Published in Statistica Sinica
Journal version available at https://doi.org/10.5705/ss.202019.0346

Some first results on the consistency of spatial regression with partial differential equation regularization

Eleonora Arnone[1], Alois Kneip[2], Fabio Nobile[3], Laura M. Sangalli[1]

July 8, 2020

[1] MOX– Dipartimento di Matematica, Politecnico di Milano
[2] Universität Bonn
[3] École polytechnique fédérale de Lausanne

Keywords: Functional Data Analysis, Smoothing, Spatial Statistics.

Abstract

We study the consistency of the estimator in spatial regression with partial differential equation (PDE) regularization. This new smoothing technique allows to accurately estimate spatial fields over complex two-dimensional domains, starting from noisy observations; the regularizing term involves a PDE that formalizes problem specific information about the phenomenon at hand. Differently from classical smoothing methods, the solution of the infinite-dimensional estimation problem cannot be computed analytically. An approximation is obtained via the finite element method, considering a suitable triangulation of the spatial domain. We first consider the consistency of the estimator in the infinite-dimensional setting. We then study the consistency of the finite element estimator, resulting from the approximated PDE. We study the bias and variance of the estimators, with respect to the sample size and to the value of the smoothing parameter. Some final simulation studies provide numerical evidence of the rates derived for the bias, variance and mean square error.

1 Introduction

In this work we study the consistency of Spatial Regression with Partial Differential Equation regularization (SR-PDE) [see, e.g., 3, 4]. This regularized least-square method defines a new class of bivariate smoothers that has a number of important advantages with respect to classical smoothers, such as smoothing splines and thin-plate-splines, whose properties have been thoroughly studied in a well established literature [see, e.g., 15, and references therein]. The regularizing term in SR-PDE enables the inclusion of problem-specific information, appropriately formalized in terms of a partial differential equation (PDE), that describes to some extent the phenomenon under study. PDEs are a very powerful tool to model complex phenomena behaviors, and they are extensively used in most fields of sciences and engineering. This makes SR-PDE broadly

applicable to the analysis of spatially distributed data in varied contexts [see, e.g., the applications in 3, 5]. In particular, the regularizing term in SR-PDE can include general linear second-order PDEs, involving space-varying second, first and zero order differential operators (instead of the simple differential operators, constant over space, typically considered by classical smoothers), as well as space-varying forcing terms. This highly flexible and rich modelling of the space variation enables the analysis of a huge variety of anisotropic and non-stationary phenomena. Furthermore, SR-PDE efficiently handles data distributed over domains having complex shapes, such as domains with strong concavities or holes [see, e.g., 30, 31]; this is a crucial feature whenever the shape of the domain influences the problem at hand. Another important advantage of SR-PDE over the classical smoothers is the possibility of imposing conditions that the field must satisfy at the boundaries of the domain of interest, concerning the value of the field and/or of its normal derivative [31, 4, 3]; this feature is fundamental in many applications to obtain meaningful estimates [see, e.g., 3, 31].

Such high flexibility comes at the price of a higher analytic complexity of these smoothers. The solution to the estimation problem cannot be computed analytically, but can only be characterized in a variational form. An approximated solution can be obtained via a mixed finite element approach, after introducing a suitable triangulation of the spatial domain of interest.

Unfortunately, when considering the consistency of such estimators, because of the unavailability of an explicit closed form solution of the infinite-dimensional estimation problem, it is not possible to leverage on the arguments used to prove the consistency of thin-plate-splines and of smoothing splines [see, e.g., 9, 10, 11, 19, 22].

It should be pointed out that the problem of analyzing data spatially distributed over irregularly shaped two-dimensional domains has recently attracted an increasing interest, and other regularized least-square smoothers have been proposed that can tackle this issue, such as bivariate splines over triangulations [see, e.g., 24, 18, 13, 25], soap film smoothing [36], and low-rank thin-plate spline approximations [34, 32]. All these methods have isotropic and stationary regularizing terms; bivariate splines over triangulations can include high order derivatives. With the exception of soap film smoothing, that can comply with some simple types of boundary conditions, the remaining methods do not possess this ability. The asymptotic properties of bivariate splines over triangulations are investigated in [25], but differently from SR-PDE, the finite-dimensional estimator, based on bivariate splines, is directly considered. To the best of our knowledge, no results on large sample properties is available for any of the other methods.

We here aim at proving the consistency of SR-PDE estimators, addressing both the case of the estimator solution of the infinite-dimensional estimation problem and the case of the finite element estimator. The paper is organized as follows. Section 2 briefly reviews the SR-PDE infinite-dimensional estimation problem, introduced in [4] and [3], and Section 4 outlines its discretized version. Sections 3 and 5 contain the main contributions of this work. In particular, in Section 3, we study the bias of the infinite-dimensional estimator in the L^2 and H^2 spatial norms; we moreover investigate the convergence of the variance of the infinite-dimensional SR-PDE estimator, when n goes to infinity. Thanks to the rates obtained for the bias and the variance, we can prove the consistency of the estimator. Furthermore, we show that the Mean Square Error (MSE) of the estimator achieves near-optimal rates of convergence. In Section 5 we then focus on the finite element estimator. Unfortunately, because of the non-conforming discretization approach used, it is not possible to derive the consistency of the finite element estimator from the one of the infinite-dimensional estimator. Nev-

ertheless, we are able to prove the consistency of the finite element estimator, in the discrete ℓ^2 norm over the data locations, when the triangulation vertices coincide with the data locations, under some simplifying hypotheses. Moreover, we show that the finite element estimator achieves optimal rates of convergence. Section 6 provides numerical evidence for the convergence rates obtained in the previous sections. Finally, Section 7 outlines future research directions.

2 Spatial regression with partial differential equation regularization: infinite-dimensional estimation problem

We briefly review the SR-PDE infinite-dimensional estimation problem, introduced in [4] and [3]. Consider a bounded domain $\Omega \subset \mathbb{R}^2$, with boundary $\partial\Omega \in C^2(\mathbb{R}^2)$. Consider n observations $z_i \in \mathbb{R}$, for $i = 1, \dots, n$, located at points $\mathbf{p}_i = (x_i, y_i) \in \Omega$. Assume that:

$$z_i = f_0(\mathbf{p}_i) + \varepsilon_i$$

where $f_0 : \Omega \to \mathbb{R}$ is the field we wish to estimate, and ε_i are independent errors with zero mean and finite variance σ^2. Assume that partial problem-specific information is available, which can be formalized in terms of a PDE $Lf = u$ modeling to some extents the phenomenon under study. Specifically, relying on the problem-specific information, we can assume that the misfit $\|Lf_0 - u\|_{L^2}$ is small, though we do not require it to be 0. Here, L is a general linear second order differential operator, that can for instance include second, first and zero order differential terms:

$$L(\mathbf{p})f = -\text{div}\left(\mathbf{K}(\mathbf{p})\nabla f\right) + \mathbf{b}(\mathbf{p}) \cdot \nabla f + c(\mathbf{p})f$$

where $\mathbf{K}(\cdot) : \Omega \to \mathbb{R}^{2\times 2}$ is a space-varying symmetric and positive-definite diffusion tensor, $\mathbf{b}(\cdot) : \Omega \to \mathbb{R}^2$ is a space-varying transport vector, and $c(\cdot) : \Omega \to \mathbb{R}_+$ is a space-varying reaction coefficient. The forcing term $u(\cdot) \in L^2(\Omega)$ can either be the null function $u = 0$ (so-called homogeneous case), or $u \neq 0$ (non-homogeneous case). Assume that the problem-specific knowledge concerns as well the behavior of the field f_0 at the boundary of the domain. Various types of boundary conditions may be considered, involving the value of the field, and/or of its normal derivatives, at the boundary $\partial\Omega$ of the domain of interest. In this work, we shall focus on Dirichlet boundary conditions. Specifically, we assume to know the value of the field at the boundary: $f_0 \,|_{\partial\Omega} = \gamma$, where $\gamma(\cdot)$ can either be the null function $\gamma = 0$ (homogeneous condition) or $\gamma \neq 0$ (non-homogeneous condition).

Denote by $H^k(\Omega)$ the Sobolev space of functions in $L^2(\Omega)$ with derivatives up to the k-th order in $L^2(\Omega)$, equipped with the norm $\|v\|_{H^k} = (\sum_{|\alpha|\leq k} \|D^\alpha v\|^2_{L^2})^{\frac{1}{2}}$. Define the affine space

$$V_\gamma = \{v \in H^2(\Omega) : v|_{\partial\Omega} = \gamma\}.$$

SR-PDE solves the following estimation problem:

$$\hat{f} = \underset{f \in V_\gamma}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{p}_i) - z_i)^2 + \lambda_n \int_\Omega (Lf - u)^2 . \tag{1}$$

The estimation functional in (1) trades off a data fidelity criterion, the sum of square errors, and a model fidelity criterion, the differential regularization, defined as the L^2-norm, over the spatial domain of interest, of the misfit with respect to the governing PDE. The smoothing parameter $\lambda_n > 0$ controls the relative weight of these two criteria.

The methodology is very flexible. The three terms in the differential operator L enable the modeling of various forms of anisotropy and non-stationarity in the field. The diffusion term $-\text{div}(K\nabla f)$ induces a smoothing in all the directions. If the diffusion matrix K is a multiple of the identity matrix I, the diffusion term has an isotropic smoothing effect, otherwise it implies an anisotropic smoothing with a preferential direction that corresponds to the first eigenvector of the diffusion tensor K. The degree of anisotropy induced by the diffusion tensor K is controlled by the ratio between its first and second eigenvalue. The transport term $\mathbf{b}\cdot\nabla f$ induces a smoothing only in the direction specified by the transport vector $\mathbf{b}$, with an intensity that depends on the length of $\mathbf{b}$. The reaction term cf has a shrinkage effect, since penalization of the L^2 norm of f induces a shrinkage of the field to zero. Moreover, since K, $\mathbf{b}$ and c can vary over space, the effects here described are non-stationary. Such flexibility is even further increased by the presence of possibly non-homogeneous forcing terms $u \in L^2(\Omega)$. SR-PDE can be seen as an extension of the more classical smoothing techniques to the anisotropic and non-stationary case. In particular, it includes as a special case the isotropic and stationary regularization of the Laplacian of the field considered in [30] and [31], when no problem-specific information is available (setting $\mathbf{K}$ to the the identity matrix, $\mathbf{b} = \mathbf{0}$, $c = 0$, so that $L = \Delta$, and a null forcing term $u = 0$).

It should be stressed that, in the estimation problem (1), the estimator $\hat{f}$ is searched in a general Sobolev space of functions (with boundary conditions). Specifically, the search is not restricted to the space of the solutions of the differential equation $Lf = u$. Indeed, as described above, we do not assume that the true f_0 satisfies the PDE in the regularization. Rather, we assume that PDE carries partial information about the true f_0, and hence we use the PDE to regularize the estimate. As a consequence, we are not interested in searching for the solution of the PDE that is closest to the data. In fact, in the following sections, we study the asymptotic properties of the estimators letting the smoothing parameter λ_n go to zero when n goes to infinity. That is, the influence of the regularizing term decreases as n increases. This is a natural setting to consider for the models here considered, since, the more the observations, the less the need to regularize the estimate.

2.1 Illustrative problem

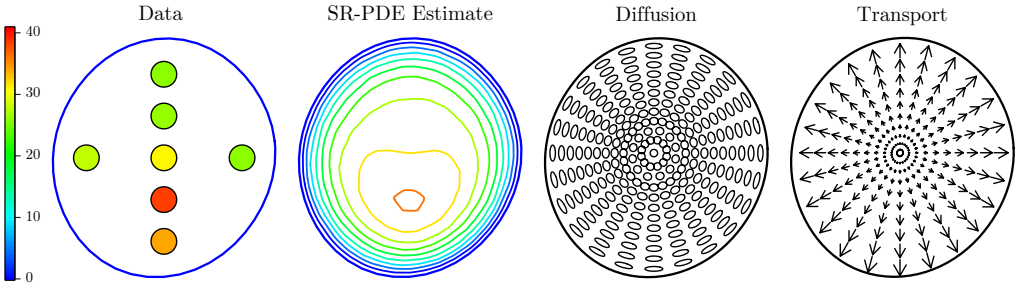


Figure 1: From left to right: ECD data on the artery cross-section; corresponding estimate of the blood-flow velocity field obtained by SR-PDE; non-stationary anisotropic diffusion tensor field K used in SR-PDE estimate; non-stationary transport field $\mathbf{b}$ used in SR-PDE estimate.

As an illustrative example, [3] considers the problem of estimating the blood-flow velocity field in a cross-section of a human carotid artery, starting from eco-color

doppler data and magnetic resonance imaging data. Figure 1, top row, left panel, shows the reconstruction of the cross-section of the common carotid artery of one of the patients in the study; this quasi-circular section is obtained from segmentation of magnetic resonance imaging data. The same figure also displays the spatial locations of seven beams where the blood-flow velocity is measured, via eco-color doppler acquisitions. The color of the beam refers to the mean blood velocity measured over the beam at systolic peak. Starting from the observations over the seven beams, the field f_0 of the blood flow velocity at systolic peak, over the entire cross-section of the carotid, must be estimated. In this applied problem, there are known conditions that the field must satisfy at the boundaries of the domain of interest, i.e., at the arterial wall. In fact, the physics of the problem implies that blood flow velocity is zero at the arterial wall, due to the friction between blood cells and the arterial wall. The influence of the shape of the domain, and the presence of specific boundary conditions, hinder the applicability of classical smoothing methods, and more generally, of classical methods for spatial data analysis. In fact, as mentioned in the Introduction, classical techniques for field estimation naturally work on tensorized domains and do not accurately deal with bounded domains, when the shape of the domain is important for the behavior of the phenomenon under study; moreover, classical techniques cannot naturally comply with specific conditions at the boundary of the domain of interest, such as those here specified. Furthermore, [4] shows that isotropic and stationary smoothers return non-physiological estimates of the blood flow velocity field, due to the cross shaped pattern of the observations. On the other hand, we can here profit of a detailed problem-specific information about the phenomenon under study. There is in fact a vast literature devoted to the study of fluid dynamics and hemodynamics; see, e.g., [17]. This information can be conveniently translated into a PDE, that describes, in a idealized setting, the main features of the velocity field. In particular, as detailed in [3], we can here consider the operator L that includes a nonstationary anisotropic diffusion tensor that smooths the observations in the tangential direction of concentric circles (see Figure 1, third panel) and a nonstationary transport field that smooths the observations in the radial direction, from the center of the section to the boundary (Figure 1, forth panel); the reaction term and the forcing term are instead not required by this application. The second panel of Figure 1, displays the corresponding estimate of the blood-flow velocity field. Suitably incorporating the problem-specific information on the phenomenon under study, SR-PDE returns a realistic estimate of the blood flow, that is not affected by the cross-shaped pattern of the observations and displays physiological and smooth isolines.

2.2 Well-posedness of SR-PDE estimation problem and linearity of the estimator

[3] and [4] show that, under regularity conditions on L, u and γ, the estimator $\hat{f}$ in (1) is unique. The required regularity conditions on the parameters are such that the operator L has so-called H^2-smoothing properties, i.e., L is such that for all $u \in L^2(\Omega)$ the solution of the problem $Lf = u$ with Dirichlet boundary conditions belongs to $H^2(\Omega)$.

Solving (1) is equivalent to finding $\hat{f}$ such that

$$\lambda \int_\Omega (L\hat{f} - u)Lv + \frac{1}{n}\sum_{i=1}^{n} \hat{f}(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^{n} z_i v(\mathbf{p}_i) \qquad \forall v \in V_0. \tag{2}$$

This formulation highlights the linearity of the estimator $\hat{f}$ in the observations z_i. Thanks

to the linearity of equation (2) in $\hat{f}$, we can write $\hat{f}$ as

$$\hat{f} = \hat{f}^* + \hat{w}$$

where $\hat{f}^*$ and $\hat{w}$ solve, respectively,

$$\lambda \int_\Omega (L\hat{f}^* - u)Lv + \frac{1}{n}\sum_{i=1}^n \hat{f}^*(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^n f_0(\mathbf{p}_i)v(\mathbf{p}_i) \qquad \forall v \in V_0$$

$$\lambda \int_\Omega L\hat{w}Lv + \frac{1}{n}\sum_{i=1}^n \hat{w}(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^n \varepsilon_i v(\mathbf{p}_i) \qquad \forall v \in V_0;$$

equivalently,

$$\hat{f}^* = \operatorname*{argmin}_{f \in V_\gamma} \left[\frac{1}{n}\sum_{i=1}^n (f(\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda_n \int_\Omega (Lf - u)^2 \right] \tag{3}$$

$$\hat{w} = \operatorname*{argmin}_{w \in V_0} \left[\frac{1}{n}\sum_{i=1}^n (w(\mathbf{p}_i) - \varepsilon_i)^2 + \lambda_n \int_\Omega (Lw)^2 \right]. \tag{4}$$

The minimization problem (3) involves the true values of the field without any observational noise, $f_0(\mathbf{p}_i)$, and the non-homogeneous regularization term (i.e., with forcing term u) with non-homogeneous boundary conditions. The minimization problem (4) instead involves pure noise data, ε_i, and an homogeneous regularization term (i.e, with no forcing term) with homogeneous boundary conditions. The minimizer $\hat{f}^*$ is deterministic, while the minimizer $\hat{w}$ is such that $\mathbf{E}[\hat{w}] = 0$. Consequently we have that: $\mathbb{E}[\hat{f}] = \hat{f}^*$ and $\mathbb{V}\mathrm{ar}(\hat{f}) = \mathbb{V}\mathrm{ar}(\hat{w})$. Thanks to this fact, we can split the analysis of the bias and the variance of the estimator: when studying the bias we can focus on the minimization problem (3), while when studying the variance we focus on the minimization problem (4).

3 Consistency of SR-PDE estimator: infinite-dimensional problem

As in [9, 10], in order to prove the consistency of the estimator, we make some assumptions on how the points $\mathbf{p}_i$ fill in the domain Ω as n goes to infinity. Denote by $F_n(\mathbf{p})$ the bivariate cumulative distribution function of the probability measure that assigns mass n^{-1} to each point $\mathbf{p}_i$. Let F be the limiting distribution of the sequence $\{F_n\}$. Define $d_n = \sup_{\mathbf{p}\in\Omega} |F(\mathbf{p}) - F_n(\mathbf{p})|$. Note that when $\Omega = [0,1]^d$ and F is the uniform measure, d_n is the so-called star-discrepancy [see, e.g., 28].

Assumption 1. *The sequence $\{F_n\}$ converges uniformly to a cumulative distribution function F having density $\mathsf{f} \in C^\infty(\bar{\Omega})$, with respect to the d-dimensional Lebesgue measure, such that, for all $\mathbf{p} \in \Omega$, $0 < \kappa_1 \leq \mathsf{f}(\mathbf{p}) \leq \kappa_2 < \infty$, for some constants κ_1, κ_2.*

Assumption 2. *λ_n is such that $\lim_{n\to\infty} d_n \lambda_n^{-1} = \lim_{n\to\infty} \lambda_n = 0$.*

The following result holds [see 10].

Lemma 1. *Under Assumption 1, if $\partial\Omega \in C^2$, for all $h, g \in H^2(\Omega)$ there exists a constant $c > 0$ such that*

$$\left| \int_\Omega hg\mathsf{f}d\mathbf{p} - \frac{1}{n}\sum_{i=1}^n h(\mathbf{p}_i)g(\mathbf{p}_i) \right| = \left| \int_\Omega hgd(F - F_n) \right| \leq c d_n \|h\|_{H^2} \|g\|_{H^2}. \tag{5}$$

The proof of this Lemma is rather involved. It is based on results from functional analysis and measure theory. We refer the reader to the original paper for the details.

3.1 Convergence of the bias term: infinite-dimensional estimator

In this section we study the bias of the estimator

$$\mathscr{B} = f_0 - \mathbb{E}\left[\hat{f}\right] = f_0 - \hat{f}^*$$

with respect to the number of observations n and the smoothing parameter λ_n. Theorem 1 gives the rates for the bias, when f_0 has different Sobolev regularities, $f_0 \in H^2(\Omega)$ and $f_0 \in H^4(\Omega)$. In the proof of the theorem we make use of fractional Sobolev spaces $H^\theta(\Omega)$, with non-integer $\theta > 0$; the space $H^\theta(\Omega)$ can be defined as the interpolation space between $H^k(\Omega)$ and $L^2(\Omega)$ with k an integer larger than θ [see, e.g., 26]. Moreover, we consider the L^*, the adjoint operator of L, defined as

$$L^*\hat{g} = -\mathrm{div}(\mathbf{K}\nabla\hat{g}) - \mathbf{b}\cdot\nabla\hat{g} + (c - \mathrm{div}(\mathbf{b}))\hat{g}. \tag{6}$$

Theorem 1. *Under Assumptions 1 and 2, for n sufficiently large, if $f_0 \in H^2(\Omega)$ and $f_0\,|_{\partial\Omega} = \gamma$, then*

$$\|\mathscr{B}\|_{L^2} \leq C\sqrt{\lambda_n}, \tag{7}$$

with C independent of n and λ_n. Moreover, if $Lf_0 - u \in H^2(\Omega)$, then

$$\|\mathscr{B}\|_{L^2} = O(\lambda_n^{5/8}) \quad \text{and} \quad \|\mathscr{B}\|_{H^2} = O(\lambda_n^{1/8}). \tag{8}$$

Finally, if in addition $(Lf_0 - u)|_{\partial\Omega} = 0$, then

$$\|\mathscr{B}\|_{L^2} = O(\lambda_n) \quad \text{and} \quad \|\mathscr{B}\|_{H^2} = O(\sqrt{\lambda_n}). \tag{9}$$

Proof. To lighten the notation we write $\lambda = \lambda_n$. Solving the minimization problem 3 is equivalent to finding $\hat{f}^*$ such that

$$\lambda \int_\Omega (L\hat{f}^* - u)Lv + \frac{1}{n}\sum_{i=1}^n \hat{f}^*(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^n f_0(\mathbf{p}_i)v(\mathbf{p}_i) \qquad \forall v \in V_0. \tag{10}$$

Let us rewrite equation (10) in terms of $\mathscr{B}$. To this aim we subtract the quantity $\lambda \int Lf_0 Lv$ on both sides of (10), getting

$$\lambda \int_\Omega L\mathscr{B}Lv = \lambda \int_\Omega (Lf_0 - u)Lv - \frac{1}{n}\sum_{i=1}^n \mathscr{B}(\mathbf{p}_i)v(\mathbf{p}_i),$$

then we add $\int \mathscr{B}v\,dF$ on both sides obtaining

$$\lambda \int_\Omega L\mathscr{B}Lv + \int_\Omega \mathscr{B}v\,dF = \lambda \int_\Omega (Lf_0 - u)Lv + \int_\Omega \mathscr{B}v\,dF - \frac{1}{n}\sum_{i=1}^n \mathscr{B}(\mathbf{p}_i)v(\mathbf{p}_i).$$

The equation above holds for all $v \in V_0$; in particular we can set $v = \mathscr{B}$. Thanks to Assumption 1, we hence get:

$$\lambda \|L\mathscr{B}\|_{L^2}^2 + \kappa_1 \|\mathscr{B}\|_{L^2}^2 \leq \lambda \int_\Omega (Lf_0 - u)L\mathscr{B} + \left\{\int_\Omega \mathscr{B}^2 d(F - F_n)\right\}. \tag{11}$$

Now, thanks to (5), we can write $\left|\int_\Omega \mathscr{B}^2 d(F - F_n)\right| \leq cd_n \|\mathscr{B}\|_{H^2}^2$. Moreover, thanks to the H^2-regularity, we have that the norm $\|Lv\|_{L^2(\Omega)}$ is equivalent to the norm $\|v\|_{H^2(\Omega)}$

for any $v \in V_0$ [see, e.g., 16]. Since $b \in V_0$, being $f_0 \in V_\gamma$, there exists a constant c_L, depending only on Ω and L, such that $c_L \|\mathscr{B}\|_{H^2}^2 \leq \|L\mathscr{B}\|_{L^2}^2$. Using these two inequalities in (11), we obtain

$$c_L\lambda \|\mathscr{B}\|_{H^2}^2 + \kappa_1 \|\mathscr{B}\|_{L^2}^2 \leq \lambda \int_\Omega (Lf_0 - u)L\mathscr{B} + cd_n \|\mathscr{B}\|_{H^2}^2 .$$

Thanks to Assumption 2, for n large enough that $d_n\lambda^{-1} \leq \frac{c_L}{2c}$, we get

$$\frac{c_L\lambda}{2} \|\mathscr{B}\|_{H^2}^2 + \kappa_1 \|\mathscr{B}\|_{L^2}^2 \leq \lambda \int_\Omega (Lf_0 - u)L\mathscr{B}. \tag{12}$$

Moreover,

$$\begin{aligned} \lambda \int_\Omega (Lf_0 - u)L\mathscr{B} &\leq \frac{\lambda}{2}\left(\frac{2}{c_L}\|Lf_0 - u\|_{L^2}^2 + \frac{c_L}{2}\|L\mathscr{B}\|_{L^2}^2\right) \\ &\leq \frac{\lambda}{c_L}\|Lf_0 - u\|_{L^2}^2 + \frac{\lambda c_L}{4}\|\mathscr{B}\|_{H^2}^2 . \end{aligned} \tag{13}$$

The equation above, together with equation (12), leads to (7).

From (12), if $Lf_0 - u \in H^2(\Omega)$, using, in order, the first Green's identity, Hölder inequality, trace theorems, interpolation between Sobolev spaces (see [26]), and Young inequality, we get:

$$\frac{c_L\lambda}{2} \|\mathscr{B}\|_{H^2}^2 + \kappa_1 \|\mathscr{B}\|_{L^2}^2 \leq \lambda \int_\Omega L^*(Lf_0 - u)\mathscr{B} + \lambda \int_{\partial\Omega} (\mathbf{K}\nabla\mathscr{B})\cdot n(Lf_0 - u) \tag{14}$$

$$\begin{aligned} &\leq \lambda \|L^*(Lf_0 - u)\|_{L^2} \|\mathscr{B}\|_{L^2} + \lambda \|\mathbf{K}\|_{L^\infty(\partial\Omega)} \|\nabla\mathscr{B}\cdot n\|_{L^2(\partial\Omega)} \|Lf_0 - u\|_{L^2(\partial\Omega)} \\ &\leq \lambda \|L^*(Lf_0 - u)\|_{L^2} \|\mathscr{B}\|_{L^2} + c\lambda \|\mathscr{B}\|_{H^{3/2}(\Omega)} \|Lf_0 - u\|_{H^1(\Omega)} \\ &\leq \lambda \|L^*(Lf_0 - u)\|_{L^2} \|\mathscr{B}\|_{L^2} + c\lambda \|\mathscr{B}\|_{L^2}^{1/4} \|\mathscr{B}\|_{H^2}^{3/4} \|Lf_0 - u\|_{H^1(\Omega)} \\ &\leq \frac{\lambda^2}{2\kappa_1}\|L^*(Lf_0 - u)\|_{L^2}^2 + \frac{\kappa_1}{2}\|\mathscr{B}\|_{L^2}^2 + \end{aligned} \tag{15}$$

$$+ \frac{\kappa_1}{8}\|\mathscr{B}\|_{L^2}^2 + \frac{3c_L\lambda}{8}\|\mathscr{B}\|_{H^2}^2 + \frac{c^2\lambda^{5/4}}{2\kappa_1^{1/4}c_L^{3/4}}\|Lf_0 - u\|_{H^1(\Omega)}^2$$

where c is a constant independent of n and λ and that changes from line to line. From (15) we can write

$$\frac{c_L\lambda}{8}\|\mathscr{B}\|_{H^2}^2 + \frac{3\kappa_1}{8}\|\mathscr{B}\|_{L^2}^2 \leq \frac{\lambda^2}{2\kappa_1}\|L^*(Lf_0 - u)\|_{L^2}^2 + \frac{c^2\lambda^{5/4}}{2\kappa_1^{1/4}c_L^{3/4}}\|Lf_0 - u\|_{H^1(\Omega)}^2 .$$

Since $Lf_0 \in H^2(\Omega)$ and $u \in H^2(\Omega)$, both $\|L^*(Lf_0 - u)\|_{L^2}^2$ and $\|Lf_0 - u\|_{H^1(\Omega)}^2$ are finite, thus we get the rates in (8). Finally, when $(Lf_0 - u)|_{\partial\Omega} = 0$, from (14) we get the rates in (9).

□

3.2 Convergence of the variance term: infinite-dimensional estimator

We here study the variance of the estimator $\hat{f}$ with respect to n and λ_n.

Theorem 2. *For all $0 < \delta \leq \frac{1}{2}$ and n sufficiently large*

$$\mathbb{V}ar_{L^2}(\hat{f}) = \mathbb{E}\left(\|\hat{w}\|_{L^2}^2\right) = O\left(\frac{\sigma^2}{n\lambda_n^{1/2+\delta}}\right). \tag{16}$$

with a constant which diverges to $+\infty$ when $\delta \to 0$.

Proof. To lighten the notation we write $\lambda = \lambda_n$. Solving the minimization problem (4) is equivalent to finding $\hat{w}$ such that

$$\lambda \int_\Omega L\hat{w} Lv + \frac{1}{n}\sum_{i=1}^n \hat{w}(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^n \varepsilon_i v(\mathbf{p}_i) \qquad \forall v \in V_0$$

or equivalently

$$\lambda \int_\Omega L\hat{w} Lv + \int_\Omega \hat{w} v\, dF = \frac{1}{n}\sum_{i=1}^n \varepsilon_i v(\mathbf{p}_i) + \int_\Omega \hat{w} v\, d(F - F_n) \qquad \forall v \in V_0 \tag{17}$$

Define the following inner product on V_0

$$(v_1, v_2)_\lambda = \lambda \int_\Omega Lv_1 Lv_2 + \int_\Omega v_1 v_2\, dF$$

which is equivalent to the H^2 inner product, and denote by $\|\cdot\|_\lambda$ the norm induced by this inner product $(\cdot,\cdot)_\lambda$. Since the norms $\|L\cdot\|_{L^2}$ and $\|\cdot\|_{H^2}$ are equivalent on V_0, there exist a constant c_L such that

$$\|v\|_{H^2}^2 \le \frac{1}{c_L}\|Lv\|_{L^2}^2 \le \frac{1}{c_L\lambda}\left(\lambda\|Lv\|_{L^2}^2 + \int_\Omega v^2\, dF\right) = \frac{1}{c_L\lambda}\|v\|_\lambda^2 . \tag{18}$$

Define T, T_1 and T_2 as follows:

$$T_1(v) = \int_\Omega \hat{w} v\, d(F - F_n) \qquad T_2(v) = \frac{1}{n}\sum_{i=1}^n \varepsilon_i v(\mathbf{p}_i) \qquad T(v) = T_1(v) + T_2(v).$$

Thanks to the Sobolev embedding theorems [see, e.g., 26, Theorem 9.8], for each $\delta > 0$, we have $T \in \left(H^{1+2\delta}(\Omega)\right)^*$, where $\left(H^{1+2\delta}(\Omega)\right)^*$ denotes the dual space of $H^{1+2\delta}(\Omega)$. Therefore we can rewrite equation (17) as

$$(\hat{w}, v)_\lambda = T(v) \qquad \forall v \in V_0.$$

We have that

$$\|\hat{w}\|_\lambda = \sup_{v\in V_0}\frac{(\hat{w}, v)_\lambda}{\|v\|_\lambda} = \sup_{v\in V_0}\frac{T(v)}{\|v\|_\lambda} \le \sup_{v\in V_0}\frac{T_1(v)}{\|v\|_\lambda} + \sup_{v\in V_0}\frac{T_2(v)}{\|v\|_\lambda}. \tag{19}$$

For the first term on the right hand side of (19), thanks to equations (5) and (18), we have

$$\sup_{v\in V_0}\frac{T_1(v)}{\|v\|_\lambda} \le c d_n \sup_{v\in V_0}\frac{\|v\|_{H^2}\|\hat{w}\|_{H^2}}{\|v\|_\lambda} \le \tilde{c} d_n \lambda^{-1}\|\hat{w}\|_\lambda . \tag{20}$$

For the second term on the right hand side of (19), setting $\theta = 1 + 2\delta$, so that $T_2 \in \left(H^\theta(\Omega)\right)^*$ for $1 < \theta \le 2$, we have

$$\begin{aligned}
\sup_{v\in V_0}\frac{T_2(v)}{\|v\|_\lambda} &\le \sup_{v\in V_0}\frac{\|T_2\|_{(H^\theta)^*}\|v\|_{H^\theta}}{\|v\|_\lambda} \le c\sup_{v\in V_0}\frac{\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*}\left(\lambda^{\theta/4}\|v\|_{H^2}^{\theta/2}\|v\|_{L^2}^{1-\theta/2}\right)}{\|v\|_\lambda} \\
&\le c\sup_{v\in V_0}\frac{\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*}\left(\frac{\theta}{2}\sqrt{\lambda}\|v\|_{H^2} + \frac{2-\theta}{2}\|v\|_{L^2}\right)}{\|v\|_\lambda} \\
&= c\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*}\sup_{v\in V_0}\frac{\left(\frac{\theta}{2}\sqrt{\lambda}\|v\|_{H^2} + \frac{2-\theta}{2}\|v\|_{L^2}\right)}{\|v\|_\lambda} \\
&\le c\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*}
\end{aligned}$$

where the last inequality is true thanks to equation (18) and to the fact that

$$\|v\|_{L^2}^2 \leq \frac{1}{\kappa_1}\int_\Omega v^2 dF \leq \frac{1}{\kappa_1}\|v\|_\lambda^2 .$$

From equation (20) we have

$$\|\hat{w}\|_\lambda \leq c_1 d_n \lambda^{-1}\|\hat{w}\|_\lambda + c_2\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*} .$$

Moreover, thanks to Assumption 2, we have that $d_n\lambda^{-1} = o(1)$. Therefore, the first part in the right hand side of the above equation can be absorbed in the second term in the right hand sid, so that, for n sufficiently large,

$$\|\hat{w}\|_\lambda \leq c\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*} .$$

By squaring and taking the expected values of both terms of the above inequality, we have

$$\mathbb{E}(\|\hat{w}\|_\lambda^2) \leq c\lambda^{-\theta/2}\mathbb{E}(\|T_2\|_{(H^\theta)^*}^2) \tag{21}$$

To conclude the proof, it remains to show that $\mathbb{E}(\|T_2\|_{(H^\theta)^*}^2) \leq \frac{c\sigma^2}{n}$. From the definition of T_2 we can write

$$T_2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_i \delta_{\mathbf{p}_i}$$

where $\delta_{\mathbf{p}_i}$ is the Dirac delta in $\mathbf{p}_i$. Thanks to Sobolev embedding theorems, $\delta_{\mathbf{p}_i} \in (H^\theta(\Omega))^*$. We denote with $(\cdot,\cdot)_{\theta,*}$ the inner product in $(H^\theta(\Omega))^*$. Recalling that the errors ε_i are uncorrelated, with zero mean and constant variance σ^2, we have

$$\begin{aligned}
\mathbb{E}\left(\|T_2\|_{(H^\theta)^*}^2\right) &= \mathbb{E}\left((T_2,T_2)_{\theta,*}\right) = \mathbb{E}\left(\frac{1}{n^2}\sum_{i,j=1}^n \varepsilon_i\varepsilon_j(\delta_{\mathbf{p}_i},\delta_{\mathbf{p}_j})_{\theta,*}\right) \\
&= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\left(\varepsilon_i\varepsilon_j\right)(\delta_{\mathbf{p}_i},\delta_{\mathbf{p}_j})_{\theta,*} \\
&= \frac{1}{n^2}\sum_{i=1}^n \sigma^2\|\delta_{\mathbf{p}_i}\|_{(H^\theta)^*}^2 \leq \frac{c\sigma^2}{n}
\end{aligned}$$

where $c = \max_{i=1,\ldots,n}\|\delta_{\mathbf{p}_i}\|_{(H^\theta)^*}^2 < \infty$. From the previous equation and from (21), we have

$$\mathbb{E}(\|\hat{w}\|_\lambda^2) \leq \frac{c\lambda^{-\theta/2}\sigma^2}{n}.$$

Finally, thanks to Assumption 1, we have that $\|w\|_{L^2} \leq \kappa_1^{-1}\|w\|_\lambda$, where κ_1 does not depend on λ nor on n. This fact and the above equation lead to (16).

□

3.3 Convergence of the MSE: infinite-dimensional estimator

We finally consider the MSE of the estimator in the L^2 norm, i.e.,

$$\mathrm{MSE}_{L^2}(\hat{f}) = \left\|\mathrm{bias}(\hat{f})\right\|_{L^2}^2 + \mathbb{V}\mathrm{ar}_{L^2}(\hat{f}).$$

The following theorem shows that the estimator $\hat{f}$ is consistent and that its MSE nearly achieves the optimal rate of convergence for non-parametric estimators [33], considering different Sobolev regularities of the true unknown field, $f_0 \in H^2(\Omega)$ and $f_0 \in H^4(\Omega)$. Precisely, the theorem shows that the MSE achieves the optimal rates, but for an infinitesimal δ, as small as desired.

Theorem 3. *If $f_0 \in H^2(\Omega)$ and $f_0 \mid_{\partial\Omega} = \gamma$, setting $\lambda_n = n^{-2/3}$ we have*

$$MSE_{L^2} = O\left(n^{-\frac{2}{3}(1-\delta)}\right) \tag{22}$$

for δ as small as desired. If, in addition, $Lf_0 - u \in H^2(\Omega)$ and $Lf_0 - u|_{\partial\Omega} = 0$, setting $\lambda_n = n^{-2/5}$ we have

$$MSE_{L^2} = O\left(n^{-\frac{4}{5}(1-\delta/2)}\right) \tag{23}$$

for δ as small as desired.

Proof. Thanks to Theorems 1 and 2, if $f_0 \in H^2(\Omega)$ and $f_0 \mid_{\partial\Omega} = \gamma$, we have

$$\mathrm{MSE}_{L^2}(\hat{f}) = O(\lambda_n) + O\left(\frac{\sigma^2}{n\lambda_n^{1/2+\delta}}\right)$$

that is minimized when $\lambda_n = n^{-2/3}$, leading to (22). Moreover, thanks to equation (9), if $Lf_0 - u \in H^2(\Omega)$ and $Lf_0 - u|_{\partial\Omega} = 0$, we have

$$\mathrm{MSE}_{L^2}(\hat{f}) = O\left(\lambda_n^2\right) + O\left(\frac{\sigma^2}{n\lambda_n^{1/2+\delta}}\right)$$

that is minimized when $\lambda_n = n^{-2/5}$, leading to (23). □

Remark 1. *As highlighted in Section 2, we do not assume that the PDE in the regularizing term describes perfectly the phenomenon under study. Hence, we do not assume that the true f_0 is a solution of PDE. On the other hand, if our problem knowledge was indeed complete, and the PDE in the regularizing term offered a perfect description of the unknown field, being $Lf_0 = u$, we would expect to gain both in terms of estimation error and in terms of rate of convergence of the MSE. Indeed, from equations* (12) *and* (13)*, we get*

$$\|\mathscr{B}\|_{L^2}^2 \leq \frac{\lambda}{c_L \kappa_1} \|Lf_0 - u\|_{L^2}^2$$

meaning that the L^2-norm of the bias is proportional to $\|Lf_0 - u\|_{L^2}$. This means that, as expected, the closer f_0 is to the solution of the PDE, the smaller the bias. In particular, if $\|Lf_0 - u\|_{L^2} = 0$, the L^2-norm of the bias is zero. In such case, the best rate for the MSE of the estimator is achieved for a constant $\lambda_n = \lambda$ for all n, and this rate is the optimal rate of convergence for parametric estimators:

$$MSE_{L^2}(\hat{f}) = O(n^{-1}).$$

4 Numerical solution of SR-PDE estimation problem

The SR-PDE estimator defined in (1) cannot be computed analytically. [3] shows that that solving (1) is equivalent to solving the following coupled system of PDEs

$$\begin{cases} L\hat{f} = u + \hat{g} & \text{in } \Omega \\ \hat{f} = \gamma & \text{on } \partial\Omega \end{cases} \qquad \begin{cases} L^*\hat{g} = -\frac{1}{\lambda_n}\sum_{i=1}^n (\hat{f} - z_i)\delta_{\mathbf{p}_i} & \text{in } \Omega \\ \hat{g} = 0 & \text{on } \partial\Omega \end{cases} \tag{24}$$

where $\hat{g}$ represents the misfit of the penalized PDE, i.e., $\hat{g} = L\hat{f} - u$, and L^* is the adjoint operator of L defined in (6). This reformulation of the problem introduces homogeneous

Dirichlet boundary conditions on $\hat{g}$, although we do not require that f_0 satisfies the boundary conditions $Lf_0 - u|_{\partial\Omega} = g_0|_{\partial\Omega} = 0$. Notice, however, that when $Lf_0 - u|_{\partial\Omega} = 0$ we obtain the best rate of convergence in Theorem 1.

The reformulation (24) of the estimation problem (1) is very convenient as it can be easily discretized by the finite element method. We briefly recall the discretization [see, e.g. 3, for details]. For simplicity of exposition, assume here that Ω is a convex polygonal domain. Let T_h be a triangulation of the domain Ω, where h is the maximum length of the edges in the triangulation, and define the finite element space of piecewise polynomial functions of degree r over the triangulation

$$V^r_{h,\gamma} = \left\{ v \in C^0(\bar{\Omega}) : v|_{\partial\Omega} = \gamma_h \;\; v|_\tau \in \mathbb{P}^r(\tau) \;\; \forall \tau \in T_h \right\}$$

where γ_h is the interpolant of γ in the space of piecewise continuous polynomial functions of degree r over $\partial\Omega$. Call $\xi_1, \ldots, \xi_{N_h}$ the interior nodes of the triangulation, which correspond to the interior vertices of the triangulation for linear finite elements. Let $\psi_1, \ldots, \psi_{N_h}$ be the associated finite element basis, that is $\psi_i \in V^r_{h,0}$ and $\psi_i(\xi_j) = \delta_{ij}$. Each $f \in V^r_{h,0}$ can be written as

$$f(x,y) = \boldsymbol{\psi}(x,y)^T \mathbf{f}$$

where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{N_h})^T$ is the vector of the basis functions, and $\mathbf{f} = (f(\xi_1), \ldots, f(\xi_{N_h}))^T$ is the vector of coefficients. Define the bilinear form $a(\cdot,\cdot)$, associated to the operator L, as

$$a(\hat{f}, \psi) = \int_\Omega \left(\mathbf{K} \nabla \hat{f} \cdot \nabla \psi + \mathbf{b} \cdot \nabla \hat{f} \psi + c \hat{f} \psi \right).$$

Define the $(N_h \times N_h)$ matrices $A_{ij} = a(\psi_j, \psi_i)$ and $M_{ij} = \int_\Omega \psi_i \psi_j$, and the $(n \times N_h)$ matrix of the evaluation of the basis functions at the data locations $\Psi_{ij} = \psi_j(\mathbf{p}_i)$. In addition, let $\boldsymbol{\psi}^D = (\psi^D_1, \ldots, \psi^D_{N^D_h})^T$ be the vector of basis functions associated to the boundary of the domain, and define: $A^D_{ij} = a(\psi^D_j, \psi_i)$, $\Psi^D_{ij} = \psi^D_j(\mathbf{p}_i)$ and $\boldsymbol{\gamma}$ the evaluation of the boundary condition γ at the boundary nodes. The coupled system of PDEs (24) is then discretized as follows

$$\begin{bmatrix} \Psi^T\Psi/n & \lambda_n A^T \\ A & -M \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \Psi^T\mathbf{z}/n - \Psi^T\Psi^D\boldsymbol{\gamma}/n \\ \mathbf{u} - A^D\boldsymbol{\gamma} \end{bmatrix}. \tag{25}$$

[4] shows that, under regularity conditions on L, there exists $h_0 > 0$ such that for every $h \leq h_0$, the solution of the discretized problem (25) is unique. The required regularity conditions on the parameters of L are such that for every $u \in L^p(\Omega)$ there exists a unique solution of the differential problem $Lf = u$ in the Sobolev space $W^{2,p}(\Omega)$, for some $p > 2$, where $W^{2,p}(\Omega)$ is the space of functions in $L^p(\Omega)$ with derivatives up to the 2-th order in $L^p(\Omega)$.

The finite element SR-PDE estimator is thus obtained as

$$\hat{f}_h = \boldsymbol{\psi}^T \hat{\mathbf{f}} + (\boldsymbol{\psi}^D)^T \boldsymbol{\gamma}. \tag{26}$$

The estimator has a penalized regression form. In particular

$$\hat{\mathbf{f}} = \left(\Psi^T\Psi + n\lambda_n P \right)^{-1} \Psi^T \mathbf{z}$$

where $P = A^T M^{-1} A$ is a discretization of the regularizing term. The fitted values $\hat{\mathbf{z}}$ can be obtained as

$$\hat{\mathbf{z}} = \Psi\hat{\mathbf{f}} + \Psi^D\boldsymbol{\gamma} = S\mathbf{z} + \mathbf{r}$$

where the smoothing matrix $S \in \mathbb{R}^{n\times n}$ and the vector $\mathbf{r}$ are given by

$$S = \Psi\left(\Psi^T\Psi + n\lambda_n P\right)^{-1}\Psi^T$$
$$\mathbf{r} = \Psi\left(\Psi^T\Psi + n\lambda_n P\right)^{-1}\left\{n\lambda_n PA^{-1}\mathbf{u} - \left(\Psi^T\Psi^D + n\lambda_n A^T M^{-1} A^D\right)\boldsymbol{\gamma}\right\}.$$

5 Consistency of SR-PDE estimator: finite element estimation problem

We are now interested in proving the consistency of the finite element SR-PDE estimator. This estimator is not a direct discretization of the infinite-dimensional SR-PDE estimator, defined in (1). As described above, the finite element estimator is based on the discretization (25) of the reformulation (24), that consists of a coupled system of second order differential problems, instead of the original fourth order problem in (1). Since all the results in Section 3 are based on the forth order problem (1), it is unfortunately not possible to derive the consistency of the finite element SR-PDE estimator from the consistency of the infinite dimensional SR-PDE estimator.

The consistency of the finite element estimator is studied in the following discrete semi-norm, defined for any function $v_h \in V_h$ as

$$\|v_h\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} v_h^2(\mathbf{p}_i).$$

This norm is an approximation of the L^2-norm, computed at the data locations.

For simplicity, we restrict our attention to the following case.

Assumption 3. *The differential operator L is self-adjoint, i.e.,* $Lf = -\text{div}\left(\mathbf{K}\nabla f\right) + cf$.

Assumption 4. *The discretization is based on linear finite elements on a constrained Delaunay triangulation of* $\mathbf{p}_1, \ldots, \mathbf{p}_n$.

See [21] for Delaunay triangulations. We also make an additional assumption on how the data locations fill the domain Ω. This assumption ensures good properties of the finite element basis. Given a family of triangulations $\{T_h\}_{h>0}$, let h_K and ρ_K be respectively the diameter (longest edge) and the radius of the inscribed circle of the triangle $K \in T_h$. The family $\{T_h\}_{h>0}$ is said to be *shape regular* if there exists σ_0 such that $\sigma_K = \frac{h_K}{\rho_K} \leq \sigma_0$ for all h and for all $K \in T_h$. Moreover, the family $\{T_h\}_{h>0}$ is said to be *quasi-uniform* if it is shape regular and there exists $c > 0$ such that $h_K \geq ch$ for all h and for all $K \in T_h$.

Assumption 5. *The points* $\mathbf{p}_1, \ldots, \mathbf{p}_n$ *are such that the constrained Delaunay triangulation* T_h *on these points is a quasi-uniform triangulation.*

5.1 Convergence of the bias term: finite element estimator

In this section we consider the bias of the finite element estimator

$$\mathscr{B}_h = f_0 - \mathbb{E}(\hat{f}_h)$$

and study its n-norm with respect to the number of observations n and the smoothing parameter λ_n.

Theorem 4. *Under Assumptions 4 and 5, for n sufficiently large, if* $f_0 \in W^{2,p}(\Omega)$ *for* $p > 2$, $f_0 \,|_{\partial\Omega} = \gamma$, $g_0 = Lf_0 - u \in H^1(\Omega)$ *and* $g_0|_{\partial\Omega} = 0$, *then*

$$\|\mathscr{B}_h\|_n^2 \leq C\left(\tfrac{1}{n} + \lambda_n\right). \tag{27}$$

Proof. [4] shows that, if $f_0 \in W^{2,p}(\Omega)$ for $p > 2$, $f_0 \,|_{\partial\Omega} = \gamma$, $g_0 = Lf_0 - u \in H^1(\Omega)$ and $g_0|_{\partial\Omega} = 0$, there exists $h_0 > 0$ such that for every $h \leq h_0$

$$\|\mathscr{B}_h\|_n^2 \leq C\left\{h^2\left[(1+\lambda_n)\|f_0\|_{W^{2,p}}^2 + \|Lf_0 - u\|_{H^1}^2\right] + \lambda_n \|Lf_0 - u\|_{L^2}^2\right\}. \tag{28}$$

Under Assumption 4, we have that $h^2 \approx \frac{1}{n}$. Thus, for n sufficiently large, we obtain (27). □

Note that the result in Theorem 4 is sub-optimal in λ_n with respect to the rate in Theorem 1.

5.2 Convergence of the variance term: finite element estimator

Here we focuss on $\mathbb{C}\text{ov}(\hat{\mathbf{z}}) = \sigma^2 SS^T$ and consider its n-norm

$$\|\mathbb{C}\text{ov}(\hat{\mathbf{z}})\|_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}\text{ar}(\mathbf{z}_i) = \frac{\sigma^2}{n}\text{Tr}(SS^T). \tag{29}$$

We are thus interested in studying the eigenvalues of the matrix SS^T. Under Assumption 4, the matrix Ψ concides with I_n, the identity matrix in $\mathbb{R}^{n\times n}$, and $S = (I_n + n\lambda_n P)^{-1}$. Therefore, we are interested in studying the eigenvalues of the penalty matrix P. Before giving the result on the variance of the finite element estimator, we need the following Lemmas.

Lemma 2. *Suppose that* H *is a* $n \times n$ *positive semi-definite symmetric matrix and* C *a* $n \times n$ *matrix. Let* $\ell_k(A)$ *denote the kth smallest eigenvalue of a positive semi-definite symmetric matrix* A. *Then, for each* $k = 1, \ldots, n$,

$$\ell_1(H)\ell_k(CC^T) \leq \ell_k(CHC^T) \leq \ell_n(H)\ell_k(CC^T).$$

Proof. See [27]. □

Lemma 3. *Let* $\{\zeta_{k,h}\}_{k=1}^n$ *be the eigenvalues of the penalty matrix* P, *ordered such that* $0 < \zeta_{1,h} \leq \cdots \leq \zeta_{n,h}$. *Under Assumptions 3, 4 and 5,*

$$\zeta_{k,h} = O(k^2 h^2). \tag{30}$$

Proof. Since $P = A^T M^{-1} A = A^T M^{-1} M M^{-1} A$, we can study the eigenvalues of P starting from the eigenvalues of $M^{-1}A$ and M. Denote by $\{\mu_{k,h}\}$ the eigenvalues of M and by $\{\eta_{k,h}\}$ the eigenvalues of $M^{-1}A$. From Lemma 2 we have

$$\mu_{1,h}\eta_{k,h}^2 \leq \zeta_{k,h} \leq \mu_{n,h}\eta_{k,h}^2 \tag{31}$$

Consider the problem of finding the eigenfunctions and the eigenvalues η_k of:

$$\begin{cases} Lv = \eta v & \text{in } \Omega \subset \mathbb{R}^2 \\ v = \gamma & \text{on } \partial\Omega \end{cases} \tag{32}$$

For self-adjoint operators L, the eigenvalues η_k are infinite, they belong to $(a, +\infty)$ for some $a > 0$ and $\eta_k \sim k$ [see, e.g., 1, 8].

The finite element discretization of (32) on the triangulation T_h leads to the following generalized eigenvalue problem $A\mathbf{v}_h = \eta_h M\mathbf{v}_h$. Since M is invertible, this is a classic eigenvalue problem, and it is equivalent to finding the eigenvalues of $M^{-1}A$. In particular we have $\eta_{k,h} \to \eta_k$ for $h \to 0$ [see, e.g., 7, 23]. More precisely $\eta_k \leq \eta_{k,h} \leq \eta_k + ch^2$ [see 6, Theorem 10.4]. Therefore, for k (and thus n) sufficiently large we have:

$$\eta_{k,h} = O(k). \tag{33}$$

Regarding the eigenvalues of M, thanks to Assumption 5, we have that for each k in $\{1, \ldots, n\}$

$$c_1 h^2 \leq \mu_{k,h} \leq c_2 h^2 \tag{34}$$

[see, e.g., 29, 12].
From (31), (33) and (34) we conclude that $\{\zeta_{k,h}\}$ increase with the same rate of $\{\eta_{k,h}^2\}$ with respect to k, leading to (30). □

Theorem 5. *Under Assumptions 3, 4 and 5, for n sufficiently large*

$$\|\mathbb{C}ov(\hat{\mathbf{z}})\|_n = O\left(\frac{1}{n\sqrt{\lambda_n}}\right). \tag{35}$$

Proof. Under Assumption 4, $S = (I_n + n\lambda_n P)^{-1}$. Thus, S has eigenvalues $1/(1+n\lambda_n\zeta_i)$, where ζ_i are the eigenvalues of P. The trace of SS^T is hence given by

$$\mathrm{Tr}(SS^T) = \sum_{i=1}^{n} \left(\frac{1}{1+n\lambda_n\zeta_i}\right)^2. \tag{36}$$

From equations (29) and (36), thanks to Lemma 3 we have

$$\|\mathbb{C}\mathrm{ov}(\hat{\mathbf{z}})\|_n \approx \sigma^2 h^2 \sum_{k=1}^{n} \left(\frac{1}{1+\lambda_n k^2}\right)^2 \approx \sigma^2 h^2 \int_1^n \left(\frac{1}{1+\lambda_n t^2}\right)^2 dt = O\left(\frac{h^2}{\sqrt{\lambda_n}}\right).$$

Recalling in addition that, under Assumption 4, $h^2 \approx n^{-1}$ we obtain equation (35). □

5.3 Convergence of the MSE: finite element estimator

The following theorem shows that the finite element estimator is consistent and its MSE achieves the optimal rate of convergence for non-parametric estimators for $H^2(\Omega)$ functions.

Theorem 6. *Under Assumptions 3, 4 and 5, for n sufficiently large, setting $\lambda_n = n^{-2/3}$ we have*

$$MSE_n(\hat{f}_h) = O\left(n^{-\frac{2}{3}}\right).$$

Proof. Thanks to Theorems 4 and 5 we have

$$\mathrm{MSE}_{L^2}(\hat{f}_h) = O(\lambda_n) + O\left(\frac{\sigma^2}{n\sqrt{\lambda_n}}\right)$$

that is minimized when $\lambda_n = n^{-2/3}$, leading to $\mathrm{MSE}_n(\hat{f}_h) = O\left(n^{-2/3}\right)$. □

Remark 2. *As for the infinite-dimensional estimator, we have studied the properties of $\hat{f}_h$ without assuming that the true f_0 satisfies the regularizing PDE. From equation* (28) *we see that, if $Lf_0 - u = 0$ the bias does not vanish, due to the discretization error. However, when $Lf_0 - u = 0$,*

$$\|\mathscr{B}_h\|_n^2 \leq ch^2 \approx cn^{-1}$$

where c is a constant the does not depend on n nor on λ_n. In such case, setting $\lambda_n = \lambda$, constant for all n, as for the infinite-dimensional case, we achieve again a parametric rate of convergence for the MSE: $MSE(\hat{f}_h) = O(n^{-1})$.

6 Numerical simulations

In this section we provide numerical evidence of the convergence rates obtained in Sections 3 and 5 for the bias and the variance of the estimators, in a simple setting. We consider the convergence in the L^2 and on the discrete n-norm, both with respect to n and λ_n. In particular, in Section 6.1 we report some simulations that illustrate the rate of convergence for the infinite-dimensional estimator, when the discretization size is small; moreover, in Section 6.2, we report some simulations that illustrate the rate of convergence for the finite element estimator, when the mesh is constrained to the data locations.

In all simulations, the domain Ω is a circle with radius $R = 1$, the differential operator L is the laplacian, and the forcing term u is equal to zero. All the rates of convergence are illustrated with a log-log plot with λ_n or n on the x-axis, and the error on the y-axis.

To illustrate the different convergence rates achieved for functions with different regularities, we consider three test functions f_0. The first test function is

$$f_{0,1}(x,y) = \left[1 - (x^2 + y^2)\right]^3.$$

This test function vanishes on $\partial\Omega$; moreover, it is such that $\Delta f_{0,1}|_{\partial\Omega} = 0$. The second test function is

$$f_{0,2}(x,y) = \left[1 - (x^2 + y^2)\right]^2.$$

Alike the previous test function, $f_{0,2}$ vanishes on $\partial\Omega$, but in this case $\Delta f_{0,2}|_{\partial\Omega} \neq 0$. The third test function is

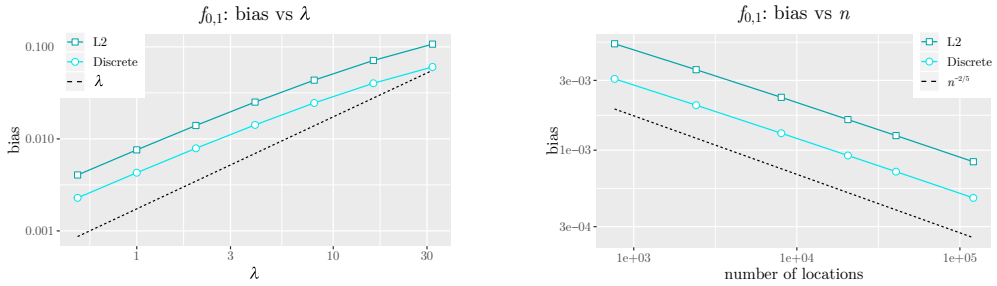
$$f_{0,3}(x,y) = \left[1 - \sqrt{x^2 + y^2}\right].$$

As the previous, $f_{0,3}$ vanishes on $\partial\Omega$; for such function $\Delta f_{0,3}|_{\partial\Omega} \neq 0$ and $\Delta f_{0,3} \in L^2(\Omega)$, but $f_{0,3} \notin H^2(\Omega)$.

6.1 Simulations with fine and fixed triangulation

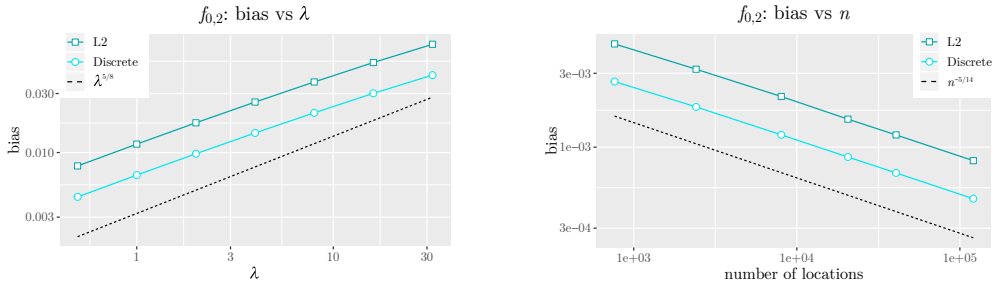
In this section we aim at illustrating the rate of convergence for the infinite-dimensional estimator. To this aim, we consider a very fine discretization, consisting of a Delaunay triangulation with $N = 123103$ nodes; we hence sample n data locations, with $n \leq N$, from the nodes of the mesh.

We first consider the bias term. For this reason, we sample data from the test functions $f_{0,1}$, $f_{0,2}$ and $f_{0,3}$ without adding any noise. We first look at the bias with respect to the smoothing parameter λ_n, when the number of observations n is fixed. In this case, we consider an observation for each interior node in the triangulation. Figures 2a, 3a and 4a show the bias decay in the three cases corresponding to the three test functions.



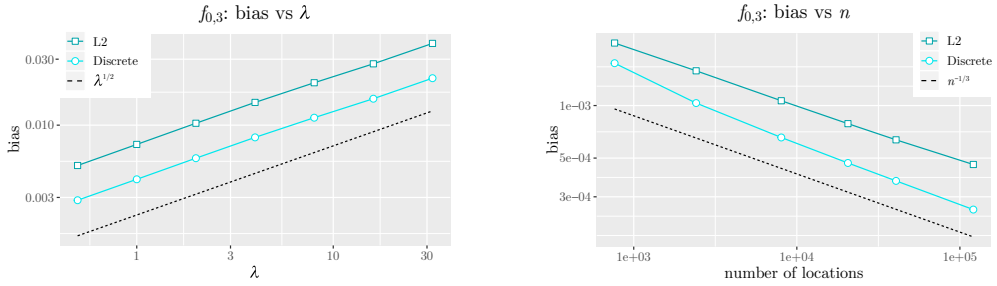
(a) (b)

Figure 2: Test function $f_{0,1}$ without noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the bias of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the bias with respect to the number of points n, with $\lambda_n = n^{-2/5}$.



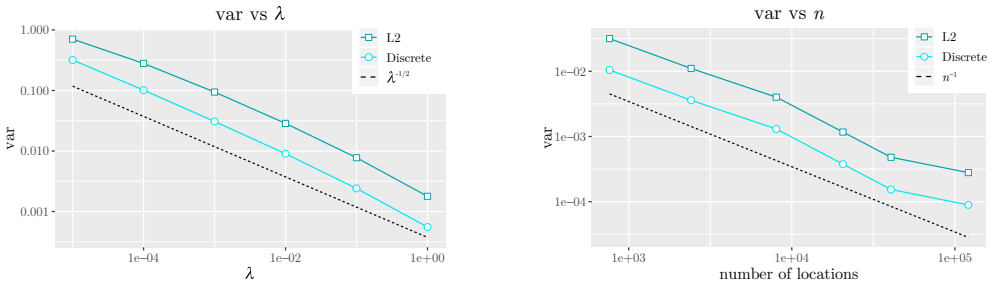
(a) (b)

Figure 3: Test function $f_{0,2}$ without noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the bias of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the bias with respect to the number of points n, with $\lambda_n = n^{-4/7}$.



(a) (b)

Figure 4: Test function $f_{0,3}$ without noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the bias of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the bias with respect to the number of points n, with $\lambda_n = n^{-2/3}$.



(a) (b)

Figure 5: Pure gaussian noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the variance of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the variance with respect to the number of data locations n, with $\lambda_n = 1$.

For the first test function $f_{0,1}$, the bias reaches the expected rate of convergence of λ_n for small values of the parameter (see Figure 2a). For $f_{0,2}$ the bias decays as $\lambda_n^{5/8}$ (see Figure 3a) and for $f_{0,3}$ as $\lambda_n^{1/2}$ (see Figure 4a), as expected.

We then look at the bias when the number of observations n increases (up to the number of interior nodes of the triangulation) and the smoothing parameter λ_n varies as a power of n. In particular, for the three test functions $f_{0,1}$, $f_{0,2}$ and $f_{0,3}$, we set λ_n proportional to $n^{-2/5}$, $n^{-4/7}$ and $n^{-2/3}$ respectively, that are the optimal ones in the minimization of the MSE, according to Theorem 3. Figure 2b, 3b and 4b show that the theoretical rates are indeed achieved in all three cases.

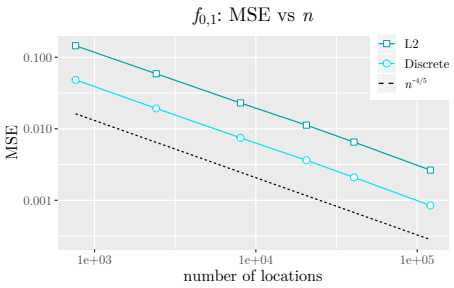
To illustrate the rate for the variance term in Theorem 2, we consider pure noise data. Specifically, we sample data as gaussian random noise with variance $\sigma^2 = 1$ in all the interior nodes of the mesh. We solve the estimation problem with $\lambda_n = 1, 10^{-1}, \ldots, 10^{-5}$, and we repeat the simulation fifty times to compute the mean of the error. The results are shown in Figure 5a. As expected, the square of the L^2 and of the discrete norm increase as $\lambda_n^{-1/2}$.

To illustrate the rate for the variance with respect to the number of observations, we proceed as for the bias, but considering an increasing number of observations n. We solve the estimation problem with a fixed $\lambda_n = 1$. We compute the mean over fifty replicates. The results are shown in Figure 5b. As expected, the square of the L^2 and of the discrete norm decay as n^{-1}.

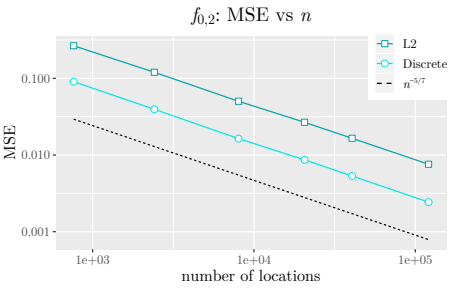
Finally, we want to illustrate the rate for MSE with respect to the number of observations. To this end, we consider the same simulation setting considered for the bias, increasing the sample size n and taking λ_n proportional to $n^{-2/5}$, $n^{-4/7}$ and $n^{-2/3}$, for the three test functions $f_{0,1}$, $f_{0,2}$ and $f_{0,3}$, respectively. We sample the data adding a gaussian random noise with variance $\sigma^2 = 1$. We compute the mean estimate over fifty simulation replicates. Figures 6a to 6c show the obtained results. As expected from Theorem 3, the L^2 and the discrete norm decay as $n^{-4/5}$, $n^{-5/7}$ and $n^{-2/3}$ respectively, for the three test functions.

6.2 Simulations with constrained triangulations

We now consider different Delaunay triangulations of Ω with an increasing number of nodes N, approximately equal to 200, 800, 3000, 12000, 50000, and 200000 interior



(a) Test function $f_{0,1}$, $\lambda_n = n^{-2/5}$.



(b) Test function $f_{0,2}$, $\lambda_n = n^{-4/7}$.

(c) Test function $f_{0,3}$, $\lambda_n = n^{-2/3}$.

Figure 6: Test function $f_{0,1}$, $f_{0,2}$ and $f_{0,3}$ with gaussian noise; fine and fixed triangulation; data sampled at an increasing number of interior node. Convergence rates of the MSE with respect to the number of points n.

nodes. We sample the three test functions $f_{0,1}, f_{0,2}, f_{0,3}$ at all mesh nodes, so that $n = N$ and the mesh nodes coincide with data locations (constrained triangulation), as in Section 5. We want to illustrate the rates derived for the finite element estimator in Section 5, and to show that, unlike for the infinite-dimensional estimator, it is not possible to improve the rate when the true function f_0 has regularity higher than H^2.

We first consider the bias with respect to the number of observations n. To this end, we consider the true data without noise. For all the three test functions we set λ_n proportional to $n^{-2/3}$, as in Theorem 6. Figure 7 shows that the bias is proportional to $n^{-1/3}$ (that is, to $\sqrt{\lambda_n}$), for all the three functions $f_{0,1}$, $f_{0,2}$, $f_{0,3}$.

We then look at the variance term. We consider pure noise data, sampling a gaussian random noise with variance $\sigma^2 = 1$ at each interior mesh node. We solve the estimation problem with $\lambda_n = 1$. We compute the mean over fifty replicates. The results are shown in Figure 8. As expected from Theorem 5, both the L^2 and the discrete norm decay as n^{-1}.

To illustrate the result in Theorem 6, we study the MSE with respect to the number of observations. We sample from $f_{0,1}, f_{0,2}, f_{0,3}$, adding a gaussian random noise with variance $\sigma^2 = 1$ (at the interior nodes). We solve the estimation problem with λ proportional to $n^{-2/3}$ for all three test functions. The results are shown in Figure 9. As expected from Theorem 6, both the L^2 and the discrete norm decay as $n^{-2/3}$.

7 Discussion and future work

We proved the consistency of the infinite-dimensional SR-PDE estimator, for a general differential operator L with H^2 regularity, when exact Dirichlet boundary conditions

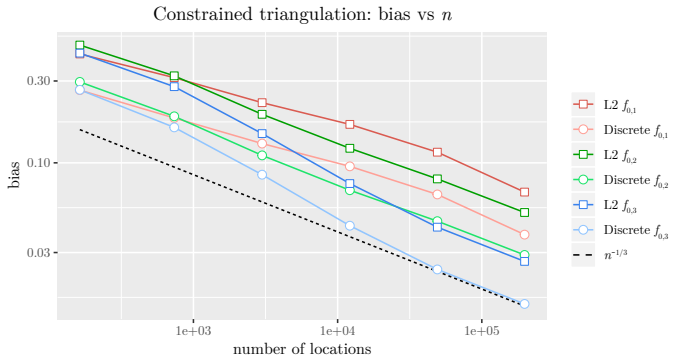


Figure 7: Three test functions without noise; constrained triangulations ($N = n$) of an increasing number of data locations. Convergence rates of the bias of the finite element estimator with respect to the number of observations n, with $\lambda_n = n^{-2/3}$. In green the rate for $f_{0,1}$, in blue for $f_{0,2}$, in black for $f_{0,3}$.

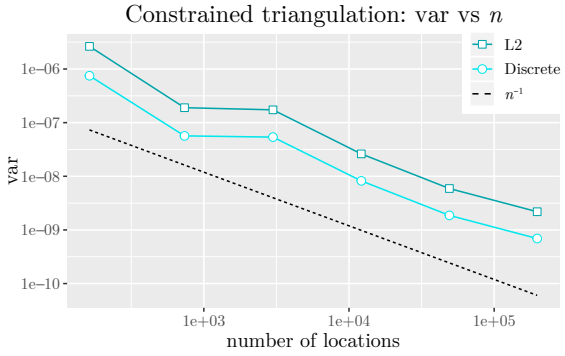


Figure 8: Pure Gaussian noise; constrained triangulations ($N = n$) of an increasing number of data locations. Convergence rates of the variance of the finite element estimator with respect to the number of obervations n, with $\lambda_n = 1$.

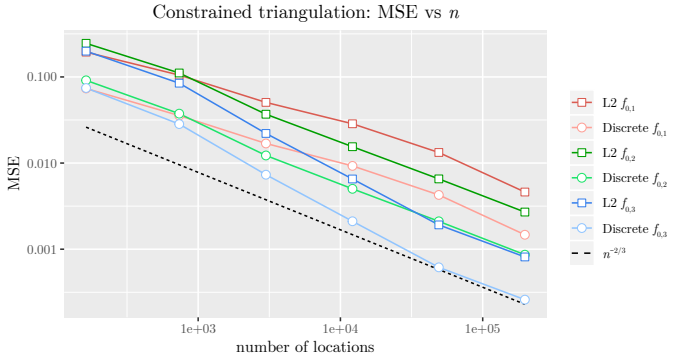


Figure 9: Three test functions with Gaussian noise; constrained triangulations ($N = n$) of an increasing number of data locations. Convergence rates of the MSE of the finite element estimator with respect to the number of observations n, with $\lambda_n = n^{-2/3}$. In green the rate for $f_{0,1}$, in blue for $f_{0,2}$, in black for $f_{0,3}$.

are imposed on $\partial\Omega$. The exact boundary conditions on f are sufficient to prove that the MSE of the estimator achieves the near-optimal rate of convergence when the true

function $f_0 \in H^2$. Moreover, when exact Dirichlet boundary conditions are also available on $g_0 = Lf_0 - u$, it is possible to improve such rate, achieving the near-optimal rate of convergence for H^4 functions. In future research, we will aim at proving the consistency under more general mixed boundary conditions on the function and possibly demonstrating that the estimators attain exactly the optimal rate.

We also proved the consistency of the finite element SR-PDE estimator. In this case we restricted the attention to self-adjoint operators L, i.e., $\mathbf{b} = \mathbf{0}$, meaning that no unidirectional smoothing is considered. However, based on the results obtained for the infinite dimensional estimator, we conjecture that the finite element estimator is consistent also in the case $\mathbf{b} \neq \mathbf{0}$. Moreover, to obtain the rate on the bias, we assumed exact boundary conditions on g_0. However, [4] show numerically that the extra error incurred if g_0 does not satisfy the imposed condition is of the same order of the bias. In addition, the rate of convergence for the bias of the finite element estimator derived in Theorem 4, is shown to be suboptimal by the numerical simulation in Section 6. As a future work, we aim to improve the rate derived in Theorem 4. To derive the rate for the variance of the finite element estimator we assumed that the triangulation is constrained to the data locations. We are currently working to relax such assumption. This would in turn allow us to relax the infill properties required on the data locations.

The infinite dimensional SR-PDE estimation problem could also be solved using different numerical approaches and bases. In [35], for instance, we use isogeometric analysis based on Non-Uniform Rational B-Splines (NURBS). The study of the consistency of the corresponding estimators is an interesting direction of future research.

Moreover, we intend to explore the consistency of the SR-PDE estimators for space-time data defined in [2]. Furthermore, we will also investigate the consistency of SR-PDE estimators over two-dimensional manifold domains defined in [14], and of SR-PDE estimators in three-dimensional domains. These studies will require different arguments then those used in this work. For instance, in the three-dimensional case, Lemma 1 does not hold, and alternative ways to control the difference between F and F_n should be sought.

Finally, it would be very interesting to prove the consistency of SR-PDE estimators in the more complex semi-parametric setting considered in [31], where also space-varying covariate information is included, following a generalized additive model where $z_i = \mathbf{q}_i^T \boldsymbol{\beta} + f_0(\mathbf{p}_i) + \varepsilon_i$, and $\mathbf{q}_i$ denotes the covariates observed at $\mathbf{p}_i$. In particular, a very interesting development of this work would consist in proving the consistency and asymptotic normality of the regression coefficients $\boldsymbol{\beta}$. A similar problem, in the simpler case of univariate smoothing splines, was considered by [20].

References

[1] S. Agmon. *Lectures on Elliptic Boundary Value Problems*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 2010.

[2] E. Arnone, L. Azzimonti, F. Nobile, and L.M. Sangalli. Modeling spatially dependent functional data via regression with differential regularization. *J. Multivariate Anal.*, 170:275–295, 2019.

[3] L. Azzimonti, L.M. Sangalli, P. Secchi, M. Domanin, and F. Nobile. Blood flow velocity field estimation via spatial regression with PDE penalization. *Journal of the American Statistical Association*, 110(511):1057–1071, 2015.

[4] L. Azzimonti, F. Nobile, L.M. Sangalli, and P. Secchi. Mixed finite elements for spatial regression with pde penalization. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):305–335, 2014.

[5] M.S. Bernardi, M. Carey, J.O. Ramsay, and L.M. Sangalli. Modeling spatial anisotropy via regression with partial differential regularization. *Journal of Multivariate Analysis*, 167:15–30, 2018.

[6] D. Boffi. Finite element approximation of eigenvalue problems. *Acta Numerica*, 19:1–120, 005 2010.

[7] D. Boffi, F. Gardini, and L. Gastaldi. *Some Remarks on Eigenvalue Approximation by Finite Elements*, pages 1–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[8] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[9] D.D. Cox. Asymptotics for m-type smoothing splines. *Annals of Statististics*, 11(2):530–551, 1983.

[10] D.D. Cox. Multivariate smoothing spline functions. *SIAM J. Numer. Anal.*, 21(4):798–813, 1984.

[11] F. Cucker and D.X. Zhou. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge Monographs Applied & Computational Mathematics. Cambridge University Press, 2007.

[12] A. Ern and J.L. Guermond. *Theory and Practice of Finite Elements*. Applied Mathematical Sciences. Springer New York, 2013.

[13] B. Ettinger, S. Guillas, and M.-J. Lai. Bivariate splines for ozone concentration forecasting. *Environmetrics*, 23(4):317–328, 2012.

[14] B. Ettinger, S. Perotto, and L.M. Sangalli. Spatial regression models over two-dimensional manifolds. *Biometrika*, 103(1):71–88, 2016.

[15] R.L. Eubank. *Nonparametric regression and spline smoothing*. CRC press, 1999.

[16] L.C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998.

[17] L. Formaggia, A. Quarteroni, and A. Veneziani. *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*, volume 1. Springer Science & Business Media, 2010.

[18] S. Guillas and M.-J. Lai. Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics*, 22(4):477–497, 2010.

[19] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York, 2002.

[20] N.E. Heckman. Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 48(2):244–248, 1986.

[21] Ø. Hjelle and M. Dæhlen. *Triangulations and applications*. Springer Science & Business Media, 2006.

[22] J.Z. Huang. Local asymptotics for polynomial spline regression. *Ann. Statist.*, 31(5):1600–1635, 2003.

[23] W.G. Kolata. Approximation in variationally posed eigenvalue problems. *Numerische Mathematik*, 29(2):159–171, 1978.

[24] M.-J. Lai and L.L. Schumaker. *Spline functions on triangulations*, volume 110. Cambridge University Press, 2007.

[25] M.-J. Lai and L. Wang. Bivariate penalized splines for regression. 23, 07 2013.

[26] J.L. Lions and E. Magenes. *Non-homogeneous Boundary Value Problems and Applications*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer, 1972.

[27] L.-Z. Lu and C.E.M. Pearce. Some new bounds for singular values and eigenvalues of matrix products. *Annals of Operations Research*, 98(1):141–148, 2000.

[28] H. Niederreiter. *Random Number Generation and quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[29] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2008.

[30] T. Ramsay. Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):307–319, 2002.

[31] L.M. Sangalli, J.O. Ramsay, and T.O. Ramsay. Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):681–703, 2013.

[32] L.A.S. Scott-Hayward, M.L. MacKenzie, C.R. Donovan, C.G. Walker, and E. Ashe. Complex region spatial smoother (cress). *Journal of Computational and Graphical Statistics*, 23(2):340–360, 2014.

[33] C.J. Stone. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4):1040–1053, 1982.

[34] H. Wang and M.G. Ranalli. Low-rank smoothing splines on complicated domains. *Biometrics*, 63(1):209–217, 2007.

[35] M. Wilhelm, L. Dedè, L.M. Sangalli, and P. Wilhelm. IGS: an IsoGeometric approach for smoothing on surfaces. *Comput. Methods Appl. Mech. Engrg.*, 302:70–89, 2016.

[36] S.N. Wood, M.V. Bravington, and S.L. Hedley. Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:931–955, 2008.