SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 13, No. 3, pp. 957–979

Regularization for the Approximation of Functions by Mollified Discretization Methods*

Marc Hoffmann† and Camille Pouchol‡

Abstract. Some prominent discretization methods such as finite elements provide a way to approximate a function of d variables from n values it takes on the nodes x_i of the corresponding mesh. The accuracy is $n^{-s_a/d}$ in L^2-norm, where s_a is the order of the underlying method. When the data are measured or computed with systematical experimental noise, some statistical regularization might be desirable, with a smoothing method of order s_r (like the number of vanishing moments of a kernel). This idea is behind the use of some regularized discretization methods, whose approximation properties are the subject of this paper. We decipher the interplay of s_a and s_r for reconstructing a smooth function on regular bounded domains from n measurements with noise of order σ. We establish that for certain regimes with small noise σ depending on n, when $s_a > s_r$, statistical smoothing is not necessarily the best option and *no regularization* is more beneficial than *statistical regularization.* We precisely quantify this phenomenon and show that the gain can achieve a multiplicative order $n^{(s_a-s_r)/(2s_r+d)}$. We illustrate our estimates by numerical experiments conducted in dimension $d=1$ with $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements.

Key words. mollified basis, discretization, nonparametric smoothing, finite elements

MSC codes. 62-08, 62C99, 62G05

DOI. 10.1137/24M1678210

1. Introduction.

1.1. Motivation.

Let Ω be a smooth bounded connected open subset of $\mathbb{R}^d$ for some $d \geqslant 1$. We are interested in reconstructing a smooth function

$$f : \overline{\Omega} \to \mathbb{R}$$

from its values on a fixed design given by n points $x_i \in \overline{\Omega}$. These values are moreover corrupted by noise. The points x_i should be thought of as forming a mesh of the set $\overline{\Omega}$.

We focus on reconstruction methods that rely on *regularized basis functions* or *mollified basis functions,* as introduced in [BS77, Tho77]. Specifically, we are concerned with the case where functions are naturally (according to some given discretization procedure) represented as a linear combination of basis functions ϕ_i: in other words, the function f is approximated by

$$\sum_{i=1}^{n} f(x_i)\, \phi_i. \tag{1.1}$$

*Received by the editors July 18, 2024; accepted for publication (in revised form) April 16, 2025; published electronically July 8, 2025.
https://doi.org/10.1137/24M1678210

†Université Paris Dauphine-PSL and Institut Universitaire de France, Paris, France (hoffmann@ceremade.dauphine.fr).
‡MAP5, UMR 8145, Université Paris Cité, Paris, France (camille.pouchol@u-paris.fr).

Typical examples include discretization of PDEs, where the ϕ_i are, e.g., basis functions associated to $\mathbb{P}_k$ finite elements [Tho07, QQ09]. Informally, given a partition of $\overline{\Omega}$, the ϕ_i form a basis of the space of continuous functions on $\overline{\Omega}$ whose restriction to each piece of the partition is a polynomial of degree k.

We will use the shorthand notation $u \lesssim v$ (or $v \gtrsim u$) whenever there exists a constant $C > 0$ independent of n, σ, and β (see below for a precise definition of the bandwidth parameter β and the noise level σ) such that $u \leqslant Cv$ for all $n > 0$, $\sigma > 0$ and $\beta > 0$. We will write $A \sim B$ whenever $A \lesssim B$ and $B \lesssim A$ hold simultaneously. We also find it convenient to introduce a *discretization parameter* $h > 0$ satisfying

$$h \sim n^{-1/d}.$$

Of course, one could simply set $h = n^{-1/d}$, but in applications such as finite elements, there is a natural parameter h which matches $n^{-1/d}$ up to multiplicative constants only.

In the setting of (1.1), one typically has an estimate of the form

$$\sup_{f \in \mathcal{F}} \left\| f - \sum_{i=1}^{n} f(x_i)\, \phi_i \right\| \sim h^{s_a} \sim n^{-s_a/d},$$

where $\|\cdot\|$ stands for the $L^2(\Omega)$-norm, $s_a > 0$ for the order of the approximation method, and $\mathcal{F}$ for a class of sufficiently smooth functions. In practice, because of measurement, numerical, or roundoff errors, the sum $\sum_{i=1}^{n} f(x_i)\, \phi_i$ is rather given by

$$\sum_{i=1}^{n} (y_\sigma)_i\, \phi_i,$$

with

$$(y_\sigma)_i = f(x_i) + \sigma \xi_i, \;\; i = 1, \ldots, n,$$

where our noise model is given by the ξ_i, assumed to be independent random variables, centered with unit variance, so that the parameter $\sigma \geqslant 0$ quantifies the noise level as the common standard deviation to each measurement error.

A common and standard approach in alleviating the corresponding error is to operate some linear regularization on the data given in the form $\sum_{i=1}^{n} (y_\sigma)_i\, \phi_i$, like, e.g., convolution or projection onto low dimensional vector spaces. By *regularization,* we mean that we are given a family of linear operators $(R_\beta)_{\beta \geqslant 0}$ indexed by a smoothing parameter $\beta \geqslant 0$ such that $R_0 = \mathrm{Id}$ and *regularization order* s_r. These typically satisfy estimates of the form

$$\sup_{f \in \mathcal{F}} \|R_\beta f - f\| \sim \beta^{s_r},$$

where, as before, $\mathcal{F}$ is a class of sufficiently smooth functions. This leads to estimators of the form

$$R_\beta \left(\sum_{i=1}^{n} (y_\sigma)_i\, \phi_i \right) = \sum_{i=1}^{n} (y_\sigma)_i\, R_\beta \phi_i,$$

and these natural candidate estimators for approximating f are therefore based on the finite-dimensional subspace generated by the n mollified basis functions $R_\beta \phi_i$.

Of course, there are many other, potentially better, estimators for reconstructing f from the data $(y_\sigma)_i$ without necessarily relying on regularized basis functions. There is immense literature on the subject in the field of nonparametric statistics; see, e.g., the textbooks [GKK+02, Tsy08]. That our estimators are linear in particular means that one cannot hope for better approximation properties than those imposed by the Kolmogorov-n-width of the class $\mathcal{F}$ [DL93, LvGM96].

Our main reason for sticking to this rigid reconstruction framework is that mollifying basis functions is actually quite common practice: such an approach dates back to the works [BS77, Tho77] for parabolic equations. Indeed, these can lead to improved convergence estimates, and more pragmatically, they tend to stabilize the output. Extensions of this framework to hyperbolic equations also exist [ML78, CLSS03], and these methods are still of current interest for applications [FOC21]. The so-called Reproducing Kernel Element Method introduced in the series of papers [LHL+04, LLH+04, LLSJ+04, SJLLL04] also relies on similar ideas; see Chapter 6 of [LL07].

However, to the best of our knowledge, the analysis of such methods does not include statistical errors such as the $\sigma\xi_i$ that are quantified by the standard deviation parameter σ, to be compared with n or h. A natural question is therefore to understand how the presence of noise (i.e., $\sigma > 0$ in our model) impacts the previous analysis. In particular, can we optimally quantify the interplay between σ and n (or equivalently between σ and the mesh size h)? In other words, how can we best mollify basis functions in the presence of noise, if mollifying is needed at all? This is the topic of the paper.

1.2. Main results. Given the setting and methodology described above, the overarching goal is to choose a regularization parameter β appropriately as a function of the other parameters (i.e., s_a, s_r, σ, n, d), so that the reconstruction error when regularizing at the order β converges to 0 as fast as possible as the number of observed data n grows to infinity. Here the reconstruction error is defined by

$$e(\beta,\sigma,n) := \sup_{f\in\mathcal{F}_s} \mathbb{E}\left[\left\| f - R_\beta\left(\sum_{i=1}^n (y_\sigma)_i\,\phi_i\right)\right\|^2\right]^{1/2}, \tag{1.2}$$

where $\mathbb{E}[\cdot]$ denotes mathematical expectation w.r.t. the error distributions $(\xi_i)_{1\leqslant i\leqslant n}$ and $\mathcal{F}_s$ is a smoothness class of order $s > 0$ in L^2 (a Sobolev ball; see the precise definition (2.5)). It is common statistical knowledge (see, e.g., [Tsy08], [GN90]) that a good choice for β as a function of other parameters is given by

$$\beta^\star(\sigma,n) \sim \sigma^{2/(2s_r+d)} n^{-1/(2s_r+d)} \tag{1.3}$$

as soon as $s \geqslant s_r$. The purpose of this work is to compare the common regularizing strategy given by (1.3) to the possible strategy of not regularizing at all (i.e., when $\beta = 0$ with the convention $R_0 = \mathrm{Id}$).

Remark 1. It would be natural to compare any of these two strategies to the *best possible* regularizing strategy, obtained by taking the infimum with respect to $\beta > 0$. In the modeling context of this article, we did not tackle this more general question; at the moment, our lower estimates for $e(\beta,\sigma,h)$ are not sufficient to answer this question in full generality.

We focus on a sufficiently simple and tractable setting as follows:

- We consider functions with sufficiently many derivatives vanishing on the boundary of Ω, thus avoiding inessential boundary issues.
- We quantify smoothness with a number of derivatives in L^2, hence considering Sobolev balls in $H_0^s(\Omega)$.
- We quantify estimation and reconstruction in integrated L^2-error loss.
- We restrict regularization to the case of convolution with a kernel possessing vanishing moment properties.

The modeling framework developed in the present work could also serve as a stepping stone to analyze similar issues in the context of ill-posed inverse problems, i.e., when one has access to (noisy approximations of) $Af(x_i)$ with A a given compact operator from some Hilbert space to $L^2(\Omega)$. When A is associated to an underlying PDE, discretization is naturally involved, while regularization becomes necessary to cope not only with measurement errors but also with the ill-posed nature of the problem [EHN96, K+11].

A general estimate. We gather our two main results by means of informal statements; the precise hypotheses are to be found in section 2. Our first result gives precise estimate of the error as a function of all the parameters.

Theorem 1.1. *The error defined in* (1.2) *satisfies*

$$e(\beta,\sigma,n) \lesssim \sigma \min(\beta^{-1}n^{-1/d},1)^{d/2} + n^{-s_a/d} + \beta^{s_r}.$$

In particular

$$\inf_{\beta>0} e(\beta,\sigma,n) \lesssim \begin{cases} \sigma + n^{-s_a/d} & \textit{if } \sigma \lesssim n^{-s_r/d}, \\ \sigma^{2s_r/(2s_r+d)} n^{-s_r/(2s_r+d)} + n^{-s_a/d} & \textit{otherwise.} \end{cases}$$

These two estimates are given in Proposition 3.2. The first estimate, valid in the regime $\sigma \lesssim n^{-s_r/d}$, is obtained in the limit $\beta \to 0$. This is consistent with what can be achieved by *not regularizing;* see Proposition 3.1. The second estimate, valid in the regime $\sigma \gtrsim n^{-s_r/d}$, is obtained by choosing β according to (1.3).

The effect of not regularizing versus regularizing via (1.3). In order to compare the effect of not regularizing versus regularizing via (1.3), we need lower bounds. We explicitly compare σ and n by writing

$$\sigma = \sigma(n) \sim n^{-\lambda/d} \sim h^{\lambda}.$$

The parameter $\lambda \geqslant 0$ quantifies the noise level, with $\lambda = 0$ corresponding to the largest possible noise level, i.e., when σ is of order 1. In this setting, the two errors we are interested in are given by

$$e_{\mathrm{reg}}(n) := e(\beta^\star(\sigma(n),n),\sigma(n),n)$$

and

$$e_{\text{noreg}}(n) := e(0, \sigma(n), n),$$

corresponding to *regularizing* via (1.3), or *not regularizing* at all, i.e., ignoring the possible effect of the noise, deemed sufficiently negligible. Theorem 1.1 establishes the existence of two regimes, depending on the relative positions of s_a and s_r.

Before discussing the two scenarios $s_a \leqslant s_r$ and $s_a > s_r$, let us introduce the following nonstandard threshold, which happens to rule the interplay between the different parameters:

$$\lambda_M := s_a + \frac{d}{2}\left(\frac{s_a}{s_r} - 1\right). \tag{1.4}$$

In the case where $s_a \leqslant s_r$, it is always at least as good to regularize by means of the rule (1.3); see Proposition 4.4. This is a rather intuitive result, since regularization in this case is of higher order and hence cannot jeopardize the approximation property associated to discretization. Figure 1.1 summarizes the corresponding estimates. It depicts the order of convergence to 0 of $e_{\text{reg}}(n)$ and $e_{\text{noreg}}(n)$, respectively, as a function of $\lambda \geqslant 0$. When only an upper bound for the error is available, which corresponds to a lower bound for the order of convergence, we use dashed lines instead of solid lines.

The interesting situation is when $s_a > s_r$, in which case we uncover regimes when the option *not to regularize* is actually better! More precisely, we obtain the following regimes depending on λ_M, as follows.

Theorem 1.2. *Assume that $s_a > s_r$. We have*

$$\begin{cases} e_{\text{reg}}(n) \sim n^{-\frac{2\lambda+d}{2s_r+d}\frac{s_r}{d}} \text{ and } e_{\text{noreg}}(n) \sim n^{-\frac{\lambda}{d}} & \text{if } \lambda \leqslant s_a, \\ e_{\text{reg}}(n) \sim n^{-\frac{2\lambda+d}{2s_r+d}\frac{s_r}{d}} \text{ and } e_{\text{noreg}}(n) \sim n^{-\frac{s_a}{d}} & \text{if } s_a < \lambda < \lambda_M, \\ e_{\text{reg}}(n) \lesssim n^{-\frac{s_a}{d}} \text{ and } e_{\text{noreg}}(n) \sim n^{-\frac{s_a}{d}} & \text{if } \lambda \geqslant \lambda_M. \end{cases}$$

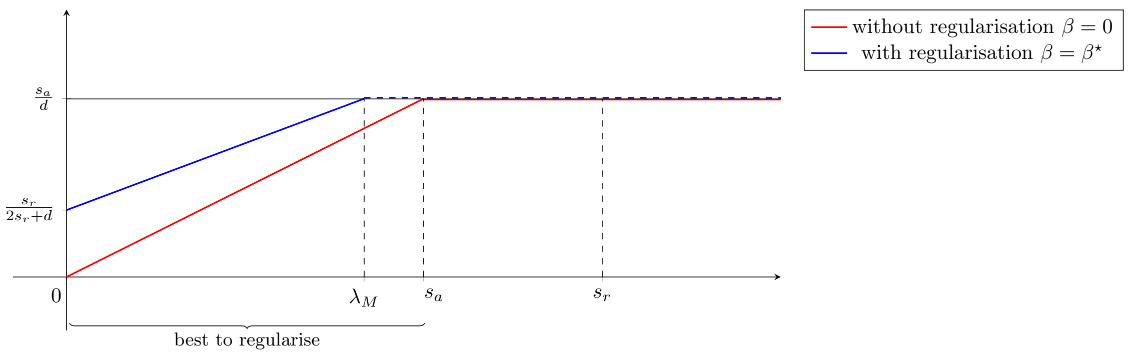

Figure 1.1. Case $s_a \leqslant s_r$. *For $\lambda \geqslant 0$, plot of the order of convergence of $e_{\text{noreg}}(n)$ and $e_{\text{reg}}(n)$ towards 0, as given by Theorem 1.2. Solid lines are used for the exact order of convergence; dashed lines are for mere lower bounds. In red, the function is $\lambda \mapsto \frac{1}{d}\min(\lambda, s_a)$, and in blue $\lambda \mapsto \frac{2\lambda+d}{2s_r+d}\frac{s_r}{d}$. Parameters for this figure are chosen to be $d = 2$, $s_a = 2$, $s_r = 3$, for which $\lambda_M = \frac{5}{3}$.*

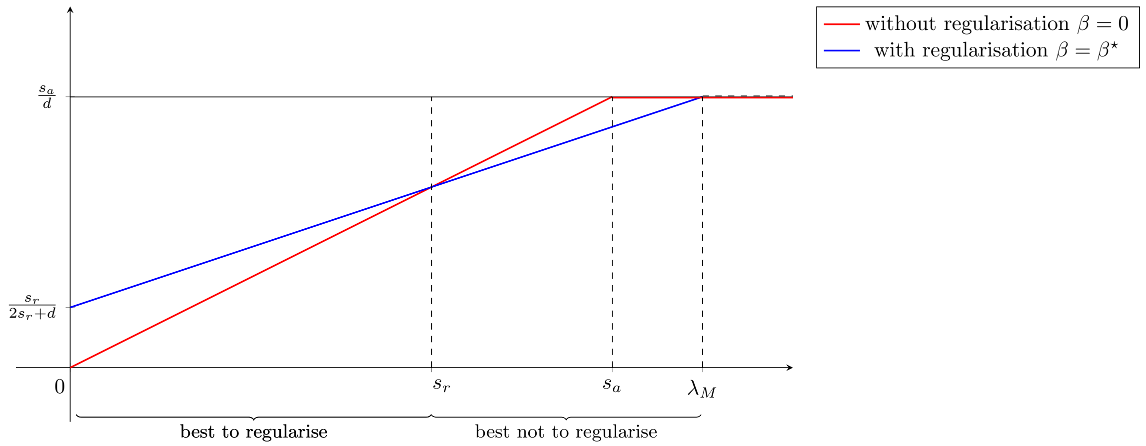

Figure 1.2. Case $s_a > s_r$. *For* $\lambda \geqslant 0$, *we display a plot of the order of convergence of* $e_{\text{noreg}}(n)$ *and* $e_{\text{reg}}(n)$ *towards* 0, *as given by Theorem* 1.2. *In red, the function is* $\lambda \mapsto \frac{1}{d}\min(\lambda, s_a)$, *and in blue* $\lambda \mapsto \frac{2\lambda+d}{2s_r+d}\frac{s_r}{d}$. *Parameters for this figure are chosen to be* $d=2$, $s_a=3$, $s_r=2$, *for which* $\lambda_M=\frac{7}{2}$.

Hence, for $0 \leqslant \lambda < \lambda_M$, Theorem 1.2 yields

$$e_{\text{reg}}(n) \sim n^{-\frac{1}{d}\min(\lambda, s_a)}, \qquad e_{\text{noreg}}(n) \sim n^{-\frac{2\lambda+d}{2s_r+d}\frac{s_r}{d}}.$$

The proofs are given in Proposition 4.5. Figure 1.2 gives a schematic description of the situation, with the same rules as in Figure 1.1.

Several remarks are in order when it comes to the case where $s_a > s_r$:

(1) Theorem 1.2 suggests the following alternative when having to choose between *not regularizing* versus *regularizing* through (1.3): *regularize* through (1.3) whenever $\lambda < s_r$, but *do not regularize* whenever $s_r < \lambda < \lambda_M$.

(2) It is easily seen that the highest gain in not regularizing occurs for $\lambda = s_a$, the value for which

$$e_{\text{reg}}(n) \sim n^{-\frac{2s_a+d}{2s_r+d}\frac{s_r}{d}}, \qquad e_{\text{noreg}}(n) \sim n^{-\frac{s_a}{d}}.$$

One can hence gain up to the order $\frac{s_a - s_r}{2s_r+d}$.

(3) We also have dependence of our estimates with respect to the dimension d. In the limit $d \to \infty$, the regime where *regularizing* through (1.3) is optimal reduces to the single value $\lambda \in \{0\}$, whereas the regime where *not regularizing* is better becomes $\lambda \in (0, \frac{1}{2}(\frac{s_a}{s_r} - 1))$. Even though the gain in *not regularizing* through (1.3) vanishes in the limit $d \to \infty$, as the maximal gain $\frac{s_a - s_r}{2s_r+d}$ converges to 0, the relative gain, in terms of order of convergence, tends to $\frac{s_a}{s_r} > 1$ as $d \to \infty$.

(4) Theorem 1.2 is, for instance, relevant to the work [FOC21], where finite element methods of order up to $s_a = 4$ are regularized with nonnegative kernels, whose order cannot exceed (and actually equals) $s_r = 2$.

Organization of the paper. In section 2, we lay out the mathematical framework and provide all the hypotheses required for our main results Theorem 1.1 and 1.2 to hold. Section 3 gathers upper bounds for the errors either with $\beta = 0$ or with fixed $\beta > 0$, which leads to Theorem 1.1. We then compare the two main strategies, thanks to lower bounds at fixed β; these results are developed in section 4 and yield Theorem 1.2. Finally, section 5 is devoted to numerical experiments confirming our theoretical results, by means of examples in dimension $d = 1$.

2. Mathematical framework.

We work in an arbitrary fixed dimension $d \in \mathbb{N}^*$, with Ω a smooth bounded connected open subset of $\mathbb{R}^d$. We let $H^s(\Omega)$ denote the fractional Sobolev space of order $s \geqslant 0$, endowed with its natural norm $\|\cdot\|_s$ that corresponds (for $s \in \mathbb{N}$) to functions having s distributional derivatives in $L^2(\Omega)$. The $L^2(\Omega)$-norm is written $\|\cdot\|$ (rather than $\|\cdot\|_0$), with corresponding inner product $\langle\cdot,\cdot\rangle$. For basic definitions and results on fractional Sobolev spaces, we refer the reader to the classical review paper [DNPV12].

We let $H_0^s(\Omega)$ denote the closure of the space $C_c^\infty(\Omega)$ of infinitely differentiable compactly supported functions for the $\|\cdot\|_s$-norm. We recall the following characterization of $H_0^s(\Omega)$ for any $s > \frac{1}{2}$ such that $s - \frac{1}{2} \notin \mathbb{N}$: these are exactly the functions $f \in H^s(\Omega)$ such that the normal derivatives $\frac{\partial^j f}{\partial \nu^j} = 0$ vanish on $\partial\Omega$ for all $0 \leqslant j < s - \frac{1}{2}$; see Theorem 11.5 of [LM12].

2.1. Statistical model and sampling.

We wish to reconstruct (equivalently estimate nonparametrically) a function $f \in H_0^s(\Omega)$ for $s > d/2$, from n noisy measurements on a fixed design of n points $x_i \in \overline{\Omega}$, with $i = 1, \dots, n$. Thanks to the Sobolev injection $H^s(\Omega) \hookrightarrow C^0(\overline{\Omega})$ valid for $s > d/2$ [DNPV12], the sampled values $f(x_i)$ are well defined. We correspondingly define a sampling operator

$$E_n : f \in H_0^s(\Omega) \longmapsto (f(x_i))_{1 \leqslant i \leqslant n} \in \mathbb{R}^n. \tag{2.1}$$

Our noisy measurements are given by the vector $y_\sigma \in \mathbb{R}^n$ via the data

$$(y_\sigma)_i = f(x_i) + \sigma\xi_i = (E_n f)_i + \sigma\xi_i, \quad i = 1, \dots, n.$$

Here, measurement noise is modeled by independent random variables $\sigma\xi_i$, $i = 1, \dots, n$, where the ξ_i are centered with unit variance.

2.2. Discretization.

Recall that the variable h is related to n by $h \sim n^{-1/d}$. We sometimes prefer to give our estimates in terms of h rather than n, since the parameters h and β are homogeneous and therefore naturally compare. We assume that we are given a discretization operator $P_n : \mathbb{R}^n \to L^2(\Omega)$ defined by means of *basis functions* $\phi_i \in C^0(\overline{\Omega})$, $i = 1, \dots, n$, via the reconstruction formula

$$\forall z \in \mathbb{R}^n, \qquad P_n z = \sum_{i=1}^n z_i \phi_i. \tag{2.2}$$

Throughout the paper, we will assume that the basis functions are positive in a neighborhood of size about h around x_i, and vanish outside of a larger neighborhood still of size about h. Our precise hypothesis reads as follows: there exist $m > 0$, $C > c > 0$ independent of i and n such that

$$\phi_i(x_i + hz) \begin{cases} \geqslant m & \text{for } |z| \leqslant c, \\ = 0 & \text{for } |z| \geqslant C. \end{cases} \tag{2.3}$$

All the symbols $\lesssim$ and $\sim$ below should also be understood to be uniform with respect to $i=1,\dots,n$.

In particular, (2.3) ensures the inclusion $B(x_i,ch)\subset \mathrm{supp}(\phi_i)\subset B(x_i,Ch)$, where $B(x_0,r) = \{x\in\overline{\Omega},\ |x-x_0|\leqslant r\}$ denotes the closed Euclidean ball with center x_0 and radius $r\geqslant 0$. We moreover assume

$$\|\phi_i\|_{L^\infty(\Omega)}\lesssim 1, \tag{2.4}$$

which in turn entails the estimates[1]

$$\|\phi_i\|\sim h^{d/2} \text{ and } \|\phi_i\|_{L^1(\Omega)}\sim h^d. \tag{H_ϕ}$$

Note that the estimates (H_ϕ) are those essential for our results. We introduce the sufficient hypotheses (2.3) and (2.4) explicitly because they are more easily checked in practice.

We will use Sobolev balls as smoothness classes:

$$\mathcal{F}_s := \{f\in H_0^s(\Omega),\ \|f\|_s\leqslant 1\}, \tag{2.5}$$

where, without loss of generality given the linearity of the problem, we considered the unit ball instead of a ball of a given radius.

We have a natural notion of accuracy of reconstruction that combine both the discretization operator P_n defined in (2.2) and the sampling operator E_n defined in (2.1). Given some order s_a, we will be interested in the assumption, for a given $s\geqslant s_a$,

$$\|P_nE_nf-f\|\lesssim \|f\|_s\, h^{s_a} \tag{2.6}$$

for every $f\in H_0^s(\Omega)$. We will also be interested in the case where the above is sharp, namely that for a given $s\geqslant s_a$,

$$\sup_{f\in\mathcal{F}_s}\|P_nE_nf-f\|\sim h^{s_a}. \tag{H_a}$$

It is known that under fairly general hypotheses, $\mathbb{P}_k$ finite elements satisfy (2.6) for all $s\geqslant s_a$, with $s_a=k+1$; see, for instance, [Tho07].

2.3. Regularization.

Pick a smooth and compactly supported kernel K over $\mathbb{R}^d$ that satisfies in particular

$$\int_{\mathbb{R}^d} K(x)\,dx = 1.$$

We let $K_\beta := \beta^{-d}K(\beta^{-1}\cdot)$ and note that

$$\|K_\beta\|_{L^1(\mathbb{R}^d)}\lesssim 1,\quad \|K_\beta\|_{L^2(\mathbb{R}^d)}\lesssim \beta^{-d/2}.$$

[1] For the lower bounds, the first inequality of (2.3) entails $\int_\Omega|\phi_i(x)|^2\,dx\geqslant m^2|B(x_i,ch)|\sim h^d$, $\int_\Omega|\phi_i(x)|\,dx\geqslant m|B(x_i,ch)|\sim h^d$, which shows $\|\phi_i\|\gtrsim h^{d/2}$, and $\|\phi_i\|_{L^1(\Omega)}\gtrsim h^d$. The uniform compact support given by (2.3) combined with (2.4) leads to $\|\phi_i\|^2=\int_\Omega|\phi_i(x)|^2\,dx\lesssim|B(x_i,Ch)|\sim h^d$, $\|\phi_i\|_{L^1(\Omega)}=\int_\Omega|\phi_i(x)|\,dx\lesssim|B(x_i,Ch)|\sim h^d$.

For a function $f \in L^2(\Omega)$, we define the convolution

$$\forall x \in \mathbb{R}^d, \quad (K_\beta * f)(x) = \int_\Omega K_\beta(x-y) f(y)\, dy.$$

We also assume that K reproduces moments up to the degree $s_r - 1 \in \mathbb{N}^*$ but does not reproduce at least one moment of degree s_r; i.e., using the notation $|r| = r_1 + \cdots + r_d$ for $r = (r_1, \dots, r_d) \in \mathbb{N}^d$,

$$\begin{aligned} &\forall r \in \mathbb{N}^d,\ 1 \leqslant |r| \leqslant s_r - 1, && \int_{\mathbb{R}^d} x_1^{r_1} \dots x_d^{r_d} K(x)\, dx = 0, \\ &\exists r \in \mathbb{N}^d,\ |r| = s_r, && \int_{\mathbb{R}^d} x_1^{r_1} \dots x_d^{r_d} K(x)\, dx \neq 0. \end{aligned} \tag{2.7}$$

Under the above assumptions and if $s \geqslant s_r$ is such that $s - \frac{1}{2}$ is not an integer,[2] we have

$$\sup_{f \in \mathcal{F}_s} \|K_\beta * f - f\| \lesssim \beta^{s_r}. \tag{2.8}$$

In fact, the estimate above is sharp thanks to the assumption that K does not reproduce at least one moment of degree s_r. In other words, for $s \geqslant s_r$ such that $s - \frac{1}{2}$ is not an integer, we have

$$\sup_{f \in \mathcal{F}_s} \|K_\beta * f - f\| \sim \beta^{s_r}. \tag{H_r}$$

Although these estimates are common, one is usually interested in the upper bound (2.8), with $\Omega = \mathbb{R}^d$ and integer parameter s. For completeness, we thus provide a proof of (H_r) in our setting whenever $s - \frac{1}{2}$ is not an integer, which we postpone to Appendix A.

Remark 2. Many common kernels (integrating to 1) are nonnegative ($K \geqslant 0$) and symmetric ($K(x) = K(-x)$ for all $x \in \mathbb{R}^d$). The nonnegativity assumption prevents numerical instabilities. However, these kernels are of order $s_r = 2$ and not more since they are such that $\int_{\mathbb{R}^d} |x|^2 K(x)\, dx \neq 0$ and hence do not reproduce all moments of order 2.

Recapitulating our assumptions. From now on, we always assume that (H_ϕ), (H_a), and (H_r) hold. Some of the results will in fact require weaker hypotheses, for instance in the form of upper bounds $\lesssim$ rather than asymptotic equivalence $\sim$. The proof of each specific result will make it clear what is actually necessary for the claimed statement to hold.

2.4. Reconstruction errors.

The case with no regularization. The first estimator is given by $P_n y_\sigma$. The corresponding error is

$$e_{\mathrm{noreg}}(\sigma, h) := \sup_{f \in \mathcal{F}_s} \mathbb{E}\left[\|P_n y_\sigma - f\|^2\right]^{1/2}. \tag{2.9}$$

[2]This technical restriction is due to some inessential subtleties related to fractional Sobolev spaces $H_0^s(\Omega)$; see the reference [LM12].

The case with regularization. The second estimator consists in adding regularization, the estimator being now given by $K_\beta * (P_n y_\sigma)$. The corresponding error is

$$e_{\mathrm{optreg}}(\sigma,h) := \inf_{\beta>0} e(\beta,\sigma,h),$$

with

$$e(\beta,\sigma,h) := \sup_{f\in\mathcal{F}_s} \mathbb{E}\left[\|K_\beta * P_n y_\sigma - f\|^2\right]^{1/2}.$$

3. Upper estimates.

3.1. The case with no regularization ($\beta=0$).

We first analyze the error $e_{\mathrm{noreg}}(\sigma,h)$, associated to the estimator $P_n y_\sigma$ when no regularization is involved.

PROPOSITION 3.1. *Assume that $s \geqslant s_a$. We have*

$$e_{\mathrm{noreg}}(\sigma,h) \lesssim \sigma + h^{s_a}.$$

Proof. Writing

$$P_n y_\sigma - f = (P_n y_\sigma - P_n E_n f) + (P_n E_n f - f),$$

and since the random variables ξ_i are centered with unit variance, we obtain

$$\begin{aligned}\mathbb{E}\left[\|P_n y_\sigma - f\|^2\right] &= \mathbb{E}\left[\|P_n(y_\sigma - y)\|^2\right] + \|P_n E_n f - f\|^2\\ &= \sigma^2\,\mathbb{E}\left[\left\|\sum_{i=1}^n \xi_i\phi_i\right\|^2\right] + \|P_n E_n f - f\|^2\\ &= \sigma^2\sum_{i=1}^n \|\phi_i\|^2 + \|P_n E_n f - f\|^2.\end{aligned}$$

Owing to $\|\phi_i\|^2 \lesssim h^d$, which follows from ($H_\phi$), we derive

$$\mathbb{E}\left[\|P_n y_\sigma - f\|^2\right] \lesssim \sigma^2 h^d n + \|P_n E_n f - f\|^2 \sim \sigma^2 + \|P_n E_n f - f\|^2.$$

Using Assumption (2.6), taking the square root and the supremum over $f\in\mathcal{F}_s$, we obtain the result. ∎

3.2. The case with regularization.

We now study $e_{\mathrm{optreg}}(\sigma,h)$, associated with the estimator $K_\beta * P_n y_\sigma$. Note first that Young's inequality yields, for all $f\in L^2(\Omega)$,

$$\|K_\beta * f\| \leqslant \|K_\beta\|_{L^1(\mathbb{R}^d)}\|f\|_{L^2(\Omega)}, \quad \text{and} \quad \|K_\beta * f\| \leqslant \|K_\beta\|_{L^2(\mathbb{R}^d)}\|f\|_{L^1(\Omega)}.$$

PROPOSITION 3.2. *Assume that $s \geqslant \max(s_r, s_a)$. We have*

$$e(\beta,\sigma,h) \lesssim \sigma\min(\beta^{-1}h,1)^{d/2} + h^{s_a} + \beta^{s_r}. \tag{3.1}$$

In particular

$$e_{\mathrm{optreg}}(\sigma,h) \lesssim \sigma\min(1,\sigma^{-1}h^{s_r})^{d/(2s_r+d)} + h^{s_a}.$$

The above alternative is obtained by letting $\beta\to 0$ and $\beta=\beta^\star(\sigma,h)$, respectively, with

$$\beta^\star(\sigma,h) \sim \sigma^{2/(2s_r+d)} h^{d/(2s_r+d)}.$$

Proof. In the same way as in the proof of Proposition 3.1, we have

$$\mathbb{E}\left[\|K_\beta * P_n y_\sigma - f\|^2\right] = \sigma^2 \sum_{i=1}^n \|K_\beta * \phi_i\|^2 + \|K_\beta * P_n E_n f - f\|^2 .$$

The second term may be estimated thanks to (2.6) and (2.8). This yields

$$\begin{aligned}\|K_\beta * P_n E_n f - f\| &\leqslant \|K_\beta * (P_n E_n f - f)\| + \|K_\beta * f - f\| \\ &\leqslant \|K_\beta\|_{L^1(\mathbb{R}^d)} \|P_n E_n f - f\| + \|K_\beta * f - f\| \\ &\lesssim \|P_n E_n f - f\| + \|K_\beta * f - f\| \\ &\lesssim \|f\|_{s_a} h^{s_a} + \|f\|_{s_r} \beta^{s_r}.\end{aligned}$$

Two upper bounds may be derived for the first term, by means of two applications of Young's inequality, together with (H_ϕ), namely

$$\|K_\beta * \phi_i\|^2 \leqslant \|K_\beta\|^2_{L^1(\mathbb{R}^d)} \|\phi_i\|^2 \lesssim \|\phi_i\|^2 \lesssim h^d$$

and

$$\|K_\beta * \phi_i\|^2 \leqslant \|K_\beta\|^2_{L^2(\mathbb{R}^d)} \|\phi_i\|^2_{L^1(\Omega)} \lesssim \beta^{-d} \|\phi_i\|^2_{L^1(\Omega)} \lesssim \beta^{-d} h^{2d}.$$

The first choice leads to the following bound, valid for any $\beta > 0$:

$$e(\beta,\sigma,h) \lesssim \sigma + h^{s_a} + \beta^{s_r}. \tag{3.2}$$

The second choice leads to

$$e(\beta,\sigma,h) \lesssim \sigma \beta^{-d/2} h^{d/2} + h^{s_a} + \beta^{s_r}. \tag{3.3}$$

Combining the two estimates, we obtain (3.1). Let us now minimize (3.2) and (3.3) with respect to β. For $\beta \lesssim h$, (3.2) is sharper, while for $\beta \gtrsim h$ (3.3) prevails. For (3.2) we achieve a minimum or order $\sigma + h^{s_a}$ by letting $\beta \to 0$. Taking derivatives, (3.3) is minimal for $\beta \sim \beta^\star(\sigma,h)$ with corresponding minimum of order $\sigma^{2s_r/(2s_r+d)} h^{ds_r/(2s_r+d)} + h^{s_a}$. Finally, $\beta^\star(\sigma,h) \lesssim h$ if and only if $\sigma \lesssim h^{s_r}$, from which we infer

$$\begin{aligned} e_{\mathrm{optreg}}(\sigma,h) &= \inf_{\beta>0} \sup_{f\in\mathcal{F}_s} \mathbb{E}\left[\|K_\beta * P_n y_\sigma - f\|_2^2\right]^{1/2} \\ &\lesssim \begin{cases} \sigma + h^{s_a} & \text{if } \sigma \lesssim h^{s_r}, \\ \sigma^{2s_r/(2s_r+d)} h^{ds_r/(2s_r+d)} + h^{s_a} & \text{else} \end{cases} \\ &= \sigma \min(1, \sigma^{-1} h^{s_r})^{d/(2s_r+d)} + h^{s_a}. \end{aligned}$$ ∎

Remark 3. As the proofs show, all the derived upper bounds are also valid in the stronger form where all errors are defined with expectation $\mathbb{E}$ and supremum $\sup_{f\in\mathcal{F}_s}$ swapped.

4. Regularization versus no regularization.

We wish to compare the effect of *not regularizing* (*i.e.* $\beta = 0$) versus *regularizing*, with the choice

$$\beta = \beta^\star(\sigma, h) \sim \sigma^{2/(2s_r+d)} h^{d/(2s_r+d)}. \tag{4.1}$$

In order to do so, we establish lower bounds for $e_{\text{noreg}}(\sigma, h)$ defined in (2.9) and

$$e_{\text{reg}}(\sigma, h) := e(\beta^*(\sigma, h), \sigma, h).$$

4.1. Estimates from below.

Lemma 4.1. *Assume $s \geqslant s_a$. We have*

$$e_{\text{noreg}}(\sigma, h) \sim \sigma + h^{s_a}.$$

Proof. Recall the identity

$$e_{\text{noreg}}(\sigma, h)^2 = \sigma^2 \sum_{i=1}^n \|\phi_i\|^2 + \sup_{f \in \mathcal{F}_s} \|P_n E_n f - f\|^2,$$

from which the result follows thanks to (H_ϕ) and (H_a). ∎

Observe in particular the identity

$$e(\beta, \sigma, h)^2 = \sigma^2 \sum_{i=1}^n \|K_\beta * \phi_i\|^2 + \sup_{f \in \mathcal{F}_s} \|K_\beta * P_n E_n f - f\|^2.$$

Lemma 4.2. *Let $s \geqslant \max(s_r, s_a)$. There exists $c > 0$ such that*

$$e(\beta, \sigma, h) \gtrsim \beta^{s_r/2}(\beta^{s_r/2} - c\, h^{s_a/2}).$$

Proof. For $f \in \mathcal{F}_s$, we write

$$\begin{aligned}
\|K_\beta * P_n E_n f - f\|^2 &= \|K_\beta * (P_n E_n f - f) + (K_\beta * f - f)\|^2 \\
&= \|K_\beta * (P_n E_n f - f)\|^2 + 2\langle K_\beta * (P_n E_n f - f), (K_\beta * f - f)\rangle \\
&\quad + \|K_\beta * f - f\|^2 \\
&\geqslant 2\langle K_\beta * (P_n E_n f - f), (K_\beta * f - f)\rangle + \|K_\beta * f - f\|^2 \\
&\geqslant -2\|K_\beta * (P_n E_n f - f)\|\,\|K_\beta * f - f\| + \|K_\beta * f - f\|^2.
\end{aligned}$$

To further estimate the first term on the right-hand side, we may use both (2.6) and (2.8) to obtain

$$\begin{aligned}
\|K_\beta * (P_n E_n f - f)\|\,\|K_\beta * f - f\| &\lesssim \|P_n E_n f - f\|\,\|K_\beta * f - f\| \\
&\lesssim \|f\|_s h^{s_a} \|f\|_s \beta^{s_r} \leqslant h^{s_a} \beta^{s_r}.
\end{aligned}$$

The second term is dealt with by (H_r). As a result, we may find $c > 0$ such that

$$\sup_{f \in \mathcal{F}_s} \|K_\beta * P_n E_n f - f\|^2 \gtrsim -c\, h^{s_a} \beta^{s_r} + \beta^{2s_r} = \beta^{s_r}(\beta^{s_r} - c\, h^{s_a}),$$

and finally (upon changing the constant c),

$$e(\beta,\sigma,h) \gtrsim \beta^{s_r/2}(\beta^{s_r/2} - c\,h^{s_a/2}). \qquad \blacksquare$$

Remark 4. A more comprehensive understanding of lower bounds for errors at fixed $\beta > 0$ would notably require lower estimates for the norms $\|K_\beta * \phi_i\|$. We were only able to establish such estimates under restrictive assumptions, namely when $K \geqslant 0$ and assuming $\beta = \beta(h) = o(h)$. Since this result is only partial and does not happen to be necessary for the comparison between the two analyzed strategies (not regularizing or regularizing through (1.3)), we delay these estimates until Appendix B.

In order to compare $e_{\text{noreg}}(\sigma,h)$ and $e_{\text{reg}}(\sigma,h) := e(\beta^*(\sigma,h),\sigma,h)$ as functions of the noise level σ, we let

$$\sigma = \sigma(h) \sim h^\lambda, \quad \text{with} \quad \lambda \geqslant 0.$$

It follows that

$$\beta^\star(\sigma,h) \sim \sigma^{2/(2s_r+d)} h^{d/(2s_r+d)} \sim h^{\frac{2\lambda+d}{2s_r+d}}$$

now only depends on h. For conciseness, we write $\beta^\star(h) = h^{\frac{2\lambda+d}{2s_r+d}}$.

Both errors now depend on h solely; abusing notation slightly, we write $e_{\text{noreg}}(h)$ and $e_{\text{reg}}(h)$, respectively. Under (H_a) and according to Lemma 4.1,

$$e_{\text{noreg}}(h) \sim \sigma(h) + h^{s_a} \sim \begin{cases} h^\lambda & \text{if } \lambda \leqslant s_a, \\ h^{s_a} & \text{if } \lambda > s_a \end{cases} \qquad (4.2)$$
$$= h^{\min(\lambda, s_a)}.$$

We will also need the following elementary useful facts:

$$\beta^\star(h)^{s_r} \lesssim \sigma(h) \iff h \lesssim \beta^\star(h) \iff \lambda \leqslant s_r,$$

and

$$h^{s_a} \lesssim \beta^\star(h)^{s_r} \iff \lambda \leqslant \lambda_M,$$

where

$$\lambda_M := s_a + \frac{d}{2}\left(\frac{s_a}{s_r} - 1\right).$$

We note that if $s_a \leqslant s_r$, we may have that λ_M is negative and $\lambda_M \leqslant s_a \leqslant s_r$, whereas if $s_a > s_r$, then $s_r < s_a < \lambda_M$ always holds.

PROPOSITION 4.3. *Assume that* $s \geqslant \max(s_r, s_a)$. *We have*

$$\begin{cases} e_{\text{reg}}(h) \sim h^{\frac{2\lambda+d}{2s_r+d}s_r} & \text{if } \lambda < \lambda_M, \\ e_{\text{reg}}(h) \lesssim h^{s_a} & \text{if } \lambda \geqslant \lambda_M. \end{cases}$$

Proof. Returning to the estimate of Proposition 3.2, we have

$$\begin{aligned} e_{\mathrm{reg}}(h) &\lesssim \sigma(h)\min(\beta^\star(h)^{-1}h,1)^{d/2} + h^{s_a} + \beta^\star(h)^{s_r} \sim \min(\sigma(h),\beta^\star(h)^{s_r}) + h^{s_a} + \beta^\star(h)^{s_r} \\ &\lesssim h^{s_a} + \beta^\star(h)^{s_r}, \end{aligned}$$

and we infer

$$e_{\mathrm{reg}}(h) \lesssim \begin{cases} h^{\frac{2\lambda+d}{2s_r+d}s_r} & \text{if } \lambda \leqslant \lambda_M, \\ h^{s_a} & \text{if } \lambda > \lambda_M. \end{cases}$$

It remains to show

$$e_{\mathrm{reg}}(h) \gtrsim h^{\frac{2\lambda+d}{2s_r+d}s_r}$$

whenever $\lambda < \lambda_M$. This is a consequence of Lemma 4.2, which gives

$$e_{\mathrm{reg}}(h) \gtrsim \beta^\star(h)^{s_r/2}(\beta^\star(h)^{s_r/2} - c\,h^{s_a/2}) \gtrsim \beta^\star(h)^{s_r} = h^{\frac{2\lambda+d}{2s_r+d}s_r}$$

since $h^{s_a} \lesssim \beta^\star(h)^{s_r}$ under the assumption $\lambda < \lambda_M$. ∎

Remark 5. We do not know whether the tighter estimate $e_{\mathrm{reg}}(h) \sim h^{s_a}$ is valid for $\lambda > \lambda_M$ under our set of hypotheses, or if additional realistic assumptions can be made to establish it.

We now highlight situations where it is *strictly* more advantageous *not to regularize* via the rule (4.1) and ignore the effect of regularization.

4.2. The case when $s_a \leqslant s_r$. Our first result is that such a scenario does not occur whenever $s_a \leqslant s_r$.

PROPOSITION 4.4. *If $s_a \leqslant s_r \leqslant s$, then*

$$\begin{cases} e_{\mathrm{reg}}(h) \sim h^{\frac{2\lambda+d}{2s_r+d}s_r} \text{ and } e_{\mathrm{noreg}}(h) \sim h^{\lambda} & \text{if } \lambda < \lambda_M, \\ e_{\mathrm{reg}}(h) \lesssim h^{s_a} \text{ and } e_{\mathrm{noreg}}(h) \sim h^{\lambda} & \text{if } \lambda_M \leqslant \lambda < s_a, \\ e_{\mathrm{reg}}(h) \lesssim h^{s_a} \text{ and } e_{\mathrm{noreg}}(h) \sim h^{s_a} & \text{if } \lambda \geqslant s_a. \end{cases}$$

In particular, it is strictly better to regularize through (4.1) *whenever $\lambda < s_a$, in which case we have*

$$e_{\mathrm{reg}}(h) = o(e_{\mathrm{noreg}}(h)).$$

It is at least as good to regularize through (4.1) *whenever $\lambda \geqslant s_a$, in which case we have*

$$e_{\mathrm{reg}}(h) \lesssim e_{\mathrm{noreg}}(h).$$

Proof. Recall that $\lambda_M \leqslant s_a \leqslant s_r$. All cases are obtained by combining Proposition 4.3 with the estimate (4.2). ∎

4.3. The case when $s_a > s_r$. In that case, it is indeed possible to find situations where it becomes *strictly* more advantageous *not to regularize* via the rule (4.1) and ignore the effect of the regularization, a perhaps surprising result.

PROPOSITION 4.5. *If $s \geqslant s_a > s_r$, then*

$$\begin{cases} e_{\text{reg}}(h) \sim h^{\frac{2\lambda+d}{2s_r+d}s_r} \text{ and } e_{\text{noreg}}(h) \sim h^{\lambda} & \text{if } \lambda \leqslant s_a, \\ e_{\text{reg}}(h) \sim h^{\frac{2\lambda+d}{2s_r+d}s_r} \text{ and } e_{\text{noreg}}(h) \sim h^{s_a} & \text{if } s_a < \lambda < \lambda_M, \\ e_{\text{reg}}(h) \lesssim h^{s_a} \text{ and } e_{\text{noreg}}(h) \sim h^{s_a} & \text{if } \lambda \geqslant \lambda_M. \end{cases}$$

In particular, it is strictly better to regularize through (4.1) whenever $\lambda < s_r$, in which case we have

$$e_{\text{reg}}(h) = o(e_{\text{noreg}}(h)).$$

It is strictly better not to regularize through (4.1) whenever $s_r < \lambda < \lambda_M$, in which case we have

$$e_{\text{noreg}}(h) = o(e_{\text{reg}}(h)).$$

Finally, it is better to regularize through (4.1) whenever $\lambda \geqslant \lambda_M$, in which case we have

$$e_{\text{reg}}(h) \lesssim e_{\text{noreg}}(h).$$

Proof. Recall that $s_r < s_a < \lambda_M$. Again, all cases are obtained by combining Proposition 4.3 with the estimate (4.2). ■

As mentioned earlier, we can go further and estimate the level of noise with highest gain *in not regularizing* compared to *regularizing* in the regime $s_a > s_r$. We find it more transparent to express this gain in terms of sampling size n rather than in terms of the mesh size h, since n may be regarded as the actual cost of measuring f over the design x_i, $i = 1, \dots, n$. Recall that $\sigma(h) = \sigma(n) \sim n^{-\frac{\lambda}{d}}$ with $\lambda \geqslant 0$. The highest gain happens when $\lambda = s_a$, for which

$$e_{\text{reg}}(n) \sim h^{\frac{2s_a+d}{2s_r+d}s_r} \sim n^{-\frac{2s_a+d}{2s_r+d}\frac{s_r}{d}}, \quad e_{\text{noreg}}(n) \sim h^{s_a} \sim n^{-\frac{s_a}{d}}.$$

One can hence gain up to a polynomial (in n) factor of order $\frac{s_a - s_r}{2s_r+d}$ which vanishes for large d. This is consistent with the condition $s > d/2$, which somehow enforces f to be smoother as d increases.

5. Numerical simulations.

5.1. Setting.

We work in dimension $d = 1$ with $\Omega = (0,1)$. We are mostly interested in situations where regularization might be detrimental, i.e., when $s_r < s_a$. Hence, we choose kernels of order $s_r = 1$ and $s_r = 2$, respectively, and approximation methods of order $s_a = 2$ and $s_a = 3$. These are defined below.

Regularization. We consider two kernels K and H, given by

$$K := \mathbb{1}_{[0,1]}, \qquad \text{and } H := \frac{1}{2}\mathbb{1}_{[-1,1]}.$$

They satisfy $s_r = 1$ and $s_r = 2$, respectively. The kernel K is not standard: it is not centered—hence its low order of convergence. We make this rather artificial choice in order to better illustrate our results which are most visible when the gap $s_a - s_r$ gets larger, especially in small dimensions.

Approximation. We consider $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements. We make sure to be consistent with our choice that n represents the number of basis functions. In doing so, the definitions below slightly differ from usual definitions, which have h rather than n as the defining parameter.

$\mathbb{P}_1$ *finite elements.* We let $n \geqslant 3$ be given. We define $h := \frac{1}{n-1}$, and for $i = 1, \dots, n$, we denote $x_i := (i-1)h$. Defining the shape function

$$\forall x \in \mathbb{R}, \qquad \varphi(x) := (1 - |x|)\mathbb{1}_{[-1,1]}(x),$$

the basis functions are then given as follows for $i = 1, \dots, n$:

$$\forall x \in [0,1], \qquad \phi_i(x) = \varphi\Big(\frac{x - x_i}{h}\Big).$$

These basis functions clearly satisfy (2.3) and (2.4), so that (H_ϕ) holds. Furthermore, the approximation operator $P_n E_n$ associated to $\mathbb{P}_1$ finite elements satisfies (2.6) with $s_a = 2$.

$\mathbb{P}_2$ *finite elements.* We let $n \geqslant 3$ be an odd integer. We define $h := \frac{2}{n-1}$, and for $i = 1, \dots, n$, we set $x_i := (i-1)\frac{h}{2}$. Defining the shape functions

$$\forall x \in [0,1], \qquad \varphi(x) := (1 - |x|)(1 - 2|x|)\mathbb{1}_{[-1,1]}(x), \quad \psi(x) := (1 - 4x^2)\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(x),$$

the basis functions are then given as follows for $i = 1, \dots, n$:

$$\forall x \in [0,1], \qquad \phi_i(x) = \begin{cases} \varphi\left(\dfrac{x - x_i}{h}\right) & \text{if } i \text{ is odd,} \\ \psi\left(\dfrac{x - x_i}{h}\right) & \text{if } i \text{ is even.} \end{cases}$$

These basis functions clearly satisfy (2.3) and (2.4), so that (H_ϕ) holds. The approximation operator $P_n E_n$ associated to $\mathbb{P}_2$ finite elements satisfies (2.6) with $s_a = 3$.

5.2. Methodology. In order to illustrate our theoretical results, we aim at computing, for a given function $f \in H_0^s(\Omega)$ with $s > d/2 = 1/2$, the two errors for various noise levels $\sigma = \sigma(n) = n^{-\lambda/d}$, which corresponds to $\sigma = \sigma(h) \sim h^\lambda$. More precisely, we are interested in finding how quickly the two errors of interest $e_{\text{noreg}}(\sigma(n), n)$ and $e(\beta^\star(n), \sigma(n), n)$ vanish as n grows, as a function of the noise level defined by the parameter λ. Recall that regularization is made with a parameter β chosen to be $\beta^\star(n)$ given by

$$\beta^\star(\sigma(n), n) \sim \sigma(n)^{\frac{2}{2s_r+d}} n^{-\frac{1}{2s_r+d}} = n^{-\frac{1}{d}\frac{2\lambda+d}{2s_r+d}}.$$

In order to compute these expectations without sampling, we use the exact relations

$$e_{\text{noreg}}(\sigma, n)^2 = \sigma^2 \sum_{i=1}^n \|\phi_i\|^2 + \|P_n E_n f - f\|^2,$$
$$e(\beta, \sigma, n)^2 = \sigma^2 \sum_{i=1}^n \|K_\beta * \phi_i\|^2 + \|K_\beta * P_n E_n f - f\|^2.$$

Mathematically, for a given choice of λ, a given error is of order $n^{-\gamma(\lambda)}$, and our goal is to estimate the function γ over a given interval for λ. Hence, for a given choice of approximation method, we have a function $\lambda \mapsto \gamma_{\text{noreg}}(\lambda)$ defined by

$$\mathbb{E}\left[\|P_n y_\sigma - f\|^2\right]^{1/2} \sim n^{-\gamma_{\text{noreg}}(\lambda)}, \tag{5.1}$$

and for a given choice of kernel and approximation method, we have a function $\lambda \mapsto \gamma_{\text{reg}}(\lambda)$ defined by

$$\mathbb{E}\left[\left\|K_{\beta^\star(n)} * P_n y_\sigma - f\right\|^2\right]^{1/2} \sim n^{-\gamma_{\text{reg}}(\lambda)}. \tag{5.2}$$

From our theoretical results, recalling that $d=1$, we expect

$$\forall \lambda \geqslant 0, \quad \gamma_{\text{noreg}}(\lambda) = \min(\lambda, s_a), \qquad \text{and} \qquad \forall \lambda \in [0, \lambda_M[, \quad \gamma_{\text{reg}}(\lambda) = \frac{2\lambda+1}{2s_r+1} s_r.$$

For $\lambda \geqslant \lambda_M$, we recall our upper bound for the error which translates into the lower bound $\gamma_{\text{reg}}(\lambda) \geqslant s_a$. When plotting functions γ_{noreg} and γ_{reg}, we pay specific attention to the regime

$$\lambda \in [0, \lambda_M] = \left[0,\, s_a + \frac{1}{2}\left(\frac{s_a}{s_r} - 1\right)\right],$$

but we will consider the larger interval $[0,5]$; the latter contains $[0,\lambda_M]$ in all cases. Numerical simulations are carried out with values of λ ranging over $[0,5]$ with step size equal to 0.5; hence 11 values of λ are used.

When estimating the error without regularization, we consider both $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements. When estimating the error with regularization, we consider all 4 possible scenarios, corresponding to choosing the kernel to be either K or H, and the approximation method to be either $\mathbb{P}_1$ or $\mathbb{P}_2$ finite elements. Note that only in the case where the kernel is H and with $\mathbb{P}_1$ finite elements does one have $s_r = s_a$; in all other cases $s_r < s_a$.

Estimating orders of convergence. For a fixed choice of $\lambda \geqslant 0$, we must evaluate how quickly a given error tends to 0 as a function of n. In order to do so, we choose $n \in \{63, 125, 251, 501, 1001\}$ and compute the slope of both errors in log-log scale.

Estimating norms. For a given draw, L^2-norms $\|\cdot\|$ are estimated by Simpson's rule with 10^5 points in order to ensure accurate estimates that do not compete with the expected orders of convergence.

Estimating convolutions. Let us stress that other integrals are involved in the process of computing the error when regularization is involved. Those are inherent in evaluating the convolution $K_\beta * P_n y_\sigma$, which in turn boils down to evaluating all functions $K_\beta * \phi_i$, $i=1,\dots,n$. In order for these computations to not impact the orders of convergence, we analytically rather than numerically compute these functions. This is possible for our choices of kernels and finite element functions.

In practice, however, these integrals would be computed with errors. All other things being equal, these errors can only further reduce the quality of regularizing compared to not regularizing.

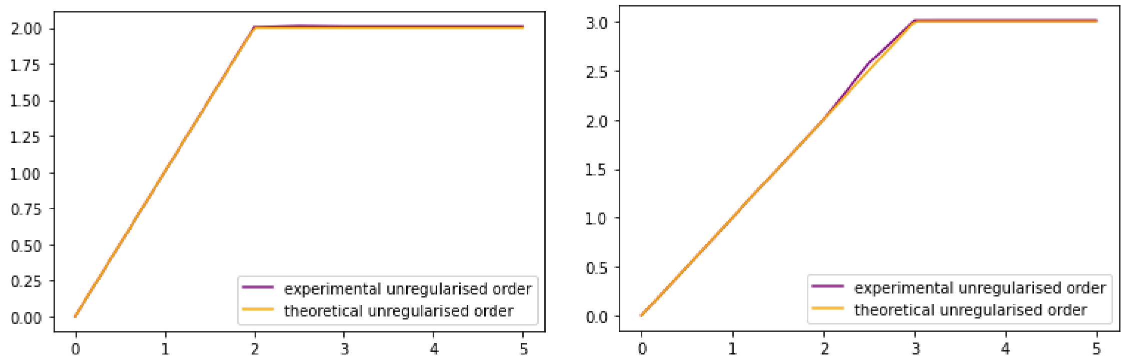

Figure 5.1. No regularization: *Plot of* $\lambda \mapsto \gamma_{noreg}(\lambda)$ *defined by* (5.1). *The left panel shows the case of* $\mathbb{P}_1$ *finite elements; the right panel shows that of* $\mathbb{P}_2$ *finite elements. In both cases, the theoretical curve* $\lambda \mapsto \min(\lambda, s_a)$ *is plotted in orange against the numerically obtained curve, in purple.*

5.3. Numerical results. For all numerical experiments, we choose

$$f : x \mapsto (1-x)^3 \sin^3(4x), \tag{5.3}$$

which satisfies $f \in H_0^3(\Omega)$ (in fact, $f \in H_0^s(\Omega)$ if and only if $s \leqslant \frac{7}{2}$), so that we will always have $s \geqslant s_r$ as well as $s \geqslant s_a$.

$\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements. The orders of convergence obtained numerically match the theoretical ones, as shown by Figure 5.1. One indeed expects the function $\gamma_{\text{noreg}}(\lambda) : \lambda \mapsto \min(\lambda, s_a)$ and this is exactly what is found.

The kernel K with $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements. In this case, $s_r = 1$, with either $s_a = 2$ or $s_a = 3$. In the case of $\mathbb{P}_1$ finite elements, one has $\lambda_M = \frac{5}{2}$, while in the second $\lambda_M = 4$. The orders of convergence obtained numerically are a good match to the theoretical ones, as shown by Figure 5.2. The match is almost perfect in the $\mathbb{P}_1$ case. In the second case of $\mathbb{P}_2$ finite elements, discrepancies may be observed as λ approaches and exceeds $\lambda_M = 4$, which could be due to numerical errors taking over in this regime.

The kernel H with $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements. In this case, $s_r = 2$, with either $s_a = 2$ or $s_a = 3$. In the first case, one has $\lambda_M = s_r = s_a = 2$, while in the second $\lambda_M = \frac{13}{4}$. The orders of convergence obtained numerically are an excellent match to the theoretical ones, as shown by Figure 5.3. In both cases, there is little difference between the two theoretical curves, making it more difficult to clearly distinguish the numerically built curve from the two theoretical ones.

Example of a reconstruction. In this 1-dimensional setting, the actual improvement obtained by *not regularizing* whenever this is superior to the regularization by (1.3) is hardly visible at the level of reconstructions.

Yet, we provide an example in the case of $\mathbb{P}_1$ finite elements with kernel equal to K, which exacerbates the sought-for effect since $s_r = 1$, $s_a = 3$. We take $n = 63$, $\lambda = \frac{3}{2}$, thus falling in the regime $s_r = 1 < \lambda < \lambda_M = 4$ where theory predicts *not regularizing* should be better than *regularizing* by (1.3).

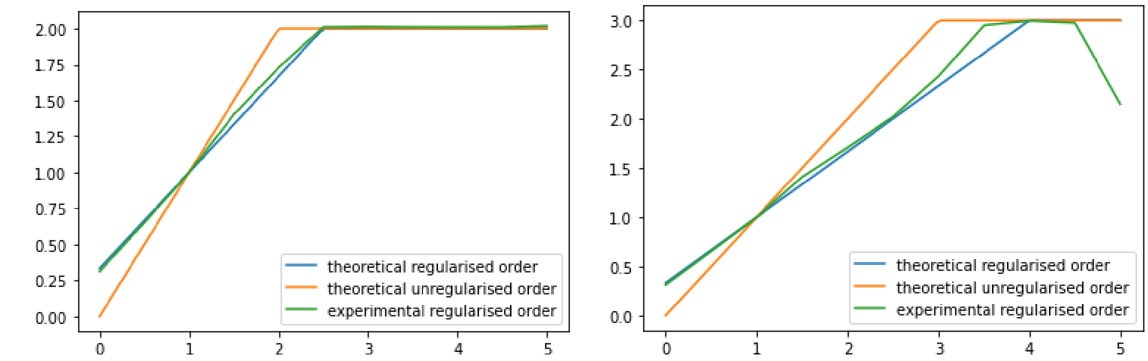

Figure 5.2. Regularization with kernel K: *Plot of* $\lambda \mapsto \gamma_{noreg}(\lambda)$ *defined by* (5.1)*, with regularization through the kernel* K*. The left panel shows the case of* $\mathbb{P}_1$ *finite elements* ($\lambda_M = \frac{5}{2}$)*; the right panel shows that of* $\mathbb{P}_2$ *finite elements* ($\lambda_M = 4$)*. In both cases, the theoretical curves without regularization* $\lambda \mapsto \min(\lambda, s_a)$ ($\lambda \geqslant 0$) *and with regularization* $\lambda \mapsto \frac{1}{3}(2\lambda+1)$ ($0 \leqslant \lambda < \lambda_M$) *are plotted in orange and blue, respectively. The numerically obtained curve (with regularization through* K*) is plotted in green.*

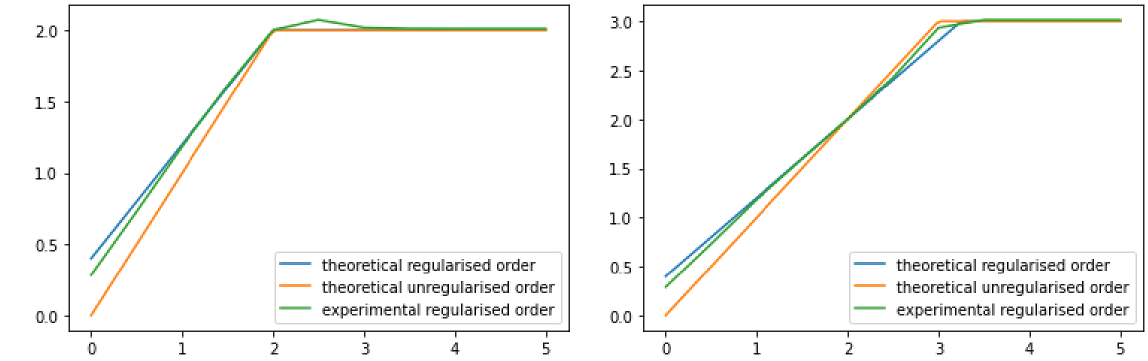

Figure 5.3. Regularization with kernel H: *Plot of* $\lambda \mapsto \gamma_{reg}(\lambda)$ *defined by* (5.1)*, with regularization through the kernel* H*. The left panel shows the case of* $\mathbb{P}_1$ *finite elements* ($\lambda_M = 2$)*; the right panel shows that of* $\mathbb{P}_2$ *finite elements* ($\lambda_M = \frac{13}{4}$)*. In both cases, the theoretical curves without regularization* $\lambda \mapsto \min(\lambda, s_a)$ ($\lambda \geqslant 0$) *and with regularization* $\lambda \mapsto \frac{2}{5}(2\lambda+1)$ ($0 \leqslant \lambda < \lambda_M$) *are plotted in orange and blue, respectively. The numerically obtained curve (with regularization through* H*) is plotted in green.*

As Figure 5.4 shows, both approximations are almost undistinguishable and equally fail to successfully approximate f where the function is close to 0. On other portions, both approximations perform much better, with the regularized version almost always slightly worse, because of an offset to the right.

Appendix A. A proof of the estimate (H_r). We here prove the estimate (H_r). We let $s \geqslant s_r$.

Lower bound. We start with the easiest part, namely the lower bound, which comes from the assumption that K does not reproduce one moment of order s_r, which we denote by $P(x) = x_1^{r_1} \dots x_d^{r_d}$ with $r_1 + \dots + r_d = s_r$. Without loss of generality, we may assume that

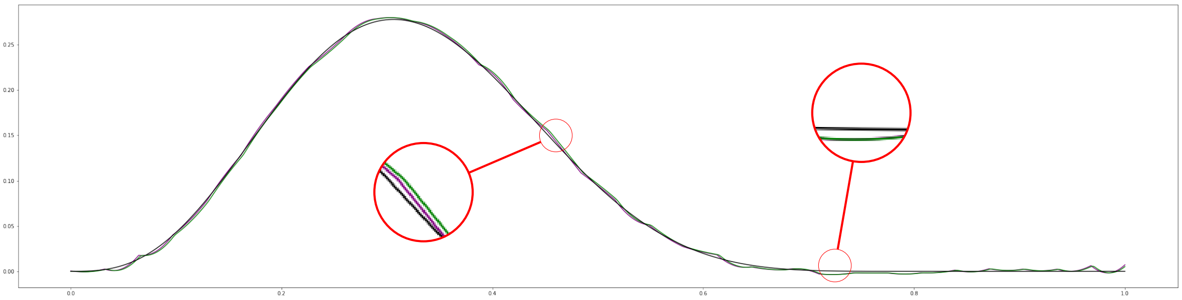

Figure 5.4. *Example of a reconstruction from the function f given by (5.3) (in black), shown against $P_n y_\sigma$ (in purple) and $K_\beta * P_n y_\sigma$ (in green) with $n = 63$, $\beta = \beta^\star(\sigma(n), n)$, $\sigma(n) = n^{-\lambda}$, $\lambda = \frac{3}{2}$. $\mathbb{P}_2$ finite elements are used, and convolution is performed with the kernel K. All functions are evaluated on a very fine grid.*

$0 \in \Omega$ and we consider the function $f = \chi P$ where $\chi \in C_c^\infty(\Omega)$ equals 1 in a neighborhood of 0. Hence we have $f \in H_0^s(\Omega)$.

For β small enough and a sufficiently small neighborhood of 0 (which we denote by Ω_0 and assume to satisfy $\overline{\Omega_0} \subset \Omega$) we may, since K has compact support, ensure that

$$\forall x \in \Omega_0, \quad \mathrm{supp}(K) \subset \beta^{-1}(x - \Omega) \quad \text{and} \quad (K_\beta * f)(x) - f(x) = (K_\beta * P)(x) - P(x).$$

Hence we may write for all $x \in \Omega_0$

$$\begin{aligned}(K_\beta * P)(x) - P(x) &= \int_{\beta^{-1}(x-\Omega)} K(u)P(x - \beta u)\, du - P(x) \\ &= \int_{\mathrm{supp}(K)} K(u)P(x - \beta u)\, du - P(x).\end{aligned}$$

When expanding the product $P(x-\beta u) = (x_1 - \beta u_1)^{r_1} \dots (x_d - \beta u_d)^{r_d}$ and integrating against K, all terms but two vanish since K reproduces moments up to order $s_r - 1$, and we are left with

$$\begin{aligned}(K_\beta * P)(x) - P(x) &= \int_{\mathrm{supp}(K)} K(u)(P(x) + (-1)^{s_r}\beta^{s_r}P(u))\, du - P(x) \\ &= P(x)\left(\int_{\mathrm{supp}(K)} K(u)\, du - 1\right) + (-1)^{s_r}\beta^{s_r}\int_{\mathrm{supp}(K)} K(u)P(u)\, du \\ &= (-1)^{s_r}\beta^{s_r}\int_{\mathrm{supp}(K)} K(u)P(u)\, du,\end{aligned}$$

where the last constant appearing is nonzero by assumption. As a result, we may write

$$\|K_\beta * f - f\| \geqslant \|K_\beta * f - f\|_{L^2(\Omega_0)} = \|K_\beta * P - P\|_{L^2(\Omega_0)} \gtrsim \beta^{s_r}.$$

Upon changing f to $\frac{f}{\|f\|_s}$, we thus have found some $f \in \mathcal{F}_s$ such that $\|K_\beta * f - f\| \gtrsim \beta^{s_r}$, and it follows that

$$\sup_{f \in \mathcal{F}_s} \|K_\beta * f - f\| \gtrsim \beta^{s_r}.$$

Upper bound. Now let $f \in H_0^s(\Omega)$. We start with the case of $\Omega = \mathbb{R}^d$. For $x \in \mathbb{R}^d$, we have

$$(K_\beta * f)(x) - f(x) = \int_{\mathbb{R}^d} (f(x-\beta y) - f(x))\, \beta^{-d} K(\beta^{-1} y)\, dy = \int_{\mathbb{R}^d} (f(x-\beta y) - f(x))\, K(y)\, dy.$$

Since K has $s_r - 1$ vanishing moments, we may replace $f(x)$ by the Taylor polynomial of f of order $s_r - 1$ at the point x, evaluated at $-\beta y$, which we denote by $P_{s_r-1}(x, -\beta y)$. Hence we find

$$(K_\beta * f)(x) - f(x) = \int_{\mathbb{R}^d} (f(x-\beta y) - P_{s_r-1}(x, -\beta y))\, K(y)\, dy.$$

Next, we apply Minkowski's integral inequality (for the Lebesgue measure dx and the measure $|K(y)|\, dy$) to obtain

$$\begin{aligned}
\|K_\beta * f - f\|^2_{L^2(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (f(x-\beta y) - P_{s_r-1}(x, -\beta y))\, K(y)\, dy \right|^2 dx \\
&\leqslant \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-\beta y) - P_{s_r-1}(x, -\beta y)|\, |K(y)|\, dy \right)^2 dx \\
&\leqslant \left(\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-\beta y) - P_{s_r-1}(x, -\beta y)|^2\, dx \right)^{1/2} |K(y)|\, dy \right)^2 \\
&= \left(\int_{\mathbb{R}^d} \|f(\cdot - \beta y) - P_{s_r-1}(\cdot, -\beta y)\|_{L^2(\mathbb{R}^d)}\, |K(y)|\, dy \right)^2.
\end{aligned}$$

We may now use the estimate for the remainder term in the Taylor expansion [dTGCV20], which for a function in $f \in H^s(\mathbb{R}^d)$ reads

$$\forall z \in \mathbb{R}^d, \qquad \|f(\cdot - z) - P_{s_r-1}(\cdot, z)\|_{L^2(\mathbb{R}^d)} \lesssim |z|^s \|f\|_{H^s(\mathbb{R}^d)}.$$

We end up with

$$\|K_\beta * f - f\|_{L^2(\mathbb{R}^d)} \lesssim \beta^s \left(\int_{\mathbb{R}^d} |y|^s |K(y)|\, dy \right) \|f\|^2_{H^s(\mathbb{R}^d)} \lesssim \beta^s \|f\|_{H^s(\mathbb{R}^d)} \lesssim \beta^{s_r} \|f\|_{H^s(\mathbb{R}^d)},$$

where we used $s \geqslant s_r$. The result is proved for $\Omega = \mathbb{R}^d$. It remains to consider the case where Ω is a smooth domain. For $f \in H_0^s(\Omega)$, its extension $\tilde{f}$ by 0 to the whole of $\mathbb{R}^d$ satisfies $\tilde{f} \in H^s(\mathbb{R}^d)$, and the extension mapping $f \mapsto \tilde{f}$ is continuous from $H_0^s(\Omega)$ to $H^s(\mathbb{R}^d)$ since $s - \frac{1}{2}$ is not an integer, by Theorem 11.4 of [LM12]. Hence, $\|\tilde{f}\|_{H^s(\mathbb{R}^d)} \lesssim \|f\|_s$ and we may use the above estimate to obtain

$$\|K_\beta * \tilde{f} - \tilde{f}\|_{L^2(\mathbb{R}^d)} \lesssim \|\tilde{f}\|_{H^s(\mathbb{R}^d)}\, \beta^{s_r} \lesssim \|f\|_s\, \beta^{s_r}.$$

This in turn leads to a bound for the error between $K_\beta * f$ and f in $L^2(\Omega)$,

$$\|K_\beta * f - f\| \leqslant \|K_\beta * \tilde{f} - \tilde{f}\|_{L^2(\mathbb{R}^d)} \lesssim \|f\|_s\, \beta^{s_r},$$

and concludes the proof.

Appendix B. Further estimates.

Lemma B.1. *Assume that* (2.3) *holds and that* $K \geqslant 0$*. Then if* $\beta = \beta(h) = o(h)$*, there holds* $\|K_\beta * \phi_i\| \gtrsim h^{d/2}$.

Proof. Let $0 < r < c$ be fixed with c given by (2.3). We also pick $M > 0$ such that $K(z) = 0$ for $|z| > M$. Let us evaluate $(K_\beta * \phi_i)(x)$ for $x \in B(x_i, rh) \cap \Omega$. For any $y \in \Omega$, we shall prove that $K_\beta(x-y) > 0 \implies y \in B(x_i, ch)$. Indeed, the first condition imposes $|y - x| \leqslant M\beta$; hence $|y - x_i| \leqslant |y - x| + |x - x_i| \leqslant M\beta + rh \leqslant ch$ for h small enough since $\beta = o(h)$. Owing to $K \geqslant 0$, this allows us to write for $x \in B(x_i, rh)$

$$(K_\beta * \phi_i)(x) = \int_\Omega K_\beta(x-y)\phi_i(y)\,dy \geqslant m \int_\Omega K_\beta(x-y)\,dy = m \int_{\beta^{-1}(x-\Omega)} K(z)\,dz.$$

For a given $x \in B(x_i, rh)$, $\beta^{-1}(x - \Omega)$ contains a ball of the form $\{z \in \mathbb{R}^d, |z| \leqslant \varepsilon\beta^{-1}h\}$ for some ε small enough, and since $1 = o(\beta^{-1}h)$, the latter ball contains the support of K for h small enough, leading to

$$(K_\beta * \phi_i)(x) = m \int_{\beta^{-1}(x-\Omega)} K(z)\,dz \gtrsim m \int_{\mathbb{R}^d} K(z)\,dz \gtrsim 1,$$

where $\gtrsim$ is uniform with respect to $x \in B(x_i, rh)$. We conclude that

$$\|K_\beta * \phi_i\|^2 \geqslant \int_{B(x_i, rh)} |K_\beta * \phi_i(x)|^2\,dx \gtrsim |B(x_i, rh)| \sim h^d. \qquad \blacksquare$$

Lemma B.2. *Under the assumptions of Lemma* B.1*, if* $\beta(h) = o(h)$*, we have*

$$e(\beta(h), \sigma, h) \gtrsim \sigma.$$

Proof. This is a direct consequence of Lemma B.1, since one then has

$$e(\beta, \sigma, h)^2 \geqslant \sigma^2 \sum_{i=1}^n \|K_\beta * \phi_i\|^2 \gtrsim \sigma^2 \sum_{i=1}^n h^d = \sigma^2 n h^d \sim \sigma^2. \qquad \blacksquare$$

REFERENCES

[BS77] J. H. Bramble and A. H. Schatz, *Higher order local accuracy by averaging in the finite element method*, Math. Comp., 31 (1977), pp. 94–111, https://doi.org/10.1090/S0025-5718-1977-0431744-9.

[CLSS03] B. Cockburn, M. Luskin, C.-W. Shu, and E. Süli, *Enhanced accuracy by post-processing for finite element methods for hyperbolic equations*, Math. Comp., 72 (2003), pp. 577–606, https://doi.org/10.1090/S0025-5718-02-01464-3.

[DL93] R. A. DeVore and G. G. Lorentz, *Constructive Approximation*, Grundlehren Math. Wiss. 303, Springer, 1993.

[DNPV12] E. Di Nezza, G. Palatucci, and E. Valdinoci, *Hitchhiker's guide to the fractional Sobolev spaces*, Bull. Sci. Math., 136 (2012), pp. 521–573, https://doi.org/10.1016/j.bulsci.2011.12.004.

[dTGCV20] F. del Teso, D. Gómez-Castro, and J. L. Vázquez, *Estimates on translations and Taylor expansions in fractional Sobolev spaces*, Nonlinear Anal., 200 (2020), 111995, https://doi.org/10.1016/j.na.2020.111995.

[EHN96] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. 375, Springer, 1996.

[FOC21] E. Febrianto, M. Ortiz, and F. Cirak, *Mollified finite element approximants of arbitrary order and smoothness*, Comput. Methods Appl. Mech. Engrg., 373 (2021), 113513, https://doi.org/10.1016/j.cma.2020.113513.

[GKK+02] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer, 2002.

[GN90] G. K. Golubev and M. Nussbaum, *A risk bound in Sobolev class regression*, Ann. Statist., 18 (1990), pp. 758–778, https://doi.org/10.1214/aos/1176347624.

[K+11] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Appl. Math. Sci. 120, Springer, 2011.

[LHL+04] W. K. Liu, W. Han, H. Lu, S. Li, and J. Cao, *Reproducing kernel element method. Part* I*: Theoretical formulation*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 933–951.

[LL07] S. Li and W. K. Liu. *Meshfree Particle Methods*, Springer, 2007.

[LLH+04] S. Li, H. Lu, W. Han, W. K. Liu, and D. C. Simkins, *Reproducing kernel element method part* II*: Globally conforming I^m/C^n hierarchies*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 953–987.

[LLSJ+04] H. Lu, S. Li, D. C. Simkins, Jr, W. K. Liu, and J. Cao, *Reproducing kernel element method part* III*: Generalized enrichment and applications*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 989–1011.

[LM12] J. L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications, Volume* 1, Grundlehren Math. Wiss. 181, Springer, 2012.

[LvGM96] G. G. Lorentz, M. von Golitschek, and Y. Makovoz. *Constructive Approximation: Advanced Problems*, Grundlehren Math. Wiss. 304, Springer, 1996.

[ML78] M. S. Mock and P. D. Lax, *The computation of discontinuous solutions of linear hyperbolic equations*, Comm. Pure Appl. Math., 31 (1978), pp. 423–430.

[QQ09] A. Quarteroni and S. Quarteroni, *Numerical Models for Differential Problems*, Springer, 2009.

[SJLLL04] D. C. Simkins, Jr., S. Li, H. Lu, and W. K. Liu, *Reproducing kernel element method. Part* IV*: Globally compatible C^n $(n \geqslant 1)$ triangular hierarchy*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1013–1034.

[Tho77] V. Thomée, *High order local approximations to derivatives in the finite element method*, Math. Comp., 31 (1977), pp. 652–660, https://doi.org/10.1090/S0025-5718-1977-0438664-4.

[Tho07] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer Ser. Comput. Math. 25, Springer, 2007.

[Tsy08] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer, 2008.