n^{-1}\|(\boldsymbol{K}_{\ell})^{-1}\|^2_{\mathrm{op}}\|\boldsymbol{\eta}_\ell\|^2\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}\leq a_\ell\right\}}\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|\geq\frac{\tau_\ell}{2}\right\}}\\ &\lesssim n^{-1}\ell^{2\nu}\|\boldsymbol{\eta}_\ell\|^2\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\|\geq\frac{\tau_\ell}{2}\right\}}\end{aligned}$$

where we successively used (5.8) and (5.9) in the same way as for the term IV, the last inequality being obtained thanks to Assumption 1. By Assumption 1 again, since

$$\|\boldsymbol{K}_\ell \boldsymbol{f}_\ell\| \leq \|\boldsymbol{K}_\ell\|_{\mathrm{op}} \|\boldsymbol{f}_\ell\| \leq Q_1(K)\ell^\nu \|\boldsymbol{f}_\ell\|$$

we derive

$$\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell \boldsymbol{f}_\ell\| \geq \frac{\tau_\ell}{2}\right\}} \leq \mathbf{1}_{\left\{\|\boldsymbol{f}_\ell\| \geq Q_1(K)^{-1}\frac{\tau_\ell}{2}\ell^\nu\right\}} = \mathbf{1}_{\left\{\|\boldsymbol{f}_\ell\| \geq c\ell^{\nu+(d-1)/2} n^{-1/2}(\log n)^{1/2}\right\}}$$

for some constant c that depends on $Q_1(K)$ and the pre-factor μ_0 in the choice of τ_ℓ only. Since $\mathbb{E}[\|\boldsymbol{\eta}_\ell\|^2] = |\Lambda_\ell| \lesssim \ell^{d-1}$, we infer, for any $1 \leq k \leq L$

$$\begin{aligned}
\mathbb{E}[V] &\lesssim n^{-1} \sum_{\ell=1}^{L} \ell^{2\nu+d-1} \mathbf{1}_{\left\{\|\boldsymbol{f}_\ell\| \geq c\,\ell^{\nu+(d-1)/2} n^{-1/2}(\log n)^{1/2}\right\}} \\
&\lesssim n^{-1}\Big(\sum_{\ell=1}^{k} \ell^{2\nu+d-1} + \sum_{\ell=k+1}^{L} n(\log n)^{-1}\|\boldsymbol{f}_\ell\|^2\Big) \\
&\leq n^{-1}k^{2\nu+d} + (\log n)^{-1}\sum_{\ell>k}\|\boldsymbol{f}_\ell\|^2 \\
&\lesssim n^{-1}k^{2\nu+d} + (\log n)^{-1}\|f\|^2_{\mathcal{W}^s} k^{-2s}.
\end{aligned}$$

The admissible choice $k = \lfloor \big(n(\log n)^{-1/2}\big)^{1/(2(s+\nu)+d)} \rfloor \wedge (\delta^2)^{-1/(2\nu+d-1)}$ yields

$$\begin{aligned}
\mathbb{E}[V] &\lesssim n^{-1}k^{\nu+d} + k^{-2s} \\
&\lesssim \big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} + k^{-2s} \\
&\lesssim \big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} \bigvee (\delta^2)^{2s/(2\nu+d-1)}
\end{aligned} \tag{5.13}$$

uniformly in $f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)$. □

- *The term VI.* We further bound the term VI via

$$VI \leq VIII + IX,$$

with

$$\begin{aligned}
VIII &= \sum_{\ell=1}^{L} \|\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell} \mathbf{1}_{\left\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\right\}}, \\
IX &= \sum_{\ell=1}^{L} \|\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell\|^2 \mathbf{1}_{\mathcal{A}_\ell\cap\mathcal{B}_\ell} \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell \boldsymbol{f}_\ell\| < \frac{\tau_\ell}{2}\right\}}.
\end{aligned}$$

On $\mathcal{A}_\ell$, we have

$$\|\delta(\boldsymbol{K}_{\delta,\ell})^{-1}\dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell\|^2 \lesssim \delta^2\,\kappa_\ell^2 \|\dot{\boldsymbol{B}}_\ell\|^2_{\mathrm{op}} \|\boldsymbol{f}_\ell\|^2$$

hence

$$\begin{aligned}\mathbb{E}\big[VIII\big] &\lesssim \delta^2 \sum_{\ell=1}^{L} \kappa_\ell^2 \|\boldsymbol{f}_\ell\|^2 \mathbb{E}\big[\|\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}^2 \mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\}}\big] \\ &\leq \delta^2 \sum_{\ell=1}^{L} \kappa_\ell^2 \|\boldsymbol{f}_\ell\|^2 \mathbb{E}\big[\|\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}}^4\big]^{1/2} \mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\big)^{1/2} \\ &\lesssim \delta^2 \sum_{\ell=1}^{L} \kappa_\ell^2 \|\boldsymbol{f}_\ell\|^2 |\Lambda_\ell| \delta^{c_0\rho^2|\Lambda_\ell|^2/2\lambda_0^2} \\ &\lesssim |\log\delta| \max_{1\leq \ell\leq L} \delta^{c_0\rho^2|\Lambda_\ell|^2/2\lambda_0^2} \|f\|_{\mathbb{H}}^2\end{aligned}$$

applying successively Cauchy-Schwarz, the moment bound (5.4) and Lemma 2. Indeed, since $a_\ell = \rho/\kappa_\ell$, by definition of κ_ℓ in (2.6), we infer

$$\begin{aligned}\mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\big) &\leq \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > |\Lambda_\ell|^{-1/2}\tfrac{\rho}{\kappa_\ell}\delta^{-1}\big) \\ &= \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > \tfrac{\rho}{\lambda_0}|\log\delta|^{1/2}\big) \\ &\leq \exp\big(-c_0\tfrac{\rho^2}{\lambda_0^2}|\log\delta||\Lambda_\ell|^2\big) = \delta^{c_0\rho^2|\Lambda_\ell|^2/\lambda_0^2}\end{aligned} \tag{5.14}$$

by (5.6) of Lemma 3 since $\frac{\rho}{\lambda_0}|\log\delta|^{1/2} \geq \beta_0$ for sufficiently small λ_0 thanks to the assumption $\delta \leq \delta_0 < 1$. Finally, since Λ_ℓ is non-empty, by taking λ_0 sufficiently small, we conclude

$$\mathbb{E}\big[VIII\big] \lesssim \delta^2 \tag{5.15}$$

uniformly in $f \in \mathcal{W}^s(M)$. We now turn to the term IX. Observe first that

$$\mathbf{1}_{\mathcal{B}_\ell}\mathbf{1}_{\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\| < \frac{\tau_\ell}{2}\}} \leq \mathbf{1}_{\{n^{-1/2}\|\boldsymbol{\eta}_\ell\| \geq \frac{\tau_\ell}{2}\}}. \tag{5.16}$$

We reproduce the steps we used for the term $VIII$, replacing the event $\{\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\}$ by $\{n^{-1/2}\|\boldsymbol{\eta}_\ell\| \geq \frac{\tau_\ell}{2}\}$. We obtain

$$\mathbb{E}\big[IX\big] \lesssim \delta^2 \sum_{\ell=1}^{L} \kappa_\ell^2 \|\boldsymbol{f}_\ell\|^2 |\Lambda_\ell| \mathbb{P}\big(n^{-1/2}\|\boldsymbol{\eta}_\ell\| \geq \tfrac{\tau_\ell}{2}\big)^{1/2}.$$

By definition of τ_ℓ in (2.7) and Lemma 5.6, we have

$$\begin{aligned}\mathbb{P}\big(n^{-1/2}\|\boldsymbol{\eta}_\ell\| > \tfrac{\tau_\ell}{2}\big) &= \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\boldsymbol{\eta}_\ell\| > \tfrac{\mu_0}{2}(\log n)^{1/2}\big) \\ &\leq \exp\big(-c_1\tfrac{\mu_0^2}{4}\log n\big) = n^{-c_1\mu_0^2/4}\end{aligned} \tag{5.17}$$

since $\frac{\mu_0}{2}(\log n)^{1/2} \geq \beta_1$ for large enough μ_0. It follows that

$$\mathbb{E}\big[IX\big] \lesssim |\log\delta| \|f\|_{\mathbb{H}}^2 n^{-c_1\mu_0^2/4} \lesssim n^{-1}|\log\delta| \tag{5.18}$$

by taking μ_0 sufficiently large. The bound is uniform in $f \in \mathcal{W}^s(M)$. Putting together the estimates (5.15) and (5.18), we derive

$$\mathbb{E}\big[VI\big] \lesssim \delta^2 + n^{-1}|\log\delta| \tag{5.19}$$

for large enough n, uniformly in $f \in \mathcal{W}^s(M)$. □

• *The term VII.* The arguments needed here are quite similar to those we used for the term VI. On $\mathcal{A}_\ell$, we have

$$\|n^{-1/2}(\boldsymbol{K}_{\delta,\ell})^{-1}\boldsymbol{\eta}_\ell\|^2 \le n^{-1}\kappa_\ell^2\|\boldsymbol{\eta}_\ell\|^2,$$

hence, using (5.16), the fact that $\mathbb{E}\big[\|\boldsymbol{\eta}_\ell\|^2\big] = |\Lambda_\ell| \lesssim \ell^{d-1}$ together with $\kappa_\ell \le n^{1/2}$ by definition (2.6), we successively obtain

$$\begin{aligned}
\mathbb{E}\big[VII\big] &\le n^{-1}\sum_{\ell=1}^{L}\kappa_\ell^2\mathbb{E}\big[\|\boldsymbol{\eta}_\ell\|^2\big]\Big(\mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\big) + \mathbb{P}\big(n^{-1/2}\|\boldsymbol{\eta}_\ell\| > \tfrac{\tau_\ell}{2}\big)\Big) \\
&\lesssim \max_{1\le\ell\le L}\big\{\mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} > a_\ell\big) + \mathbb{P}\big(n^{-1/2}\|\boldsymbol{\eta}_\ell\| > \tfrac{\tau_\ell}{2}\big)\big\}\sum_{\ell=1}^{L}\ell^{d-1} \\
&\lesssim L^{d-1}\big(\delta^{c_0\rho^2/\lambda_0^2} + n^{-c_1\mu_0^2/4}\big)
\end{aligned}$$

where we applied (5.14) and (5.17) to obtain the last inequality. The choice of L in (3.2) leads to

$$\mathbb{E}\big[VII\big] \lesssim (\delta^2)^{-\frac{d-1}{2\nu+d-1}+\frac{c_0\rho^2}{\lambda_0^2}} + n^{\frac{1}{2\nu+d}-\frac{c_1\mu_0^2}{4}} \lesssim \delta^2 \bigvee n^{-1} \tag{5.20}$$

by taking λ_0 sufficiently small and μ_0 sufficiently large. □

• *The term I, conclusion.* We put together the estimates (5.12), (5.13), (5.19) and (5.20). We obtain

$$\mathbb{E}\big[I\big] \lesssim \big(\delta^2|\log\delta|\big)^{1\bigwedge 2s/(2\nu+d-1)} \bigvee \big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} \tag{5.21}$$

uniformly in $f \in \mathcal{W}^s(M)$. □

• *The term II.* We claim the following inequality

$$\mathbf{1}_{\mathcal{A}^c} \le \mathbf{1}_{\big\{\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}} \ge \frac{\kappa_\ell}{2}\big\}} + \mathbf{1}_{\big\{\|\boldsymbol{K}_{\delta,\ell}-\boldsymbol{K}_\ell\|_{\mathrm{op}} \ge \kappa_\ell^{-1}\big\}}, \tag{5.22}$$

a consequence of the following elementary lemma

Lemma 4. *Let A and B be two bounded operators with bounded inverse. If $\|B^{-1}\| \ge \kappa$ for some $\kappa > 0$, then either $\|A^{-1}\| \ge \kappa/2$ or $\|A - B\| \ge 1/\kappa$.*

Proof of Lemma 4. Write $B = A+\xi$. Assume that $\|A^{-1}\| < \kappa/2$. By the triangle inequality, $\|(A+\xi)^{-1} - A^{-1}\| \geq \kappa/2$. We proceed by contradiction: suppose that $\|\xi\| \leq 1/\kappa$. Then we have $\|A^{-1}\xi\| \leq \|A^{-1}\|\|\xi\| \leq 1/2 < 1$ and a standard Neumann series argument entails

$$\begin{aligned}\|(A+\xi)^{-1} - A^{-1}\| &= \|(\mathrm{Id} + A^{-1}\xi)^{-1}A^{-1} - A^{-1}\| \\ &= \|\sum_{i\geq 1}(-1)^i (A^{-1})^{i+1}\xi^i\| \\ &\leq \sum_{i\geq 1}\|A^{-1}\|^{i+1}\|\xi\|^i \\ &< \frac{\kappa}{2}\sum_{i\geq 1}\Big(\frac{\kappa}{2}\Big)^i\Big(\frac{1}{\kappa}\Big)^i = \frac{\kappa}{2},\end{aligned}$$

a contradiction. □

By Assumption 1, we have $\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}} \leq Q_2(K)\ell^\nu$. Therefore

$$\mathbf{1}_{\left\{\|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}} \geq \frac{\kappa_\ell}{2}\right\}} \leq \mathbf{1}_{\left\{\ell \geq c\left(\delta^2|\log\delta|\right)^{1/(2\nu+d-1)} \wedge\, n^{1/(2\nu)}\right\}}$$

for some constant c that depends on $Q_2(K)$ and λ_0 only. For the second term in the right-hand side of (5.22), we apply by Lemma 2 in the same way as we obtained (5.15) for the term $VIII$. We derive

$$\begin{aligned}\mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \geq \kappa_\ell^{-1}\big) &= \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \geq \mu_0^{-1}|\log\delta|^{1/2}\big) \\ &\leq \exp\big(-\tfrac{c_0}{\mu_0^2}|\log\delta||\Lambda_n|^2\big) = \delta^{\,c_0|\Lambda_\ell|^2/\mu_0} \leq \delta^{\,c_0/\mu_0}\end{aligned}$$

for large enough μ_0. Therefore

$$\begin{aligned}\mathbb{E}\big[II\big] &\leq \sum_{\ell=1}^{L}\|\boldsymbol{f}_\ell\|^2\Big(\mathbf{1}_{\left\{\ell \geq c\left(\delta^2|\log\delta|\right)^{1/(2\nu+d-1)} \wedge\, n^{1/(2\nu)}\right\}} + \mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\|_{\mathrm{op}} \geq \kappa_\ell^{-1}\big)\Big) \\ &\lesssim \big(n^{-s/\nu}\bigvee\big(\delta^2|\log\delta|\big)^{2s/(2\nu+d-1)}\big)\|f\|^2_{\mathcal{W}^s} + \|f\|^2_{\mathbb{H}}\,\delta^{\,c_0/\mu_0}.\end{aligned}$$

We finally obtain

$$\begin{aligned}\mathbb{E}\big[II\big] &\lesssim \big(\delta^2\log\delta^{-1}\big)^{2s/(2\nu+d-1)} + \delta^2 + n^{-s/\nu} \\ &\lesssim \big(\delta^2|\log\delta|\big)^{1\,\wedge\, 2s/(2\nu+d-1)}\bigvee\big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} \qquad (5.23)\end{aligned}$$

uniformly in $f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)$. □

• *The term III.* Obviously, the decomposition (5.5) entails

$$\mathbf{1}_{\mathcal{B}^c} = \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell + n^{-1/2}\boldsymbol{\eta}_\ell\| < \tau_\ell\right\}} \leq \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\| \leq 2\tau_\ell\right\}} + \mathbf{1}_{\left\{n^{-1/2}\|\boldsymbol{\eta}_\ell\| > \tau_\ell\right\}}.$$

On the one hand, we have

$$\|\boldsymbol{K}_\ell\boldsymbol{f}_\ell\| \geq \|(\boldsymbol{K}_\ell)^{-1}\|^{-1}_{\mathrm{op}}\|\boldsymbol{f}_\ell\| \geq Q_2(K)^{-1}\ell^{-\nu}\|\boldsymbol{f}_\ell\|$$

by Assumption 1. By definition of τ_ℓ in (2.7) it follows that, for any $1 \le k \le L$,

$$\begin{aligned}\sum_{\ell=1}^{L} \|\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell \boldsymbol{f}_\ell\| \le 2\tau_\ell\right\}} &\le \sum_{\ell=1}^{L} \|\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\left\{\|\boldsymbol{f}_\ell\| \le 2Q_2(K)^{-1}\ell^\nu \tau_\ell\right\}} \\ &\lesssim \sum_{\ell=1}^{k} \ell^{2\nu}\tau_\ell^2 + \sum_{\ell=k+1}^{L} \|\boldsymbol{f}_\ell\|^2 \\ &\lesssim (n^{-1}\log n) \sum_{\ell=1}^{k} \ell^{2\nu+d-1} + \|f\|_{\mathcal{W}^s}^2 k^{-2s} \\ &\lesssim (n^{-1}\log n)\, k^{2\nu+d} + \|f\|_{\mathcal{W}^s}^2 k^{-2s}.\end{aligned}$$

The choice $k = \lfloor \big(n^{1/2}(\log n)^{-1/2}\big)^{1/(2(s+\nu)+d)} \rfloor$ yields

$$\sum_{\ell=1}^{L} \|\boldsymbol{f}_\ell\|^2 \mathbf{1}_{\left\{\|\boldsymbol{K}_\ell \boldsymbol{f}_\ell\| \le 2\tau_\ell\right\}} \lesssim \big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} \tag{5.24}$$

uniformly in $f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)$. On the other hand, by (5.17), we have

$$\sum_{\ell=1}^{L} \|\boldsymbol{f}_\ell\|^2 \mathbb{P}\big(n^{-1/2}\|\boldsymbol{\eta}_\ell\| > \tau_\ell\}\big) \lesssim \|f\|_{\mathbb{H}}^2 n^{-c_1\mu_0^2/4} \lesssim n^{-1}$$

by taking μ_0 large enough, uniformly in $f \in \mathcal{W}^s(M)$. Combining this last estimate with (5.24) we infer

$$\mathbb{E}\big[III\big] \lesssim \big(n^{-1}\log n\big)^{2s/(2(s+\nu)+d)} + n^{-1} \tag{5.25}$$

uniformly in $f \in \mathcal{W}^s(M), K \in \mathcal{G}^\nu(Q)$. □

Proof of Theorem 1, completion. It remains to piece together the estimates (5.7), (5.21), (5.23) and (5.25). □

5.3. *Proof of Corollary 1*

It suffices to prove that (3.7). Let $\beta > 0$. We have

$$\mathbb{P}\Big(|\Lambda_\ell|^{-1} \sum_{\lambda \in \Lambda_\ell} \eta_{n,\lambda}^2 \geqslant \beta^2\Big) \le \sum_{\lambda \in \Lambda_\ell} \mathbb{P}\big(|\eta_{n,\lambda}| \ge \beta\big) \le c\,\mathbb{P}\big(|\eta_{n,\lambda}| \ge \beta\big),$$

where $c = \max_{\ell \ge 1} |\Lambda_\ell|$ is finite by (3.8). Also,

$$n^{-1/2}\eta_{n,\lambda} = n^{-1} \sum_{i=1}^{n} \big(g_\lambda(Z_i) - \mathbb{E}[g_\lambda(Z_i)]\big)$$

is the empirical mean of centred and independent random variables that satisfy

$$\big|g_\lambda(Z_i) - \mathbb{E}[g_\lambda(Z_i)]\big| \le 2 \max_{\ell \ge 1, |\lambda| = \ell} \sup_{x \in \mathcal{D}} |g_\lambda(x)| = c'$$

which is finite by (3.8). By Hoeffding inequality, it follows that

$$\begin{aligned}\mathbb{P}\big(|\eta_{n,\lambda}| \geq \beta\big) &= \mathbb{P}\Big(\big|n^{-1}\sum_{i=1}^{n}\big(g_\lambda(Z_i) - \mathbb{E}[g_\lambda(Z_i)]\big)\big| \geq n^{-1/2}\beta\Big) \\ &\leq \exp\big(-2(c')^{-2}\beta^2\big) \leq \exp\big(-2(c'c)^{-2}\beta^2|\Lambda_\ell|\big)\end{aligned}$$

and (3.7) is proved with $c_1 = (c'c)^2$ and $\beta_1 = c^{-5/2}\log(2c)(c')^{-1}$ for instance.

5.4. Proof of Theorem 2

The lower bound in the case $\delta = 0$ is classical (Nussbaum and Pereverzev [30]) and will not decrease for increasing noise levels δ or $n^{-1/2}$ whence it suffices to provide the case which formally corresponds to observing Kf without noise while K remains unknown.

Preliminaries: a Bayesian inequality

For every $\ell \geq 1$, denote by $\mathcal{M}_\ell$ the set of $|\Lambda_\ell| \times |\Lambda_\ell|$ matrices. We denote by $\mathcal{M}_\ell^\nu(Q)$ the subset of $\mathcal{M}_\ell$ of matrices $\boldsymbol{K}_\ell$ such that

$$\|\boldsymbol{K}_\ell\|_{\mathrm{op}} \leq Q_2\ell^{-\nu} \;\text{ and }\; \|(\boldsymbol{K}_\ell)^{-1}\|_{\mathrm{op}} \leq Q_1\ell^{\nu}.$$

Define

$$\boldsymbol{K}_\ell^0 = c_1\ell^{-\nu}\boldsymbol{I}_\ell \tag{5.26}$$

where $\boldsymbol{I}_\ell$ denotes the identity in $\mathcal{M}_\ell$ and $c_1 > 0$ is such that

$$1/Q_1 < c_1 < Q_2$$

so that $\boldsymbol{K}_\ell^0 \in \mathcal{M}_\ell^\nu(Q)$. We assume a Bayesian approach and pick $\boldsymbol{K}_\ell$ at random, with

$$\boldsymbol{K}_\ell = \boldsymbol{K}_\ell^0 + c_2\,\delta\dot{\boldsymbol{W}}_\ell,$$

for some $c_2 > 0$ and where $\dot{\boldsymbol{W}}_\ell$ is an independent copy of $\dot{\boldsymbol{B}}_\ell$. Define $\boldsymbol{g}_\ell = (1\;0\;\cdots\;0)^T$ as the first canonical (column) vector in $\mathbb{R}^{|\Lambda_\ell|}$. Define also

$$\boldsymbol{\vartheta} = -(\boldsymbol{K}_\ell^0)^{-1}(\boldsymbol{K}_\ell - \boldsymbol{K}_\ell^0)(\boldsymbol{K}_\ell^0)^{-1}\boldsymbol{g}_\ell \tag{5.27}$$

and

$$\boldsymbol{X} = -(\boldsymbol{K}_\ell^0)^{-1}(\boldsymbol{K}_{\delta,\ell} - \boldsymbol{K}_\ell^0)(\boldsymbol{K}_\ell^0)^{-1}\boldsymbol{g}_\ell. \tag{5.28}$$

Lemma 5. *There exists a constant c_3 depending on ν, Q and c_2 only such that*

$$\inf_T \mathbb{P}\big(\delta^{-2}\ell^{-4\nu}|\Lambda_\ell|^{-1}\|T(\boldsymbol{X}) - \boldsymbol{\vartheta}\|^2 \geq c_3\big) \geq \tfrac{1}{2}, \tag{5.29}$$

where the infimum is taken among all estimators T based on the observation $\boldsymbol{X}$.

Proof of Lemma 5. We have $\boldsymbol{X} = \boldsymbol{\vartheta} + \boldsymbol{\varepsilon}$, with

$$\boldsymbol{\vartheta} = -(\boldsymbol{K}_\ell^0)^{-1} c_2 \delta \dot{\boldsymbol{W}} (\boldsymbol{K}_\ell^0)^{-1} \boldsymbol{g}_\ell \text{ and } \boldsymbol{\varepsilon} = -(\boldsymbol{K}_\ell^0)^{-1} \delta \dot{\boldsymbol{B}} (\boldsymbol{K}_\ell^0)^{-1} \boldsymbol{g}_\ell.$$

By construction, $\boldsymbol{\vartheta}$ and $\boldsymbol{\varepsilon}$ are two independent Gaussian random vectors. More precisely, by definition of $\boldsymbol{g}_\ell$ and with obvious notation, we have

$$\boldsymbol{\vartheta} \sim \mathcal{N}\big(0, \delta^2 c_2^2 c_1^{-4} \ell^{4\nu} \boldsymbol{I}_\ell\big) \text{ and } \boldsymbol{\varepsilon} \sim \mathcal{N}\big(0, \delta^2 c_1^{-4} \ell^{4\nu} \boldsymbol{I}_\ell\big).$$

It readily follows that the posterior law of $\boldsymbol{\vartheta}$ given $\boldsymbol{X}$ is

$$\mathcal{L}(\boldsymbol{\vartheta} \,|\, \boldsymbol{X}) = \mathcal{N}\Big(\frac{c_2^2}{1+c_2^2} \boldsymbol{X}, \delta^2 \frac{c_2^2}{1+c_2^2} c_1^{-4} \ell^{4\nu} \boldsymbol{I}_\ell\Big).$$

Now, for $c_3 > 0$, define

$$H_\delta(c_3, \boldsymbol{x}) = \mathbf{1}_{\{\delta^{-2}\ell^{-4\nu}|\Lambda_\ell|^{-1}\|\boldsymbol{x}\|^2 \geq c_3\}} \text{ for } \boldsymbol{x} \in \mathbb{R}^{|\Lambda_\ell|}.$$

Setting $z(\boldsymbol{X}) = T(\boldsymbol{X}) - \mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}]$, we have

$$\begin{aligned} \mathbb{E}\big[H_\delta\big(c_3, T(\boldsymbol{X}) - \boldsymbol{\vartheta}\big) \,|\, \boldsymbol{X}\big] &= \mathbb{E}\big[H_\delta\big(c_3, z(\boldsymbol{X}) + \mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \boldsymbol{\vartheta}\big) \,|\, \boldsymbol{X}\big] \\ &\geq \mathbb{E}\big[H_\delta\big(c_3, \mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \boldsymbol{\vartheta}\big) \,|\, \boldsymbol{X}\big] \end{aligned}$$

where we used a version of Anderson's Lemma given in Lemma 10.2 in [19] p. 157. Indeed, the law of $\mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \vartheta$ has a centrally symmetric density and the function H_δ is nonnegative, centrally symmetric, satisfies $H_\delta(0) = 0$ and the sets $\{\boldsymbol{x},\ H_\delta(c_3, \boldsymbol{x}) < c\}$ are convex for any $c > 0$.

Now, $\|\mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \boldsymbol{\vartheta}\|^2$ has a χ^2-distribution with $|\Lambda_\ell|$ degrees of freedom, up to a scaling factor of order $\delta^2 \ell^{4\nu}$. This means that the sequence of random variables $\delta^{-2}\ell^{-4\nu}|\Lambda_\ell|^{-1}\|\mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \boldsymbol{\vartheta}\|^2$ is bounded below in probability in $\ell \geq 1$ and $\delta > 0$. Since $\mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \boldsymbol{\vartheta}$ is moreover independent of $\boldsymbol{X}$, it follows that there exists c_3 independent of δ and ℓ such that

$$\mathbb{E}\big[H_\delta\big(c_3, \mathbb{E}[\boldsymbol{\vartheta} \,|\, \boldsymbol{X}] - \boldsymbol{\vartheta}\big) \,|\, \boldsymbol{X}\big] \geq \tfrac{1}{2}.$$

Integrating with respect to $\boldsymbol{X}$, we obtain (5.29) and the result follows. □

Proof of Theorem 2

We assume with no loss of generality that $2\nu + d - 1 \geq 2s$. (Otherwise, the lower bound δ trivially follows from the parametric case.) Let $\Pi^{s,\nu}(M, Q_1)$ denote the set of sequences $\pi = (\pi_\ell)_{\ell \geq 1}$ satisfying

$$\sum_{\ell \geq 1} \pi_\ell^2 \ell^{2(s+\nu)} \leq \frac{M^2}{Q_1^2}. \tag{5.30}$$

For $\pi \in \Pi^{s,\nu}(M, Q_1)$ and $K \in \mathcal{G}^\nu(Q)$, define f via its coordinates in H_ℓ by

$$\boldsymbol{f}_\ell = \pi_\ell \boldsymbol{K}_\ell^{-1} \boldsymbol{g}_\ell, \ \ \ell \geq 1,$$

where $\boldsymbol{g}_\ell$ is an arbitrary vector in $\mathbb{R}^{|\Lambda_\ell|}$ with $\|\boldsymbol{g}_\ell\| = 1$ (fixed in the sequel). Then

$$\sum_{\ell\geq 1} \ell^{2s}\|\pi_\ell \boldsymbol{K}_\ell^{-1}\boldsymbol{g}_\ell\|^2 \leq \sum_{\ell\geq 1} \pi_\ell^2 \|\boldsymbol{K}_\ell^{-1}\|_{\mathrm{op}}^2 \|\boldsymbol{g}_\ell\|^2 \leq Q_1^2 \sum_{\ell\geq 1} \pi_\ell^2 \ell^{2(s+\nu)} \leq M^2$$

since $\pi \in \Pi^{s,\nu}(M, Q_1)$. Therefore $f \in \mathcal{W}^s(M)$. It follows that for an arbitrary estimator $\widehat{f}$, we have

$$\begin{aligned}
&\sup_{f\in\mathcal{W}^s(M), K\in\mathcal{G}^\nu(Q)} \mathbb{E}\Big[\big\|\widehat{f} - f\big\|_{\mathbb{H}}^2\Big] \\
&= \sup_{f\in\mathcal{W}^s(M), K\in\mathcal{G}^\nu(Q)} \sum_{\ell\geq 1} \mathbb{E}\Big[\big\|\widehat{\boldsymbol{f}}_\ell - \boldsymbol{f}_\ell\big\|^2\Big] \\
&\geq \sup_{\pi\in\Pi^{s,\nu}(M,Q_1), K\in\mathcal{G}^\nu(Q)} \sum_{\ell\geq 1} \mathbb{E}\Big[\big\|\widehat{\boldsymbol{f}}_\ell - \pi_\ell \boldsymbol{K}_\ell^{-1}\boldsymbol{g}_\ell\big\|^2\Big].
\end{aligned}$$

Lemma 6. *There exist a choice of $\boldsymbol{g}_\ell$ with $\|\boldsymbol{g}_\ell\| = 1$ and constants c_4, c_5 (depending on s, ν, M, Q) such that for any $\pi \in \Pi^{s,\nu}(M, Q_1)$, if $|\Lambda_\ell|^{1/2}\delta \leq c_4\,\ell^{-\nu}$, we have*

$$\inf_{\widehat{\boldsymbol{f}}_\ell} \sup_{K\in\mathcal{G}^\nu(Q)} \mathbb{E}\big[\|\widehat{\boldsymbol{f}}_\ell - \pi_\ell \boldsymbol{K}_\ell^{-1}\boldsymbol{g}_\ell\|^2\big] \geq c_5\,\delta^2 \ell^{4\nu+d-1}\pi_\ell^2 \tag{5.31}$$

where the infimum is taken over all estimators and provided $\delta > 0$ is sufficiently small.

With (5.31), we easily conclude: Define $L = \lfloor c_6 \delta^{-2/(2\nu+d-1)} \rfloor$ with $c_6 > 0$. For $1 \leq \ell \leq L$, the assumption $|\Lambda|^{1/2}\delta \leq c_4\,\ell^{-\nu}$ of Lemma 6 is satisfied by picking $c_6 > 0$ sufficiently small and we have

$$\begin{aligned}
&\sup_{\pi\in\Pi^{s,\nu}(M,Q_1), K\in\mathcal{G}^\nu(Q)} \sum_{\ell\geq 1} \mathbb{E}\Big[\big\|\widehat{\boldsymbol{f}}_\ell - \pi_\ell \boldsymbol{K}_\ell^{-1}\boldsymbol{g}_\ell\big\|^2\Big] \\
&\geq c_5\delta^2 \sup_{\pi\in\Pi^{s,\nu}(M,Q_1)} \sum_{\ell=1}^{L} \ell^{4\nu+d-1}\pi_\ell^2 \\
&\geq c_5\delta^2 \tfrac{M^2}{Q_1^2} L^{2\nu+d-1-2s} \geq\; c_5 c_6^{2\nu+d-1-2s} \tfrac{M^2}{Q_1^2} \delta^{2s/(2\nu+d-1)}
\end{aligned}$$

thanks to the admissible choice π specified by $\pi_\ell^2 = \ell^{-2(\nu+s)}M^2/Q_1^2$ if $\ell = L$ and 0 otherwise. Theorem 2 follows. It remains to prove Lemma 6.

Proof of Lemma 6. In view of (5.31), we may (and will) assume that $\pi_\ell = 1$. We rely on the notation and definition of the preliminaries. Observe first that

$$\begin{aligned}
&\inf_{\widehat{\boldsymbol{f}}_\ell} \sup_{K\in\mathcal{G}^\nu(Q)} \mathbb{E}\big[\|\widehat{\boldsymbol{f}}_\ell - \boldsymbol{K}_\ell^{-1}\boldsymbol{g}_\ell\|^2\big] \\
&= \inf_{\widehat{\boldsymbol{f}}_\ell} \sup_{K\in\mathcal{G}^\nu(Q)} \mathbb{E}\big[\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\|^2\big].
\end{aligned}$$

where $\boldsymbol{K}^0$ is defined in (5.26). Put $v_{\delta,\ell} = \delta^2 \ell^{4\nu+d-1}$. For any $c > 0$, by Chebyshev inequality, we have

$$\begin{aligned} c^2 v_{\delta,\ell}^{-2} \inf_{\widehat{\boldsymbol{f}}_\ell} \sup_{K\in\mathcal{G}^\nu(Q)} \mathbb{E}\big[\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\|^2\big] \\ \geq \inf_{\widehat{\boldsymbol{f}}_\ell} \sup_{K\in\mathcal{G}^\nu(Q)} \mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\| \geq c\, v_{\delta,\ell}\big). \end{aligned} \tag{5.32}$$

We adopt the same Bayesian approach as in the preliminaries and consider $\boldsymbol{K}_\ell$ as a random matrix with distribution such that

$$\boldsymbol{K}_\ell = \boldsymbol{K}_\ell^0 + c_2\, \delta \dot{\boldsymbol{W}}_\ell, \tag{5.33}$$

where $\dot{\boldsymbol{W}}_\ell$ is an independent copy of $\dot{\boldsymbol{B}}_\ell$ and $c_2 > 0$ is to be specified later. Using the randomisation (5.33) on $\boldsymbol{K}_\ell$, the right-hand side in (5.32) is now bigger than

$$\inf_{\widehat{\boldsymbol{f}}_\ell} \mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\| \geq c\, v_{\delta,\ell}\big) - \mathbb{P}\big(\boldsymbol{K}_\ell \notin \mathcal{M}_\ell^\nu(Q)\big). \tag{5.34}$$

Let us first show that

$$\inf_{\widehat{\boldsymbol{f}}_\ell} \mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\| \geq c\, v_{\delta,\ell}\big) \tag{5.35}$$

is bounded below for an appropriate choice of $c > 0$. Introduce the event

$$\mathcal{A}_\delta = \big\{Q_1 \ell^\nu c_2 \delta \|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}} \leq \rho\big\}$$

for some $0 < \rho < 1$. Observe that $\|(\boldsymbol{K}_\ell^0)^{-1} c_2 \delta \dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}} \leq \rho$ on $\mathcal{A}_\delta$, therefore, by an usual Neuman series argument, we have the decomposition

$$\begin{aligned} &\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1} \\ &= -\,(\boldsymbol{K}_\ell^0)^{-1}(c_2\delta\dot{\boldsymbol{W}}_\ell)(\boldsymbol{K}_\ell^0)^{-1} + \sum_{n\geq 2} (-1)^n \big((\boldsymbol{K}_\ell^0)^{-1} c_2 \dot{\boldsymbol{W}}_\ell\big)^n (\boldsymbol{K}_\ell^0)^{-1} \end{aligned}$$

Applying the vector $\boldsymbol{g}_\ell = (1, 0, \ldots, 0)$ and setting

$$\boldsymbol{\zeta}_{\delta,\ell} = \sum_{n\geq 2} (-1)^n \big((\boldsymbol{K}_\ell^0)^{-1} c_2 \dot{\boldsymbol{W}}_\ell\big)^n (\boldsymbol{K}_\ell^0)^{-1} \boldsymbol{g}_\ell,$$

we obtain the decomposition

$$\begin{aligned} \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell &= -\,(\boldsymbol{K}_\ell^0)^{-1}(c_2\delta\dot{\boldsymbol{W}}_\ell)(\boldsymbol{K}_\ell^0)^{-1}\boldsymbol{g}_\ell + \boldsymbol{\zeta}_{\delta,\ell} \\ &= \boldsymbol{\vartheta} + \boldsymbol{\zeta}_{\delta,\ell}, \end{aligned}$$

where $\boldsymbol{\vartheta}$ is defined in (5.27). We derive, for any $c > 0$

$$\begin{aligned} &\mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\| \geq c\, v_{\delta,\ell}\big) \\ &\geq \mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - (\boldsymbol{\vartheta} + \boldsymbol{\zeta}_{\delta,\ell})\| \geq c\, v_{\delta,\ell} \text{ and } \mathcal{A}_\delta\big) \\ &\geq \mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \boldsymbol{\vartheta}\| \geq \tfrac{1}{2} c\, v_{\delta,\ell} \text{ and } \mathcal{A}_\delta \text{ and } \|\boldsymbol{\zeta}_{\delta,\ell}\| \leq \tfrac{1}{2} c\, v_{\delta,\ell}\big) \end{aligned}$$

by the triangle inequality. We claim that for any $\varepsilon > 0$, there exists a choice of sufficiently small c_2 such that for any $c > 0$:

$$\limsup_{\delta\to 0} \mathbb{P}\big(\mathcal{A}_\delta \ \text{ and } \ \|\boldsymbol{\zeta}_{\delta,\ell}\| \leq \tfrac{1}{2} c\, v_{\delta,\ell}\big) \geq 1 - \varepsilon. \tag{5.36}$$

Let us admit temporarily (5.36). For such a choice, we thus have

$$\begin{aligned} &\mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \big(\boldsymbol{K}_\ell^{-1} - (\boldsymbol{K}_\ell^0)^{-1}\big)\boldsymbol{g}_\ell\| \geq c\, v_{\delta,\ell}\big) \\ \geq &\mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \boldsymbol{\vartheta}\| \geq \tfrac{1}{2} c\, v_{\delta,\ell}\big) - \varepsilon. \end{aligned}$$

Let us now look at an apparently different problem: we want to estimate $\boldsymbol{\vartheta}$ from our observation $\boldsymbol{K}_{\delta,\ell}$, or equivalently, from the observation

$$-(\boldsymbol{K}_\ell^0)^{-1}(\boldsymbol{K}_{\delta,\ell} - \boldsymbol{K}_\ell^0)(\boldsymbol{K}_\ell^0)^{-1}.$$

The choice $\boldsymbol{g}_\ell = (1\ 0\ \cdots\ 0)^T$ entails that $-(\boldsymbol{K}_\ell^0)^{-1}(\boldsymbol{K}_{\delta,\ell} - \boldsymbol{K}_\ell^0)(\boldsymbol{K}_\ell^0)^{-1}\boldsymbol{g}_\ell$ is a sufficient statistic, but this last quantity is precisely $\boldsymbol{X}$ defined in (5.28). Thus, without loss of generality, $\widehat{\boldsymbol{f}}_\delta$ can be taken as an estimator of the form $T(\boldsymbol{X})$. By Lemma 5, we know that $v_{\delta,\ell}$ is a lower bound for estimating $\boldsymbol{\vartheta}$.

More specifically, by taking c such that $c \leqslant 2\sqrt{c_3}$, we have

$$\mathbb{P}\big(\|\widehat{\boldsymbol{f}}_\ell - \boldsymbol{\vartheta}\| \geq \tfrac{1}{2} c\, v_{\delta,\ell}\big) - \varepsilon \geq \tfrac{1}{2} - \varepsilon \geq \tfrac{1}{4}$$

say, since the choice of ε is arbitrary, and (5.35) follows. It remains to prove (5.36).

First, we have that $|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}$ is bounded in probability by Lemma 2 in $\ell \geq 1$. Since $|\Lambda_\ell|^{1/2}\delta \leq c_4\ell^\nu$ by assumption, we also have that $\ell^\nu\delta\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}$ is bounded in probability, hence the probability of $\mathcal{A}_\delta$ can be taken arbitrarily close to 1 by taking c_2 sufficiently small. Moreover, on $\mathcal{A}_\delta$, we have

$$\begin{aligned} \|\zeta_{\delta,\ell}\| &\leq Q_1\ell^\nu \sum_{n\geq 2} \big(Q_1\ell^\nu c_2\delta\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}\big)^n \\ &\leq (1-\rho)^{-1} c_2^2\, Q_1^3\delta^2\ell^{3\nu}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}^2 \\ &\leq (1-\rho)^{-1} c_2^2\, Q_1^3\delta\ell^{2\nu}|\Lambda_\ell|^{1/2}c_4|\Lambda_\ell|^{-1}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}^2 \end{aligned}$$

where we again used the fact that $|\Lambda_\ell|^{1/2}\delta \leq c_4\ell^{-\nu}$ by assumption. The claim follows from the fact that $|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}$ is bounded in probability. Hence (5.36) and (5.35) is proved.

In order to complete the proof of Lemma 6, we need to check that the term $\mathbb{P}\big(\boldsymbol{K}_\ell \notin \mathcal{M}_\ell^\nu(Q)\big)$ can be taken arbitrarily small when bounding (5.32) below by (5.34). We have

$$\begin{aligned} &\mathbb{P}\big(\boldsymbol{K}_\ell \notin \mathcal{M}_\ell^\nu(Q)\big) \\ &\leq \mathbb{P}\big(\|\boldsymbol{K}_\ell\|_{\mathrm{op}} > Q_2\ell^{-\nu}\big) + \mathbb{P}\big(\|\boldsymbol{K}_\ell^{-1}\|_{\mathrm{op}} > Q_1\ell^\nu\big). \end{aligned} \tag{5.37}$$

For the first term in the right-hand side of (5.37), we have

$$\mathbb{P}\big(\|\boldsymbol{K}_\ell\|_{\mathrm{op}} > Q_2\ell^{-\nu}\big) \le \mathbb{P}\big(\|c_2\delta\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}} > Q_2\ell^{-\nu} - \|\boldsymbol{K}_\ell^0\|_{\mathrm{op}}\big)$$
$$\le \mathbb{P}\big(\|c_2\delta\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}} > (Q_2 - c_1)\ell^{-\nu}\big).$$

The last term can be rewritten as

$$\mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}} > (Q_2 - c_1)c_2^{-1}\ell^{-\nu}|\Lambda_\ell|^{-1/2}\delta^{-1}\big).$$

For the second term in the right-hand side of (5.37), thanks to the property $\|\boldsymbol{K}_\ell^{-1}\|_{\mathrm{op}} \le \big(c_1\ell^{-\nu} - \|c_2\delta\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}\big)^{-1}$ we derive

$$\mathbb{P}\big(\|\boldsymbol{K}_\ell^{-1}\|_{\mathrm{op}} > Q_1\ell^{\nu}\big)$$
$$\le \mathbb{P}\big(|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}} > (c_1 - Q_1^{-1})c_2^{-1}\ell^{-\nu}|\Lambda_\ell|^{-1/2}\delta^{-1}\big).$$

By assumption, we have that $\ell^{-\nu}|\Lambda_\ell|^{-1/2}\delta^{-1}$ is bounded away from zero. Since $|\Lambda_\ell|^{-1/2}\|\dot{\boldsymbol{W}}_\ell\|_{\mathrm{op}}$ is tight in $\ell \ge 1$, we can conclude by taking c_2 sufficiently small. The proof of Lemma 6 is complete. □

Acknowledgements

The research of M. Hoffmann is partly supported by the French Agence Nationale de la Recherche (Blanc SIMI 1 2011 project CALIBRATION). The research of D. Picard is partly supported by the French Agence Nationale de la Recherche (ANR-09-BLAN-0128 PARCIMONIE). We are grateful to J. Rousseau and A. Tsybakov, as well as anonymous referees for helpful comments.

References

[1] F. Abramovich and B.W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85:115–129, 1998. MR1627226

[2] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C.R. Math. Acad. Sci. Paris*, 334:495–500, 2002. MR1890640

[3] T. Cai. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.*, 27:898–924, 1999. MR1724035

[4] T. Cai and H. Zou. A data-driven block thresholding approach to wavelet estimation. *Ann. Statist.*, 37:569–595, 2009. MR2502643

[5] L. Cavalier and N.W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361, 2005. MR2158113

[6] L. Cavalier and M. Raimondo. Wavelet deconvolution with noisy eigenvalues. *IEEE Transactions on signal processing*, 55:2414–2424, 2007. MR1500172

[7] L. Cavalier and M. Raimondo. Wavelet deconvolution with noisy eigenvalues. *IEEE Trans. SIgnal Processing*, 55:2414–2424, 2007. MR1500172

[8] L. Cavalier and A.B. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probab. Theory Relat. Fields*, 123:323–254, 2002. MR1918537

[9] F. Comte and C. Lacour. Data driven density estimation in presence of unknown convolution operator. *J. Royal Stat. Soc., Ser B.*, 73:601–627, 2011. MR2853732

[10] J.G. McNally, C. Preza, J.A. Conchello and L.J. Thomas. Artifacts in computational optical-sectioning microscopy. *J. Opt. Soc. Am. A*, 11:1056–1067, 1994.

[11] K.R. Davidson and S.J. Szarek. *Local operator theory, random matrices and Banach spaces. In Handbook on the Geometry of Banach Spaces 1.* North-Holland, Amsterdam, (w. b. johnson and j. lindenstrauss, eds.) edition, 2001. MR1863696

[12] D. Donoho. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2:101–126, 1995. MR1325535

[13] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90:1200–1224, 1995. MR1379464

[14] S. Efromovich and V. Kolchinskii On inverse problems with unknown operators. *IEEE Transf. Inf. Theory*, 47:2876–2894, 2001. MR1872847

[15] S.F. Gibson and F. Lanni. Diffraction by a circular aperture as a model for three-dimensional optical microscopy. *J. Opt. Soc. Am. A*, 6, 1357–1367.

[16] D.M. Healy, H. Hendriks and P.T. Kim. Spherical deconvolution. *J. Multivariate Anal.*, 67, 1–22. MR1659108

[17] A. Cohen, M. Hoffmann and M. Reiss. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Numer. Anal.*, 42:1479–1501, 2004. MR2114287

[18] M. Hoffmann and M. Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.*, 36:310–336, 2008. MR2387973

[19] I.A. Ibragimov and R.Z Hasminskii. *Statistical Estimation. Asymptotic Theory.* Springer-Verlag, 1981. MR0620321

[20] J. Johannes. Deconvolution with unknown error distribution. *Ann. Statist.*, 37:2301–2323, 2009. MR2543693

[21] J. Johannes and M. Schwarz. Adaptive circular deconvolution by model selection under unknown error distribution. arXiv:0912.1207, 2012.

[22] J. Johannes and M. Schwarz. Adaptive gaussian inverse regression with partially unknown operator. arXiv:1204.1226v1 [math.ST], 2012.

[23] H.E. Keller. *Handbook of Biological Confocal Microscopy.* Plenum Press, New York, 2nd edition edition, 1995. chapter Objective lenses for confocal microscopy.

[24] P. Hall, G. Kerkyacharian and D. Picard. On the minimax optimality of block thresh- olded wavelet estimators. *Statist. Sinica*, 9:33–50, 1999. MR1678880

[25] Koo Kim and Luo. Weyl eigenvalue asymptotics and sharp adaptation on vector bundles. *Journal of Multivariate Analysis*, page 1962–1978, 2009.

[26] P.T. Kim and Y.Y. Koo Optimal spherical deconvolution. *J. Multivariate Anal.*, 80:21–42, 2002. MR1889831

[27] P. Massart. *Concentration inequalities and model selection*. Ecole d'Eté de Probabilités de Saint-Flour XXXIII (Jean Picard ed.), Lecture Notes in Mathematics 1986. Springer, 2007. MR2319879

[28] M.H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.*, 7:307–330, 1997. MR1460203

[29] G. Kerkyacharian, T.M. Pham Ngoc and D. Picard. Localized deconvolution on the sphere. *Ann. Statist.*, 39:1042–1068, 2011. MR2816347

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

[31] P. Pankajakshan, L. Blanc-Féraud, B. Zhang, Z. Kam, J-C. Olivo-Marin and J. Zerubia. Parametric blind deconvolution for confocal laser scanning microscopy (clsm)-proof of concept. *Projet INRIA Ariana, research report 6493*, 2008.

[32] G. Kerkyacharian, G. Kyriazis, E. Le Pennec, P. Petrushev and D. Picard. Inversion of noisy radon transform by svd based needlets. *Appl. Comput. Harmonic Anal.*, 28:24–45, 2010. MR2563258

[33] 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. MR2088290

[34] G. Reiner, C. Cremer, S. Hell and E.H.K. Stelzer. Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *J. Microscopy*, 169:391–405, 1993.

[35] P.A. Stokseth. Properties of a defocused optical system. *J. Opt. Soc. Am. A*, 59:1314–1321, 1969.

[36] A.B. Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C.R. Acad. Sci. Paris Sér. I Math.*, 330:835–840, 2000. MR1769957

[37] A.C.M. van Rooij and F.H. Ruymgaart. Regularized deconvolution on the circle and the sphere. *In Nonparametric Functional Estimation and Related Topics (Spetses, 1990). NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, 335:679–690, 1990. MR1154359

[38] N.J. Vilenkin. *Fonctions spéciales et théorie de la représentation des groupes*. Monographies Universitaires de Mathématiques, 33. Dunod, Paris, 1969. MR0243143