Estimation of linear operators from scattered impulse responses

Jérémie Bigot* Paul Escande† Pierre Weiss‡

December 7, 2017

Abstract

We provide a new estimator of integral operators with smooth kernels, obtained from a set of scattered and noisy impulse responses. The proposed approach relies on the formalism of smoothing in reproducing kernel Hilbert spaces and on the choice of an appropriate regularization term that takes the smoothness of the operator into account. It is numerically tractable in very large dimensions. We study the estimator's robustness to noise and analyze its approximation properties with respect to the size and the geometry of the dataset. In addition, we show minimax optimality of the proposed estimator.

Keywords: Integral operator, scattered approximation, estimator, convergence rate, numerical complexity, radial basis functions, Reproducing Kernel Hilbert Spaces, minimax.

AMS classifications: 47A58, 41A15, 41A25, 68W25, 62H12, 65T60, 94A20.

Acknowledgments

The authors wish to acknowledge the excellent reviewing work for this paper. Important technical inconsistencies were pointed out in the first version of the paper, which helped improving the manuscript substantially. The great care taken here has become very rare and we are truly indebted to the reviewers. The authors are also grateful to Bruno Torrésani and Rémi Gribonval for their interesting comments on a preliminary version of this paper. The PhD degree of Paul Escande has been supported by the MODIM project funded by the PRES of Toulouse University and the Midi-Pyrénées Région. This work was partially supported by the OPTIMUS project from RITC.

*Institut de Mathématiques de Bordeaux et CNRS, IMB-UMR5251, Université de Bordeaux, jeremie.bigot@math.u-bordeaux.fr

†Département d'Ingénierie des Systèmes Complexes (DISC), Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France, paul.escande@gmail.com

‡Institut des Technologies Avancées en Sciences du Vivant, ITAV-USR3505 and Institut de Mathématiques de Toulouse, IMT-UMR5219, CNRS and Université de Toulouse, Toulouse, France, pierre.armand.weiss@gmail.com

1 Introduction

Let $H : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ denote a linear integral operator defined for all $u \in L^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by:

$$Hu(x) = \int_{\mathbb{R}^d} K(x, y)u(y)dy, \tag{1.1}$$

where $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, is the operator kernel. Given a set of functions $(u_i)_{1\leq i\leq n}$, the problem of operator identification consists of recovering H from the knowledge of $f_i = Hu_i + \epsilon_i$, where ϵ_i is an unknown perturbation.

This problem arises in many fields of science and engineering such as mobile communication [20], imaging [15] and geophysics [3]. Many different reconstruction approaches have been developed, depending on the operator's regularity and the set of test functions (u_i). Assuming that H has a bandlimited Kohn-Nirenberg symbol and that its action on a Dirac comb is known, a few authors proposed extensions of Shannon's sampling theorem [20, 21, 30, 16]. Another recent trend is to assume that H can be decomposed as a linear combination of a small number of elementary operators. When the operators are fixed, recovering H amounts to solving a linear system. The work [7] analyzes the conditioning of this linear system when H is a matrix applied to a random Gaussian vector. When the operator can be sparsely represented in a dictionary of elementary matrices, compressed sensing theories can be developed [31]. Finally, in astrophysics, a few authors considered interpolating the coefficients of a few known impulse responses (also called Point Spread Functions, PSF) in a well chosen basis [15, 25, 6]. This strategy corresponds to assuming that $u_i = \delta_{y_i}$ and it is often used when the PSFs are compactly supported and have smooth variations. Notice that in this setting, each PSFs is known *independently* of the others, contrarily to the work [30].

This last approach is particularly effective in large scale imaging applications due to two useful facts. First, representing the impulse responses in a small dimensional basis allows reducing the number of parameters to identify. Second, there now exist efficient interpolation schemes based on radial basis functions. Despite its empirical success, this method still lacks of solid mathematical foundations and many practical questions remain open:

- Under what hypotheses on the operator H can this method be applied?
- What is the influence of the geometry of the set $(y_i)_{1\leq i\leq n}$?
- Is the reconstruction stable to the pertubations $(\epsilon_i)_{1\leq i\leq n}$? If not, how to make robust reconstructions, tractable in very large scale problems?
- What theoretical guarantees can be provided in this challenging setting?

The objective of this work is to address the above mentioned questions. We design a robust algorithm applicable in large scale applications. It yields a finite dimensional operator estimator of H allowing for fast matrix-vector products, which are essential for further processing. The theoretical convergence rate of the estimator as the number of observations increases is studied thoroughly.

The outline of this paper is as follows. We first specify the problem setting precisely in Section 2. We then describe the main outcomes of our study in Section 3. We provide a detailed explanation of the numerical algorithm in Section 4. Finally, the proofs of the main results are given in Section 5.

2 Problem setting

Throughout the paper, $\Omega \subset \mathbb{R}^d$ will denote a bounded, open and connected set, with Lipschitz continuous boundary.

The value of a function f at x is denoted $f(x)$, while the i-th value of a vector $v \in \mathbb{R}^N$ is denoted $v[i]$. The (i,j)-th element of a matrix A is denoted $A[i,j]$. The Sobolev space $H^s(\Omega)$ is defined for s in $\mathbb{N}$ by

$$H^s(\Omega) = \left\{ u \in L^2(\Omega), \partial^\alpha u \in L^2(\Omega), \text{ for all multi-index } \alpha \in \mathbb{N}^d \ s.t. \ |\alpha| = \sum_{i=1}^{d} \alpha[i] \leq s \right\}. \quad (2.1)$$

The space $H^s(\Omega)$ can be endowed with a norm $\|u\|_{H^s(\Omega)} = \left(\sum_{|\alpha|\leq s} \|\partial^\alpha u\|^2_{L^2(\Omega)}\right)^{1/2}$ and the semi-norm $|u|_{H^s(\Omega)} = \left(\sum_{|\alpha|=s} \|\partial^\alpha u\|^2_{L^2(\Omega)}\right)^{1/2}$. In addition, we will use the Beppo-Levi semi-norm defined by $|u|^2_{BL^s(\Omega)} = \sum_{|\alpha|=s} \frac{s!}{\alpha_1!\alpha_2!\ldots\alpha_d!} \|\partial^\alpha u\|^2_{L^2(\Omega)}$ and the Beppo-Levi semi-inner product defined by

$$\langle f, g \rangle_{BL^s(\Omega)} = \sum_{|\alpha|=s} \frac{s!}{\alpha_1!\alpha_2!\ldots\alpha_d!} \langle \partial^\alpha f, \partial^\alpha g \rangle_{L^2(\Omega)}. \quad (2.2)$$

Let a and b denote two functions depending on a parameter u living in a set U. The notation $a(u) \lesssim b(u)$ means that there exists a constant $c > 0$ such that $a(u) \leq cb(u)$ for all $u \in U$, with c independent of the parameters u. The notation $a(u) \asymp b(u)$ means that a and b are equivalent, i.e. there exists $0 < c \leq C$ such that $ca(u) \leq b(u) \leq Ca(u)$.

The Beppo-Levi and the Sobolev semi-norms are equivalent over the space $H^s(\mathbb{R}^d)$:

$$|u|^2_{BL^s(\Omega)} \asymp |u|^2_{H^s(\Omega)}. \quad (2.3)$$

2.1 The sampling model

An integral operator can be represented in many different ways. A key representation in this paper is the Space Varying Impulse Response (SVIR) $S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ by:

$$S(x,y) = K(x+y, y). \quad (2.4)$$

The impulse response or Point Spread Function (PSF) at location $y \in \mathbb{R}^d$ is defined by $S(\cdot, y)$.

The main purpose of this paper is the reconstruction of the SVIR of an operator from the observation of a few impulse responses $S(\cdot, y_i)$ at scattered (but known) locations $(y_i)_{1\leq i\leq n}$ in a

set Ω. In applications, the PSFs $S(\cdot, y_i)$ can only be observed through a projection onto an N dimensional linear subspace V_N. We assume that the linear subspace V_N reads

$$V_N = \operatorname{span}\left(\phi_k, 1 \leq k \leq N\right), \tag{2.5}$$

where $(\phi_k)_{k\in\mathbb{N}}$ is an orthonormal basis of $L^2(\mathbb{R}^d)$. In addition, the data is often corrupted by noise and we therefore observe a set of N dimensional vectors $(F_i^\epsilon)_{1\leq i\leq n}$ defined for all $k \in \{1, \ldots, N\}$ by

$$F_i^\epsilon[k] = \langle S(\cdot, y_i), \phi_k\rangle + \epsilon_i[k],\ 1 \leq i \leq n, \tag{2.6}$$

where ϵ_i is a random vector with independent and identically distributed (iid) components with zero mean and finite variance σ^2. For (2.6) to be well defined, S should be sufficiently smooth and we will provide precise regularity conditions in the next section.

Since impulse responses are observed on a bounded set Ω, we can only expect reconstructing S faithfully on $\mathbb{R}^d \times \Omega$ and not on the whole space $\mathbb{R}^d \times \mathbb{R}^d$. Hence the objective of this work is to define an estimator $\hat{H}$ with kernel $\hat{K}$ close to H with respect to the Hilbert-Schmidt norm defined by:

$$\|\hat{H} - H\|_{HS}^2 := \int_\Omega \int_{\mathbb{R}^d} |\hat{K}(x, y) - K(x, y)|^2 \, dx \, dy. \tag{2.7}$$

Controlling the Hilbert-Schmidt norm allows controlling the action of $\hat{H}$ on functions compactly supported on Ω. Indeed, for a function $u \in L^2(\mathbb{R}^d)$ with $\operatorname{supp}(u) \subseteq \Omega$, we get - using Cauchy-Schwarz inequality:

$$\begin{aligned}
\|\hat{H}u - Hu\|_{L^2(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} (\hat{K}(x, y) - K(x, y))u(y) \, dy \right)^2 dx \\
&\leq \int_{\mathbb{R}^d} \|\hat{K}(x, \cdot) - K(x, \cdot)\|_{L^2(\Omega)}^2 \|u\|_{L^2(\mathbb{R}^d)}^2 \, dx \\
&= \|\hat{H} - H\|_{HS}^2 \|u\|_{L^2(\mathbb{R}^d)}^2.
\end{aligned}$$

2.2 Space varying impulse response regularity

The SVIR encodes the impulse response variations in the y direction, instead of the $(x - y)$ direction for the kernel representation, see Figure 1 for a 1D example. It is convenient since in many applications, the smoothness of S in the x and y directions is driven by different physical phenomena. For instance in astrophysics, the regularity of $S(\cdot, y)$ depends on the optical system, while the regularity of $S(x, \cdot)$ may depend on exteriors factors such as atmospheric turbulence or weak gravitational lensing [6]. This property will be expressed through the specific regularity assumptions of S defined hereafter.

First we will make use of the following functional space.

Definition 2.1. The space $\mathcal{E}^r(\mathbb{R}^d)$ (also denoted $\mathcal{E}^r$) is defined, for all $r \in \mathbb{R}$ and $r > \frac{d}{2}$, as the set of functions $u \in L^2(\mathbb{R}^d)$ such that:

$$\|u\|_{\mathcal{E}^r(\mathbb{R}^d)}^2 = \sum_{k\in\mathbb{N}} w[k] |\langle u, \phi_k\rangle|^2 < +\infty, \tag{2.8}$$

where $w : \mathbb{N} \to \mathbb{R}_+^*$ is a weight sequence satisfying $w[k] \gtrsim (1+k^2)^{r/d}$.

Remark 2.1. This definition is introduced in reference to the Sobolev spaces $H_\Delta^m(\mathbb{R}^d)$ of functions with m derivatives in $L^2(\mathbb{R}^d)$ supported on a compact set Δ. This space can be defined - alternatively to equation (2.1) - by:

$$H_\Delta^m(\mathbb{R}^d) = \left\{ u \in L^2(\mathbb{R}^d), \sum_{\lambda \in \Lambda} 2^{2m|\lambda|} |\langle u, \psi_\lambda \rangle|^2 < +\infty \right\}, \tag{2.9}$$

where $(\psi_\lambda)_{\lambda \in \Lambda}$ is a wavelet basis with at least $m+1$ vanishing moments (see e.g. [24, Chapter 9]) and $\lambda = (j,k)$ is a scale-space parameter.

Remark 2.2. Definition 2.1 encompasses many other spaces. For instance, it allows choosing a basis $(\phi_k)_{k \in \mathbb{N}}$ that is best adapted to the impulse responses at hand, by using principal component analysis, as was proposed in a few applied papers [18, 4].

The following definition gathers all the assumptions made on the operators. It will be used throughout the paper.

Definition 2.2. Let A_1 and A_2 be positive constants. Set $r > \frac{d}{2}$ and $s > \frac{d}{2}$. The ball $\mathcal{E}^{r,s}(A_1, A_2)$ is defined as the set of linear integral operators H with SVIR S belonging to $L^2(\mathbb{R}^d \times \mathbb{R}^d)$ with:

Smooth variations
$$\int_{x \in \mathbb{R}^d} \|S(x,\cdot)\|^2_{H^s(\mathbb{R}^d)} dx \leq A_1 \tag{2.10}$$

Impulse response regularity
$$\int_{y \in \mathbb{R}^d} \|S(\cdot,y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)} dy \leq A_2 \tag{2.11}$$

Let us comment on these assumptions:

- Equation (2.11) means that $S(\cdot, y)$ belong to $\mathcal{E}^r(\mathbb{R}^d)$ for a.e. $y \in \mathbb{R}^d$.
- Similarly, assumption (2.10) means that $S(x,\cdot)$ is in $H^s(\mathbb{R}^d)$ for a.e. $x \in \mathbb{R}^d$. The hypothesis $s > d/2$ ensures existence of a *continuous* representant of $S(x,\cdot)$ for a.e. x, by Sobolev embedding theorems [34, Thm.2, p.124]. This regularity condition will allow the use of fine approximation results based on radial basis functions [2].
- The two regularity conditions are sufficient for the sampling procedure (2.6) to be well defined. Lemma 5.3 indeed indicates that the functions $y \mapsto \langle S(\cdot,y), \phi_k \rangle$ are in $H^s(\mathbb{R}^d)$ for all $k \in \mathbb{N}$. By Sobolev embedding theorems [34, Thm.2, p.124], there exists a continuous representant of these functions and hence, we can give a meaning to $\langle S(\cdot,y), \phi_k \rangle$.
- In the particular case where $\mathcal{E}^r(\mathbb{R}^d) = H^r(\mathbb{R}^d)$ the space $\mathcal{E}^{r,s}$ is the mixed-Sobolev space $H^{r,s}(\mathbb{R}^d \times \mathbb{R}^d)$ [22, 29, 42].

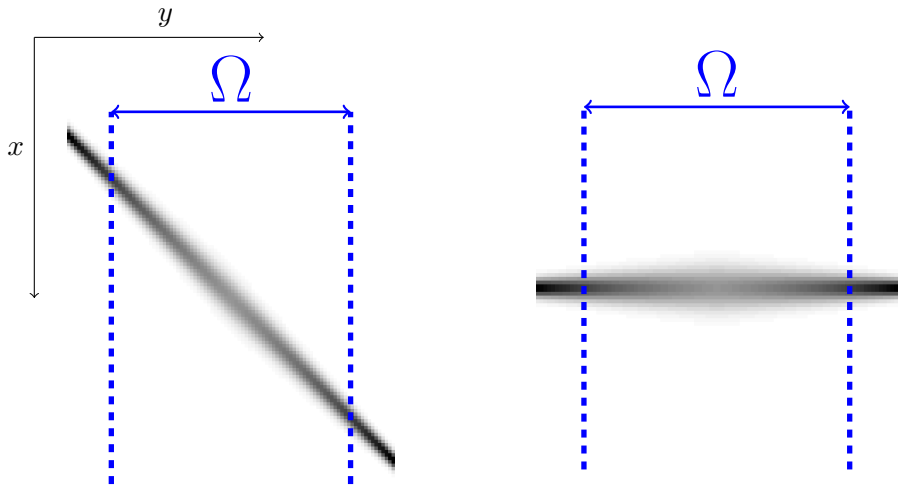


Figure 1: Illustration of the two representations of an integral operator considered in this paper. The kernel is defined by $K(x,y) = \frac{1}{\sqrt{2\pi}\sigma(y)} \exp\left(-\frac{1}{2\sigma(y)^2}|x-y|^2\right)$ for all $(x,y) \in \mathbb{R}^2$, where $\sigma(y) = 0.05\,(1 + 2\min(y, 1-y))$ for $y \in \Omega = [0,1]$. Left: kernel representation (see equation (1.1)). Right: SVIR representation (see equation (2.4)).

3 Main results

3.1 Construction of an estimator

Let $F : \mathbb{R}^d \to \mathbb{R}^N$ denote the vector-valued function representing the impulse responses coefficients (IRC) in basis $(\phi_k)_{k\in\mathbb{N}}$:

$$F(y)[k] = \langle S(\cdot, y), \phi_k \rangle. \tag{3.1}$$

Based on the observation model (2.6), a natural approach to estimate the SVIR, consists in constructing an estimate $\hat{F} : \mathbb{R}^d \to \mathbb{R}^N$ of F. The estimated SVIR is then defined as

$$\hat{S}(x,y) = \sum_{k=1}^{N} \hat{F}(y)[k]\phi_k(x), \text{ for } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{3.2}$$

Definition 2.2 motivates the introduction of the following space.

Definition 3.1 (Space $\mathcal{H}$ of IRC)**.** The space $\mathcal{H}(\mathbb{R}^d)$ of admissible IRC is defined as the set of vector-valued functions $G : \mathbb{R}^d \to \mathbb{R}^N$ such that

$$\|G\|^2_{\mathcal{H}(\mathbb{R}^d)} = \alpha \int_{y\in\mathbb{R}^d} \sum_{k=1}^{N} w[k]\,|G(y)[k]|^2\,dy + (1-\alpha) \sum_{k=1}^{N} |G(\cdot)[k]|^2_{BL^s(\mathbb{R}^d)} < +\infty, \tag{3.3}$$

where $\alpha \in [0,1)$ allows to balance the smoothness in each direction.

The following result is straightforward (the proof is similar to that of Lemma 5.2).

Lemma 3.1. *Operators in $\mathcal{E}^{r,s}(A_1, A_2)$ have an IRC belonging to $\mathcal{H}(\mathbb{R}^d)$.*

To construct an estimator of F, we propose to define $\hat{F}_\mu$ as the minimizer of the following optimization problem:

$$\hat{F}_\mu = \underset{F \in \mathcal{H}(\mathbb{R}^d)}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|F_i^\epsilon - F(y_i)\|_{\mathbb{R}^N}^2 + \mu \|F\|_{\mathcal{H}(\mathbb{R}^d)}^2, \tag{3.4}$$

where $\mu > 0$ is a regularization parameter.

Remark 3.1. The proposed formulation can be interpreted with the formalism of regression and smoothing in vector-valued Reproducing Kernel Hilbert Spaces (RKHS) [26, 27]. The space $\mathcal{H}(\mathbb{R}^d)$ can be shown to be a vector-valued Reproducing Kernel Hilbert Space (RKHS). The formalism of vector-valued RKHS has been developed for the purpose of multi-task learning, and its application to operator estimation appears to be novel.

3.2 Mixed-Sobolev space interpretation

The problem formulation (3.4) might seem abstract at first sight. In this section we show that it encompasses the formalism of mixed-Sobolev spaces [22, 29, 42] and that the proposed methodology can be interpreted in terms of SVIR instead of IRC.

Lemma 3.2. *Suppose $H \in \mathcal{E}^{r,s}(A_1, A_2)$. In the specific case where $(\phi_k)_{k\in\mathbb{N}}$ is a wavelet or a Fourier basis and $N = +\infty$, The cost function in Problem* (3.4) *is equivalent to*

$$\frac{1}{n} \sum_{i=1}^{n} \left\| F_i^\epsilon - (\langle S(\cdot, y_i), \phi_k \rangle)_{1 \le k \le N} \right\|_2^2 + \mu \left(\alpha \int_{\mathbb{R}^d} \|S(\cdot, y)\|_{H^r(\mathbb{R}^d)}^2 dy + (1-\alpha) \int_{\mathbb{R}^d} |S(x, \cdot)|_{BL^s(\mathbb{R}^d)}^2 dx \right). \tag{3.5}$$

Proof. The proof is straightforward once showing the results in Lemma 5.3. □

This formulation is quite intuitive: the data fidelity term allows finding a TVIR that is close to the observed data, the first regularization term allows smoothing the additive noise on the acquired PSFs and the second one interpolates the missing data.

3.3 Numerical complexity

Thanks to the results in [27], computing $\hat{F}_\mu$ amounts to solving a *finite-dimensional* system of linear equations. However, for an arbitrary orthonormal basis $(\phi_k)_{k\in\mathbb{N}}$, and without any further assumptions on the kernel of the RKHS $\mathcal{H}(\mathbb{R}^d)$, evaluating $\hat{F}_\mu$ leads to the resolution of a *full* $nN \times nN$ linear system, which is untractable for large N and n.

With the specific choice of norm introduced in Definition 3.1, the problem simplifies to the resolution of N systems of equations of size $n \times n$. This step is investigated in details in Section 4. In this paragraph we gather the results describing the numerical complexity of the method.

Proposition 3.1. *The solution of* (3.4) *can be computed in no more than $O(Nn^3)$ operations for any choice of basis $(\phi_k)_{k\in\mathbb{N}}$.*

Proof. See Section 4. □

In addition, if the weight function w is piecewise constant, some $n \times n$ matrices are identical, allowing to compute an LU factorization once for all and using it to solve many systems. This yields the following result.

Proposition 3.2. *In the specific case where $(\phi_k)_{k\in\mathbb{N}}$ is a wavelet basis, it is natural to set function w as a constant over each wavelet subband [24, Thm. 9.4]. Then, the solution of* (3.4) *can be computed in no more than $O\left(\frac{\log(N)}{d}n^3 + Nn^2\right)$ operations.*

Proof. See Section 4. □

Finally let us remark that for well chosen bases $(\phi_k)_{k\in\mathbb{N}}$ the impulse responses can be well approximated using a small number N of atoms. Such instances of bases include Fourier bases, wavelet bases with appropriate properties and the basis formed with the principal components of the impulse responses. This makes the method tractable even in very large scale applications.

To conclude this paragraph, let us mention that the representation of an operator of type (3.2) can be used to evaluate matrix-vector products rapidly. We refer the interested reader to [12] for more details.

3.4 Convergence rates

The convergence of the proposed estimator with respect to the number n of observations is captured by the theorems of this section. We show that the approximation efficiency of our method depends on the geometry of the set of data locations, and - in particular - on the fill and separation distances defined below.

Definition 3.2 (Fill distance)**.** The fill distance of $Y = \{y_1, \ldots, y_n\} \subset \Omega$ is defined as:

$$h_{Y,\Omega} = \sup_{y\in\Omega} \min_{1\leq j\leq n} \|y - y_j\|_2. \tag{3.6}$$

This is the distance for which any $y \in \Omega$ is at most at a distance $h_{Y,\Omega}$ of Y. It can also be interpreted as the radius of the largest ball with center in Ω that does not intersect Y.

Definition 3.3 (Separation distance)**.** The separation distance of $Y = \{y_1, \ldots, y_n\} \subset \Omega$ is defined as:

$$q_{Y,\Omega} = \frac{1}{2} \min_{i\neq j} \|y_i - y_j\|_2. \tag{3.7}$$

This quantity gives the maximal radius $r > 0$ such that all balls $\{y \in \mathbb{R}^d : \|y - y_j\|_2 < r\}$ are disjoint.

The following condition plays a key role in our analysis [28, 32].

Definition 3.4 (Quasi-uniformity condition)**.** A set of data locations $Y = \{y_1, \ldots, y_n\} \subset \Omega$ is said to be quasi-uniform with respect to a constant $B > 0$ if

$$q_{Y,\Omega} \leq h_{Y,\Omega} \leq B q_{Y,\Omega}. \tag{3.8}$$

Remark 3.2. Our main theorems will be stated under a quasi-uniformity condition of the sampling set. It is likely that this hypothesis can be refined using more stable reconstruction schemes as is commonly done in the reconstruction of bandlimited functions [13].

Theorem 3.1. *Assume that $H \in \mathcal{E}^{r,s}(A_1, A_2)$ and that its SVIR S is sampled using model* (2.6) *under the quasi-uniformity condition given in Definition 3.4. Then the estimating operator $\hat{H}$ with SVIR $\hat{S}$ defined in equation* (3.2) *satisfies the following inequality*

$$\mathbb{E}\left(\|H - \hat{H}\|_{HS}^2\right) \lesssim N^{-\frac{2r}{d}} + (N\sigma^2 n^{-1})^{\frac{2s}{2s+d}}(1-\alpha)^{-\frac{2s+2d}{2s+d}}, \tag{3.9}$$

for $\mu \propto (N\sigma^2 n^{-1})^{\frac{2s}{2s+d}}(1-\alpha)^{\frac{-d}{2s+d}}$. This inequality holds uniformly on the ball $\mathcal{E}^{r,s}(A_1, A_2)$.

Proof. See Section 5. □

In applications where the user can choose the number of observations N (e.g. if it is sufficiently large), the upper-bound (3.9) can be optimized with respect to N.

Corollary 3.1. *Assume that $H \in \mathcal{E}^{r,s}(A_1, A_2)$ and that its SVIR S is sampled using model* (2.6) *under the quasi-uniformity condition given in Definition 3.4. Then the estimator $\hat{H}$ with SVIR $\hat{S}$ defined in equation* (3.2) *satisfies the following inequality*

$$\mathbb{E}\left(\|H - \hat{H}\|_{HS}^2\right) \lesssim (\sigma^2 n^{-1}(1-\alpha)^{-(d/s+1)})^{\frac{2q}{2q+d}}, \tag{3.10}$$

with the relation $1/q = 1/r + 1/s$, for $\mu \propto (\sigma^2 n^{-1})^{\frac{2q}{2q+d}}(1-\alpha)^{\frac{-d}{2s+d}}$ and $N \propto (\sigma^2 n^{-1})^{-\frac{dq}{r(2q+d)}}(1-\alpha)^{\frac{(d^2+sd)q}{rs(2q+d)}}$. This inequality holds uniformly on the ball $\mathcal{E}^{r,s}(A_1, A_2)$.

Proof. See Section 5. □

Corollary 3.1 gives some insights on the estimator behavior. In particular:

- It provides an explicit way of choosing the value of the regularization parameter μ: it should decrease as the number of observations increases.
- If the number of observations n is small, it is unnecessary to project the impulse responses on a high dimensional basis (i.e. N large). The basic reason is that not enough information has been collected to reconstruct the fine details of the kernel.
- The optimal value of α in the corollary is $\alpha = 0$, suggesting that the best option is to not use the additional regularizer $\alpha \int_{y\in\mathbb{R}^d} \sum_{k=1}^N w[k]\,|G(y)[k]|^2\,dy$. This phenomenon is due to a rough upper-bound in the proof. Unfortunately, we did not manage to obtain finer estimates of some eigenvalues in the proof. From a practical perspective, we observed a good behavior of this additional term in our numerical experiments and therefore decided to present the theory including this regularizer.

Finally, to conclude this section on convergence rates, it is shown that, under mild assumptions on the basis $(\phi_k)_{k\geq 1}$, the rate of convergence $(\sigma^2 n^{-1})^{\frac{2q}{2q+d}}$ in inequality (3.10) is optimal in the case of Gaussian noise and for the expected Hilbert-Schmidt norm $\mathbb{E}\left\|H-\hat{H}\right\|_{HS}^2$. Optimality of the rate of convergence (3.10) has to be understood in the minimax sense as classically done in the literature on nonparametric statistics (we refer to [35] for a detailed introduction to this topic). For simplicity, this optimality result is stated in the case where the domain $\Omega=[0,1]^d$ is the d-dimensional hypercube.

Theorem 3.2. *Let H be a linear operator belonging to $\mathcal{E}^{r,s}(A_1,A_2)$. Define q by $1/q=1/r+1/s$. Suppose that the weights in Definition 2.1 satisfy $w[k]\leq c_1(1+k^2)^{r/d}$ for all $k\in\mathbb{N}$ and some constant $c_1>0$. Assume that the PSF locations $y_1,\ldots,y_n$ satisfy the quasi-uniformity condition given in Definition 3.4. Assume that the random values $(\epsilon_i[k])_{i,k}$ in the observation model* (2.6) *are iid Gaussian with zero mean and variance σ^2.*

Then, there exists a constant $c_0>0$ such that

$$\inf_{\hat{H}}\sup_{H\in\mathcal{E}^{r,s}(A_1,A_2)}\mathbb{E}\left\|\hat{H}-H\right\|_{HS}^2\geq c_0(\sigma^2 n^{-1})^{\frac{2q}{2q+d}}, \tag{3.11}$$

where the above infimum is taken over all possible estimators $\hat{H}$ (linear integral operators) with SVIR $\hat{S}\in L^2(\mathbb{R}^d\times\mathbb{R}^d)$ defined as a measurable function.

Proof. See Section 5. □

Remark 3.3. In this paper we only study the robustness of the method towards perturbations over the discretization of the impulse responses. It is also of great interest to study the behavior of the method with respect to to jitter errors, i.e. what happens if the impulse responses are sampled at perturbed positions y_i' instead of the exact y_i? This question is left aside in this paper but the analysis in [14] suggests than one can also prove some robustness of the method with respect to those type of perturbations.

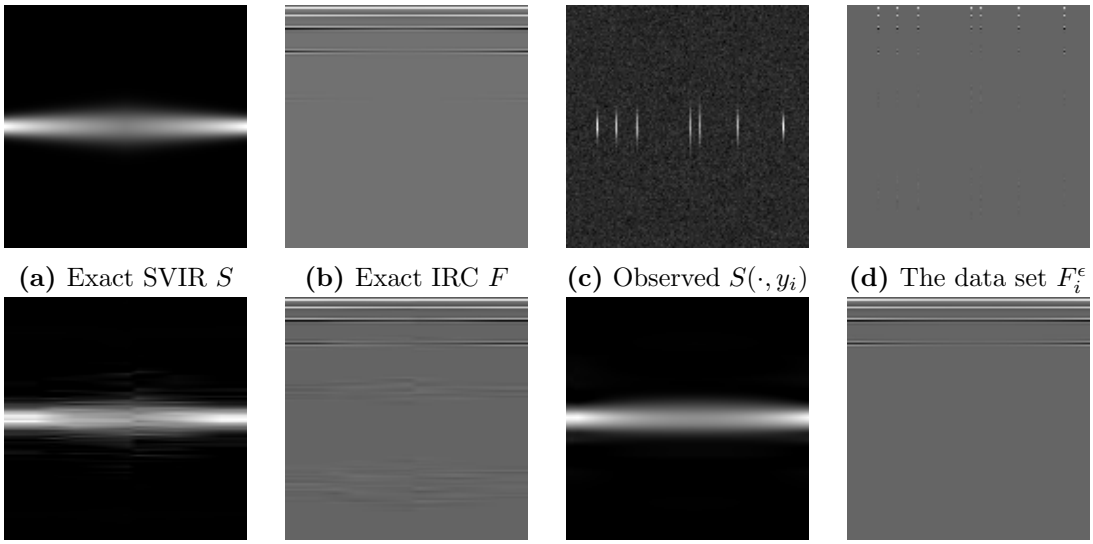
3.5 Illustrations and numerical experiments

In this section, we highlight the main ideas of the paper through two numerical experiments.

A 1D estimation problem In the first experiment, we wish to reconstruct the operator H with kernel $K(x,y)=\frac{1}{\sqrt{|2\pi\Sigma(y)|}}\exp\left(-\langle\Sigma(y)^{-1}(y-x),y-x\rangle\right)$, with diagonal covariance matrices $\Sigma(y)=\sigma(y)\mathrm{Id}$ where $\sigma(y)=1+2\max(1-y,y)$ for $y\in[0,1]$. The SVIR (Space Varying Impulse Response) and the IRC (Impulse Response Coefficients) of this kernel are shown in Fig. 2 (a) and (b). Here, we projected the impulse responses on a discrete orthogonal wavelet basis. Notice how the information is compacted, in (b) compared to (a): most of the information is concentrated on just a fews rows.

In Fig. 2 (c), we show the 7 impulse responses that are used to estimate the kernel. In Fig. 2 (d), we show their projection on the orthogonal wavelet basis. The problem studied in this paper is to estimate the SVIR in (a) from the data in (d). Given the noisy dataset, the

proposed algorithm simultaneously interpolates along rows and denoises along columns to obtain the results in Figure 2 (e-h). Notice how the regularization in the vertical direction ($\alpha > 0$) allows improving the estimator: the result in (g) is very similar to (a).



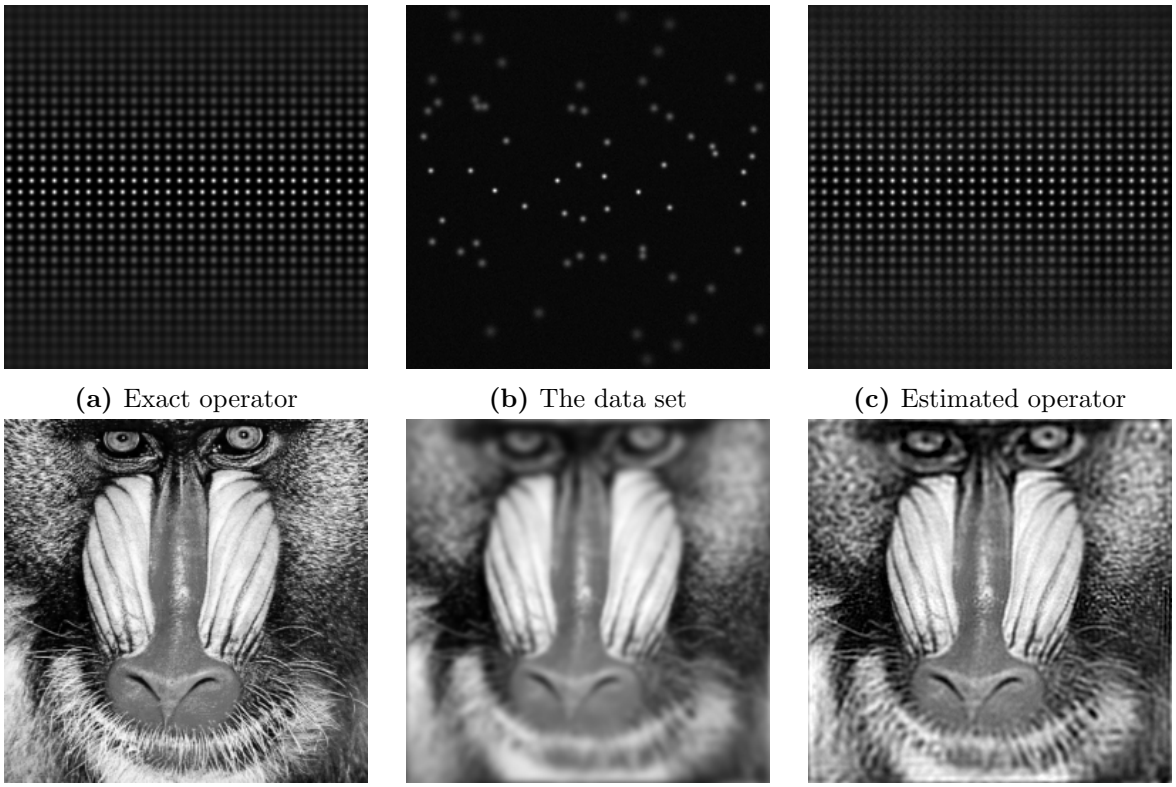
(a) Exact SVIR S **(b)** Exact IRC F **(c)** Observed $S(\cdot, y_i)$ **(d)** The data set F_i^ϵ

(e) Estimated SVIR $\widehat{S}$– without denoising ($\alpha = 0$) **(f)** Estimated IRC $\widehat{F}$– without denoising ($\alpha = 0$) **(g)** Estimated SVIR $\widehat{S}$– with denoising ($\alpha = 0.3$) **(h)** Estimated IRC $\widehat{F}$– with denoising ($\alpha = 0.3$)

Figure 2: Illustration of the methodology and its results on a 1D estimation problem.

A 2D deblurring problem In this experiment, we show how the proposed ideas allow estimating a blur operator in imaging and then use this estimate to deblur images. The results are displayed in Fig. 3. In Fig. 3 (a), an operator H is applied to a 2D Dirac comb, providing an idea of the operator's shape: each impulse response is an isotropic Gaussian with variance $\sigma(y_1, y_2)$ varying along the vertical direction only (namely $\sigma(y_1, y_2) = 1 + 2\max(1 - y_1, y_1)$ for $(y_1, y_2) \in [0, 1]^2$). In Fig. 3 (b), we show a set of noisy impulse responses that will be used to perform the estimation. Since the impulse response are near compactly supported, we can isolate each of them in the image to perform the estimation. Here the projection basis (ϕ_k) is simply the canonical basis. In Fig. 3 (c), we show the estimated operator $\widehat{H}$ through its action on a Dirac comb. The estimation seems faithful to the exact operator in (a).

To validate the findings, we perform a deblurring experiment. An sharp image in (d) is blurred with the exact operator H in (a), and some white Gaussian noise is added. Then, using the operator $\widehat{H}$ estimated in (c), we deblur the image with a total variation regularized inverse problem [33]. As can be seen, the image is significantly sharper, despite some ringing appearing in the bottom.



(a) Exact operator

(b) The data set

(c) Estimated operator

(d) Sharp image
256×256

(e) Degraded image
pSNR = 19.17dB

(f) Restored image
pSNR = 21.20dB

Figure 3: A 2D estimation used to deblur images.

4 Radial basis functions implementation

The objective of this section is to provide a fast algorithm to solve Problem (3.4) and to prove Propositions 3.1 and 3.2. A key observation is provided below.

Lemma 4.1. *For $k \in \{1, \dots, N\}$, the function $\hat{F}(\cdot)[k]$ is the solution of the following variational problem:*

$$\min_{f \in H^s(\mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^{n} (F_i^\epsilon[k] - f(y_i))^2 + \mu \left(\alpha w[k] \|f\|^2_{L^2(\mathbb{R}^d)} + (1-\alpha) |f|^2_{BL^s(\mathbb{R}^d)} \right). \tag{4.1}$$

Proof. It suffices to remark that Problem (3.4) consists of solving N independent sub-problems. □

We now focus on the resolution of Sub-problem (4.1) which completely fits the framework of radial basis function approximation. In the sequel, we gather a few important results related to radial basis functions that will be used to construct the algorithm.

4.1 Standard approximation results in RKHS

A nice way to introduce radial basis functions is through the theory of reproducible kernel Hilbert spaces (RKHS). We recall the basic definitions and a few key results regarding RKHS. Most of them can be found in the book of Wendland [39].

Definition 4.1 (Positive definite function)**.** A continuous function $\rho : \mathbb{R}^d \to \mathbb{C}$ is called positive semi-definite if, for all $n \in \mathbb{N}$, all sets of pairwise distinct centers $Y = \{y_1, \dots, y_n\} \subset \mathbb{R}^d$, and all $\alpha \in \mathbb{C}^n$, the quadratic form

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \bar{\alpha}_k \rho(y_j - y_k) \tag{4.2}$$

is nonnegative. It is called positive definite if (4.2) is positive for all $\alpha \neq 0$ and all sets of pairwise distinct locations Y.

Definition 4.2 (Reproducing kernel)**.** Let $\mathcal{G}$ denote a Hilbert space of real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$ endowed with a scalar product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$. A function $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called reproducing kernel for $\mathcal{G}$ if

1. $\Phi(\cdot, y) \in \mathcal{G}, \quad \forall y \in \mathbb{R}^d$,
2. $f(y) = \langle f, \Phi(\cdot, y) \rangle_{\mathcal{G}}$, for all $f \in \mathcal{G}$ and all $y \in \mathbb{R}^d$.

Theorem 4.1 (RKHS)**.** *Suppose that $\mathcal{G}$ is a Hilbert space of functions $f : \mathbb{R}^d \to \mathbb{R}$. Then the following statements are equivalent:*

1. *the point evaluations functionals δ_y are continuous for all $y \in \mathbb{R}^d$.*
2. *$\mathcal{G}$ has a reproducing kernel.*

A Hilbert space satisfying the properties above is called a Reproducing Kernel Hilbert Space (RKHS).

The Fourier transform of a function $f \in L^1(\mathbb{R}^d)$ is defined by

$$\mathcal{F}[f](\xi) = (2\pi)^{-d/2} \int_{x \in \mathbb{R}^d} f(x) e^{-i\langle x, \xi \rangle} dx, \tag{4.3}$$

and the inverse transform by

$$\mathcal{F}^{-1}[\widehat{f}](x) = (2\pi)^{-d/2} \int_{\xi \in \mathbb{R}^d} \widehat{f}(\xi) e^{i\langle x, \xi \rangle} d\xi. \tag{4.4}$$

The Fourier transform can be extended to $L^2(\mathbb{R}^d)$ and to $\mathcal{S}'(\mathbb{R}^d)$ the space of tempered distributions.

Theorem 4.2 ([39, Theorem 10.12])**.** *Suppose that $\rho \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ is a real-valued positive definite function. Define $\mathcal{G} = \left\{ f \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) : \mathcal{F}[f]/\sqrt{\mathcal{F}[\rho]} \in L^2(\mathbb{R}^d) \right\}$ equipped with*

$$\langle f, g \rangle_{\mathcal{G}} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\mathcal{F}[f](\xi)\overline{\mathcal{F}[g](\xi)}}{\mathcal{F}[\rho](\xi)} d\xi. \tag{4.5}$$

Then $\mathcal{G}$ is a real Hilbert space with inner-product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ and reproducing kernel Φ defined as $\Phi(x, y) = \rho(x - y)$ for all $x, y \in \mathbb{R}^d$.

This theorem is a consequence of Sobolev embedding theorems [1]. In the following, we will make the abuse to call $\rho : \mathbb{R}^d \to \mathbb{R}$ the reproducing kernel of an Hilbert space $\mathcal{G}$. It should be understood as: the reproducing kernel of $\mathcal{G}$ is $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined as $\Phi(x, y) = \rho(x - y)$ for all $x, y \in \mathbb{R}^d$.

Theorem 4.3. *Let $\mathcal{G}$ be an RKHS with positive definite reproducing kernel $\rho : \mathbb{R}^d \to \mathbb{R}$. Let $(y_1, \ldots, y_n)$ denote a set of points in $\mathbb{R}^d$ and $z \in \mathbb{R}^n$ denote a set of altitudes. The solution of the following approximation problem*

$$\min_{u \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} (u(y_i) - z[i])^2 + \frac{\mu}{2} \|u\|_{\mathcal{G}}^2 \tag{4.6}$$

can be written as:

$$u(x) = \sum_{i=1}^{n} c[i] \rho(x - y_i), \tag{4.7}$$

where vector $c \in \mathbb{R}^n$ is the unique solution of the following linear system of equations

$$(G + n\mu \mathrm{Id})c = z \text{ with } G[i, j] = \rho(y_i - y_j). \tag{4.8}$$

It is shown in [39], that the condition number of G depends on the ratio $h_{Y,\Omega}/q_{Y,\Omega}$. For numerical reasons it might therefore be useful to implement thinning methods in order to discard locations that are too close to each other without creating larger gaps, if possible [10, 17, 11].

4.2 Application to our problem

Let us now show how the above results help solving Problem (4.1).

Proposition 4.1. *Let $\mathcal{G}_k$ be the Hilbert space of functions $f : \mathbb{R}^d \to \mathbb{R}$ such that $|f|^2_{BL^s(\mathbb{R}^d)} + \|f\|^2_{L^2(\mathbb{R}^d)} < +\infty$, equipped with the inner product:*

$$\langle f, g\rangle_{\mathcal{G}_k} = (1-\alpha)\,\langle f, g\rangle_{BL^s(\mathbb{R}^d)} + \alpha w[k]\langle f, g\rangle^2_{L^2(\mathbb{R}^d)}. \tag{4.9}$$

Then $\mathcal{G}_k$ is an RKHS and its scalar product reads

$$\langle f, g\rangle_{\mathcal{G}_k} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\mathcal{F}[f](\xi)\overline{\mathcal{F}[g](\xi)}}{\mathcal{F}[\rho_k](\xi)} d\xi, \tag{4.10}$$

where the reproducing kernel ρ_k, is defined by:

$$\mathcal{F}[\rho_k](\xi) = \left((1-\alpha)\|\xi\|^{2s} + \alpha w[k]\right)^{-1}. \tag{4.11}$$

Proof. The proof is a direct application of the different results stated previously. □

The Fourier transform $\mathcal{F}[\rho_k]$ is radial, so that ρ_k is radial too and the resolution of (4.1) fits the formalism of radial basis functions interpolation/approximation [5].

Remark 4.1. For some applications, it makes sense to set $w[k] = 0$ for some values of k. For instance, if $(\phi_k)_{k\in\mathbb{N}}$ is a wavelet basis, then it is usually good to set $w[k] = 0$ when k is the index of a scaling wavelet. In that case, the theory of conditionally positive definite kernels should be used instead of the one above. We do not detail this aspect since it is well described in standard textbooks [39, 5].

The whole procedure computing $\widehat{F}$ is presented in Algorithm 1. The principle of the algorithm is derived from Lemma 4.1 showing that computing $\widehat{F}$ solution of (3.4), boils down to solving N independent sub-systems. Each sub-system computes $\widehat{F}(\cdot)[k]$ and according to Proposition 4.1 it falls in the formalism of RKHS with an explicit definition of the kernel. Therefore, in virtue of Theorem 4.3 each function $\widehat{F}(\cdot)[k]$ can be computed by solving a $n \times n$ linear system. The resolution of the linear systems can accelerated using LU decompositions. Hence, it starts with a preprocessing step.

The associated $\widehat{S}$ can be recovered for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ through

$$\widehat{S}(x, y) = \sum_{k=1}^{N} \widehat{F}(y)[k]\phi_k(x) = \sum_{k=1}^{N}\sum_{i=1}^{n} c_k[i]\rho_k(y - y_i)\phi_k(x). \tag{4.12}$$

Before being able to use $\widehat{S}$ for subsequent numerical algorithms, the IRC $\widehat{F}$ might have to be discretized or sampled. The complexity of this step is not comprised in Proposition 3.1 and depends on the discretization procedure. In many cases, $\widehat{F}(y)[k]$ has to be evaluated on a Cartesian grid. This step can be performed efficiently by using nonuniform fast Fourier transforms or multipole methods [39].

Algorithm 1 Computation of $\widehat{F}$

Input:

Weight vector $w \in \mathbb{R}^N$

Regularity $s \in \mathbb{N}$

PSF locations $Y = \{y_1, \dots, y_n\} \in \mathbb{R}^{d \times n}$

Observed data $(F_i^\epsilon)_{1 \leq i \leq n}$, where $F_i^\epsilon \in \mathbb{R}^N$

Output:

The IRC estimator $\widehat{F}$

Algorithm:

Identify the $m \leq N$ weights of identical values in vector $w \in \mathbb{R}^N$. ▷ $O(N)$

for Each unique weight ω **do** ▷ $O(mn^3)$

  Compute matrix G from formula (4.8) with ρ_ω defined in (4.11).

  Compute a LU decomposition of $M_\omega = (G + n\mu\mathrm{Id}) = L_\omega U_\omega$.

end for

for $k = 1$ to N **do** ▷ $O(Nn^2)$

  Identify the value ω such that $w[k] = \omega$.

  Set $z = (F_i^\epsilon[k])_{1 \leq i \leq n}$.

  Solve the linear system $L_\omega U_\omega c_k = z$.

  Possibly reconstruct $\hat{F}$ by (see equation (4.7))

$$\hat{F}(y)[k] = \sum_{i=1}^{n} c_k[i] \rho_\omega(y - y_i).$$

end for

5 Proofs of the main results

First we prove Theorem 3.1 about the convergence rate of the quadratic risk $\mathbb{E}\|H - \hat{H}\|_{HS}^2$.

5.1 Operator norm risk

To analyse the theoretical properties of a given estimator of the operator H, we introduce the quadratic risk defined as:

$$R(\hat{H}, H) = \mathbb{E}\left\|\hat{H} - H\right\|_{HS}^2, \tag{5.1}$$

where $\hat{H}$ is the operator associated to the SVIR $\hat{S}$ defined in (3.2). The above expectation is taken with respect to the distribution of the observations in (2.6). Notice that $\|H\|_{HS} = \|K\|_{L^2(\mathbb{R}^d\times\Omega)} = \|S\|_{L^2(\mathbb{R}^d\times\Omega)}$. From this observation we get that:

$$\begin{aligned} R(\hat{H}, H) &= \mathbb{E}\left\|\hat{H} - H\right\|_{HS}^2 \\ &\leq 2\left(\|H - H_N\|_{HS}^2 + \mathbb{E}\left\|H_N - \hat{H}\right\|_{HS}^2\right) \\ &= 2\left(\underbrace{\|S - S_N\|_{L^2(\mathbb{R}^d\times\Omega)}^2}_{\epsilon_d(N)} + \underbrace{\mathbb{E}\left\|S_N - \hat{S}\right\|_{L^2(\mathbb{R}^d\times\Omega)}^2}_{\epsilon_e(n)}\right), \end{aligned} \tag{5.2}$$

where H_N is the operator associated to the SVIR S_N defined by $S_N(x,y) = \sum_{k=1}^N F(y)[k]\phi_k(x)$ and $\hat{H}$ the estimating operator associated to the SVIR $\hat{S}$ as in (3.2).

In equation (5.2), the risk is decomposed as the sum of two terms $\epsilon_e(n)$ and $\epsilon_d(N)$ (standard bias/variance decomposition in statistics). The first one $\epsilon_d(N)$ is the error introduced by the discretization step. The second term $\epsilon_e(N)$ is the quadratic risk between S_N and the estimator $\hat{S}$. In the next sections, we provide upper-bounds for $\epsilon_d(N)$ and $\epsilon_e(n)$.

5.2 Discretization error ϵ_d

The discretization error $\epsilon_d(N)$ can be controlled using the standard approximation result below (see e.g. [23, Theorem 9.1, p. 503]).

Theorem 5.1. *There exists a universal constant $c > 0$ such that for all $f \in \mathcal{E}^r(\mathbb{R}^d)$ the following estimate holds*

$$\|f - f_N\|_2^2 \leq c\|f\|_{\mathcal{E}^r(\mathbb{R}^d)}^2 N^{-2r/d}, \tag{5.3}$$

with $f_N = \sum_{k=1}^N \langle f, \phi_k\rangle\phi_k$.

Corollary 5.1. *Under the assumption $H \in \mathcal{E}^{r,s}(A_1, A_2)$, the discretization error satisfies:*

$$\epsilon_d(N) \lesssim N^{-2r/d}. \tag{5.4}$$

Proof. By assumption (2.11), $S(\cdot, y) \in \mathcal{E}^r(\mathbb{R}^d)$ for almost every $y \in \Omega$. Therefore, by Theorem 5.1:

$$\|S(\cdot, y) - S_N(\cdot, y)\|^2_{L^2(\mathbb{R}^d)} \leq cN^{-2r/d} \|S(\cdot, y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)} \tag{5.5}$$

Finally:

$$\begin{aligned}
\|S - S_N\|^2_{L^2(\mathbb{R}^d \times \Omega)} &= \int_{y \in \Omega} \|S(\cdot, y) - S_N(\cdot, y)\|^2_{L^2(\mathbb{R}^d)} \, dy \\
&\leq c \left(\int_{y \in \Omega} \|S(\cdot, y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)} dy \right) N^{-2r/d} \\
&\leq c A_2 N^{-2r/d}
\end{aligned}$$

□

5.3 Estimation error ϵ_e

This section provides an upper-bound on the estimation error

$$\epsilon_e(n) = \mathbb{E} \left\| S_N - \hat{S} \right\|^2_{L^2(\mathbb{R}^d \times \Omega)}. \tag{5.6}$$

This part is significantly harder than the rest of the paper. Let us begin with a simple remark.

Lemma 5.1. *The estimation error satisfies*

$$\epsilon_e(n) = \mathbb{E} \left\| F - \hat{F} \right\|^2_{\mathbb{R}^N \times L^2(\Omega)}. \tag{5.7}$$

Proof. Since $(\phi_k)_{1 \leq k \leq N}$ is an orthonormal basis, Parseval's theorem gives

$$\begin{aligned}
\|S_N - \hat{S}\|^2_{L^2(\mathbb{R}^d \times \Omega)} &= \int_\Omega \int_{\mathbb{R}^d} \left(S_N(x, y) - \hat{S}(x, y) \right)^2 dx dy \\
&= \int_\Omega \int_{\mathbb{R}^d} \left(\sum_{k=1}^N (F(y)[k] - \hat{F}(y)[k]) \phi_k(x) \right)^2 dx dy \\
&= \int_\Omega \sum_{k=1}^N (F(y)[k] - \hat{F}(y)[k])^2 dy \\
&= \sum_{k=1}^N \|F(\cdot)[k] - \hat{F}(\cdot)[k]\|^2_{L^2(\Omega)} =: \|F - \hat{F}\|^2_{\mathbb{R}^N \times L^2(\Omega)}.
\end{aligned} \tag{5.8}$$

□

By Lemma 4.1 the estimator defined in (3.4) can be decomposed as N independent estimators. Lemma 5.2 below provides a convergence rate for each of them. This result is strongly related to the work in [37] on smoothing splines. Unfortunately, we cannot directly apply the results in [37] to our setting since the kernel defined in (4.11) is not that of a thin-plate smoothing spline.

Lemma 5.2. *Suppose that $\Omega \subset \mathbb{R}^d$ is a bounded connected open set in $\mathbb{R}^d$ with Lipschitz continuous boundary. Let $Y = \{y_1, \dots, y_n\} \subset \Omega$ be a quasi-uniform sampling set of PSF locations. Recall that $\|f\|_{\mathcal{G}_k(\mathbb{R}^d)} = (1-\alpha)|f|_{BL^s(\mathbb{R}^d)} + \alpha w[k]\|f\|_{L^2(\mathbb{R}^d)}$, for all $k \geq 1$. Then, each function $\hat{F}(\cdot)[k]$ solution of Problem (4.1) satisfies:*

$$\mathbb{E}\|\hat{F}(\cdot)[k] - F(\cdot)[k]\|^2_{L^2(\Omega)} \lesssim \mu(1-\alpha)^{-1}\|F(\cdot)[k]\|^2_{\mathcal{G}_k(\mathbb{R}^d)} + n^{-1}\sigma^2 \left[(1-\alpha)\mu\right]^{-\frac{d}{2s}} (1-\alpha)^{-1}, \quad (5.9)$$

provided that $n\mu^{d/2s} \geq 1$.

Proof. In order to prove the upper-bound (5.9), we first decompose the expected squared error $\mathbb{E}\|\hat{F}(\cdot)[k] - F(\cdot)[k]\|^2_{L^2(\Omega)}$ into bias and variance terms:

$$\mathbb{E}\|\hat{F}(\cdot)[k] - F(\cdot)[k]\|^2_{L^2(\Omega)} \leq 2\left(\underbrace{\|\hat{F}^0(\cdot)[k] - F(\cdot)[k]\|^2_{L^2(\Omega)}}_{\text{Bias term}} + \underbrace{\mathbb{E}\|\hat{F}^0(\cdot)[k] - \hat{F}(\cdot)[k]\|^2_{L^2(\Omega)}}_{\text{Variance term}}\right), \quad (5.10)$$

where $\hat{F}^0(\cdot)[k]$ is the solution of the noise-free problem

$$\hat{F}^0(\cdot)[k] = \underset{f \in H^s(\mathbb{R}^d)}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}(F(y_i)[k] - f(y_i))^2 + \mu\left(\alpha w[k]\|f\|^2_{L^2(\mathbb{R}^d)} + (1-\alpha)|f|^2_{BL^s(\mathbb{R}^d)}\right). \quad (5.11)$$

We then treat the bias and variance terms separately.

Control of the bias The bias control relies on sampling inequalities in Sobolev spaces. They first appeared in [9] to control the norm of functions in Sobolev spaces with scattered zeros. They have been generalized in different ways, see e.g. [40] and [2]. In this paper, we will use the following result from [2].

Theorem 5.2 ([2, Theorem 4.1])**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded connected open set with Lipschitz continuous boundary, and $p, q, x \in [1, +\infty]$ be given. Let s be a real number such that $s \geq d$ if $p = 1$, $s > d/p$ if $1 < p < \infty$ or $s \in \mathbb{N}^*$ if $p = \infty$. Furthermore, let $l_0 = s - d(1/p - 1/q)_+$ and $\gamma = \max(p, q, x)$ where $(\cdot)_+ = \max(0, \cdot)$.*

Then there exist two positive constants η_s (depending on Ω and s) and C (depending on Ω, n, s, p, q and x) satisfying the following property: for any finite set $Y \subset \bar{\Omega}$ (or $Y \subset \Omega$ if $p = 1$ and $s = d$) such that $h_{Y,\Omega} \leq \eta_s$, for any $u \in W^{s,p}(\Omega)$ and for any $l = 0, \dots, \lceil l_0 \rceil - 1$, we have

$$\|u\|_{W^{l,q}(\Omega)} \leq C\left(h^{s-l-d(1/p-1/q)_+}_{Y,\Omega}|u|_{W^{s,p}(\Omega)} + h^{d/\gamma - l}_{Y,\Omega}\|u|_Y\|_x\right), \quad (5.12)$$

where $\|u|_Y\|_x = (\sum_{i=1}^n u(y_i)^x)^{1/x}$. If $s \in \mathbb{N}^*$ this bound also holds with $l = l_0$ when either $p < q < \infty$ and $l_0 \in \mathbb{N}$ or $(p, q) = (1, \infty)$ or $p \geq q$.

The above theorem is the key to obtain Proposition 5.1 below.

Proposition 5.1. *Set $0 \leq \alpha < 1$ and let $\mathcal{G}_k(\Omega)$ be the RKHS with norm defined by $\|\cdot\|^2_{\mathcal{G}_k(\Omega)} = (1-\alpha)|\cdot|^2_{BL^s(\Omega)} + \alpha w[k]\|\cdot\|^2_{L^2(\Omega)}$. Let $u \in H^s(\Omega)$ denote a target function and $Y = \{y_1, \ldots, y_n\} \subset \Omega$ a data site set. Let f_μ denote the solution of the following variational problem*

$$f_\mu = \operatorname*{arg\,min}_{f \in \mathcal{G}(\mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^{n} (u(y_j) - f(y_j))^2 + \mu \|f\|^2_{\mathcal{G}_k(\mathbb{R}^d)}. \tag{5.13}$$

Then

$$\|f_\mu - u\|_{L^2(\Omega)} \leq C \left((1-\alpha)^{-1/2} h^s_{Y,\Omega} + h^{d/2}_{Y,\Omega} \sqrt{n\mu} \right) \|u\|_{\mathcal{G}_k(\mathbb{R}^d)}, \tag{5.14}$$

where C is a constant depending only on Ω and s and $h_{Y,\Omega}$ is the fill distance defined in 3.2.

Proof. By applying the Sobolev sampling inequality of Theorem 5.2 for $p = q = x = 2$, $l = 0$, we get

$$\|v\|_{L^2(\Omega)} \leq C \left(h^s_{Y,\Omega} |v|_{H^s(\Omega)} + h^{d/2}_{Y,\Omega} \left(\sum_{i=1}^{n} v(y_i)^2 \right)^{1/2} \right), \tag{5.15}$$

for all $v \in H^s$. This inequality applied to function $v = f_\mu - u$ yields

$$\|f_\mu - u\|_{L^2(\Omega)} \leq C \left(h^s_{Y,\Omega} |f_\mu - u|_{H^s(\Omega)} + h^{d/2}_{Y,\Omega} \left(\sum_{i=1}^{n} (f_\mu(y_i) - u(y_i)^2 \right)^{1/2} \right). \tag{5.16}$$

The remaining task is to bound $|f_\mu - u|_{H^s(\Omega)}$ and $\left(\sum_{i=1}^n (f_\mu(y_i) - u(y_i))^2\right)^{1/2}$ by $\|u\|_{\mathcal{G}_k(\mathbb{R}^d)}$. To this end, let us define two functionals $f \mapsto E(f) = \frac{1}{n}\sum_{i=1}^n (u(y_j) - f(y_j))^2$ and $f \mapsto J(f) = \|f\|^2_{\mathcal{G}_k(\mathbb{R}^d)}$. We will use the function $f_0 : \Omega \to \mathbb{R}$ defined as the solution of

$$f_0 = \operatorname*{arg\,min}_{\substack{f \in \mathcal{G}_k(\mathbb{R}^d) \\ f(y_j) = u(y_j)}} \|f\|^2_{\mathcal{G}_k(\mathbb{R}^d)}, \tag{5.17}$$

We notice that the set $\{f \in \mathcal{G}_k(\mathbb{R}^d) \,|\, \forall 1 \leq j \leq n,\, f(y_j) = u(y_j)\}$ is non-empty, convex and closed. Furthermore, the squared norm $\|\cdot\|^2_{\mathcal{G}_k(\mathbb{R}^d)}$ is strictly convex. Those two facts imply that the function f_0 is uniquely determined.

Since f_μ is the minimizer of (5.13), it satisfies

$$E(f_\mu) + \mu J(f_\mu) \leq E(f_0) + \mu J(f_0). \tag{5.18}$$

In addition $E(f_0) = 0$ and $J(f_0) \leq J(u)$. Therefore we have the following sequence of inequalities:

$$E(f_\mu) + \mu J(f_\mu) \leq E(f_0) + \mu J(f_0) = \mu J(f_0) \leq \mu J(u). \tag{5.19}$$

Hence,

$$E(f_\mu) = \frac{1}{n}\sum_{i=1}^{n}(u(y_j) - f_\mu(y_j))^2 \leq \mu\|u\|^2_{\mathcal{G}_k(\mathbb{R}^d)}. \tag{5.20}$$

To finish, the triangle inequality yields $|f_\mu - u|_{H^s(\Omega)} \leq |f_\mu|_{H^s(\Omega)} + |u|_{H^s(\Omega)}$. The equivalence between the Sobolev semi-norm $|\cdot|_{H^s}$ and the Beppo-Levi semi-norm $|\cdot|_{BL^s}$ yields

$$\begin{aligned}(1-\alpha)|f_\mu|^2_{H^s(\Omega)} &\lesssim \|f_\mu\|^2_{\mathcal{G}_k(\mathbb{R}^d)} \leq \|u\|^2_{\mathcal{G}_k(\mathbb{R}^d)},\\ (1-\alpha)|u|^2_{H^s(\Omega)} &\lesssim \|u\|^2_{\mathcal{G}_k(\mathbb{R}^d)}.\end{aligned} \tag{5.21}$$

Replacing bounds (5.21) and (5.20) in the sampling inequality (5.16) completes the proof of Proposition 5.1. □

Applying Proposition 5.1 to $\hat{F}^0(\cdot)[k]$, we get

$$\|\hat{F}^0(\cdot)[k] - F(\cdot)[k]\|^2_{L^2(\Omega)} \leq C\left((1-\alpha)^{-1/2}h^s_{Y,\Omega} + \sqrt{\mu n}h^{d/2}_{Y,\Omega}\right)^2\|F(\cdot)[k]\|^2_{\mathcal{G}_k(\mathbb{R}^d)}. \tag{5.22}$$

The trick is now to use the quasi-uniformity condition to control $h^s_{Y,\Omega}$ and $\sqrt{\mu n}h^{d/2}_{Y,\Omega}$. This is achieved using the following proposition.

Proposition 5.2 ([39, Proposition 14.1] or [37])**.** *Let $Y = \{y_1, \ldots, y_n\} \subset \Omega$ be a quasi-uniform set with respect to B. Then, there exist constants $c > 0$ and $C > 0$ depending only on d, Ω and B such that,*

$$cn^{-1} \leq h^d_{Y,\Omega} \leq Cn^{-1}. \tag{5.23}$$

Condition $n\mu^{d/2s} \geq 1$ combined with the right-hand-side of (5.23) yields $h^d_{Y,\Omega} \leq C\mu^{d/2s}$, so that $h^s_{Y,\Omega} \lesssim \sqrt{\mu}$. Similarly, the right-hand-side of (5.23) yields $\sqrt{\mu n}h^{d/2}_{Y,\Omega} \lesssim \sqrt{\mu}$. Hence

$$\|\hat{F}^0(\cdot)[k] - F(\cdot)[k]\|^2_{L^2(\Omega)} \lesssim (1-\alpha)^{-1}\mu\|F(\cdot)[k]\|^2_{\mathcal{G}_k(\mathbb{R}^d)}. \tag{5.24}$$

Control of the variance The variance term is treated following arguments similar to those in [37]. However, the change of kernel needs additional treatment. First of all, note that due to the linearity of the estimators of Problem (4.1) (that can be seen from equation (4.8)), we have $\hat{F}^0_\mu(\cdot)[k] - \hat{F}_\mu(\cdot)[k] = f^\eta_k$ with $\eta \in \mathbb{R}^n$ defined as $\eta[i] = \epsilon_i[k]$ and

$$f^\eta_k = \operatorname*{arg\,min}_{f\in H^s(\mathbb{R}^d)} \frac{1}{n}\sum_{i=1}^{n}(f(y_i) - \eta[i])^2 + \mu\left(\alpha w[k]\|f\|^2_{L^2(\mathbb{R}^d)} + (1-\alpha)|f|^2_{BL^s(\mathbb{R}^d)}\right). \tag{5.25}$$

We therefore need to estimate $\mathbb{E}\|f^\eta_k\|^2_{L^2(\Omega)}$. From Theorem 5.2 applied with $p = q = x = 2$ and $l = 0$ we obtain that for $u \in H^s(\Omega)$

$$\|u\|_{L^2(\Omega)} \leq C\left(h^s_{Y,\Omega}|u|_{H^s(\Omega)} + h^{d/2}_{Y,\Omega}\|u|_Y\|_2\right).$$

Using the above inequality together with Proposition 5.2, we get that

$$\|f_k^\eta\|^2_{L^2(\Omega)} \leq 2C\left(h^{2s}_{Y,\Omega}|f_k^\eta|^2_{H^s(\Omega)} + n^{-1}\sum_{i=1}^n f_k^\eta(y_i)^2\right).$$

As in [37], let us define the $n \times n$ symmetric matrix $\tilde{\Gamma}$ such that

$$\langle \tilde{\Gamma} z, z\rangle = \min_{\substack{u\in BL^s(\mathbb{R}^d)\\ u(y_i)=z[i]}} (1-\alpha)|u|^2_{H^s(\mathbb{R}^d)} + \alpha w[k]\|u\|^2_{L^2(\mathbb{R}^d)}. \tag{5.26}$$

The solution of Problem (5.26) is a spline interpolating the data $(y_i, z[i])_{i=1}^n$. Using this matrix, we can write (5.25) as:

$$\min_{z\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n (z[i]-\eta[i])^2 + \mu\langle\tilde{\Gamma}z, z\rangle,$$

see [37, 36, 38] for details. Thus, the solution $\hat{z} = (f_k^\eta(y_i))_{i=1}^n$ is obtained by:

$$\hat{z} = (\mathrm{Id} + n\mu\tilde{\Gamma})^{-1}\eta.$$

By letting $E_\mu = (\mathrm{Id} + n\mu\tilde{\Gamma})^{-1}$, we obtain

$$n^{-1}\sum_{i=1}^n f_k^\eta(y_i)^2 = n^{-1}\sum_{i=1}^n \hat{z}[i]^2 = n^{-1}\eta^T E_\mu^2\eta$$

and

$$\begin{aligned}(1-\alpha)|f_k^\eta|^2_{H^s(\mathbb{R}^d)} + \alpha w[k]\|f_k^\eta\|^2_{L^2(\mathbb{R}^d)} &= \hat{z}^T\tilde{\Gamma}\hat{z} = \eta^T E_\mu\tilde{\Gamma}E_\mu\eta\\ &= (n\mu)^{-1}\eta^T E_\mu(E_\mu^{-1} - \mathrm{Id})E_\mu\eta\\ &= (n\mu)^{-1}\eta^T(E_\mu - E_\mu^2)\eta.\end{aligned}$$

Thus

$$|f_k^\eta|^2_{H^s(\Omega)} \leq |f_k^\eta|^2_{H^s(\mathbb{R}^d)} \leq (n\mu(1-\alpha))^{-1}\eta^T(E_\mu - E_\mu^2)\eta.$$

Using the fact that η has i.i.d. components with zero mean and variance σ^2, we get that,

$$\mathbb{E}\left[n^{-1}\sum_{i=1}^n f_\lambda^\eta(y_i)^2\right] = n^{-1}\sigma^2\mathrm{Tr}(E_\mu^2),$$

and on the other hand

$$\begin{aligned}\mathbb{E}|f_k^\eta|^2_{H^s(\Omega)} &\leq (n\mu(1-\alpha))^{-1}\sigma^2(\mathrm{Tr}(E_\mu) - \mathrm{Tr}(E_\mu^2))\\ &\leq (n\mu(1-\alpha))^{-1}\sigma^2\mathrm{Tr}(E_\mu).\end{aligned}$$

We now have to focus on the estimation of both $\mathrm{Tr}(E_\mu) = \sum_{i=1}^n (1+n\mu\lambda_i(\tilde{\Gamma}))^{-1}$ and $\mathrm{Tr}(E_\mu^2) = \sum_{i=1}^n (1+n\mu\lambda_i(\tilde{\Gamma}))^{-2}$, where $\lambda_i(\tilde{\Gamma})$ is the i-th eigenvalue of $\tilde{\Gamma}$. This will be achieved by analyzing

the eigenvalues of the matrix $\tilde{\Gamma}$. This step is quite cumbersome. Fortunately, we can rely on the work of Utreras who analyzed the eigenvalues of the matrix Γ associated to thin-plate splines in [37]. Matrix Γ is defined in a similar way as (5.26):

$$\langle \Gamma z, z \rangle = \min_{\substack{u \in BL^s(\mathbb{R}^d) \\ u(y_i)=z[i]}} |u|^2_{H^s(\mathbb{R}^d)}. \tag{5.27}$$

One therefore has that $(1-\alpha) z^T \Gamma z \leq z^T \tilde{\Gamma} z$ for all $z \in \mathbb{R}^N$. Therefore the matrix $\tilde{\Gamma} - (1-\alpha)\Gamma$ is semi-definite positive. By virtue of Weyl Monotonicity Theorem [41], we get that $(1-\alpha)\lambda_i(\Gamma) \leq \lambda_i(\tilde{\Gamma})$. Hence we can bound the traces of the matrices E_μ and E_μ^2 as follows

$$\begin{aligned}
\mathrm{Tr}(E_\mu) &\leq \sum_{i=1}^{n} (1 + (1-\alpha) n \mu \lambda_i(\Gamma))^{-1}, \\
\mathrm{Tr}(E_\mu^2) &\leq \sum_{i=1}^{n} (1 + (1-\alpha) n \mu \lambda_i(\Gamma))^{-2}.
\end{aligned}$$

It is shown in [37], that $\gamma = \binom{s-1+d}{s-1}$ eigenvalues $\lambda_i(\Gamma)$ are null and the others satisfy $i^{2s/d} n^{-1} \lesssim \lambda_i(\Gamma) \lesssim i^{2s/d} n^{-1}$ for $\gamma + 1 \leq i \leq n$. Following [37], it can be shown that both traces are bounded by quantities proportional to $[(1-\alpha)\mu]^{-d/2s}$. Thus one has that

$$\mathbb{E}\|f_k^\eta\|^2_{L^2(\Omega)} \lesssim \sigma^2 (n^{-1} [(1-\alpha)\mu]^{-d/2s} + n^{-1} h_{Y,\Omega}^{2s} \mu^{-1} [(1-\alpha)\mu]^{1-d/2s}).$$

Since $\mu^{d/2s} n \geq 1$ and using Proposition 5.2 that gives $n \lesssim h_{Y,\Omega}^{-d}$ we obtain that $h_{Y,\Omega}^{2s} \mu^{-1} \lesssim 1$. Hence

$$\mathbb{E}\|f_k^\eta\|^2_{L^2(\Omega)} \lesssim \sigma^2 n^{-1} [(1-\alpha)\mu]^{-d/2s} \left(1 + h_{Y,\Omega}^{2s} [(1-\alpha)\mu]^{-1}\right) \lesssim \sigma^2 n^{-1} [(1-\alpha)\mu]^{-d/2s} (1-\alpha)^{-1},$$

which completes the proof of Lemma 5.2. □

Finally, we will need the following technical Lemma.

Lemma 5.3. *Let H be an operator in $\mathcal{E}^{r,s}(A_1, A_2)$ with SVIR S (2.4) and IRC F (3.1). We have*

$$\partial_y^\alpha \langle S(\cdot, y), \phi_k \rangle = \langle \partial_y^\alpha S(\cdot, y), \phi_k \rangle \quad \forall |\alpha| \leq s, \text{ and for a.e. } y \in \Omega, \tag{5.28}$$

$$F(\cdot)[k] : y \mapsto \langle S(\cdot, y), \phi_k \rangle \in H^s(\mathbb{R}^d) \quad \forall k \in \mathbb{N}, \tag{5.29}$$

$$\sum_{k \in \mathbb{N}} |F(\cdot)[k]|^2_{BL^s(\mathbb{R}^d)} = \int_{x \in \mathbb{R}^d} |S(x, \cdot)|^2_{BL^s(\mathbb{R}^d)} dx, \tag{5.30}$$

$$\sum_{k \in \mathbb{N}} w[k] \, \|F(\cdot)[k]\|^2_{L^2(\mathbb{R}^d)} = \int_{\mathbb{R}^d} \|S(\cdot, y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)} \, dy. \tag{5.31}$$

Proof. The first point is derived using a result in [19, Theorem 7.40]: since $S \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$, it defines a generalized function. We obtain that $\partial_y^\alpha \langle S(\cdot, y), \phi_k \rangle = \langle \partial_y^\alpha S(\cdot, y), \phi_k \rangle$ in the sense of generalized functions. Moreover,

$$\begin{aligned}\int_{\mathbb{R}^d} \left| \langle \partial_y^\alpha S(\cdot, y), \phi_k \rangle \right|^2 dy &\leq \int_{\mathbb{R}^d} \left\| \partial_y^\alpha S(\cdot, y) \right\|^2_{L^2(\mathbb{R}^d)} dy \\ &\leq A_1,\end{aligned}$$

since $S \in \mathcal{E}^{r,s}(A_1, A_2)$. Thus the equality is also valid in $L^2(\mathbb{R}^d)$ and $\partial_y^\alpha \langle S(\cdot, y), \phi_k \rangle = \langle \partial_y^\alpha S(\cdot, y), \phi_k \rangle$ almost everywhere. The second point is shown by observing that

$$\begin{aligned}&\| y \mapsto \partial_y^\alpha \langle S(\cdot, y), \phi_k) \rangle \|^2_{L^2(\mathbb{R}^d)} \\ &\qquad = \int_{\mathbb{R}^d} |\langle \partial_y^\alpha S(\cdot, y), \phi_k) \rangle|^2 dy \\ &\qquad \leq \sum_{k \in \mathbb{N}} \int_{\mathbb{R}^d} |\langle \partial_y^\alpha S(\cdot, y), \phi_k) \rangle|^2 dy \\ &\qquad = \int_{\mathbb{R}^d} \sum_{k \in \mathbb{N}} |\langle \partial_y^\alpha S(\cdot, y), \phi_k) \rangle|^2 dy \\ &\qquad = \int_{\mathbb{R}^d} \| \partial_y^\alpha S(\cdot, y) \|^2_{L^2(\mathbb{R}^d)} dy \\ &\qquad = \int_{\mathbb{R}^d} \| \partial_y^\alpha S(x, \cdot) \|^2_{L^2(\mathbb{R}^d)} dx.\end{aligned}$$

We used the Tonelli Theorem to switch the sum with the integrals and then the two integrals. Therefore

$$\begin{aligned}\| y \mapsto \langle S(\cdot, y), \phi_k) \rangle \|^2_{H^s(\mathbb{R}^d)} &\leq \sum_{|\alpha| \leq s} \| y \mapsto \partial_y^\alpha \langle S(\cdot, y), \phi_k) \rangle \|^2_{L^2(\mathbb{R}^d)} \\ &\leq \int_{\mathbb{R}^d} \sum_{|\alpha| \leq s} \| \partial_y^\alpha S(x, \cdot) \|^2_{L^2(\mathbb{R}^d)} dx \\ &= \int_{\mathbb{R}^d} \| S(x, \cdot) \|^2_{H^s(\mathbb{R}^d)} dx \\ &\leq A_2.\end{aligned}$$

The third one is straightforward once the following is shown

$$\begin{aligned}\sum_{k \in \mathbb{N}} \| \partial_y^\alpha F(\cdot)[k] \|^2_{L^2(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} \sum_{k \in \mathbb{N}} |\partial_y^\alpha F(y)[k]|^2 dy \\ &= \int_{y \in \mathbb{R}^d} \| \partial_y^\alpha S(\cdot, y) \|^2_{L^2(\mathbb{R}^d)} dy \\ &= \int_{x \in \mathbb{R}^d} \| \partial_y^\alpha S(x, \cdot) \|^2_{L^2(\mathbb{R}^d)} dx.\end{aligned}$$

Note that we switched the sum with the integral then the two integrals using the Tonelli Theorem. The last point goes as follows:

$$\sum_{k\in\mathbb{N}} w[k]\,\|F(\cdot)[k]\|^2_{L^2(\mathbb{R}^d)} = \int_{\mathbb{R}^d} \sum_{k\in\mathbb{N}} w[k]\,\left|\langle \partial_y^\alpha S(\cdot,y),\phi_k\rangle\right|^2 dy$$
$$= \int_{\mathbb{R}^d} \|S(\cdot,y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)}\,dy.$$

Note that we switched the sum and integral using the Tonelli Theorem. □

5.4 Proof of the main results

Proof of Theorem 3.1

Proof. By equation (5.2):

$$\mathbb{E}\|\hat{H}-H\|^2_{HS} \le 2(\epsilon_d(N)+\epsilon_e(n)). \tag{5.32}$$

By Corollary (5.1)

$$\epsilon_d(N) \lesssim N^{-2r/d}. \tag{5.33}$$

Now, let us control ϵ_e.

$$\epsilon_e(n) = \mathbb{E}\|\hat{F}-F\|^2_{\mathbb{R}^N\times L^2(\Omega)} \tag{5.34}$$
$$= \sum_{k=1}^N \mathbb{E}\|\hat{F}(\cdot)[k]-F(\cdot)[k]\|^2_{L^2(\Omega)} \tag{5.35}$$
$$\overset{(5.9)}{\lesssim} \sum_{k=1}^N \left(\mu(1-\alpha)^{-1}\|F(\cdot)[k]\|^2_{\mathcal{G}_k(\mathbb{R}^d)} + n^{-1}\sigma^2\left[(1-\alpha)\mu\right]^{-d/2s}(1-\alpha)^{-1}\right) \tag{5.36}$$
$$= \mu(1-\alpha)^{-1}\sum_{k=1}^N \|F(\cdot)[k]\|^2_{\mathcal{G}_k(\mathbb{R}^d)} + Nn^{-1}\sigma^2\left[(1-\alpha)\mu\right]^{-d/2s}(1-\alpha)^{-1} \tag{5.37}$$

Further calculations give

$$\sum_{k=1}^N \|F(\cdot)[k]\|^2_{\mathcal{G}_k(\mathbb{R}^d)} = (1-\alpha)\sum_{k=1}^N |F(\cdot)[k]|^2_{BL^s(\mathbb{R}^d)} + \alpha\sum_{k=1}^N w[k]\|F(\cdot)[k]\|^2_{L^2(\mathbb{R}^d)} \tag{5.38}$$
$$\le (1-\alpha)A_1 + \alpha A_2. \tag{5.39}$$

where the last inequality is derived using Lemma (5.3). Hence,

$$\epsilon_e(n) \le \mu(1-\alpha)^{-1}(A_1+A_2) + Nn^{-1}\sigma^2\left[(1-\alpha)\mu\right]^{-d/2s}(1-\alpha)^{-1}.$$

This upper bound allows to set the value of the regularization parameter μ by balancing the two terms $(1-\alpha)^{-1}\mu$ and $Nn^{-1}\sigma^2\left[(1-\alpha)\mu\right]^{-d/2s}(1-\alpha)^{-1}$:

$$(1-\alpha)^{-1}\mu \propto Nn^{-1}\sigma^2\left[(1-\alpha)\mu\right]^{-d/2s}(1-\alpha)^{-1}. \tag{5.40}$$

This yields

$$\mu \propto \left(N\sigma^2 n^{-1}\right)^{\frac{2s}{2s+d}} (1-\alpha)^{\frac{-d}{2s+d}}. \tag{5.41}$$

Plugging this value in the upper-bound of $\epsilon_e(n)$ gives

$$\mu(1-\alpha)^{-1} \propto \left(N\sigma^2 n^{-1}\right)^{\frac{2s}{2s+d}} (1-\alpha)^{-1}(1-\alpha)^{\frac{-d}{2s+d}}. \tag{5.42}$$

Hence,

$$\epsilon_e(n) \lesssim (N\sigma^2 n^{-1})^{\frac{2s}{2s+d}}(1-\alpha)^{-\frac{2s+2d}{2s+d}}. \tag{5.43}$$

□

Proof of Corollary 3.1

Proof. To obtain this bound we use Theorem 3.1 and we balance the two terms so that:

$$N^{-2r/d} \propto (N\sigma^2 n^{-1})^{\frac{2s}{2s+d}}(1-\alpha)^{-\frac{2s+2d}{2s+d}} \tag{5.44}$$

This gives the choice $N \propto (\sigma^{-2}n)^{\frac{2sd}{4rs+2rd+2sd}}(1-\alpha)^{\frac{(2s+2d)d}{4rs+2rd+2sd}}$. Replacing N by this value in bound (3.9) gives

$$\begin{aligned} N^{-2r/d} &\propto (\sigma^2 n^{-1})^{\frac{4rs}{4rs+2rd+2sd}}(1-\alpha)^{-\frac{4rd+4rs}{4rs+2rd+2sd}}, \\ &= \left(\sigma^2 n^{-1}(1-\alpha)^{-(d/s+1)}\right)^{\frac{2q}{2q+d}}. \end{aligned} \tag{5.45}$$

□

Proof of Theorem 3.2

Proof. We first need to define an appropriate wavelet basis to characterize the fact that a function belongs to the Sobolev ball (for some constant $A > 0$)

$$H^s(\Omega, A) = \left\{u \in L^2(\Omega),\ \|u\|^2_{H^s(\Omega)} \leq A\right\},$$

through its wavelet coefficients. The scaling and wavelet functions at scale j (that is at resolution level 2^j) will be denoted by ϕ_λ and ψ_λ, respectively, where the index λ summarizes both the usual scale and space parameters j and k. In other words, for $d = 1$, we set $\lambda = (j, k)$ and denote $\phi_{j,k}(\cdot) = 2^{j/2}\phi(2^j \cdot -k)$ and $\psi_{j,k}(\cdot) = 2^{j/2}\psi(2^j \cdot -k)$. For $d \geq 2$, the notation ψ_λ stands for the adaptation of scaling and wavelet functions to $\Omega = [0,1]^d$ (see [8], Chapter 2). The notation $|\lambda| = j$ will be used to denote a wavelet at scale j, where j_0 denotes the coarse level of approximation. In order to simplify the notation, as it is commonly used, we take $j_0 = 0$, and we write $(\psi_\lambda)_{|\lambda|=-1}$ for $(\phi_\lambda)_{|\lambda|=0}$. Finally, $|\lambda| < j_1$ denotes all wavelets at scales j, with

$-1 \leq j < j_1$, and we use the notation $\tilde{\psi}_\lambda$ to denote the dual wavelet basis of ψ_λ. Now, assume that a function $u \in L^2(\Omega)$ admits the wavelet decomposition

$$u(y) = \sum_{j=-1}^{+\infty} \sum_{|\lambda|=j} c[\lambda]\psi_\lambda(y) \tag{5.46}$$

where the $c[\lambda]$'s are real coefficients satisfying $c[\lambda] = \langle u, \tilde{\psi}_\lambda \rangle_{L^2(\Omega)}$. It is well known that wavelet coefficients may be used to characterize the smoothness of functions. For instance, by Theorem 3.10.5 in [8] (on the equivalence of norms between Besov and sequence of wavelet coefficients spaces) and using the fact that the Besov space $B^s_{2,2}(\Omega)$ is equal to the Sobolev space $H^s(\Omega)$ (see e.g. Remark 3.2.4 in [8]), it follows that, under appropriate assumptions on the scaling function ϕ and its dual version (see e.g. those of Theorem 3.10.5 in [8]), there exist two constants $C_1(s,\Omega) > 0$ and $C_2(s,\Omega) > 0$ (depending only on s and Ω) such that, for any u admitting the decomposition (5.46),

$$C_1(s,\Omega) \sum_{j=-1}^{+\infty} \sum_{|\lambda|=j} 2^{2js}|c[\lambda]|^2 \leq \|u\|^2_{H^s(\Omega)} \leq C_2(s,\Omega) \sum_{j=-1}^{+\infty} \sum_{|\lambda|=j} 2^{2js}|c[\lambda]|^2. \tag{5.47}$$

Throughout the proof, it is assumed that the bi-orthogonal wavelet basis is chosen such that the wavelet characterization of Sobolev norms (5.47) is satisfied. In particular, we assume that ψ possesses $s+1$ vanishing moments.

The arguments to prove the lower bound (3.11) are based on the standard Assouad's cube technique (see, e.g. [35], Chapter 2, Section 2.7.2). Assuming that the wavelets ψ_λ are extended outside Ω using the convention that $\psi_\lambda(y) = 0$ for $y \notin \Omega$, we consider the following SVIR test functions

$$S_v(x,y) = \mu_{k_1,j_1} \sum_{k=1}^{k_1} \sum_{|\lambda|<j_1} v[k,\lambda]\phi_k(x)\psi_\lambda(y), \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d,$$

where $v = (v[k,\lambda])_{k\leq k_1, |\lambda|<j_1} \in \mathcal{V} := \{1,-1\}^{k_1 2^{j_1 d}}$, and μ_{k_1,j_1} is a positive sequence of reals satisfying the condition

$$\mu_{k_1,j_1} = c k_1^{-1/2} 2^{-j_1 d/2} \min\left(k_1^{-r/d}, 2^{-j_1 s}\right), \tag{5.48}$$

for some constant $c > 0$ not depending on k_1 and j_1. Note that, for all $x \in \mathbb{R}^d$, $S_v(x,\cdot)$ is compactly supported in Ω.

Let us first discuss the choice of the constant c in (5.48). It has to be chosen such that S_v is the SVIR of an operator in $\mathcal{E}^{r,s}(A_1,A_2)$.

First, for any $v \in \mathcal{V}$ one has that:

$$\begin{aligned}\mu_{k_1,j_1}^2 \int_{\mathbb{R}^d} \sum_{j=-1}^{j_1-1} \sum_{|\lambda|=j} 2^{2js} \left(\sum_{k=1}^{k_1} v[k,\lambda]\phi_k(x) \right)^2 dx &= \mu_{k_1,j_1}^2 \sum_{j=-1}^{j_1-1} \sum_{|\lambda|=j} 2^{2js} \int_{\mathbb{R}^d} \sum_{k,l=1}^{k_1} v[k,\lambda]v[l,\lambda]\phi_k(x)\phi_l(x)dx \\ &= \mu_{k_1,j_1}^2 \sum_{j=-1}^{j_1-1} 2^{2js} \sum_{|\lambda|=j} \sum_{k=1}^{k_1} v[k,\lambda]^2 \|\phi_k\|_{L^2(\mathbb{R}^d)}^2 \\ &= \mu_{k_1,j_1}^2 \sum_{j=-1}^{j_1-1} 2^{j(2s+d)} k_1 \leq \mu_{k_1,j_1}^2 k_1 2^{j_1(2s+d)}.\end{aligned}$$

where the above equalities use the orthonormality of the basis (ϕ_k), the definition $v[k,\lambda] = \pm 1$ and the fact that the number of wavelets at scale j is 2^{jd}. Now, using that $\int_{\mathbb{R}^d} \|S_v(x,\cdot)\|_{H^s(\mathbb{R}^d)}^2 dx = \int_{\mathbb{R}^d} \|S_v(x,\cdot)\|_{H^s(\Omega)}^2 dx$, the wavelet characterization of Sobolev norms (5.47) and by the condition (5.48) on μ_{k_1,j_1}, it follows that if $c^2 \leq C_2^{-1}(s,\Omega)A_1$ then

$$\int_{x\in\mathbb{R}^d} \|S_v(x,\cdot)\|_{H^s(\mathbb{R}^d)}^2 dx \leq A_1,$$

for any $v \in \mathcal{V}$.

We now proceed to the other inequality. A key element is again that $S_v(x,\cdot)$ is compactly supported in Ω for any $x \in \mathbb{R}^d$. For any $v \in \mathcal{V}$:

$$\begin{aligned}\int_{y\in\mathbb{R}^d} \|S_v(\cdot,y)\|_{\mathcal{E}^r(\mathbb{R}^d)}^2 dy &= \int_{y\in\Omega} \|S_v(\cdot,y)\|_{\mathcal{E}^r(\mathbb{R}^d)}^2 dy \\ &= \int_{y\in\Omega} \sum_{k\in\mathbb{N}} w[k] \left|\langle S_v(\cdot,y), \phi_k\rangle\right|^2 dy \\ &= \int_{y\in\Omega} \sum_{k=1}^{k_1} w[k]\mu_{k_1,j_1}^2 \left(\sum_{j=-1}^{j_1-1} \sum_{|\lambda|=j} v[k,\lambda]\psi_\lambda(y) \right)^2 dy \\ &\leq c_1\mu_{k_1,j_1}^2 \int_{y\in\Omega} \sum_{k=1}^{k_1} (1+k^2)^{r/d} \left(\sum_{j=-1}^{j_1-1} \sum_{|\lambda|=j} v[k,\lambda]\psi_\lambda(y) \right)^2 dy,\end{aligned} \tag{5.49}$$

using the assumption that $w[k] \leq c_1(1+k^2)^{r/d}$. Let us define the set

$$I_j(y) = \{\lambda \ : \ |\lambda| = j \text{ and } \psi_\lambda(y) \neq 0\}.$$

Due to the compact support of ψ_λ the cardinality of $I_j(y)$ is bounded by a constant $D_s > 0$ that is independent of j and y. Thus using the relation $\|\psi_\lambda\|_\infty \leq C_\infty 2^{jd/2}$ (for some constant $C_\infty > 0$) for any λ at scale j, we obtain from inequality (5.49) and the fact that $\left(\int_{y\in\Omega} dy\right) = 1$ (since

$\Omega = [0,1]^d$ in this proof),

$$\begin{aligned}\int_{y\in\mathbb{R}^d} \|S_v(\cdot,y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)} dy &\leq c_1\mu^2_{k_1,j_1}\int_{y\in\Omega}\sum_{k=1}^{k_1}(1+k^2)^{r/d}\left(\sum_{j=-1}^{j_1-1}\sum_{\lambda\in I_j(y)}|\psi_\lambda(y)|\right)^2 dy\\ &\leq c_1 D_s^2 C_\infty^2\mu^2_{k_1,j_1}\sum_{k=1}^{k_1}(1+k^2)^{r/d}\left(\sum_{j=-1}^{j_1-1}2^{jd/2}\right)^2,\\ &\leq c_1 D_s^2 C_\infty^2\mu^2_{k_1,j_1}k_1^{2r/d+1}2^{j_1 d}.\end{aligned}$$

Hence, by the condition (5.48) on μ_{k_1,j_1} it follows that if $c^2 \leq A_2 c_1^{-1} D_s^{-2} C_\infty^{-2}$, then

$$\int_{y\in\mathbb{R}^d}\|S_v(\cdot,y)\|^2_{\mathcal{E}^r(\mathbb{R}^d)}dy \leq A_2,$$

for any $v\in\mathcal{V}$. Thus we have shown that if the constant c in (5.48) is chosen sufficiently small, then the operator H_v with SVIR function S_v belongs to the ball $\mathcal{E}^{r,s}(A_1,A_2)$ for any $v\in\mathcal{V}$. In the rest of the proof, the constant c is assumed to be chosen in such a manner.

In what follows, we use the notation $\mathbb{E}_{H_v}$ to denote expectation with respect to the distribution $\mathbb{P}_{H_v}$ of the random process $F^\epsilon = (F_1^\epsilon,\ldots,F_n^\epsilon)$ obtained from model (2.6) under the hypothesis that $S = S_v$ where S_v is the SVIR function of the operator H_v.

The minimax risk

$$\mathcal{R}_{\sigma^2,n} := \inf_{\hat{H}}\sup_{H\in\mathcal{E}^{r,s}(A_1,A_2)}\mathbb{E}\left\|\hat{H}-H\right\|^2_{HS}$$

can be bounded from below as follows

$$\mathcal{R}_{\sigma^2,n} \geq \inf_{\hat{H}}\sup_{v\in\mathcal{V}}\mathbb{E}_{H_v}\left\|\hat{H}-H_v\right\|^2_{HS}.$$

Since $\left\|\hat{H}-H_v\right\|^2_{HS} = \left\|\hat{S}-S_v\right\|^2_{L^2(\mathbb{R}^d\times\Omega)}$ it follows from orthonormality of the bases $(\phi_k)_{k\geq1}$ and the Riesz stability property for bi-orthogonal wavelet bases (see e.g. inequality (7.156) in [24]), that there exists a constant $c_\psi>0$ such that

$$\left\|\hat{H}-H_v\right\|^2_{HS} \geq c_\psi\sum_{k=1}^{k_1}\sum_{|\lambda|<j_1}|\hat{\alpha}[k,\lambda]-\mu_{k_1,j_1}v[k,\lambda]|^2 \text{ where } \hat{\alpha}[k,\lambda] = \int_{\mathbb{R}^d\times\Omega}\hat{S}(x,y)\phi_k(x)\psi_\lambda(y)dxdy.$$

Therefore, the minimax risk satisfies the following inequality

$$\mathcal{R}_{\sigma^2,n} \geq \inf_{\hat{H}}\sup_{v\in\mathcal{V}} c_\psi\sum_{k=1}^{k_1}\sum_{|\lambda|<j_1}\mathbb{E}_{H_v}\left|\hat{\alpha}[k,\lambda]-\mu_{k_1,j_1}v[k,\lambda]\right|^2.$$

Then, define

$$\hat{v}[k,\lambda] := \arg\min_{v\in\{-1,1\}}\left|\hat{\alpha}[k,\lambda]-\mu_{k_1,j_1}v\right|,$$

and remark that the triangular inequality and the definition of $\hat{v}[k,\lambda]$ imply that

$$\mu_{k_1,j_1}\left|\hat{v}[k,\lambda]-v[k,\lambda]\right| \leq 2\left|\hat{\alpha}[k,\lambda]-\mu_{k_1,j_1}v[k,\lambda]\right|,$$

which yields

$$\begin{aligned}\mathcal{R}_{\sigma^2,n} &\geq \inf_{\hat{H}}\sup_{v\in\mathcal{V}}\frac{c_\psi\mu_{k_1,j_1}^2}{4}\sum_{k=1}^{k_1}\sum_{|\lambda|<j_1}\mathbb{E}_{H_v}\left|\hat{v}[k,\lambda]-v[k,\lambda]\right|^2\\ &\geq \inf_{\hat{H}}\frac{c_\psi\mu_{k_1,j_1}^2}{4}\frac{1}{\#\mathcal{V}}\sum_{v\in\mathcal{V}}\sum_{k=1}^{k_1}\sum_{|\lambda|<j_1}\mathbb{E}_{H_v}\left|\hat{v}[k,\lambda]-v[k,\lambda]\right|^2,\end{aligned}$$

where $\#\mathcal{V}$ denotes the cardinality of the finite set $\mathcal{V}$.

For a given pair $[k,\lambda]$ of indices and any $v\in\mathcal{V}$, we define the vector $v^{(k,\lambda)}\in\mathcal{V}$ having all its components equal to v except the $[k,\lambda]$-th element. Moreover, to simplify the notation, we let $\sum_{k,\lambda}$ denote the summation $\sum_{k=1}^{k_1}\sum_{|\lambda|<j_1}$. Then

$$\begin{aligned}\mathcal{R}_{\sigma^2,n} &\geq \inf_{\hat{H}}\frac{c_\psi\mu_{k_1,j_1}^2}{4}\frac{1}{\#\mathcal{V}}\sum_{k,\lambda}\sum_{v\in\mathcal{V}\,:\,v[k,\lambda]=1}\left(\mathbb{E}_{H_v}\left|\hat{v}[k,\lambda]-v[k,\lambda]\right|^2+\mathbb{E}_{H_{v^{(k,\lambda)}}}\left|\hat{v}[k,\lambda]-v^{(k,\lambda)}[k,\lambda]\right|^2\right)\\ &\geq \inf_{\hat{H}}\frac{c_\psi\mu_{k_1,j_1}^2}{4}\frac{1}{\#\mathcal{V}}\sum_{k,\lambda}\sum_{v\in\mathcal{V}\,:\,v[k,\lambda]=1}\mathbb{E}_{H_v}\left(\left|\hat{v}[k,\lambda]-v[k,\lambda]\right|^2+\left|\hat{v}[k,\lambda]-v^{(k,\lambda)}[k,\lambda]\right|^2\frac{d\mathbb{P}_{H_{v^{(k,\lambda)}}}}{d\mathbb{P}_{H_v}}(F^\epsilon)\right).\end{aligned}$$

where $\frac{d\mathbb{P}_{H_{v^{(k,\lambda)}}}}{d\mathbb{P}_{H_v}}(F^\epsilon)$ is the log-likelihood ratio between the hypothesis $H_{v^{(k,\lambda)}}: S=S_{v^{(k,\lambda)}}$ and the hypothesis $H_v: S=S_v$ in model (2.6).

Since $v^{(k,\lambda)}[k,\lambda]=-v[k,\lambda]$ and $\hat{v}[k,\lambda]\in\{-1,1\}$, one has that, for any $0<\delta<1$,

$$\begin{aligned}\mathcal{R}_{\sigma^2,n} &\geq 4c_\psi\mu_{k_1,j_1}^2\frac{1}{\#\mathcal{V}}\sum_{k,\lambda}\sum_{v\in\mathcal{V}\,:\,v[k,\lambda]=1}\mathbb{E}_{H_v}\left(\min\left(1,\frac{d\mathbb{P}_{H_{v^{(k,\lambda)}}}}{d\mathbb{P}_{H_v}}(F^\epsilon)\right)\right)\\ &\geq 4\delta c_\psi\mu_{k_1,j_1}^2\frac{1}{\#\mathcal{V}}\sum_{k,\lambda}\sum_{v\in\mathcal{V}\,:\,v[k,\lambda]=1}\mathbb{P}_{H_v}\left(\frac{d\mathbb{P}_{H_{v^{(k,\lambda)}}}}{d\mathbb{P}_{H_v}}(F^\epsilon)>\delta\right),\end{aligned}\tag{5.50}$$

by Markov's inequality. Thanks to the Girsanov's formula (see e.g. Lemma A.5 in [35]), one has that, under the hypothesis that $S=S_v$ in model (2.6):

$$\log\left(\frac{d\mathbb{P}_{H_{v^{(k,\lambda)}}}}{d\mathbb{P}_{H_v}}(F^\epsilon)\right)=\sum_{i=1}^{n}\sum_{\ell=1}^{+\infty}\left(\sigma^{-1}\langle S_{v^{(k,\lambda)}}(\cdot,y_i)-S_v(\cdot,y_i),\phi_\ell\rangle\eta_{i,\ell}-\frac{\sigma^{-2}}{2}\left|\langle S_{v^{(k,\lambda)}}(\cdot,y_i)-S_v(\cdot,y_i),\phi_l\rangle\right|^2\right)$$

where the $\eta_{i,\ell}$'s are iid standard Gaussian variables. By definition of $v^{(k,\lambda)}$ and for $v[k,\lambda]=1$ one has that, for each $1\leq i\leq n$ and $1\leq k\leq k_1$,

$$\langle S_{v^{(k,\lambda)}}(\cdot,y_i)-S_v(\cdot,y_i),\phi_\ell\rangle=\begin{cases}-2\mu_{k_1,j_1}\psi_\lambda(y_i) & \text{if } \ell=k,\\ 0 & \text{otherwise.}\end{cases}$$

Therefore, the random variable $Z_{k,\lambda} := \log\left(\frac{d\mathbb{P}_{H_v(k,\lambda)}}{d\mathbb{P}_{H_v}}(F^\epsilon)\right)$ is Gaussian with mean θ_λ and variance γ_λ^2 satisfying

$$\theta_\lambda = -2\sigma^{-2}\mu_{k_1,j_1}^2 \sum_{i=1}^n \psi_\lambda^2(y_i) \text{ and } \gamma_\lambda^2 = 4\sigma^{-2}\mu_{k_1,j_1}^2 \sum_{i=1}^n \psi_\lambda^2(y_i) = -2\theta_\lambda,$$

under the hypothesis that $S = S_v$ in model (2.6). The negativity of θ_λ implies that

$$\mathbb{P}_{H_v}\left(Z_{k,\lambda} \geq 3\theta_\lambda\right) = \mathbb{P}_{H_v}\left(\frac{Z_{k,\lambda} - \theta_\lambda}{\sqrt{2|\theta_\lambda|}} \geq -\sqrt{2|\theta_\lambda|}\right) \geq \frac{1}{2},$$

by symmetry of the standard Gaussian distribution. Hence, inserting the above inequality into (5.50) with $\delta = \exp(3\theta_\lambda)$, it implies that

$$\mathcal{R}_{\sigma^2,n} \geq c_\psi \exp(3\theta_\lambda)\mu_{k_1,j_1}^2 \; k_1 C_\psi 2^{dj_1} \tag{5.51}$$

$$= c_\psi \exp(3\theta_\lambda) c^2 C_\psi \min\left(k_1^{-2r/d}, 2^{-2j_1 s}\right). \tag{5.52}$$

By setting $k_1 = k_1^{(\sigma^2,n)}$ and $j_1 = j_1^{(\sigma^2,n)}$ with

$$k_1^{(\sigma^2,n)} = \lfloor \left(\sigma^2 n^{-1}\right)^{-\frac{q}{(2q+d)r/d}} \rfloor \quad \text{and} \quad 2^{j_1^{(\sigma^2,n)}} = \lfloor \left(\sigma^2 n^{-1}\right)^{-\frac{q}{(2q+d)s}} \rfloor, \tag{5.53}$$

we get

$$\mathcal{R}_{\sigma^2,n} \geq c^2 c_\psi C_\psi \exp(3\theta_\lambda) \left(\sigma^2 n^{-1}\right)^{\frac{2q}{2q+d}}. \tag{5.54}$$

It now remains to show that the constant θ_λ is bounded from below, independently of σ and n.

The idea is to observed that $\frac{1}{n}\sum_{i=1}^n \psi_\lambda^2(y_i)$ behaves like a Riemann integral of ψ_λ and should therefore be bounded by a constant since $\|\psi_\lambda\|_2 = 1$. This statement can be proved using the following reasoning. Since vector $Y = (y_1, \ldots, y_n)$ of PSFs locations satisfies the quasi-uniformity condition $h_{Y,\Omega} \leq B q_{Y,\Omega}$, we get from Proposition 5.2 that the separation distance $q_{Y,\Omega}$ satisfies $q_{Y,\Omega}^d \geq B_1 n^{-1}$ for some constant B_1. Now, the support of wavelet ψ_λ is contained in a hypercube of volume proportional to $2^{-d|\lambda|}$. Hence, the number of locations y_i in $\mathrm{supp}(\psi_\lambda)$ is bounded above by $2^{-d|\lambda|}/q_{Y,\Omega}^d \leq B_2 n 2^{-d|\lambda|}$ for some constant B_2. To conclude, notice that $\|\psi_\lambda\|_\infty = 2^{d|\lambda|/2}\|\psi\|_\infty$, hence:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \psi_\lambda(y_i)^2 &= \frac{1}{n} \sum_{y_i \in \mathrm{supp}(\psi_\lambda)} \psi_\lambda(y_i)^2 \\
&\leq \frac{1}{n} B_2 n 2^{-d|\lambda|} \|\psi_\lambda\|_\infty^2 \\
&\leq B_2 \|\psi\|_\infty^2 =: B_3.
\end{aligned}$$

This implies that

$$\theta_\lambda \geq -2B_3 c_1, \text{ for all } \lambda < j_1^{(\sigma^2,n)}.$$

Hence, inserting the above inequality into (5.54), we finally obtain that there exists a constant $c_0 > 0$, that does not depend on $\frac{\sigma^2}{n}$, such that

$$\mathcal{R}_{\sigma^2,n} \geq c_0 \left(\sigma^2 n^{-1}\right)^{\frac{2q}{2q+d}},$$

completing the proof of the theorem. □

References

[1] R. A. Adams and J. J. Fournier. *Sobolev Spaces*, volume 140. Academic press, 2003.

[2] R. Arcangéli, M. C. L. de Silanes, and J. J. Torrens. An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing. *Numer. Math.*, 107(2):181–211, 2007.

[3] R. Bélanger-Rioux and L. Demanet. Compressed absorbing boundary conditions via matrix probing. *SIAM J. Numer. Anal.*, 53(5):2441–2471, 2015.

[4] J. Bergé, S. Price, A. Amara, and J. Rhodes. On point spread function modelling: towards optimal interpolation. *Monthly Notices of the Royal Astronomical Society*, 419(3):2356–2368, 2012.

[5] M. D. Buhmann. Radial basis functions: theory and implementations. *Cambridge Monogr. Appl. and Comput. Math.*, 12:147–165, 2003.

[6] C. Chang, P. Marshall, J. Jernigan, J. Peterson, S. Kahn, S. Gull, Y. AlSayyad, Z. Ahmad, J. Bankert, D. Bard, et al. Atmospheric point spread function interpolation for weak lensing in short exposure imaging data. *Monthly Notices of the Royal Astronomical Society*, 427(3):2572–2587, 2012.

[7] J. Chiu and L. Demanet. Matrix probing and its conditioning. *SIAM J. Numer. Anal.*, 50(1):171–193, 2012.

[8] A. Cohen. *Numerical Analysis of Wavelet Methods*, volume 32. Elsevier, 2003.

[9] J. Duchon. Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines. *RAIRO-Analyse numérique*, 12(4):325–334, 1978.

[10] N. Dyn, M. S. Floater, and A. Iske. Adaptive thinning for bivariate scattered data. *J. of Comput. and Appl. Math.*, 145(2):505–517, 2002.

[11] N. Dyn, A. Iske, and H. Wendland. Meshfree thinning of 3d point clouds. *Found. of Comput. Math.*, 8(4):409–425, 2008.

[12] P. Escande and P. Weiss. Approximation of integral operators using product-convolution expansions. *J. of Math. Imaging and Vision*, 58(3):333–348, 2017.

[13] H. G. Feichtinger, K. Gröchenig, and T. Strohmer. Efficient numerical methods in non-uniform sampling theory. *Numer. Math.*, 69(4):423–440, 1995.

[14] H. G. Feichtinger and T. Werther. *Robustness of regular sampling in Sobolev algebras*, pages 83–113. Birkhäuser Boston, 2004.

[15] M. Gentile, F. Courbin, and G. Meylan. Interpolating point spread function anisotropy. *Astronomy & Astrophysics*, 549:A1, 2013.

[16] K. Gröchenig and E. Pauwels. Uniqueness and reconstruction theorems for pseudodifferential operators with a bandlimited kohn-nirenberg symbol. *Adv. in Comput. Math.*, 40(1):49–63, 2014.

[17] A. Iske. *Multiresolution Methods in Scattered Data Modelling*, volume 37. Springer, 2004.

[18] M. Jee, J. Blakeslee, M. Sirianni, A. Martel, R. White, and H. Ford. Principal component analysis of the time-and position-dependent point-spread function of the advanced camera for surveys. *Publications of the Astronomical Society of the Pacific*, 119(862):1403, 2007.

[19] D. S. Jones. *The Theory of Generalised Functions.* Cambridge Univ. Press, 1982.

[20] T. Kailath. Sampling models for linear time-variant filters. Master's thesis, MIT, Dept. of Electrical Engineering, 1959.

[21] W. Kozek and G. E. Pfander. Identification of operators with bandlimited symbols. *SIAM J. Math. Anal.*, 37(3):867–888, 2005.

[22] J.-L. Lions and E. Magens. *Non-Homogeneous Boundary Value Problems and Applications, Vol II.* Springer, 1972.

[23] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1999.

[24] S. Mallat. *A Wavelet Tour of Signal Processing – The Sparse Way.* Third Edition. Academic Press, 2008.

[25] F. N. Mboula, J.-L. Starck, S. Ronayette, K. Okumura, and J. Amiaux. Super-resolution method using sparse regularization for point-spread function recovery. *Astronomy & Astrophysics*, 575:A86, 2015.

[26] C. A. Micchelli and M. Pontil. Kernels for multi-task learning. In *Proceedings of NIPS 2004*, 2004.

[27] C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, 2005.

[28] F. J. Narcowich and J. D. Ward. Norms of inverses and condition numbers for matrices associated with scattered data. *Journal of Approximation Theory*, 64(1):69–94, 1991.

[29] B. Opic and J. Rákosník. Estimates for mixed derivatives of functions from anisotropic Sobolev-Slobodeckij spaces with weights. *Quart. J. of Math. Oxford. Ser.*, 42(1):347–363, 1991.

[30] G. E. Pfander. Sampling of operators. *J. Fourier Anal. Appl.*, 19(3):612–650, 2013.

[31] G. E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.

[32] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3(3):251–264, 1995.

[33] J. Starck, E. Pantin, and F. Murtagh. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051–1069, 2002.

[34] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*, volume 30. Princeton Univ. Press, 2016.

[35] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[36] F. Utreras. Cross-validation techniques for smoothing spline functions in one or two dimensions. In *Smoothing Techniques for Curve Estimation*, pages 196–232. Springer, 1979.

[37] F. I. Utreras. Convergence rates for multivariate smoothing spline functions. *J. Approx. Theory*, 52(1):1–27, 1988.

[38] G. Wahba. *Convergence Rates of "Thin Plate" Smoothing Splines when the Data are Noisy*, volume 757. Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, 1979.

[39] H. Wendland. *Scattered Data Approximation*, volume 17. Cambridge Univ. Press, 2004.

[40] H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101(4):729–748, 2005.

[41] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.*, 71(4):441–479, 1912.

[42] A. Zeiser. Wavelet approximation in weighted sobolev spaces of mixed order with applications to the electronic Schrödinger equation. *Constr. Approx.*, 35(3):293–322, 2012.