SOME FIRST RESULTS ON THE CONSISTENCY OF SPATIAL REGRESSION WITH PARTIAL DIFFERENTIAL EQUATION REGULARIZATION

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

[1]*Politecnico di Milano,* [2]*Universität Bonn*
and [3]*École polytechnique fédérale de Lausanne*

Abstract: We study the consistency of the estimator in a spatial regression with partial differential equation (PDE) regularization. This new smoothing technique allows us 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. In contrast to classical smoothing methods, the solution to the infinite-dimensional estimation problem cannot be computed analytically. An approximation is obtained using 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 the value of the smoothing parameter. Lastly, simulation studies provide numerical evidence of the rates derived for the bias, variance, and mean square error.

Key words and phrases: Functional data analysis, smoothing, spatial statistics.

1. Introduction

In this study, we examine the consistency of the spatial regression with partial differential equation (SR-PDE) regularization (see, e.g., Azzimonti et al. (2015); Azzimonti et al. (2014)). This regularized least squares method defines a new class of bivariate smoothers that has a number of advantages over classical smoothers, such as smoothing splines and thin-plate splines, which have well-established properties (see, e.g., Eubank (1999) and the references therein). The regularizing term in an 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 powerful tool for modeling complex behaviors, and are used extensively in most

---

Corresponding author: Laura M. Sangalli, MOX - Dipartimento di Matematica, Politecnico di Milano, 20133 Milano, Italy. E-mail: laura.sangalli@polimi.it.

fields of science and engineering. This makes the SR-PDE broadly applicable to the analysis of spatially distributed data in varied contexts (see, e.g., the applications in Azzimonti et al. (2015); Bernardi et al. (2018)). In particular, the regularizing term in an SR-PDE can include general linear second-order PDEs. These involve 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 modeling of the space variation enables the analysis of a wide variety of anisotropic and nonstationary phenomena. Furthermore, an SR-PDE efficiently handles data distributed over domains with complex shapes, such as strong concavities or holes (see, e.g., Ramsay (2002); Sangalli, Ramsay and Ramsay (2013)). This is a crucial feature whenever the shape of the domain influences the problem at hand. Another important advantage of the SR-PDE over the classical smoothers is the possibility of imposing conditions on the value of the field and/or of its normal derivative that the field must satisfy at the boundaries of the domain of interest (Sangalli, Ramsay and Ramsay (2013); Azzimonti et al. (2014); Azzimonti et al. (2015)). This feature is fundamental in many applications to obtain meaningful estimates (see, e.g., Azzimonti et al. (2015); Sangalli, Ramsay and Ramsay (2013)).

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, and can only be characterized in a variational form. An approximated solution can be obtained using 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 the arguments used to prove the consistency of thin-plate splines and smoothing splines (see, e.g., Cox (1983, 1984); Cucker and Zhou (2007); Györfi et al. (2002); Huang (2003)).

Note that the problem of analyzing data that are spatially distributed over irregularly shaped two-dimensional domains has begun attracting interest. As such, other regularized least squares smoothers have been proposed to tackle this issue, such as bivariate splines over triangulations (see, e.g., Lai and Schumaker (2007); Guillas and Lai (2010); Ettinger, Guillas and Lai (2012); Lai and Wang (2013)), soap film smoothing (Wood, Bravington and Hedley (2008)), and low-rank thin-plate spline approximations (Wang and Ranalli (2007); Scott-Hayward et al. (2014)). These methods all have isotropic and stationary regularizing terms; bivariate splines over triangulations can include high-order derivatives. With the

exception of soap film smoothing, which complies 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 Lai and Wang (2013). However, rather than the SR-PDE, they directly consider the finite-dimensional estimator based on bivariate splines. To the best of our knowledge, no results exist on the large-sample properties of any of the other methods.

We prove the consistency of SR-PDE estimators, including the estimator solution of the infinite-dimensional estimation problem, and the case of the finite-element estimator. The remainder of the paper is organized as follows. Section 2 briefly reviews the SR-PDE infinite-dimensional estimation problem, introduced in Azzimonti et al. (2014) and Azzimonti et al. (2015), 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 also investigate the convergence of the variance of the infinite-dimensional SR-PDE estimator when n goes to infinity. Using the rates obtained for the bias and the variance, we prove the consistency of the estimator. Furthermore, we show that the mean squared error (MSE) of the estimator achieves near-optimal rates of convergence. In Section 5, we 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 that of the infinite-dimensional estimator. Nevertheless, 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 concludes the paper and discusses future research directions.

2. Spatial Regression with Partial Differential Equation Regularization: The Infinite-Dimensional Estimation Problem

We briefly review the SR-PDE infinite-dimensional estimation problem, as introduced in Azzimonti et al. (2014) and Azzimonti et al. (2015). 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 $i = 1, \ldots, n$, 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 extent, the phenomenon under study. Specifically, relying on the problem-specific information, we assume that the misfit $\|Lf_0 - u\|_{L^2}$ is small, though we do not require it to be zero. Here, L is a general linear second-order differential operator that can, for instance, include second-, first-, and zero-order differential terms:

$$L(\mathbf{p})f = -\mathrm{div}\,(\mathbf{K}(\mathbf{p})\nabla f) + \mathbf{b}(\mathbf{p}) \cdot \nabla f + c(\mathbf{p})f,$$

where $\mathbf{K}(\cdot) : \Omega \to \mathbb{R}^{2\times2}$ 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 is also related to 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 focus on Dirichlet boundary conditions. Specifically, we assume that we know the value of the field at the boundary: $f_0 \mid_{\partial\Omega} = \gamma$, 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 the functions in $L^2(\Omega)$ with derivatives up to the kth order in $L^2(\Omega)$, equipped with the norm $\|v\|_{H^k} = (\sum_{|\alpha|\leq k} \|D^\alpha v\|_{L^2}^2)^{1/2}$. Define the affine space

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

SR-PDE solves the following estimation problem:

$$\hat{f} = \operatorname*{argmin}_{f \in V_\gamma} \frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{p}_i) - z_i)^2 + \lambda_n \int_\Omega (Lf - u)^2. \qquad (2.1)$$

The estimation functional in (2.1) trades off a data fidelity criterion (the sum of the squared 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 us to model various forms of anisotropy and nonstationarity in the field. The diffusion term $-\mathrm{div}(K\nabla f)$ induces 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 eigenvalues. 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, because the penalization of the L^2 norm of f shrinks the field to zero. Moreover, because K, $\mathbf{b}$, and c can vary over space, the effects described here are nonstationary. This flexibility is increased further by the presence of the possibly non-homogeneous forcing terms $u \in L^2(\Omega)$. The SR-PDE can be viewed as an extension of the classical smoothing techniques to the anisotropic and nonstationary case. In particular, it includes as a special case the isotropic and stationary regularization of the Laplacian of the field considered in Ramsay (2002) and Sangalli, Ramsay and Ramsay (2013), when no problem-specific information is available (setting $\mathbf{K}$ to the the identity matrix, $\mathbf{b} = \mathbf{0}$, $c = 0$, such that $L = \Delta$, and a null forcing term $u = 0$).

Note that in the estimation problem given in (2.1), 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 the PDE carries partial information about the true f_0; 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 by 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 considered here, because a greater number of observations means there is less need to regularize the estimate.

2.1. Illustrative problem

As an illustrative example, Azzimonti et al. (2015) considers the problem of estimating the blood-flow velocity field in a cross-section of a human carotid artery, using eco-color doppler data and magnetic resonance imaging data. Figure
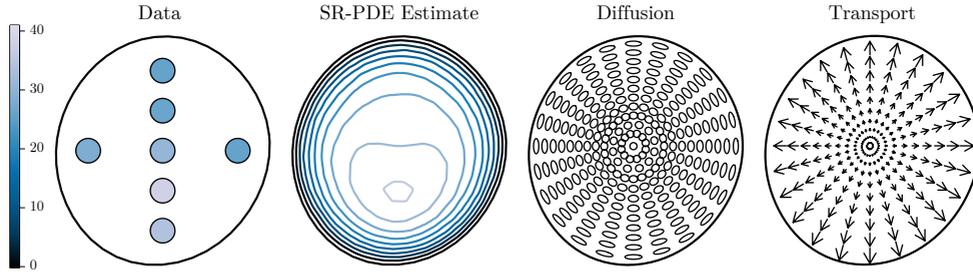
Figure 1. From left to right: ECD data on the artery cross-section; corresponding estimate of the blood-flow velocity field obtained using the SR-PDE; nonstationarity anisotropic diffusion tensor field K used in the SR-PDE estimate; nonstationarity transport field $\mathbf{b}$ used in the SR-PDE estimate.

1, top row, left panel, shows a reconstruction of the cross-section of the common carotid artery of one of the patients in the study; this quasi-circular section is obtained by segmenting the magnetic resonance imaging data. The same figure also displays the spatial locations of seven beams where the blood-flow velocity is measured using eco-color doppler acquisitions. The color of the beam refers to the mean blood velocity, measured over the beam at the systolic peak. Starting from the observations over the seven beams, we need to estimate the field f_0 of the blood-flow velocity at the systolic peak, over the entire cross-section of the carotid. In this applied problem, there are known conditions that the field must satisfy at the boundaries of the domain of interest, that is, at the arterial wall. In fact, the physics of the problem implies that the blood-flow velocity is zero at the arterial wall, owing 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, classical techniques for field estimation work naturally 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, Azzimonti et al. (2014) shows that isotropic and stationary smoothers return non-physiological estimates of the blood-flow velocity field, owing to the cross-shaped pattern of the observations. On the other hand, we can benefit from the detailed problem-specific information about the phenomenon under study. There is a vast body of literature devoted to the study of fluid dynamics and hemodynamics; see, for example, Formaggia, Quarteroni and Veneziani (2010). This information can be conveniently trans-

lated into a PDE that describes, in an idealized setting, the main features of the velocity field. In particular, as in Azzimonti et al. (2015), we 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, fourth panel); the reaction and forcing terms are 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, the 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

Azzimonti et al. (2015) and Azzimonti et al. (2014) show that, under regularity conditions on L, u, and γ, the estimator $\hat{f}$ in (2.1) is unique. The required regularity conditions on the parameters are such that the operator L has so-called H^2-smoothing properties; that is, L is such that, for all $u \in L^2(\Omega)$, the solution to the problem $Lf = u$ with Dirichlet boundary conditions belongs to $H^2(\Omega)$.

Solving (2.1) is equivalent to finding $\hat{f}$ such that

$$\lambda \int_\Omega (L\hat{f} - u)Lv + \frac{1}{n}\sum_{i=1}^n \hat{f}(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^n z_i v(\mathbf{p}_i) \quad \forall v \in V_0. \qquad (2.2)$$

This formulation highlights the linearity of the estimator $\hat{f}$ in the observations z_i. Owing to the linearity of equation (2.2) in $\hat{f}$, we can write $\hat{f}$ as

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

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

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

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

equivalently,

$$\hat{f}^* = \underset{f \in V_\gamma}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda_n \int_\Omega (Lf - u)^2 \right] \qquad (2.3)$$

$$\hat{w} = \underset{w \in V_0}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^{n} (w(\mathbf{p}_i) - \varepsilon_i)^2 + \lambda_n \int_\Omega (Lw)^2 \right]. \qquad (2.4)$$

The minimization problem (2.3) involves the true values of the field, without any observational noise, $f_0(\mathbf{p}_i)$, and the non-homogeneous regularization term (that is, with forcing term u) with non-homogeneous boundary conditions. The minimization problem (2.4) instead involves pure noise data, ε_i, and a homogeneous regularization term (i.e, with no forcing term) with homogeneous boundary conditions. The minimizer $\hat{f}^*$ is deterministic, whereas the minimizer $\hat{w}$ is such that $\mathbf{E}[\hat{w}] = 0$. Consequently, we have that: $\mathbb{E}[\hat{f}] = \hat{f}^*$ and $\mathbb{V}\mathrm{ar}(\hat{f}) = \mathbb{V}\mathrm{ar}(\hat{w})$. As a result, 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 (2.3), and when studying the variance, we focus on the minimization problem (2.4).

3. Consistency of SR-PDE Estimator: Infinite-Dimensional Problem

As in Cox (1983, 1984), in order to prove the consistency of the estimator, we make some assumptions on how the points $\mathbf{p}_i$ fill 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., Niederreiter (1992)).

Assumption 1. *The sequence $\{F_n\}$ converges uniformly to a cumulative distribution function F with 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 and 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 Cox (1984)).

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*

$$\left| \int_\Omega hg\mathsf{f}d\mathbf{p} - \frac{1}{n} \sum_{i=1}^{n} h(\mathbf{p}_i)g(\mathbf{p}_i) \right| = \left| \int_\Omega hgd(F - F_n) \right| \leq cd_n \|h\|_{H^2} \|g\|_{H^2}. \quad (3.1)$$

The proof of this lemma is rather involved. It is based on results from functional analysis and measure theory; refer 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

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

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 use fractional Sobolev spaces $H^\theta(\Omega)$, with noninteger $\theta > 0$; 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., Lions and Magenes (1972)). Moreover, we consider L^*, the adjoint operator of L, defined as

$$L^*\hat{g} = -\mathrm{div}(\mathbf{K}\nabla \hat{g}) - \mathbf{b} \cdot \nabla \hat{g} + (c - \mathrm{div}(\mathbf{b}))\hat{g}. \tag{3.2}$$

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

$$\|\mathcal{B}\|_{L^2} \le C\sqrt{\lambda_n}, \tag{3.3}$$

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

$$\|\mathcal{B}\|_{L^2} = O(\lambda_n^{5/8}) \quad and \quad \|\mathcal{B}\|_{H^2} = O(\lambda_n^{1/8}). \tag{3.4}$$

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

$$\|\mathcal{B}\|_{L^2} = O(\lambda_n) \quad and \quad \|\mathcal{B}\|_{H^2} = O(\sqrt{\lambda_n}). \tag{3.5}$$

Proof. To lighten the notation we write $\lambda = \lambda_n$. Solving the minimization problem (2.3) is equivalent to finding $\hat{f}^*$, such that

$$\lambda \int_\Omega (L\hat{f}^* - u)Lv + \frac{1}{n}\sum_{i=1}^n \hat{f}^*(\mathbf{p}_i)v(\mathbf{p}_i) = \frac{1}{n}\sum_{i=1}^n f_0(\mathbf{p}_i)v(\mathbf{p}_i) \quad \forall v \in V_0. \tag{3.6}$$

Let us rewrite equation (3.6) in terms of $\mathcal{B}$. To this end, we subtract the quantity $\lambda \int Lf_0 Lv$ on both sides of (3.6), yielding

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

Then we add $\int \mathcal{B}v\,dF$ on both sides, obtaining

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

The equation above holds for all $v \in V_0$; in particular, we can set $v = \mathcal{B}$. Hence, from Assumption 1, we have:

$$\lambda \|L\mathcal{B}\|_{L^2}^2 + \kappa_1 \|\mathcal{B}\|_{L^2}^2 \le \lambda \int_\Omega (Lf_0 - u)L\mathcal{B} + \left\{\int_\Omega \mathcal{B}^2 d(F - F_n)\right\}. \qquad (3.7)$$

Now, owing to (3.1), we can write $\left|\int_\Omega \mathcal{B}^2 d(F - F_n)\right| \le cd_n \|\mathcal{B}\|_{H^2}^2$. Moreover, from 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., Evans (1998)). Because $\mathcal{B} \in V_0$, being $f_0 \in V_\gamma$, there exists a constant c_L, depending only on Ω and L, such that $c_L \|\mathcal{B}\|_{H^2}^2 \le \|L\mathcal{B}\|_{L^2}^2$. Using these two inequalities in (3.7), we obtain

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

From Assumption 2, for n large enough that $d_n\lambda^{-1} \le c_L/2c$, we get

$$\frac{c_L\lambda}{2} \|\mathcal{B}\|_{H^2}^2 + \kappa_1 \|\mathcal{B}\|_{L^2}^2 \le \lambda \int_\Omega (Lf_0 - u)L\mathcal{B}. \qquad (3.8)$$

Moreover,

$$\lambda \int_\Omega (Lf_0 - u)L\mathcal{B} \le \frac{\lambda}{2}\left(\frac{2}{c_L}\|Lf_0 - u\|_{L^2}^2 + \frac{c_L}{2}\|L\mathcal{B}\|_{L^2}^2\right)$$
$$\le \frac{\lambda}{c_L}\|Lf_0 - u\|_{L^2}^2 + \frac{\lambda c_L}{4}\|\mathcal{B}\|_{H^2}^2. \qquad (3.9)$$

The equation above, together with equation (3.8), leads to (3.3).

From (3.8), if $Lf_0 - u \in H^2(\Omega)$, using, in order, the first Green identity, Hölder inequality, trace theorems, interpolation between Sobolev spaces (see Lions and Magenes (1972)), and Young inequality, we get:

$$\frac{c_L\lambda}{2}\|\mathcal{B}\|_{H^2}^2 + \kappa_1 \|\mathcal{B}\|_{L^2}^2$$
$$\le \lambda \int_\Omega L^*(Lf_0 - u)\mathcal{B} + \lambda \int_{\partial\Omega} (\mathbf{K}\nabla\mathcal{B})\cdot n(Lf_0 - u) \qquad (3.10)$$
$$\le \lambda \|L^*(Lf_0 - u)\|_{L^2}\|\mathcal{B}\|_{L^2} + \lambda \|\mathbf{K}\|_{L^\infty(\partial\Omega)}\|\nabla\mathcal{B}\cdot n\|_{L^2(\partial\Omega)}\|Lf_0 - u\|_{L^2(\partial\Omega)}$$
$$\le \lambda \|L^*(Lf_0 - u)\|_{L^2}\|\mathcal{B}\|_{L^2} + c\lambda \|\mathcal{B}\|_{H^{3/2}(\Omega)}\|Lf_0 - u\|_{H^1(\Omega)}$$

$$\leq \lambda \|L^*(Lf_0 - u)\|_{L^2} \|\mathcal{B}\|_{L^2} + c\lambda \|\mathcal{B}\|_{L^2}^{1/4} \|\mathcal{B}\|_{H^2}^{3/4} \|Lf_0 - u\|_{H^1(\Omega)}$$

$$\leq \frac{\lambda^2}{2\kappa_1} \|L^*(Lf_0 - u)\|_{L^2}^2 + \frac{\kappa_1}{2} \|\mathcal{B}\|_{L^2}^2 + \qquad\qquad (3.11)$$

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

where c is a constant independent of n and λ, and that changes from line to line. From (3.11), we can write

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

Because $Lf_0 \in H^2(\Omega)$ 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 (3.4). Finally, when $(Lf_0 - u)|_{\partial\Omega} = 0$, from (3.10), we get the rates in (3.5).

3.2. Convergence of the variance term: infinite-dimensional estimator

In this section, we study the variance of the estimator $\hat{f}$ with respect to n and λ_n.

Theorem 2. *For all $0 < \delta \leq 1/2$ and n sufficiently large,*

$$\mathbb{V}ar_{L^2}(\hat{f}) = \mathbb{E}\left(\|\hat{w}\|_{L^2}^2\right) = O\left(\frac{\sigma^2}{n\lambda_n^{1/2+\delta}}\right), \qquad (3.12)$$

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

Proof. To lighten the notation we write $\lambda = \lambda_n$. Solving the minimization problem (2.4) is equivalent to finding $\hat{w}$, such that

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

or equivalently,

$$\lambda \int_\Omega L\hat{w}Lv + \int_\Omega \hat{w}v\,dF = \frac{1}{n} \sum_{i=1}^n \varepsilon_i v(\mathbf{p}_i) + \int_\Omega \hat{w}v\,d(F - F_n) \quad \forall v \in V_0. \quad (3.13)$$

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 $(\cdot,\cdot)_\lambda$. Because the norms $\|L\cdot\|_{L^2}$ and $\|\cdot\|_{H^2}$ are equivalent on V_0, there exists a constant c_L, such that

$$\|v\|_{H^2}^2 \le \frac{1}{c_L}\|Lv\|_{L^2}^2 \le \frac{1}{c_L\lambda}\left(\lambda\|Lv\|_{L^2}^2 + \int_\Omega v^2\,dF\right) = \frac{1}{c_L\lambda}\|v\|_\lambda^2. \qquad (3.14)$$

Define T, T_1, and T_2 as follows:

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

From the Sobolev embedding theorems (see, e.g., Lions and Magenes (1972, Thm. 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 (3.13) as

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

We have that

$$\|\hat{w}\|_\lambda = \sup_{v\in V_0}\frac{(\hat{w},v)_\lambda}{\|v\|_\lambda} = \sup_{v\in V_0}\frac{T(v)}{\|v\|_\lambda} \le \sup_{v\in V_0}\frac{T_1(v)}{\|v\|_\lambda} + \sup_{v\in V_0}\frac{T_2(v)}{\|v\|_\lambda}. \qquad (3.15)$$

For the first term on the right-hand side of (3.15), from equations (3.1) and (3.14), we have

$$\sup_{v\in V_0}\frac{T_1(v)}{\|v\|_\lambda} \le cd_n\sup_{v\in V_0}\frac{\|v\|_{H^2}\|\hat{w}\|_{H^2}}{\|v\|_\lambda} \le \tilde{c}d_n\lambda^{-1}\|\hat{w}\|_\lambda. \qquad (3.16)$$

For the second term on the right-hand side of (3.15), setting $\theta = 1 + 2\delta$, such that $T_2 \in \left(H^\theta(\Omega)\right)^*$, for $1 < \theta \le 2$, we have

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

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

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

$$\le c\lambda^{-\theta/4}\|T_2\|_{(H^\theta)^*},$$

where the last inequality is true from equation (3.14) and 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 (3.16), 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, from Assumption 2, we have that $d_n \lambda^{-1} = o(1)$. Therefore, the first part on the right-hand side of the above equation can be absorbed in the second term on the right-hand side, such 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

$$\mathbb{E}(\|\hat{w}\|_\lambda^2) \leq c \lambda^{-\theta/2} \mathbb{E}\left(\|T_2\|_{(H^\theta)^*}^2 \right). \tag{3.17}$$

To conclude the proof, it remains to show that $\mathbb{E}(\|T_2\|_{(H^\theta)^*}^2) \leq c \sigma^2/n$. 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 $\mathbf{p}_i$. From the Sobolev embedding theorems, $\delta_{\mathbf{p}_i} \in (H^\theta(\Omega))^*$. We denote as $(\cdot, \cdot)_{\theta,*}$ 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

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

$$= \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}\left(\varepsilon_i \varepsilon_j \right) (\delta_{\mathbf{p}_i}, \delta_{\mathbf{p}_j})_{\theta,*}$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \|\delta_{\mathbf{p}_i}\|_{(H^\theta)^*}^2 \leq \frac{c \sigma^2}{n},$$

where $c = \max_{i=1,\ldots,n} \|\delta_{\mathbf{p}_i}\|_{(H^\theta)^*}^2 < \infty$. From the previous equation and from (3.17), we have

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

Finally, from Assumption 1, we have that $\|w\|_{L^2} \leq \kappa_1^{-1} \|w\|_\lambda$, where κ_1 does not depend on λ or on n. This fact and the above equation lead to (3.12).

3.3. Convergence of the MSE: infinite-dimensional estimator

We finally consider the MSE of the estimator in the L^2 norm; that is,

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

The following theorem shows that the estimator $\hat{f}$ is consistent and that its MSE nearly achieves the optimal rate of convergence for nonparametric estimators (Stone (1982)), considering different Sobolev regularities of the true unknown field, $f_0 \in H^2(\Omega)$ and $f_0 \in H^4(\Omega)$. Specifically, 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*

$$MSE_{L^2} = O\left(n^{-2(1-\delta)/3}\right), \tag{3.18}$$

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

$$MSE_{L^2} = O\left(n^{-4(1-\delta/2)/5}\right), \tag{3.19}$$

for δ as small as desired.

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

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

which is minimized when $\lambda_n = n^{-2/3}$, leading to (3.18). Moreover, from equation (3.5), if $Lf_0 - u \in H^2(\Omega)$ and $Lf_0 - u|_{\partial\Omega} = 0$, we have

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

which is minimized when $\lambda_n = n^{-2/5}$, leading to (3.19).

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 the PDE. On the other hand, if our

problem knowledge is complete, and the PDE in the regularizing term offers a perfect description of the unknown field, $Lf_0 = u$, we expect to benefit in terms of both the estimation error and the rate of convergence of the MSE. Indeed, from equations (3.8) and (3.9), we get

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

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 is the bias. In particular, if $\|Lf_0 - u\|_{L^2} = 0$, the L^2-norm of the bias is zero. In this 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 the parametric estimators:

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

4. Numerical Solution to the SR-PDE Estimation Problem

The SR-PDE estimator defined in (2.1) cannot be computed analytically. Azzimonti et al. (2015) shows that solving (2.1) is equivalent to solving the following coupled system of PDEs:

$$\begin{cases} L\hat{f} = u + \hat{g} & \text{in } \Omega \\ \hat{f} = \gamma & \text{on } \partial\Omega \end{cases} \qquad \begin{cases} L^*\hat{g} = -\frac{1}{\lambda_n} \sum_{i=1}^n (\hat{f} - z_i)\delta_{\mathbf{p}_i} & \text{in } \Omega \\ \hat{g} = 0 & \text{on } \partial\Omega, \end{cases} \qquad (4.1)$$

where $\hat{g}$ represents the misfit of the penalized PDE, that is, $\hat{g} = L\hat{f} - u$, and L^* is the adjoint operator of L defined in (3.2). 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 $Lf_0 - u|_{\partial\Omega} = g_0|_{\partial\Omega} = 0$. Note, however, that when $Lf_0 - u|_{\partial\Omega} = 0$, we obtain the best rate of convergence in Theorem 1.

The reformulation (4.1) of the estimation problem (2.1) is convenient because it can be discretized easily using the finite-element method. We briefly recall the discretization (see, e.g., Azzimonti et al. (2015) 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

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

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

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

where $\boldsymbol{\psi} = (\psi_1, \dots, \psi_{N_h})^T$ is the vector of the basis functions, and $\mathbf{f} = (f(\xi_1), \dots, f(\xi_{N_h}))^T$ is the vector of coefficients. Define the bilinear form $a(\cdot, \cdot)$, associated with 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 evaluations of the basis functions at the data locations $\Psi_{ij} = \psi_j(\mathbf{p}_i)$. In addition, let $\boldsymbol{\psi}^D = (\psi_1^D, \dots, \psi_{N_h^D}^D)^T$ be the vector of basis functions associated with the boundary of the domain, and define $A_{ij}^D = a(\psi_j^D, \psi_i)$, $\Psi_{ij}^D = \psi_j^D(\mathbf{p}_i)$, and $\boldsymbol{\gamma}$ as the evaluation of the boundary condition γ at the boundary nodes. The coupled system of PDEs (4.1) is then discretized as follows

$$\begin{bmatrix} \dfrac{\Psi^T \Psi}{n} & \lambda_n A^T \\ A & -M \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \dfrac{\Psi^T \mathbf{z}}{n} - \dfrac{\Psi^T \Psi^D \boldsymbol{\gamma}}{n} \\ \mathbf{u} - A^D \boldsymbol{\gamma} \end{bmatrix}. \tag{4.2}$$

Azzimonti et al. (2014) 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 (4.2) 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 second order in $L^p(\Omega)$.

The finite-element SR-PDE estimator is thus obtained as

$$\hat{f}_h = \boldsymbol{\psi}^T \hat{\mathbf{f}} + (\boldsymbol{\psi}^D)^T \boldsymbol{\gamma}. \tag{4.3}$$

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

$$S = \Psi \left(\Psi^T \Psi + n\lambda_n P\right)^{-1} \Psi^T$$

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

5. Consistency of SR-PDE Estimator: Finite-Element Estimation Problem

We now prove 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 (2.1). As described above, the finite-element estimator is based on the discretization (4.2) of the reformulation (4.1), which consists of a coupled system of second-order differential problems, rather than the original fourth-order problem in (2.1). Because the results in Section 3 are all based on the forth-order problem (2.1), 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; that is, $Lf = -\mathrm{div}\left(\mathbf{K}\nabla f\right) + cf$.*

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

See Hjelle and Dæhlen (2006) 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 triangu-

lations $\{T_h\}_{h>0}$, let h_K and ρ_K be the diameter (longest edge) and the radius, respectively, 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 $\sigma_K = h_K/\rho_K \leq \sigma_0$, for all h and 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 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

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

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 $p > 2$, $f_0 \mid_{\partial\Omega} = \gamma$, $g_0 = Lf_0 - u \in H^1(\Omega)$, and $g_0|_{\partial\Omega} = 0$, then*

$$\|\mathcal{B}_h\|_n^2 \leq C \left(\frac{1}{n} + \lambda_n \right). \tag{5.1}$$

Proof. Azzimonti et al. (2014) shows that if $f_0 \in W^{2,p}(\Omega)$, for $p > 2$, $f_0 \mid_{\partial\Omega} = \gamma$, $g_0 = Lf_0 - u \in H^1(\Omega)$, and $g_0|_{\partial\Omega} = 0$, there exists $h_0 > 0$ such that for every $h \leq h_0$,

$$\|\mathcal{B}_h\|_n^2 \leq C \left\{ h^2 \left[(1 + \lambda_n) \|f_0\|_{W^{2,p}}^2 + \|Lf_0 - u\|_{H^1}^2 \right] + \lambda_n \|Lf_0 - u\|_{L^2}^2 \right\}. \tag{5.2}$$

Under Assumption 4, we have that $h^2 \approx 1/n$. Thus, for n sufficiently large, we obtain (5.1).

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 focus on $\mathbb{C}\mathrm{ov}(\hat{\mathbf{z}}) = \sigma^2 S S^T$, and consider its n-norm

$$\|\mathbb{C}\mathrm{ov}(\hat{\mathbf{z}})\|_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\mathbf{z}_i) = \frac{\sigma^2}{n} \mathrm{Tr}(S S^T). \tag{5.3}$$

We are thus interested in studying the eigenvalues of the matrix $S S^T$. Under Assumption 4, the matrix Ψ coincides 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 an $n \times n$ positive semi-definite symmetric matrix, and C is an $n \times n$ 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 Lu and Pearce (2000).

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,*

$$\zeta_{k,h} = O(k^2 h^2). \tag{5.4}$$

Proof. Because $P = A^T M^{-1} A = A^T M^{-1} M M^{-1} A$, 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

$$\mu_{1,h}\eta_{k,h}^2 \leq \zeta_{k,h} \leq \mu_{n,h}\eta_{k,h}^2. \tag{5.5}$$

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

$$\begin{cases} Lv = \eta v & \text{in } \Omega \subset \mathbb{R}^2 \\ v = \gamma & \text{on } \partial\Omega. \end{cases} \tag{5.6}$$

For self-adjoint operators L, the eigenvalues η_k are infinite and belong to $(a, +\infty)$, for some $a > 0$ and $\eta_k \sim k$ (see, e.g., Agmon (2010); Brezis (2010)).

The finite-element discretization of (5.6) on the triangulation T_h leads to the generalized eigenvalue problem $A\mathbf{v}_h = \eta_h M\mathbf{v}_h$. Because M is invertible, this is a classic eigenvalue problem, and 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., Boffi, Gardini and Gastaldi (2012); Kolata (1978)). More precisely, $\eta_k \leq \eta_{k,h} \leq \eta_k + ch^2$ (see Boffi (2010, Thm. 10.4). Therefore, for k (and thus n) sufficiently large, we have

$$\eta_{k,h} = O(k). \tag{5.7}$$

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

$$c_1 h^2 \leq \mu_{k,h} \leq c_2 h^2 \tag{5.8}$$

(see, e.g., Quarteroni and Valli (2008); Ern and Guermond (2013)). From (5.5), (5.7), and (5.8) we conclude that $\{\zeta_{k,h}\}$ increases at the same rate as $\{\eta_{k,h}^2\}$ with respect to k, leading to (5.4).

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

$$\|\mathbb{C}ov(\hat{\mathbf{z}})\|_n = O\left(\frac{1}{n\sqrt{\lambda_n}}\right). \tag{5.9}$$

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

$$\text{Tr}(SS^T) = \sum_{i=1}^{n}\left(\frac{1}{1+n\lambda_n\zeta_i}\right)^2. \tag{5.10}$$

From equations (5.3) and (5.10), from Lemma 3, we have

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

Recalling that under Assumption 4, $h^2 \approx n^{-1}$, we obtain equation (5.9). ∎

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 nonparametric estimators of $H^2(\Omega)$ functions.

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

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

Proof. From Theorems 4 and 5, we have

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

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

Remark 2. As we did 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 (5.2), we see that if $Lf_0 - u = 0$, the bias does not vanish, owing to the discretization error. However, when $Lf_0 - u = 0$,

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

where c is a constant that does not depend on n or λ_n. In this case, setting $\lambda_n = \lambda$, constant for all n, as in the infinite-dimensional case, we achieve a parametric rate of convergence for the MSE: $\mathrm{MSE}(\hat{f}_h) = O(n^{-1})$.

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 simulations that illustrate the rate of convergence for the infinite-dimensional estimator when the discretization size is small. In Section 6.2, we report 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 rates of convergence are illustrated using a log-log plot, with λ_n or n on the x-axis and the error on the y-axis.

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

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

Similarly to the previous 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 before $f_{0,3}$ vanishes on $\partial\Omega$; for this 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 illustrate the rate of convergence for the infinite-dimensional estimator. To this end, we consider a very fine discretization, consisting of a Delaunay triangulation with $N = 123103$ nodes; hence, we sample n data locations, with $n \leq N$, from the nodes of the mesh.
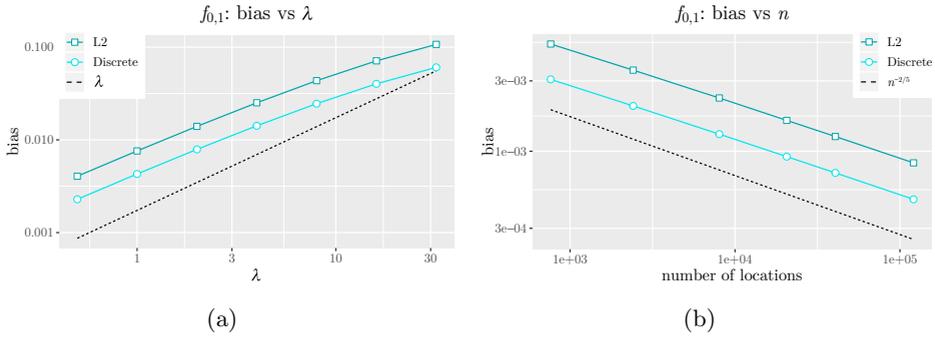
Figure 2. Test function $f_{0,1}$ without noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the bias of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the bias with respect to the number of points n, with $\lambda_n = n^{-2/5}$.
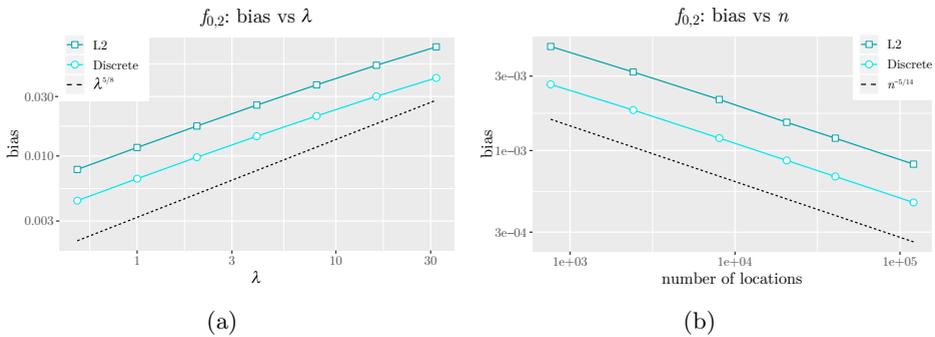

Figure 3. Test function $f_{0,2}$ without noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the bias of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the bias with respect to the number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

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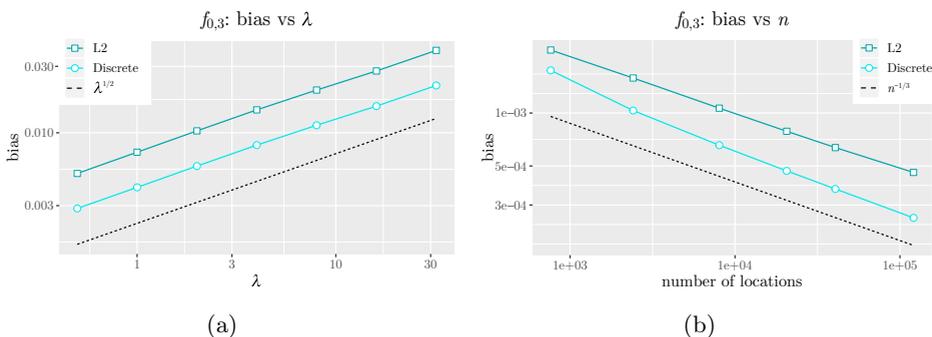
Figure 4. Test function $f_{0,3}$ without noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the bias of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the bias with respect to the number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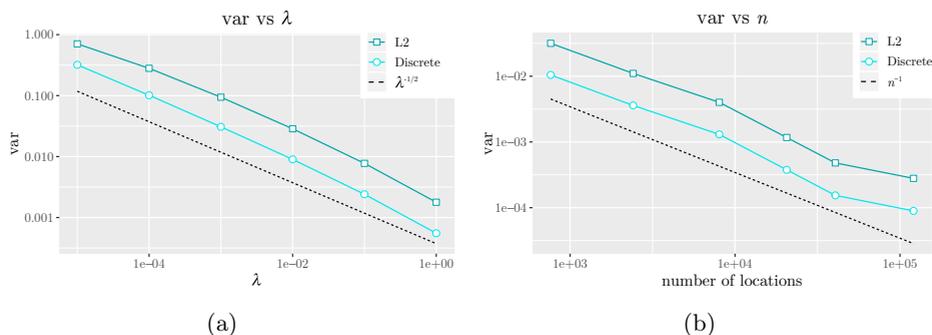

Figure 5. Pure Gaussian noise; fine and fixed triangulation. (a) Data sampled at each interior node. Convergence rates of the variance of the estimator with respect to λ_n. (b) Data sampled at an increasing number of interior nodes. Convergence rates of the variance with respect to the number of data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

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. These are the optimal settings 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 interior nodes of the mesh. We solve the estimation problem with $\lambda_n = 1, 10^{-1}, \ldots, 10^{-5}$, and we repeat the simulation 50 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 $\lambda_n^{-1/2}$.
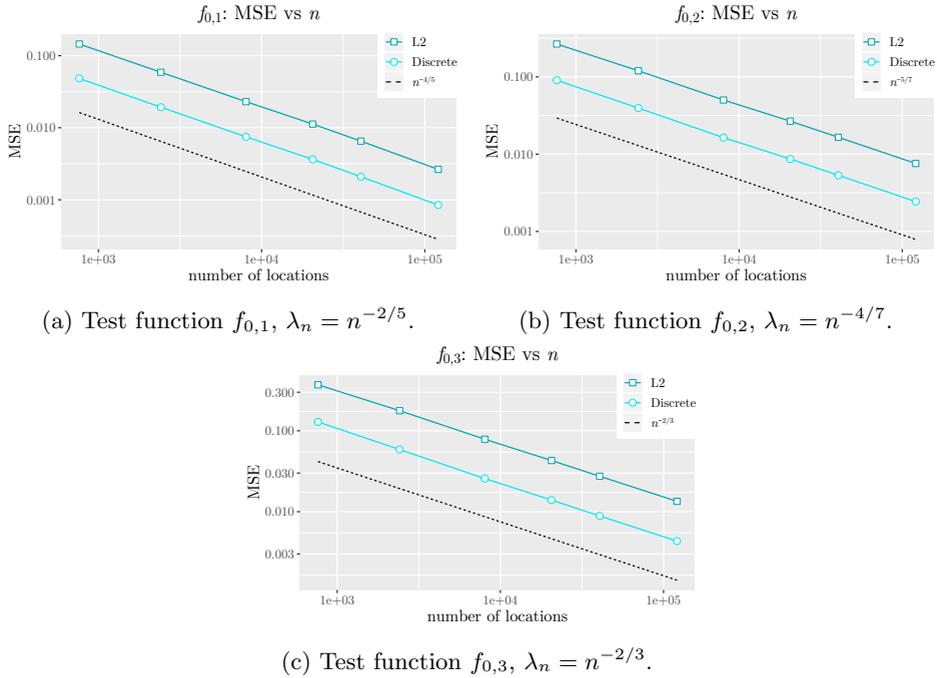
(a) Test function $f_{0,1}$, $\lambda_n = n^{-2/5}$.   (b) Test function $f_{0,2}$, $\lambda_n = n^{-4/7}$.

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

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

To illustrate the rate for the variance with respect to the number of observations, we proceed as for the bias, but consider an increasing number of observations n. We solve the estimation problem using a fixed $\lambda_n = 1$. We compute the mean over 50 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 illustrate the rate for the 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 test functions $f_{0,1}$, $f_{0,2}$, and $f_{0,3}$, respectively. We sample the data, adding Gaussian random noise with variance $\sigma^2 = 1$. We compute the mean estimate over 50 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 $n^{-2/3}$ for the three corresponding 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
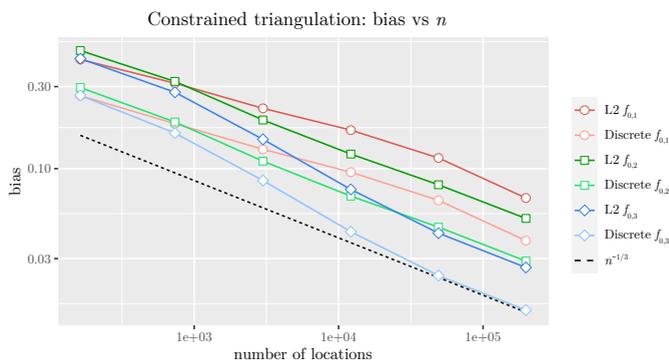
Figure 7. Three test functions without noise; constrained triangulations ($N = n$) of an increasing number of data locations. Convergence rates of the bias of the finite-element estimator with respect to the number of observations n, with $\lambda_n = n^{-2/3}$. The rates for $f_{0,1}$, $f_{0,2}$, and $f_{0,3}$ are shown in circles, squares, and diamonds, respectively.
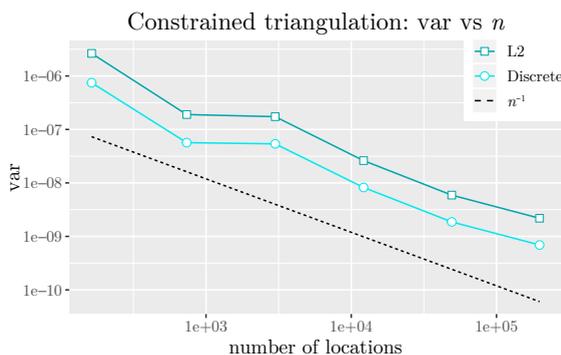

Figure 8. Pure Gaussian noise; constrained triangulations ($N = n$) of an increasing number of data locations. Convergence rates of the variance of the finite-element estimator with respect to the number of obervations n, with $\lambda_n = 1$.

200000 interior nodes. We sample the test functions $f_{0,1}, f_{0,2}$, and $f_{0,3}$ at all mesh nodes, such that $n = N$ and the mesh nodes coincide with data locations (constrained triangulation), as in Section 5. We illustrate the rates derived for the finite-element estimator in Section 5, and 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 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 $n^{-1/3}$ (i.e., to $\sqrt{\lambda_n}$) for all three functions, $f_{0,1}$, $f_{0,2}$, and $f_{0,3}$.

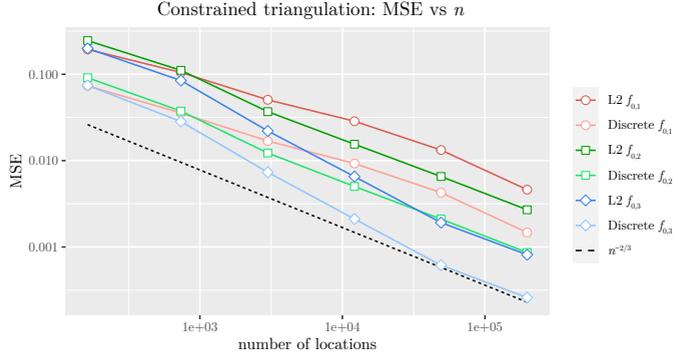We then look at the variance term. We consider pure noise data, sampling

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 $\lambda_n = n^{-2/3}$. The rates for $f_{0,1}$, $f_{0,2}$, and $f_{0,3}$ are shown in circles, squares, and diamonds, respectively.

Gaussian random noise with variance $\sigma^2 = 1$ at each interior mesh node. We solve the estimation problem using $\lambda_n = 1$. We compute the mean over 50 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 $f_{0,1}, f_{0,2}$, and $f_{0,3}$, adding 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

We here 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 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 the rate, achieving the near-optimal rate of convergence for H^4 functions. In future research, we intend to prove the consistency under more general mixed boundary conditions on the function, and possibly show that the estimators attain exactly the optimal rate.

We here also proved the consistency of the finite-element SR-PDE estimator.

In this case, we restricted our attention to self-adjoint operators L, that is, $\mathbf{b} = \mathbf{0}$, 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 also consistent for $\mathbf{b} \neq \mathbf{0}$. Moreover, to obtain the rate on the bias, we assumed exact boundary conditions on g_0. However, Azzimonti et al. (2014) show numerically that the extra error incurred if g_0 does not satisfy the imposed condition is of the same order as that 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. In 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 this assumption, which would allow us to relax the infill properties required on the data locations.

The infinite-dimensional SR-PDE estimation problem can also be solved using different numerical approaches and bases. For instance Wilhelm et al. (2016) use an isogeometric analysis based on non-uniform rational B-splines (NURBS). Examining 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 the space-time data defined in Arnone et al. (2019). Furthermore, we will investigate the consistency of SR-PDE estimators over the two-dimensional manifold domains defined in Ettinger, Perotto and Sangalli (2016), and of SR-PDE estimators in three-dimensional domains. These studies will require different arguments to those presented here. 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 Sangalli, Ramsay and Ramsay (2013). The latter setting includes space-varying covariate information, 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 $\mathbf{p}_i$. A particular interesting extension of this work would be to prove 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 Heckman (1986).

Acknowledgments

References

Agmon, S. (2010). *Lectures on Elliptic Boundary Value Problems*. AMS Chelsea Publications.

Arnone, E., Azzimonti, L., Nobile, F. and Sangalli, L. (2019). Modeling spatially dependent functional data via regression with differential regularization. *Journal of Multivariate Analysis* **170**, 275–295.

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

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

Bernardi, M., Carey, M., Ramsay, J. and Sangalli, L. (2018). Modeling spatial anisotropy via regression with partial differential regularization. *Journal of Multivariate Analysis* **167**, 15–30.

Boffi, D. (2010). Finite element approximation of eigenvalue problems. *Acta Numerica* **19**, 1–120.

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

Brezis, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York.

Cox, D. (1983). Asymptotics for M-type smoothing splines. *The Annals of Statististics* **11**, 530–551.

Cox, D. (1984). Multivariate smoothing spline functions. *SIAM Journal on Numerical Analysis* **21**, 798–813.

Cucker, F. and Zhou, D. (2007). *Learning Theory: An Approximation Theory Viewpoint*. Cambridge Monographs Applied & Computational Mathematics. Cambridge University Press.

Ern, A. and Guermond, J. (2013). *Theory and Practice of Finite Elements*. Springer, New York.

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

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

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*. CRC press.

Evans, L. (1998). *Partial Differential Equations*. American Mathematical Society.

Formaggia, L., Quarteroni, A. and Veneziani, A. (2010). *Cardiovascular Mathematics: Modeling and Simulation of the Circulatory System*. Springer, Milano.

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

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

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

Hjelle, Ø. and Dæhlen, M. (2006). *Triangulations and Applications.* Springer, Berlin, Heidelberg.

Huang, J. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics* **31**, 1600–1635.

Kolata, W. (1978). Approximation in variationally posed eigenvalue problems. *Numerische Mathematik* **29**, 159–171.

Lai, M.-J. and Schumaker, L. (2007). *Spline Functions on Triangulations.* Cambridge University Press.

Lai, M.-J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica* **23**, 1399–1417.

Lions, J. and Magenes, E. (1972). *Non-homogeneous Boundary Value Problems and Applications.* Springer, Berlin, Heidelberg.

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

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* Society for Industrial and Applied Mathematics, Philadelphia.

Quarteroni, A. and Valli, A. (2008). *Numerical Approximation of Partial Differential Equations.* Springer, Berlin, Heidelberg.

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

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

Scott-Hayward, L., MacKenzie, M., Donovan, C., Walker, C. and Ashe, E. (2014). Complex region spatial smoother (cress). *Journal of Computational and Graphical Statistics* **23**, 340–360.

Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics* **10**, 1040–1053.

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

Wilhelm, M., Dedè, L., Sangalli, L. and Wilhelm, P. (2016). IGS: An IsoGeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering* **302**, 70–89.

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

Eleonora Arnone

MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.

E-mail: eleonora.arnone@polimi.it

Alois Kneip

Universität Bonn, 53012 Bonn, Germany.

E-mail: akneip@uni-bonn.de

Fabio Nobile

CSQI - École polytechnique fédérale de Lausanne, Route Cantonale, 1015 Lausanne, Switzerland.

E-mail: fabio.nobile@epfl.ch

Laura M. Sangalli

MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.

E-mail: laura.sangalli@polimi.it