Mixed Finite Elements for spatial regression with PDE penalization

Laura Azzimonti[1] Fabio Nobile[1,2] Laura M. Sangalli[1]
Piercesare Secchi[1]

April 30, 2013

[1] MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy (e-mail: laura.azzimonti@mail.polimi.it)
[2] MATHICSE-CSQI, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Abstract

We study a class of models at the interface between statistics and numerical analysis. Specifically, we consider non-parametric regression models for the estimation of spatial fields from pointwise and noisy observations, that account for problem specific prior information, described in terms of a PDE governing the phenomenon under study. The prior information is incorporated in the model via a roughness term using a penalized regression framework. We prove the well-posedness of the estimation problem and we resort to a mixed equal order Finite Element method for its discretization. We prove the well posedness and the optimal convergence rate of the proposed discretization method. Finally the smoothing technique is extended to the case of areal data, particularly interesting in many applications.

Keywords: mixed Finite Element method, fourth order problems, non-parametric regression, smoothing.

1 Introduction

In this work we study the properties of a non-parametric regression technique for the estimation of bidimensional or three dimensional fields on bounded domains from some pointwise noisy evaluations. The technique is particularly well suited for applications in physics, engineering, biomedicine, etc., where a prior knowledge on the field might be available from physical principles and should be taken into account in the field estimation or smoothing process. We consider in particular phenomena where the field can be described by a partial differential equation (PDE) and has to satisfy some known boundary conditions.

Spatial regression with PDE penalization (SR-PDE) has been developed in [1] for the estimation of the blood velocity field on an artery section, from Echo-Doppler data. This technique has very broad applicability since PDEs are commonly used to describe phenomena behavior in many fields of physics, mechanics, biology and engineering. Many applications of particular interest can be named: the estimation of the concentration of pollutant released in water or in the air and transported by the stream or by the wind from noisy observations, the estimation of temperature or pressure fields from electronic control units or sensors in environmental sciences and many other phenomena in physical and biological sciences or engineering. In this work we focus on phenomena that are well described by linear second order elliptic PDEs, typically transport-reaction-diffusion problems.

SR-PDE uses a functional data analysis approach (see, e.g., [16]) and generalizes classical spatial smoothing techniques, such as thin-plate splines. SR-PDE, in fact, estimate the surface or the field minimizing a penalized least square functional, with the roughness penalty involving a partial differential operator. Many methods for surface estimation define the estimate as the minimizer of a penalized sum-of-square-error functional, with the penalty term involving a simple partial differential operator. Thin-plate spline smoothing, for example, penalizes an energy functional in $\mathbb{R}^2$ that involves second order derivatives. The minimizer of this functional belongs to the linear space generated by the Green's functions associated to the bilaplacian (see [21] for details). Thin-plate spline smoothing has been first extended to the case of bounded domains in [19], where the thin-plate energy is computed only over the bounded domain of interest. Since the minimizer cannot be directly characterized, it is approximated by a surface in the space of tensor product B-splines. Recently, more complex smoothing methods have been developed, that deal with general bounded domains in $\mathbb{R}^2$ and general boundary conditions. Some examples are soap-film smoothing, described in [22], and the spatial spline regression models described in [18], which generalize the Finite Element L-splines introduced in [17]. These methods estimate bidimensional surfaces on complex bounded domains penalizing the Laplace operator of the surface as a measure of the local curvature. Soap-film smoothing approximates the minimizer of the penalized least square functional with a linear combination of Green's functions of the bilaplacian on the domain of interest, centered on the vertices of a lattice. On the other hand Finite Element L-splines and spatial spline regression models solve directly the PDE associated to the penalized least square functional by means of a mixed Finite Element method.

Following the approach presented in [17] and [18], we propose to estimate the field minimizing a least square functional regularized with the L^2-norm over the domain of interest of the misfit of a second order PDE, $Lf = u$, modeling the phenomenon under study. The important novelty with respect to the methods cited above is that the problem-specific prior information, formalized in the PDE, is here used to model the phenomenon space variation. Furthermore, SR-PDE allows for important modeling flexibility in this direction, accounting for instance for space anisotropy and non-stationarity in a straightforward way. We assume here that all the parameters appearing in the operator L and the boundary conditions are known while the forcing term in the PDE is not completely determined. This approach is similar to the one used in control theory when a distributed control is considered; see for example [15]. The main analytic difference with respect with respect to classical results in control theory is that the observations are pointwise and affected by noise. For this reason it is necessary to require higher regularity to the field to ensure that the penalized least square functional is well defined.

The penalized least square functional has a unique minimum in the Sobolev space H^2 and the minimum is the solution of a fourth order problem. In order to prove the existence and the uniqueness of the estimator we resort to a mixed approach for fourth order problems, since the penalized error functional is not necessarily convex in H^2. Accordingly, a mixed equal order Finite Element method, similar to classical mixed methods described for example in [5], is used for discretizing the estimation problem. Other classical conforming and nonconforming methods (see [5] and references therein) or more recent discontinuous Galerkin methods (see, e.g., [3, 20, 12]) can be used for the discretization of the fourth order problem. However, in the specific case here considered the mixed Finite Element method is a convenient choice since the problem in exam can be written as a system of second order PDEs. Moreover the mixed approach provides also a good approximation of second order derivatives of the field that can be useful in order to compute physical quantities of interest.

The proposed mixed equal order Finite Elements discretization is known to have sub-optimal convergence rate when applied to fourth order problems with arbitrary boundary conditions and, in particular, the first order approximation might not converge to the exact solution (see, e.g., [4, 5]). However we are able to prove the optimal convergence of the proposed discretization method for the specific set of boundary conditions that are naturally associated to the smoothing problem, whenever the true underlying field satisfies exactly those conditions. The theoretical results are confirmed by numerical experiments.

The inspected convergence concerns the study of the bias of the estimator, while the study of the variance of the estimator and the convergence when the number of observations goes to infinity will be the subject of a future work. These topics are studied in the classical setting of smoothing splines (see, e.g., [6]), thin-plate splines or multidimensional smoothing splines (see, e.g., [7, 8, 13] and references therein) but they cannot be directly generalized to SR-PDE models.

The smoothing technique is also extended to the case of areal data, i.e., data that represent quantities computed on some subdomains; this data framework is frequent in many applications. For instance in the case of the driving problem considered in [1], which concerns the velocity field estimation from Echo-Doppler acquisitions, the data represent the mean velocity of blood on some subdomains on the considered artery section. The properties of the estimator in the areal setting are obtained along the same line followed for pointwise observations.

The paper is organized as follows. Section 2 introduces SR-PDE model used for pointwise observations. Section 3 proves the well-posedness of the estimation problem and Section 4 obtains a bound for the bias of the estimator. Section 5 describes the mixed Finite Element method used for the discretization of the estimation problem and proves the well-posedness of the discrete problem. Section 6 proves the convergence of the proposed mixed Finite Element method and provides a bound for the bias of the Finite Element estimator. Section 7 presents the numerical experiments supporting the theoretical results. Section 8 extends the models to the case of areal data and presents the asymptotic results in this setting. Section 9 outlines future research directions.

2 Surface estimator for pointwise data

Consider a bounded, regular, open domain $\Omega \subset \mathbb{R}^d$ with $d \leq 3$, whose boundary $\partial\Omega$ is a curve of class C^2, and a regular function $f_0 : \Omega \to \mathbb{R}$ to be estimated from noisy observations. Let z_i, for $i = 1, \ldots, n$, be n observations that represent noisy evaluations of the field f_0 at points $\mathbf{p}_i \in \Omega$. The error model that we consider for the observations is a classical additive model:

$$z_i = f_0(\mathbf{p}_i) + \epsilon_i \tag{1}$$

where ϵ_i, $i = 1, \ldots, n$, are independent errors with zero mean and constant variance σ^2. This model is a classical framework used in functional data analysis, see for example [16].

We suppose to have, in addition to the observations z_i, a physical knowledge of the phenomenon under study and that this prior knowledge can be described by means of a differential operator. Specifically, we can formalize this as a PDE that f_0 satisfies:

$$\begin{cases} Lf_0 = \tilde{u} & in\ \Omega \\ \mathcal{B}_c f_0 = h & on\ \partial\Omega \end{cases} \tag{2}$$

where the operator L and the boundary conditions are completely determined and fixed, while the forcing term $\tilde{u} = u + g_0 \in L^2(\Omega)$ is composed by a known and fixed part u and an unknown term, called g_0, that will be estimated from data. The parameters of the PDE and the boundary conditions could be as well considered partly unknown and estimated from data, but in this work we assume them to be known and fixed. We focus on second order elliptic operators, in particular L is a diffusion-transport-reaction operator

$$Lf_0 = -\mathrm{div}(\mathbf{K}\nabla f_0) + \mathbf{b} \cdot \nabla f_0 + cf_0 \tag{3}$$

with smooth and bounded parameters. The matrix $\mathbf{K} \in \mathbb{R}^{d\times d}$ is a symmetric and positive definite diffusion tensor, $\mathbf{b} \in \mathbb{R}^d$ is the transport vector and $c \geq 0$ is the reaction term. These parameters can be spatially varying in Ω; i.e., $\mathbf{K} = \mathbf{K}(\mathbf{x})$, $\mathbf{b} = \mathbf{b}(\mathbf{x})$ and $c = c(\mathbf{x})$, with $\mathbf{x} \in \Omega$. The boundary conditions of the PDE are homogeneous or non-homogeneous Dirichlet, Neumann, Robin (or mixed) conditions. All the admissible boundary conditions are summarized in

$$\mathcal{B}_c f_0 = \begin{cases} f_0 & \text{on } \Gamma_D \\ \mathbf{K}\nabla f_0 \cdot \boldsymbol{\nu} & \text{on } \Gamma_N \\ \mathbf{K}\nabla f_0 \cdot \boldsymbol{\nu} + \gamma f_0 & \text{on } \Gamma_R \end{cases} \qquad h = \begin{cases} h_D & \text{on } \Gamma_D \\ h_N & \text{on } \Gamma_N \\ h_R & \text{on } \Gamma_R \end{cases} \tag{4}$$

where $\boldsymbol{\nu}$ is the outward unit normal vector to $\partial\Omega$, $\gamma \in \mathbb{R}$ is a positive constant and $\partial\Omega = \bar{\Gamma}_D \cup \bar{\Gamma}_N \cup \bar{\Gamma}_R$, with $\Gamma_D, \Gamma_N, \Gamma_R$ not overlapping.

In what follows, we make the following assumption.

Assumption 1. $\Gamma_D \neq \emptyset$, *so that a Poincaré inequality holds, i.e.,*

$$\|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)} . \tag{5}$$

In order to estimate the field f_0, starting from the observations $z_1, \dots, z_n$ and using the a priori knowledge on the phenomenon, we propose to minimize the penalized sum-of-square-error functional

$$J(f) = \frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{p}_i) - z_i)^2 + \lambda \int_{\Omega} (Lf - u)^2 \tag{6}$$

over the set of functions $V = \{v \in H^2(\Omega) : \mathcal{B}_c v = h\}$ where $H^2(\Omega)$ is the Sobolev space of functions in $L^2(\Omega)$ with first and second derivatives in $L^2(\Omega)$; notice that the boundary conditions (4) are imposed directly in V. Even if in this case we are considering fixed and deterministic boundary conditions, when data on the boundary are available it is possible to include the uncertainty on the boundary conditions in the model including in the least square functional a dedicated regularizing term. The functional J is composed by a data fitting criterion, consisting in classical least square errors, and a model fitting criterion, formalized as a roughness term that penalizes the misfit of a PDE governing the phenomenon. Notice that by minimizing the misfit of the PDE $Lf_0 - u$, where u is the known part of the forcing term, we are actually minimizing the contribution of the unknown forcing term g_0. The contribution of the data fitting criterion and of the model fitting criterion is tuned by means of the parameter λ. A large literature is devoted to the optimal choice of this parameter; see, e.g., [14, 16] and references therein; classical methods are for example the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Generalized Cross-Validation (GCV) criterion. See [1] for details on the GCV computation in these models.

The functional $J(f)$ is well defined if $f \in H^2(\Omega)$ thanks to the embedding $H^2(\Omega) \subset C(\bar{\Omega})$ if $\Omega \subset \mathbb{R}^d$ with $d \leq 3$. For data in $\mathbb{R}^d$ with $d > 3$ one has to require more regularity in order to obtain $f \in C(\Omega)$; in particular one needs $f \in H^s(\Omega)$ with $s > d/2$; see, e.g., [11].
The estimation problem is formulated as follows.

Problem 1. *Find $\hat{f} \in V$ such that*

$$\hat{f} = \operatorname*{argmin}_{f \in V} J(f).$$

As it will be shown in the next section, this problem is well posed if we assume some regularity on the parameters of the PDE and on the domain Ω. In particular, in the case $d \leq 3$, we make the following assumption.

Assumption 2. *The parameters of the PDE are such that $\forall \tilde{u} \in L^2(\Omega)$ there exists a unique solution f_0 of the PDE (2), which moreover satisfies $f_0 \in H^2(\Omega)$.*

The Lax-Milgram theorem guarantees the existence and the uniqueness of the solution of the PDE (2) in $H^1(\Omega)$ when the parameters of the PDE $\mathbf{K}$, $\mathbf{b}$ and c satisfy some classical requests, for example $\mathbf{K}_{ij},\ \mathbf{b}_j,\ c \in L^\infty(\Omega)$, $\mathbf{K}$ is symmetric and uniformly elliptic, i.e., $\boldsymbol{\xi}^T \mathbf{K}(\mathbf{x}) \boldsymbol{\xi} \geq \alpha_K\ \forall \mathbf{x} \in \Omega$ and $\forall \boldsymbol{\xi} \in \mathbb{R}^d$, $\mathbf{b} \cdot \boldsymbol{\nu} \geq 0$ on $\Gamma_N \cup \Gamma_R$, $-1/2\mathrm{div}(\mathbf{b}(\mathbf{x})) + c(\mathbf{x}) \geq -\alpha_K / C_P$, where α_K is the ellipticity constant and C_P is the Poincaré constant, $\gamma \in L^\infty(\partial\Omega)$, $\gamma \geq 0$ and $h_D \in H^{1/2}(\partial\Omega)$, $h_N \in H^{-1/2}(\partial\Omega)$, $h_R \in H^{-1/2}(\partial\Omega)$.
To guarantee that the solution of the PDE is in $H^2(\Omega)$ we need to make further

assumptions on the parameters of the PDE and on the boundary conditions requiring extra regularity: $\mathbf{K}_{ij}$ is Lipschitz continuous, $h_D \in H^{3/2}(\partial\Omega)$, $h_N \in H^{1/2}(\partial\Omega)$, $h_R \in H^{1/2}(\partial\Omega)$. If the boundary conditions imposed are mixed, they have to satisfy some joint conditions in order not to reduce the regularity of the solution; see [11] for more details.

3 Well posedness analysis

To analyze the well-posedness of Problem 1 we introduce a new quantity $g \in \mathcal{G} = L^2(\Omega)$ that represents the misfit of the PDE in the penalizing term. This new quantity, $g \in \mathcal{G}$, is defined as $g = Lf - u$, where L is the second order elliptic operator (3), and is the classical *control* term in PDE optimal control theory.
It is useful to introduce also the space $V_0 = \{v \in V : \mathcal{B}_c v = 0\}$, which represents the space of functions in V with homogeneous boundary conditions, and the operator $B : L^2(\Omega) \to V_0$ such that $B\tilde{u}$ is the unique solution of the PDE (2) with forcing term $\tilde{u}$ and homogeneous boundary conditions, i.e., $L(B\tilde{u}) = \tilde{u}$ in Ω and $\mathcal{B}_c(B\tilde{u}) = 0$ on $\partial\Omega$. Under Assumptions 1 and 2, thanks to the well-posedness and the H^2-regularity of the PDE (2), the operator B is an isomorphism between the spaces L^2 and V_0 and the H^2-norm of Bu is equivalent to the L^2-norm of u, i.e., there exist two positive constants C_1 and C_2 such that

$$C_1 \|u\|_{L^2(\Omega)} \leq \|Bu\|_{H^2(\Omega)} \leq C_2 \|u\|_{L^2(\Omega)}. \tag{7}$$

The solution of the PDE (2) can thus be written as $f = f_b + B\tilde{u}$ where f_b is the solution of the PDE with homogeneous forcing term and non-homogeneous boundary conditions.
Existence and uniqueness of the estimator $\hat{f}$ is obtained thanks to classical results of calculus of variations. We recall here the result stated, e.g., in [15].

Theorem 1. *If the functional $J(g)$ has the form*

$$J(g) = \mathcal{A}(g, g) + \mathcal{L}g + c \tag{8}$$

where $\mathcal{A} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ is a continuous, coercive and symmetric bilinear form in $\mathcal{G}$, $\mathcal{L} : \mathcal{G} \to \mathbb{R}$ is a linear operator, c is a constant and $\mathcal{G}$ is a Hilbert space, then $\exists!\ \hat{g} \in \mathcal{G}$ such that $J(\hat{g}) = \inf_{\mathcal{G}} J(g)$.
Moreover $\hat{g}$ satisfies the following Euler-Lagrange equation:

$$(J'(\hat{g}), \varphi) = 2\mathcal{A}(\hat{g}, \varphi) + \mathcal{L}\varphi = 0 \quad \forall \varphi \in \mathcal{G}. \tag{9}$$

The existence and uniqueness of the estimator is stated in the following theorem.

Theorem 2. *Under Assumptions 1-2, the solution of Problem 1 exists and is unique.*

Proof. Thanks to the definition of g we can write f as an affine transformation of g, i.e., $f = f_b + B(u + g)$, and the functional (6) as

$$J_g(g) = J(f_b + B(u+g)) = \frac{1}{n}\sum_{i=1}^{n} \left(B(u+g)(\mathbf{p}_i) + f_b(\mathbf{p}_i) - z_i\right)^2 + \lambda \|g\|^2_{L^2(\Omega)}. \tag{10}$$

This reformulation of the functional J is very useful since we can now write J_g in the quadratic form (8) where

$$\begin{aligned}\mathcal{A}(g,\varphi) =& \frac{1}{n}\sum_{i=1}^{n} Bg(\mathbf{p}_i)B\varphi(\mathbf{p}_i) + \lambda\int_\Omega g\varphi \\ \mathcal{L}\varphi =& \frac{2}{n}\sum_{i=1}^{n} B\varphi(\mathbf{p}_i)(Bu(\mathbf{p}_i) + f_b(\mathbf{p}_i) - z_i) \\ c =& \frac{1}{n}\sum_{i=1}^{n}(Bu(\mathbf{p}_i) + f_b(\mathbf{p}_i) - z_i)^2.\end{aligned}$$

Clearly $\mathcal{A}(g,\varphi)$ is a bilinear form, since both B and the pointwise evaluation of a function are linear operators. Moreover, it is continuous in $\mathcal{G}$; indeed, thanks to the embedding $H^2(\Omega) \subset C(\bar{\Omega})$ if $\Omega \subset \mathbb{R}^d$ with $d \leq 3$ and thanks to (7) we have that

$$|Bg(\mathbf{p}_i)| \leq \|Bg\|_{C(\bar{\Omega})} \leq C\,\|Bg\|_{H^2(\Omega)} \leq C\,\|g\|_{L^2(\Omega)}\,.$$

We thus obtain that $\mathcal{A}(g,\varphi) \leq (C^2+\lambda)\,\|g\|_{L^2(\Omega)}\,\|\varphi\|_{L^2(\Omega)}$.
Finally, the operator $\mathcal{A}(g,\varphi)$ is coercive in $L^2(\Omega)$, since

$$\mathcal{A}(g,g) = \frac{1}{n}\sum_{i=1}^{n}|Bg(\mathbf{p}_i)|^2 + \lambda\int_\Omega g^2 \geq \lambda\int_\Omega g^2 = \lambda\,\|g\|^2_{L^2(\Omega)}\,.$$

Due to the fact that the bilinear form $\mathcal{A}(\cdot,\cdot)$ is continuous and coercive in $\mathcal{G} = L^2(\Omega)$, that the operator $\mathcal{L}$ is linear and that c is a constant, Theorem 1 states the existence and the uniqueness of $\hat{g} = \operatorname{argmin}_{g\in\mathcal{G}} J_g(g)$. From the bijectivity of $B : L^2(\Omega) \to V_0$ we deduce the existence and uniqueness of $\hat{f} = f_b + B(\hat{g}+u) = \operatorname{argmin}_{f\in V} J(f)$. □

The estimator $\hat{f}$ is obtained by solving:

$$\begin{cases} L\hat{f} = u + \hat{g} & \textit{in } \Omega \\ \mathcal{B}_c\hat{f} = h & \textit{on } \partial\Omega. \end{cases} \tag{11}$$

We now show that if $\hat{g}$ is smooth enough, e.g., $\hat{g} \in H^2(\Omega)$, then $\hat{g}$ solves the PDE

$$\begin{cases} L^*\hat{g} = -\frac{1}{n\lambda}\sum_{i=1}^{n}(\hat{f} - z_i)\delta_{\mathbf{p}_i} & \textit{in } \Omega \\ \mathcal{B}_c{}^*\hat{g} = 0 & \textit{on } \partial\Omega \end{cases} \tag{12}$$

where $\delta_{\mathbf{p}_i}$ is the Dirach mass located in $\mathbf{p}_i$, L^* is the adjoint operator of L

$$L^*g = -\operatorname{div}(\mathbf{K}\nabla g) - \mathbf{b}\cdot\nabla g + (c - \operatorname{div}(\mathbf{b}))g \tag{13}$$

and the "adjoint" boundary conditions are

$$\mathcal{B}_c{}^*g = \begin{cases} g & \text{on } \Gamma_D \\ \mathbf{K}\nabla g\cdot\boldsymbol{\nu} + \mathbf{b}\cdot\boldsymbol{\nu} g & \text{on } \Gamma_N \\ \mathbf{K}\nabla g\cdot\boldsymbol{\nu} + (\mathbf{b}\cdot\boldsymbol{\nu} + \gamma)g & \text{on } \Gamma_R. \end{cases} \tag{14}$$

Indeed, if we associate to any $\varphi \in \mathcal{G}$ the function $v \in V_0$ such that $v = B\varphi$ (or equivalently $Lv = \varphi$ with homogeneous boundary conditions), we have that

$$\begin{aligned}
\frac{1}{2}\left(J_g'(\hat{g}), \varphi\right) &= \int_\Omega \frac{1}{n}\sum_{i=1}^n \left(\hat{f} - z_i\right) v\delta_{\mathbf{p}_i} + \lambda \int_\Omega \hat{g} L v \\
&= \int_\Omega \frac{1}{n}\sum_{i=1}^n \left(\hat{f} - z_i\right) v\delta_{\mathbf{p}_i} + \lambda \int_\Omega L^*\hat{g} v + \int_{\partial\Omega} \hat{g}\mathcal{B}_c v + \int_{\partial\Omega} \mathcal{B}_c{}^*\hat{g} v \\
&= \int_\Omega \left[\frac{1}{n}\sum_{i=1}^n \left(\hat{f} - z_i\right) \delta_{\mathbf{p}_i} + \lambda L^*\hat{g}\right] B\varphi + \int_{\partial\Omega} \mathcal{B}_c{}^*\hat{g} B\varphi = 0 \quad \forall \varphi \in \mathcal{G}.
\end{aligned}$$

Since B is a bijection between $L^2(\Omega)$ and V_0 the latter is equivalent to

$$\int_\Omega \left[\frac{1}{n}\sum_{i=1}^n \left(\hat{f} - z_i\right) \delta_{\mathbf{p}_i} + \lambda L^*\hat{g}\right] v + \int_{\partial\Omega} \mathcal{B}_c{}^*\hat{g} v = 0 \quad \forall v \in V_0. \tag{15}$$

Choosing $v \in C^\infty$ with compact support in Ω, equation (15) implies that $\hat{g}$ is the solution in the sense of distributions of

$$L^*\hat{g} = -\frac{1}{n\lambda}\sum_{i=1}^n (\hat{f} - z_i)\delta_{\mathbf{p}_i}.$$

Choosing $v \in C^\infty(\bar{\Omega})$ with support not including any of the location points $\mathbf{p}_i$, we obtain from equation (15) the boundary conditions for $\hat{g}$ that are $\mathcal{B}_c{}^*\hat{g} = 0$ where $\mathcal{B}_c{}^*$ is defined in equation (14).
In this case, the estimator $\hat{f}$ is obtained by solving the coupled system of PDEs

$$\begin{cases} L\hat{f} = u + \hat{g} & in\ \Omega \\ \mathcal{B}_c\hat{f} = h & on\ \partial\Omega \end{cases} \qquad \begin{cases} L^*\hat{g} = -\frac{1}{n\lambda}\sum_{i=1}^n (\hat{f} - z_i)\delta_{\mathbf{p}_i} & in\ \Omega \\ \mathcal{B}_c{}^*\hat{g} = 0 & on\ \partial\Omega. \end{cases} \tag{16}$$

4 Bias of the estimator

The penalty term in the functional $J(f)$ induces a bias in the estimator $\hat{f}$ unless the unknown part of the forcing term $g_0 = 0$ and the true underlying field f_0 satisfies exactly the penalized PDE $Lf_0 = u$; we want now to quantify this bias. The estimator $\hat{f}$ is obtained as the unique minimum of the functional $J(f)$, solving the Euler-Lagrange equation (9). Thanks to the linearity of equation (9), we can write

$$\begin{aligned}
\hat{f} =& \operatorname*{argmin}_{f \in V} \left[\frac{1}{n}\sum_{i=1}^n \left(f(\mathbf{p}_i) - f_0(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega (Lf - u)^2\right] \\
&+ \operatorname*{argmin}_{w \in V_0} \left[\frac{1}{n}\sum_{i=1}^n \left(w(\mathbf{p}_i) - \epsilon_i\right)^2 + \lambda \int_\Omega (Lw)^2\right]
\end{aligned}$$

where $f_0(\mathbf{p}_i)$ is the mean value of the observation z_i located in $\mathbf{p}_i$, i.e., $\mathbb{E}[z_i] = f_0(\mathbf{p}_i)$. The first term is the deterministic part of $\hat{f}$, while the second term

$$\hat{w} = \operatorname*{argmin}_{w \in V_0} \left[\frac{1}{n}\sum_{i=1}^n \left(w(\mathbf{p}_i) - \epsilon_i\right)^2 + \lambda \int_\Omega (Lw)^2\right] \tag{17}$$

is related to the observation noise: $\hat{w}$ is in fact the minimizer of the functional when data are pure noise and the penalized PDE is homogeneous (both the forcing term and the boundary conditions are homogeneous). Notice that $\hat{w}$ is a linear function of the errors ϵ_i and for this reason it has zero mean. Indeed, $\hat{w}$ is obtained as the solution of the PDE

$$\begin{cases} L\hat{w} = \hat{g}_w & in\ \Omega \\ \mathcal{B}_c{}^*\hat{w} = 0 & on\ \partial\Omega \end{cases} \tag{18}$$

where $\hat{g}_w$ satisfies

$$\begin{cases} L^*\hat{g}_w = -\frac{1}{n\lambda}\sum_{i=1}^n(\hat{w} - \epsilon_i)\delta_{\mathbf{p}_i} & in\ \Omega \\ \mathcal{B}_c{}^*\hat{g}_w = 0 & on\ \partial\Omega. \end{cases} \tag{19}$$

Moreover, thanks to the PDE (18) we know that $\mathbb{E}[\hat{w}] = B\mathbb{E}[\hat{g}_w]$, while from the PDE (19) we obtain that

$$L^*\mathbb{E}[\hat{g}_w] + \frac{1}{n\lambda}\sum_{i=1}^n B\mathbb{E}[\hat{g}_w(\mathbf{p}_i)]\delta_{\mathbf{p}_i} = \frac{1}{n\lambda}\sum_{i=1}^n \mathbb{E}[\epsilon_i]\delta_{\mathbf{p}_i} = 0.$$

Finally, since L^*, B and the evaluation in a point are linear operators, we have that both $\hat{g}_w$ and $\hat{w}$ have zero mean. It follows then that the mean value of the estimator $\mathbb{E}[\hat{f}]$ is the minimizer of the functional when the observations are without noise and the PDE is not homogeneous (both the forcing term and the boundary conditions are not homogeneous), i.e.,

$$\mathbb{E}[\hat{f}] = \underset{f\in V}{\operatorname{argmin}}\left[\frac{1}{n}\sum_{i=1}^n (f(\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda\int_\Omega (Lf - u)^2\right]. \tag{20}$$

Notice that the sum of the Euler-Lagrange equations associated to the functionals (17) and (20) corresponds to the Euler-Lagrange equation (9) associated to the functional (6).

Since $\mathbb{E}[\hat{f}]$ is related to the bias induced by the penalizing term, we are interested in studying the error term $\mathbb{E}[\hat{f} - f_0]$; in particular it is natural to study it in the norm induced by the functional $J(f)$, i.e.,

$$\|f\|_J^2 = \frac{1}{n}\sum_{i=1}^n f(\mathbf{p}_i)^2 + \lambda\int_\Omega (Lf)^2. \tag{21}$$

Lemma 1. *The norm* (21) *of the bias of* $\hat{f}$ *is bounded by*

$$\left\|\mathbb{E}[\hat{f} - f_0]\right\|_J^2 \le 4\lambda\,\|Lf_0 - u\|_{L^2(\Omega)}^2. \tag{22}$$

Proof. In order to obtain the inequality (22) we can use the optimality (20) of $\mathbb{E}[\hat{f}]$ in the minimization of the functional with respect to any other function in V. We have in fact that

$$\left\|\mathbb{E}[\hat{f}] - f_0\right\|_J^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{f}](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda \left\|L(\mathbb{E}[\hat{f}] - f_0)\right\|_{L^2(\Omega)}^2$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{f}](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + 2\lambda \left\|L\mathbb{E}[\hat{f}] - u\right\|_{L^2(\Omega)}^2 + 2\lambda \left\|Lf_0 - u\right\|_{L^2(\Omega)}^2$$
$$\leq 2\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{f}](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda \left\|L\mathbb{E}[\hat{f}] - u\right\|_{L^2(\Omega)}^2\right] + 2\lambda \left\|Lf_0 - u\right\|_{L^2(\Omega)}^2$$
$$\leq 2\lambda \left\|Lf_0 - u\right\|_{L^2(\Omega)}^2 + 2\lambda \left\|Lf_0 - u\right\|_{L^2(\Omega)}^2 .$$

□

This result means that the estimator is asymptotically unbiased in the norm $\|\cdot\|_J$ either if $\|Lf_0 - u\|_{L^2(\Omega)} = 0$ or if $\lambda \to 0$ for $n \to +\infty$. The condition $\|Lf_0 - u\|_{L^2(\Omega)} = 0$ means that the real field f_0 is in the kernel of the penalty term, while the condition $\lambda \to 0$ for $n \to +\infty$ means that the more observations we have, the less we penalize the PDE misfit.

5 Finite Element estimator

The estimation problem presented in Section 2 is infinite dimensional and cannot be solved analytically. To reduce this infinite dimensional problem to a finite dimensional one we approximate the PDE system (16) with the Finite Element method; this method has already been used in this framework for example in [17], [18] and [10]. The Finite Element approximation of the system (16) can be regarded as a naive mixed Finite Element method for the discretization of Problem 1. More complex methods for the discretization of fourth order problems could be used: in [5], for example, some conforming and nonconforming methods for the discretization of fourth order problems are introduced, while in [3, 20, 12] more recent discontinuous Galerkin methods are described.

Let $\mathcal{T}_h$ be a regular and quasi-uniform triangulation of the domain, that for convenience we assume here to be polygonal and convex, and $h = \max_{K\in\mathcal{T}_h} \operatorname{diam}(K)$ be the characteristic mesh size (see, e.g., [2]). Notice that the mesh $\mathcal{T}_h$ can be defined independently of the location of the observations $\mathbf{p}_1, \cdots, \mathbf{p}_n$. We consider the space V_h^r of piecewise continuous polynomial functions of degree $r \geq 1$ on the triangulation

$$V_h^r = \left\{v \in C^0(\bar{\Omega}) : v|_K \in \mathbb{P}^r(K)\ \forall K \in \mathcal{T}_h\right\}$$

and $V_{h,\Gamma_D}^r = V_h^r \cap H_{\Gamma_D}^1(\Omega)$ where $H_{\Gamma_D}^1 = \left\{v \in H^1(\Omega) :\ v|_{\Gamma_D} = 0\right\}$.
In order to discretize the PDE system (16) we define the bilinear forms

$$r(g, v) = \int_\Omega gv, \qquad l(f, \psi) = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{p}_i)\psi(\mathbf{p}_i),$$

$$a(f, \psi) = \int_\Omega (\mathbf{K}\nabla f \cdot \nabla\psi + \mathbf{b} \cdot \nabla f\psi + cf\psi) + \int_{\Gamma_R} \gamma f\psi, \tag{23}$$

the latter being the bilinear form associated to the operator L; we also introduce the linear operator $F(\psi) = \int_\Omega u\psi + \int_{\Gamma_N} h_N\psi + \int_{\Gamma_R} h_R\psi$.

Let now $f_{D,h} \in V_h^r$ be a lifting of the non-homogeneous Dirichlet conditions, i.e, $f_{D,h}|_{\Gamma_D} = h_{D,h}$, where $h_{D,h}$ is the interpolant of h_D in the space of piecewise continuous polynomial functions of degree r on the Dirichlet boundary Γ_D. The Finite Element approximation of the system (16) becomes

$$\begin{cases} \frac{1}{\lambda} l(\hat{f}_h, \psi_h) + a(\psi_h, \hat{g}_h) = \frac{1}{n\lambda} \sum_{i=1}^n z_i \psi_h(\mathbf{p}_i) & \forall \psi_h \in V_{h,\Gamma_D}^r \\ a(\hat{f}_h, v_h) - r(\hat{g}_h, v_h) = F(v_h) & \forall v_h \in V_{h,\Gamma_D}^r \end{cases} \tag{24}$$

with $(\hat{f}_h - f_{D,h}, \hat{g}_h) \in V_{h,\Gamma_D}^r \times V_{h,\Gamma_D}^r$.

In this Section and in the following one, we need a slightly stronger regularity assumption on the PDE (2), in particular we need its solution to be in a Sobolev space $W^{2,p}$, where $W^{s,p}(\Omega)$ is the space of functions in $L^p(\Omega)$ with derivatives up to order s in $L^p(\Omega)$.

Assumption 3. *The parameters of the PDE are such that $\forall \tilde{u} \in L^p(\Omega)$ there exists a unique solution $f_0 \in W^{2,p}(\Omega)$, for some $p > d$.*

Lemma 2. *Under Assumption 3, there exists $h_0 > 0$ s.t. $\forall h \leq h_0$, Problem* (24) *has a unique solution.*

Proof. The proof mimics the strategy used to prove the existence and the uniqueness of the estimator at the continuos level in Theorem 2.
Let $B : L^2(\Omega) \to V_0$ be the operator defined in Section 3 such that $\psi = B\varphi_h$ is the solution of

$$a(\psi, v) = \int_\Omega \varphi_h v \quad \forall v \in H_{\Gamma_D}^1.$$

We define the operator B_h as the discretization of the operator B, i.e., $\psi_h = B_h \varphi_h \in V_{h,\Gamma_D}^r$ is the solution of

$$a(\psi_h, v_h) = \int_\Omega \varphi_h v_h \quad \forall v_h \in V_{h,\Gamma_D}^r.$$

It is easy to show that the operator B_h is stable in the L^∞-norm, i.e., $\|\psi_h\|_{L^\infty(\Omega)} \leq C \|\varphi_h\|_{L^2(\Omega)}$. We have in fact that

$$\|\psi_h\|_{L^\infty(\Omega)} \leq \|\psi - \psi_h\|_{L^\infty(\Omega)} + \|\psi\|_{L^\infty(\Omega)}.$$

Thanks to the H^2-elliptic regularity of the PDE (2) (see Assumption 2) we have that

$$\|\psi\|_{L^\infty(\Omega)} \leq C \|\psi\|_{H^2(\Omega)} \leq C \|\varphi_h\|_{L^2(\Omega)}$$

while thanks to Assumption 3 and the Sobolev inequality (see, e.g., [2])

$$\|w\|_{L^\infty(\Omega)} \leq C \|w\|_{W^{1,p}(\Omega)} \quad \forall w \in W^{1,p}(\Omega) \quad \forall p > d \tag{25}$$

where $W^{1,p}(\Omega)$ is the space of functions in $L^p(\Omega)$ with first derivatives in $L^p(\Omega)$, we obtain the bound for the error term in the L^∞-norm

$$\begin{aligned} \|\psi - \psi_h\|_{L^\infty(\Omega)} &\leq C \|\psi - \psi_h\|_{W^{1,p}(\Omega)} \leq C \inf_{v_h \in V_{h,\Gamma_D}^r} \|\psi - v_h\|_{W^{1,p}(\Omega)} \\ &\leq Ch |\psi|_{W^{2,p}(\Omega)} \leq Ch \|\varphi_h\|_{L^p(\Omega)} \leq Ch^{1+\min\left\{0, \frac{d}{p} - \frac{d}{2}\right\}} \|\varphi_h\|_{L^2(\Omega)}. \end{aligned}$$

In the last step we have used an inverse inequality; see, e.g., [9], and taking $p = 2d/(d-2)$, which is larger than d for $d \le 3$, we conclude

$$\|\psi - \psi_h\|_{L^\infty(\Omega)} \le C \|\varphi_h\|_{L^2(\Omega)}.$$

We define now the operators $\mathcal{A}_h$ and $\mathcal{L}_h$ as the discretization of the operators $\mathcal{A}$ and $\mathcal{L}$ defined in Section 3:

$$\begin{aligned}
\mathcal{A}_h(g_h, \varphi_h) &= \frac{1}{n}\sum_{i=1}^{n} B_h g_h(\mathbf{p}_i) B_h \varphi_h(\mathbf{p}_i) + \lambda \int_\Omega g_h \varphi_h \\
\mathcal{L}_h \varphi_h &= \frac{2}{n}\sum_{i=1}^{n} B_h \varphi_h(\mathbf{p}_i)(B_h u(\mathbf{p}_i) + f_{b,h}(\mathbf{p}_i) - z_i)
\end{aligned}$$

where $f_{b,h}$ is the discretization of f_b. The operator $\mathcal{A}_h$ is coercive in L^2, in fact

$$\mathcal{A}_h(g_h, g_h) = \frac{1}{n}\sum_{i=1}^{n}(B_h g_h(\mathbf{p}_i))^2 + \lambda \int_\Omega g_h^2 \ge \lambda \|g_h\|^2_{L^2(\Omega)}.$$

Thanks to the stability in the L^∞-norm of the operator B_h, both the operators $\mathcal{A}_h$ and $\mathcal{L}_h$ are continuous:

$$\begin{aligned}
\mathcal{A}_h(g_h, \varphi_h) &\le C \|B_h g_h\|_{L^\infty(\Omega)} \|B_h \varphi_h\|_{L^\infty(\Omega)} + \lambda \|g_h\|_{L^2(\Omega)} \|\varphi_h\|_{L^2(\Omega)} \\
&\le C \|g_h\|_{L^2(\Omega)} \|\varphi_h\|_{L^2(\Omega)} \\
\mathcal{L}_h(\varphi_h) &\le C \|B_h \varphi_h\|_{L^\infty(\Omega)} \le C \|\varphi_h\|_{L^2(\Omega)}.
\end{aligned}$$

Thanks to the fact that $\mathcal{A}_h$ is continuous and coercive in $L^2(\Omega)$ and $\mathcal{L}_h$ is continuous in $L^2(\Omega)$, the equation

$$2\mathcal{A}_h(g_h, \varphi_h) + \mathcal{L}_h(\varphi_h) = 0 \quad \forall \varphi_h : B_h \varphi_h = \psi_h \in V^r_{h,\Gamma_D}$$

has a unique solution $g_h \in V^r_{h,\Gamma_D}$. This equation corresponds to the first equation of the system (24).
Once $\hat{g}_h$ is known, $\hat{f}_h$ is recovered uniquely from the second equation in (24). $\square$

Let now $\{\psi_k\}_{k=1}^{N_h}$ be the Lagrangian basis of the space V^r_{h,Γ_D}, where $N_h = \dim(V^r_{h,\Gamma_D})$, and let $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_h}$ be the nodes associated to the N_h basis functions. Thanks to the Lagrangian property of the basis functions we can write a function $f \in \mathrm{span}\{\psi_1, \ldots, \psi_{N_h}\}$ as

$$f(\mathbf{x}) = \sum_{k=1}^{N_h} f(\xi_k)\psi_k(\mathbf{x}) = \mathbf{f}^T \boldsymbol{\psi}$$

where $\mathbf{f} = (f_1, \ldots, f_{N_h})^T = (f(\boldsymbol{\xi}_1), \ldots, f(\boldsymbol{\xi}_{N_h}))^T$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{N_h})^T$.
Analogously, we define the Lagrangian basis of the space $V^r_h \backslash V^r_{h,\Gamma_D}$ as $\{\psi^D_k\}_{k=1}^{N^D_h}$, where $N^D_h = \dim(V^r_h \backslash V^r_{h,\Gamma_D})$ and the nodes on the boundary Γ_D as $\boldsymbol{\xi}^D_1, \ldots, \boldsymbol{\xi}^D_{N^D_h}$. A lifting $f_{D,h}$ can be constructed in $\mathrm{span}\{\psi^D_1, \ldots, \psi^D_{N^D_h}\}$ as $f_{D,h} = \mathbf{f}^T_D \boldsymbol{\psi}^D$ where $\mathbf{f}_D = (f_D(\boldsymbol{\xi}^D_1), \ldots, f_D(\boldsymbol{\xi}^D_{N^D_h}))^T$ and $\boldsymbol{\psi}^D = (\psi^D_1, \ldots, \psi^D_{N^D_h})^T$.

The Finite Element solution $\hat{f}_h$ of the discrete counterpart of the estimation problem can thus be written as

$$\hat{f}_h = \hat{\mathbf{f}}^T \boldsymbol{\psi} + \mathbf{f}_D^T \boldsymbol{\psi}^D$$

where $\hat{\mathbf{f}}$ is the solution of the linear system

$$\begin{bmatrix} \boldsymbol{\Psi}^T\boldsymbol{\Psi}/(n\lambda) & \mathbf{A}^T \\ \mathbf{A} & -\mathbf{R} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}^T\mathbf{z}/(n\lambda) - \boldsymbol{\Psi}^T\boldsymbol{\Psi}^D\mathbf{f}_D/(n\lambda) \\ \mathbf{u} + \mathbf{h}_N + \mathbf{h}_R - \mathbf{A}^D\mathbf{f}_D \end{bmatrix}. \tag{26}$$

$\mathbf{R}_{jk} = \int_\Omega \psi_j\psi_k$ is the mass matrix, $\boldsymbol{\Psi}_{ij} = \psi_j(\mathbf{p}_i)$ and $\boldsymbol{\Psi}^D_{ij} = \psi^D_j(\mathbf{p}_i)$ are the matrices of pointwise evaluation of the basis functions, $\mathbf{A}_{jk} = a(\psi_k, \psi_j)$ and $\mathbf{A}^D_{jk} = a(\psi^D_k, \psi_j)$ are the matrices associated to the bilinear form $a(\cdot,\cdot)$. The vector $\mathbf{z} = (z_1, \ldots, z_n)$ contains the observed data while the vectors $\mathbf{u}_j = \int_\Omega u\psi_j$, $(\mathbf{h}_N)_j = \int_{\Gamma_N} h_N\psi_j$ and $(\mathbf{h}_R)_j = \int_{\Gamma_R} h_R\psi_j$ are related to the forcing term and the non homogeneous boundary conditions.
Thanks to the linearity of the estimator $\hat{f}_h$ in the observations we derive in [1] some classical inferential tools such as approximate pointwise confidence bands and prediction intervals, providing measures for uncertainty quantification within such models.

6 Bias of the Finite Element estimator

The Finite Element estimator $\hat{f}_h$ can be split, as its continuous counterpart $\hat{f}$, in two different terms $\mathbb{E}[\hat{f}_h]$ and $\hat{w}_h$ that are respectively the Finite Element approximation of $\mathbb{E}[\hat{f}]$ and $\hat{w}$. Reasoning as for the continuous problem, we can easily show that $\mathbb{E}[\hat{w}_h] = 0$. Neglecting this zero mean term, we now aim at studying the bias of the Finite Element estimator, $\mathbb{E}[\hat{f}_h - f_0]$, in the norm

$$\left|\left|\left|\mathbb{E}[\hat{f}_h - f_0]\right|\right|\right|^2 = \left\|\mathbb{E}[\hat{f}_h - f_0]\right\|_n^2 + \lambda\left[\left\|\mathbb{E}[\hat{f}_h] - f_0\right\|^2_{H^1(\Omega)} + \|\mathbb{E}[\hat{g}_h] - g_0\|^2_{L^2(\Omega)}\right] \tag{27}$$

where the norm $\|\cdot\|_n$, is the norm induced by the bilinear operator $l(\cdot,\cdot)$, defined as

$$\left\|\mathbb{E}[\hat{f}_h - f_0]\right\|_n = \frac{1}{n}\sum_{i=1}^n (\mathbb{E}[\hat{f}_h](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2.$$

Notice that the norm $|||\cdot|||$ contains both the norm $\|\cdot\|_J$ and the H^1-norm of $\hat{f}_h$. We need in fact also an explicit control on the H^1-norm of $\hat{f}_h$ to study the convergence properties of the mixed Finite Element solution of the system (24).

Remark 1. *One might be tempted to compare $\mathbb{E}[\hat{f}_h]$ to its continuous counterpart $\mathbb{E}[\hat{f}]$. However, due to the presence of $\delta_{\mathbf{p}_i}$ in the forcing term of the dual equation in* (16)*, $\mathbb{E}[\hat{f}]$ is not smooth in general. For this reason, in the error analysis proposed in this Section, we directly compare $\mathbb{E}[\hat{f}_h]$ with the true underlying field f_0, which is assumed to be sufficiently smooth.*

The convergence of the bias term is studied when $h \to 0$, fixing the number n of observations and the penalty parameter λ. Since we are considering n and λ fixed we expect to obtain an error bound that contains a term going to zero as $h \to 0$

and a term that represents the bias induced by the roughness penalty, similarly to the continuous setting in Lemma 1.

Thanks to the introduction of the adjoint variable $\hat{g}$, which represents the misfit of the PDE, the estimation Problem 1 can be reformulated as a constrained problem that is more convenient for the study of the convergence of the Finite Element estimator.

Problem 2. *Find $\hat{f} \in V$, $\hat{g} \in \mathcal{G}$ such that*

$$(\hat{f}, \hat{g}) = \underset{(f,g)\in\mathcal{W}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}(f(\mathbf{p}_i) - z_i)^2 + \lambda \int g^2$$

where $\mathcal{W}$ is the constrained space

$$\mathcal{W} = \{(f, g) \in V \times \mathcal{G} : \; Lf - u = g\}. \tag{28}$$

The constrained space $\mathcal{W}$, can be discretized as

$$\begin{aligned}\mathcal{W}_h = \{(f_h, g_h) \in V_h^r \times V_{h,\Gamma_D}^r : f_h|_{\Gamma_D} = h_{D,h} \text{ and}\\ a(f_h, v_h) - r(g_h, v_h) = F(v_h), \; \forall v_h \in V_{h,\Gamma_D}^r\},\end{aligned}$$

where $a(\cdot,\cdot)$, $r(\cdot,\cdot)$, $F(\cdot)$ and $h_{D,h}$ are defined in Section 5. The expected value of the Finite Element estimator $(\mathbb{E}[\hat{f}_h], \mathbb{E}[\hat{g}_h])$ is thus the solution of the equation

$$l(\mathbb{E}[\hat{f}_h], \psi_h) + \lambda a(\psi_h, \mathbb{E}[\hat{g}_h]) = l(f_0, \psi_h) \;\; \forall \psi_h \in V_{h,\Gamma_D}^r \tag{29}$$

in the constrained space $\mathcal{W}_h$.
The bound for the bias of the Finite Element estimator $\mathbb{E}[\hat{f}_h - f_0]$ is obtained thanks to the following Lemma and Theorem.

Lemma 3. *Let $g_0 = Lf_0 - u$. The bias of the Finite Element estimator $(\mathbb{E}[\hat{f}_h], \mathbb{E}[\hat{g}_h]) \in \mathcal{W}_h$ satisfies the inequality*

$$\begin{aligned}&\left\|f_0 - \mathbb{E}[\hat{f}_h]\right\|_n^2 + \lambda\left[\left\|f_0 - \mathbb{E}[\hat{f}_h]\right\|_{H^1(\Omega)}^2 + \|g_0 - \mathbb{E}[\hat{g}_h]\|_{L^2(\Omega)}^2\right]\\ &\leq C\Bigg\{\inf_{(\varphi_h,p_h)\in\mathcal{W}_h}\left[\|f_0 - \varphi_h\|_n^2 + \lambda\|f_0 - \varphi_h\|_{H^1(\Omega)}^2 + \lambda\|g_0 - p_h\|_{L^2(\Omega)}^2\right]\\ &+ \lambda\|g_0\|_{L^2(\Omega)}^2\Bigg\}\end{aligned} \tag{30}$$

for some constant $C > 0$ independent of h.

Proof. We set $f_h^* = \mathbb{E}[\hat{f}_h]$ and $g_h^* = \mathbb{E}[\hat{g}_h]$ and we recall that $\|\cdot\|_n$ is the norm induced by the bilinear form $l(\cdot,\cdot)$, i.e., $\|f\|_n^2 = l(f, f)$.
In order to prove Lemma 3 we can use the theory of saddle points systems. From equation (29) and the definition of $\mathcal{W}_h$ we have immediately

$$\begin{aligned}&\frac{1}{\lambda}l(\hat{f}_h^* - f_0, \psi_h) + a(\psi_h, \hat{g}_h^*) = 0 \;\; \forall \psi_h \in V_{h,\Gamma_D}^r\\ &a(\hat{f}_h^* - \varphi_h, v_h) = r(\hat{g}_h^* - p_h, v_h) \;\; \forall v_h \in V_{h,\Gamma_D}^r, \; (\varphi_h, p_h) \in \mathcal{W}_h.\end{aligned}$$

Choosing $(\varphi_h, p_h) \in \mathcal{W}_h$ we thus obtain

$$
\begin{aligned}
&\left\|\hat{f}_h^* - \varphi_h\right\|_n^2 + \lambda \left\|\hat{g}_h^* - p_h\right\|_{L^2(\Omega)}^2 = l(\hat{f}_h^* - \varphi_h, \hat{f}_h^* - \varphi_h) + \lambda r(\hat{g}_h^* - p_h, \hat{g}_h^* - p_h) \\
&= l(\hat{f}_h^* - f_0, \hat{f}_h^* - \varphi_h) + l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) + \lambda r(\hat{g}_h^* - p_h, \hat{g}_h^* - p_h) \\
&= -\lambda a(\hat{f}_h^* - \varphi_h, \hat{g}_h^*) + l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) + \lambda r(\hat{g}_h^* - p_h, \hat{g}_h^* - p_h) \\
&= l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) - \lambda r(\hat{g}_h^* - p_h, p_h) \qquad (31)
\end{aligned}
$$

since $\hat{f}_h^* - \varphi_h \in V_{h,\Gamma_D}^r$. We now bound the term $\hat{f}_h^* - \varphi_h$ in the H^1-norm using the coercivity of the bilinear form $a(\cdot,\cdot)$ in $H^1(\Omega)$:

$$
\begin{aligned}
\left\|\hat{f}_h^* - \varphi_h\right\|_{H^1(\Omega)}^2 &\leq \frac{1}{\alpha} a(\hat{f}_h^* - \varphi_h, \hat{f}_h^* - \varphi_h) = \frac{1}{\alpha} r(\hat{g}_h^* - p_h, \hat{f}_h^* - \varphi_h) \\
&\leq \frac{\sqrt{1 + C_P^2}}{\alpha} \left\|\hat{g}_h^* - p_h\right\|_{L^2(\Omega)} \left\|\hat{f}_h^* - \varphi_h\right\|_{H^1(\Omega)}
\end{aligned}
$$

where α is the coercivity constant and C_P is the constant in the Poincaré inequality (5), which holds thanks to Assumption 1. Summing this inequality to (31) we obtain

$$
\begin{aligned}
&\left\|\hat{f}_h^* - \varphi_h\right\|_n^2 + \lambda \left[\frac{\alpha^2}{4(1 + C_P^2)} \left\|\hat{f}_h^* - \varphi_h\right\|_{H^1(\Omega)}^2 + \left\|\hat{g}_h^* - p_h\right\|_{L^2(\Omega)}\right] \\
&\leq l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) - \lambda r(\hat{g}_h^* - p_h, p_h) + \frac{\lambda}{4} \left\|\hat{g}_h^* - p_h\right\|_{L^2(\Omega)}^2 \\
&\leq \frac{1}{2} \left\|f_0 - \varphi_h\right\|_n^2 + \frac{1}{2} \left\|\hat{f}_h^* - \varphi_h\right\|_n^2 + \frac{\lambda}{2} \left\|\hat{g}_h^* - p_h\right\|_{L^2(\Omega)}^2 + 2\lambda \left\|g_0 - p_h\right\|_{L^2(\Omega)}^2 \\
&+ 2\lambda \left\|g_0\right\|_{L^2(\Omega)}^2 .
\end{aligned}
$$

This inequality provides the bound

$$
\begin{aligned}
&\left\|\hat{f}_h^* - \varphi_h\right\|_n^2 + \lambda \left[\left\|\hat{f}_h^* - \varphi_h\right\|_{H^1(\Omega)}^2 + \left\|\hat{g}_h^* - p_h\right\|_{L^2(\Omega)}^2\right] \\
&\leq C \left\{\left\|f_0 - \varphi_h\right\|_n^2 + \lambda \left\|g_0 - p_h\right\|_{L^2(\Omega)}^2 + \lambda \left\|g_0\right\|_{L^2(\Omega)}^2\right\} .
\end{aligned}
$$

The final error bound (30) can now be obtained by triangular inequality and exploiting the arbitrariness of $(\varphi_h, p_h) \in \mathcal{W}_h$. □

We want now to split the error term on the constrained space $\mathcal{W}_h$ in two different errors for $\mathbb{E}[\hat{f}_h]$ and $\mathbb{E}[\hat{g}_h]$ on the space V_h^r. Assuming moreover that f_0 and g_0 are in proper Sobolev spaces $W^{s,p}(\Omega)$ we obtain the following result.

Theorem 3. *Using Finite Elements of degree r, if $f_0 \in W^{r+1,p}(\Omega)$ with $f_0|_{\Gamma_D} = h_D$ and $g_0 \in W^{r,p}(\Omega)$ with $g_0|_{\Gamma_D} = 0$ for $p > d$ then, under Assumption 3, there exists $h_0 > 0$ s.t. $\forall h \leq h_0$*

$$
\begin{aligned}
&\left\|f_0 - \mathbb{E}[\hat{f}_h]\right\|_n^2 + \lambda \left[\left\|f_0 - \mathbb{E}[\hat{f}_h]\right\|_{H^1(\Omega)}^2 + \left\|g_0 - \mathbb{E}[\hat{g}_h]\right\|_{L^2(\Omega)}^2\right] \\
&\leq C \left[h^{2r} \left(\left|f_0\right|_{W^{r+1,p}(\Omega)}^2 + \lambda \left|g_0\right|_{W^{r,p}(\Omega)}^2\right) + \lambda \left\|g_0\right\|_{L^2(\Omega)}^2\right] . \qquad (32)
\end{aligned}
$$

Proof. In order to prove the result we need to split in two parts the constrained error term

$$\inf_{(\varphi_h,p_h)\in\mathcal{W}_h}\left[\|f_0-\varphi_h\|_n^2+\lambda\|f_0-\varphi_h\|_{H^1(\Omega)}^2+\lambda\|g_0-p_h\|_{L^2(\Omega)}^2\right]$$

in inequality (30).
We fix in the following $p_h \in V^r_{h,\Gamma_D}$ and we chose $\varphi_h \in V^r_h$ that satisfies $a(\varphi_h, v_h) = r(p_h, v_h) + F(v_h)$ and $\varphi_h|_{\Gamma_D} = h_{D,h}$, so that $(\varphi_h, p_h) \in \mathcal{W}_h$. Thanks to this choice we obtain the following bound

$$\|f_0-\varphi_h\|_{H^1(\Omega)}^2 \le C\left[\|f_0-z_h\|_{H^1(\Omega)}^2+\|g_0-p_h\|_{L^2(\Omega)}^2\right] \tag{33}$$

where z_h is an arbitrary function in V^r_h such that $z_h|_{\Gamma_D} = h_{D,h}$. This inequality is obtained thanks to the fact that

$$\|f_0-\varphi_h\|_{H^1(\Omega)}^2 \le 2\|f_0-z_h\|_{H^1(\Omega)}^2+2\|z_h-\varphi_h\|_{H^1(\Omega)}^2$$

and that

$$\begin{aligned}
\alpha\|z_h-\varphi_h\|_{H^1(\Omega)}^2 &\le a(z_h-\varphi_h,z_h-\varphi_h)=a(f_0-\varphi_h,z_h-\varphi_h)+a(z_h-f_0,z_h-\varphi_h)\\
&= r(g_0-p_h,z_h-\varphi_h)+a(z_h-f_0,z_h-\varphi_h)\\
&\le \|g_0-p_h\|_{L^2(\Omega)}\|z_h-\varphi_h\|_{L^2(\Omega)}+\|z_h-f_0\|_{H^1(\Omega)}\|z_h-\varphi_h\|_{H^1(\Omega)}\\
&\le C\left[\|g_0-p_h\|_{L^2(\Omega)}+\|z_h-f_0\|_{H^1(\Omega)}\right]\|z_h-\varphi_h\|_{H^1(\Omega)}\,.
\end{aligned}$$

The term $\|f_0-\varphi_h\|_n^2$ can be bounded with the $W^{1,p}$-norm ($p > d$) of the same quantity, i.e.,

$$\|f_0-\varphi_h\|_n \le C\|f_0-\varphi_h\|_{W^{1,p}(\Omega)}\,. \tag{34}$$

We have in fact that

$$\frac{1}{n}\sum_{i=1}^{n}(f_0(\mathbf{p}_i)-\varphi_h(\mathbf{p}_i))^2 \le \max_{\mathbf{p}_i}(f_0(\mathbf{p}_i)-\varphi_h(\mathbf{p}_i))^2 \le \|f_0-\varphi_h\|_{L^\infty}^2$$

and thanks to the Sobolev inequality (25) we obtain the upper bound (34).
We define now $f_{0h} \in V^r_h$ such that $a(f_{0h}, \psi_h) - r(g_0, \psi_h) = F(\psi_h)\ \forall \psi_h \in V^r_{h,\Gamma_D}$; the error term can be split in two parts

$$\|f_0-\varphi_h\|_{W^{1,p}(\Omega)} \le \|f_0-f_{0h}\|_{W^{1,p}(\Omega)}+\|f_{0h}-\varphi_h\|_{W^{1,p}(\Omega)}\,.$$

The first term on the right-hand side of the inequality represents the $W^{1,p}$-norm of the Finite Element error of the elliptic equation. The quantity f_{0h} can in fact be seen as the Finite Element approximation of the exact solution f_0 and for this reason (see, e.g., [2])

$$\|f_0-f_{0h}\|_{W^{1,p}(\Omega)} \le C\inf_{\substack{z_h\in V^r_h\\ z_h|_{\Gamma_D}=h_{D,h}}}\|f_0-z_h\|_{W^{1,p}(\Omega)}\,.$$

The second term of the right-hand side of the inequality can be bounded by

$$\|f_{0h}-\varphi_h\|_{W^{1,p}(\Omega)} \le C\|g_0-p_h\|_{L^p(\Omega)}$$

where $p > d$. This bound is obtained thanks to the L^p-stability of the problem

$$a(f_{0,h} - \varphi_h, v_h) = r(g_0 - p_h, v_h) \quad \forall v_h \in V^r_{h,\Gamma_D},$$

see, e.g., [2]. Therefore

$$\|f_0 - \varphi_h\|_n \le C \left(\inf_{\substack{z_h \in V^r_h \\ z_h|_{\Gamma_D} = h_{D,h}}} \|f_0 - z_h\|_{W^{1,p}(\Omega)} + \|g_0 - p_h\|_{L^p(\Omega)} \right). \tag{35}$$

Collecting the bounds (33) and (35), since $L^p(\Omega) \subseteq L^2(\Omega)$ for $p \ge 2$ and Ω is bounded, we obtain for $p > d$, $d = 2, 3$ the unconstrained upper bound

$$\begin{aligned} &\left\| f_0 - \hat{f}^*_h \right\|^2_n + \lambda \left[\left\| f_0 - \hat{f}^*_h \right\|^2_{H^1(\Omega)} + \|g_0 - \hat{g}^*_h\|^2_{L^2(\Omega)} \right] \le \\ &C \left\{ \inf_{\substack{z_h \in V^r_h \\ z_h|_{\Gamma_D} = h_{D,h}}} \|f_0 - z_h\|^2_{W^{1,p}(\Omega)} + \lambda \inf_{p_h \in V^r_{h,\Gamma_D}} \|g_0 - p_h\|^2_{L^p(\Omega)} + \lambda \|g_0\|^2_{L^2(\Omega)} \right\}. \end{aligned} \tag{36}$$

The classic error bound for the interpolant $\Pi^r_h v \in V^r_h$ of $v \in W^{r+1,p}(\Omega)$ with $p > 1$:

$$\|v - \Pi^r_h v\|_{W^{k,p}(\Omega)} \le C h^{r+1-k} |v|_{W^{r+1,p}(\Omega)}, \tag{37}$$

provides

$$\inf_{\substack{z_h \in V^r_h \\ z_h|_{\Gamma_D} = h_{D,h}}} \|f_0 - z_h\|_{W^{1,p}(\Omega)} \le C h^r |f_0|_{W^{r+1,p}(\Omega)}$$

$$\inf_{p_h \in V^r_{h,\Gamma_D}} \|g_0 - p_h\|_{L^p(\Omega)} \le C h^r |g_0|_{W^{r,p}(\Omega)}.$$

□

Notice that the inequality (32) can be split in two terms, the first term of the right-hand side goes to zero for $h \to 0$ while the second term $\|g_0\|^2_{L^2(\Omega)}$ is the same bias term obtained in the error splitting (22) and goes to zero when $\lambda \to 0$.

Remark 2. *In this work we propose an equal order Finite Elements approximation for $\hat{f}$ and $\hat{g}$. Equal order Finite Elements are known to lead to sub-optimal convergence rates for the fourth order biharmonic problem (see, e.g., [4, 5]). However, here we are able to recover the optimal convergence rate thanks to the fact that the boundary conditions that are naturally associated to the smoothing problem are the same for $\hat{f}$ and $\hat{g}$. It should be noticed that the optimal convergence rate is recovered only if g_0 satisfies exactly the homogeneous Dirichlet boundary conditions on Γ_D, which might be a restrictive hypothesis. If g_0 does not satisfy the Dirichlet boundary conditions we should expect a "boundary term" decaying as $h^{1/2}$ both in two and three dimensions. Observe however that the approximation term $\lambda \inf_{p_h \in V^r_{h,\Gamma_D}} \|g_0 - p_h\|^2_{L^p(\Omega)}$ is always smaller than $\lambda \|g_0\|^2_{L^p(\Omega)}$ and for this reason the "boundary term" effect will be hidden by the bias term.*

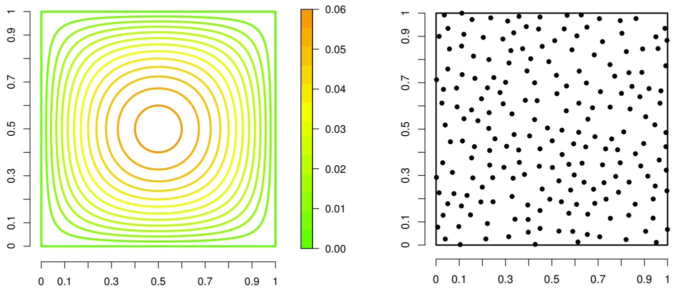

Figure 1: Left: true surface f_0 used for the simulation studies of Test 1; the image displays the isolines $(0, 0.005, 0.01, \ldots, 0.06)$. Right: location points sampled uniformly on the domain for Test 1 and 2.

7 Numerical simulations

7.1 Test 1

We propose to verify in a simple setting the convergence results shown in Section 6. We consider the bidimensional domain $\Omega = [0,1] \times [0,1]$ and we assume that the true underlying surface f_0 satisfies the following PDE

$$\begin{cases} \Delta f_0 = 2\,[x(x-1) + y(y-1)] & \text{in } \Omega \\ f_0 = 0 & \text{on } \partial\Omega \end{cases} \tag{38}$$

whose solution, $f_0 = xy(x-1)(y-1)$, is represented in Figure 1, Left. We consider the $n = 200$ observation points $\mathbf{p}_1, \ldots, \mathbf{p}_n$, represented in Figure 1, Right, and we want to test the convergence of $\|\mathbb{E}[\hat{f}_h - f_0]\|$ when $h \to 0$. For this reason we solve the estimation problem on different uniform structured meshes with size $h = 1/2, 1/4, \ldots, 1/2^9$.
We will consider different settings.

A. The observations are without noise, i.e., $z_i = f_0(\mathbf{p}_i)$, and the functional $J(f)$ penalizes the misfit of the governing PDE (38), i.e., $L = \Delta$ and $u = 2\,[x(x-1) + y(y-1)]$.

B. The observations are without noise, i.e., $z_i = f_0(\mathbf{p}_i)$, but the functional $J(f)$ penalizes the misfit of a PDE different from the governing PDE (38). In particular $L = \Delta$ but $u \neq 2\,[x(x-1) + y(y-1)]$:

   1. the penalized forcing term u is such that $g_0 = Lf_0 - u$ satisfies homogeneous Dirichlet boundary conditions on $\partial\Omega$: $u = 2(x(x-1) + y(y-1))(1 + (x(x-1)y(y-1)))$;
   2. different penalized forcing terms u are considered, such that $g_0 = Lf_0 - u$ does not satisfies homogeneous Dirichlet boundary conditions on $\partial\Omega$:
      (a) $u = (x(x-1) + y(y-1))$, which corresponds to the knowledge of the real forcing term up to a multiplying constant factor;

(b) $u = 2x(x-1)$, which corresponds to the knowledge of only a part of the real forcing term;

(c) $u = 2(x(x-1) + y(y-1)) + (x^{10} + y^{10} + (x-1)^{10} + (y-1)^{10})$, which forces g_0 to be equal to -1 on the boundary, with a relatively large boundary layer.

C. The observations are with noise, i.e., $z_i = f_0(\mathbf{p}_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and the functional $J(f)$ penalizes the misfit of the governing PDE (38), i.e., $L = \Delta$ and $u = 2\,[x(x-1) + y(y-1)]$.

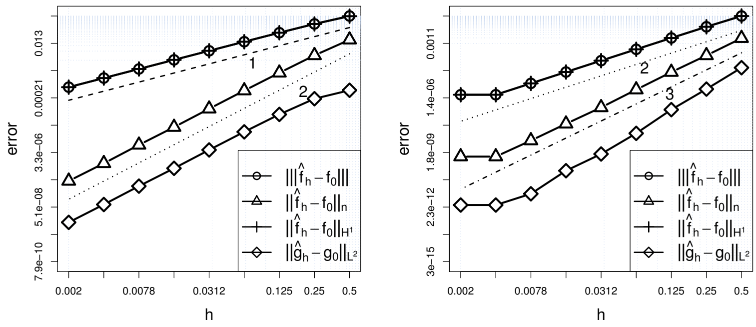

Figure 2: Test 1, case A: convergence rates of the bias of the estimator in the norms $\|\hat{f}_h - f_0\|$, $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ with $\lambda = 1$. Left: linear mixed Finite Element approximation. Right: quadratic mixed Finite Element approximation.

Case A (no bias, no noise) We solve the estimation problem both with linear and quadratic Finite Elements fixing the roughness parameter $\lambda = 1$. We recall that we are using the same order of approximation for $\hat{f}_h$ and $\hat{g}_h$. The results of the linear and the quadratic mixed Finite Element approximation are shown respectively in the left and right panel of Figure 2. In particular we show the convergence of the error $\|\hat{f}_h - f_0\|$ as well as the convergence of each individual term of the norm $\|\cdot\|$, namely $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$. Since we are considering the case of observations without noise $\mathbb{E}[\hat{f}_h - f_0] = \hat{f}_h - f_0$ and $\mathbb{E}[\hat{g}_h - g_0] = \hat{g}_h - g_0$. We notice that both with the linear and the quadratic approximation we obtain a rate of convergence equal to or higher than the expected rate for all the error terms. In particular, the H^1-norm of the error is the dominating term both in the linear and the quadratic approximation and decays as h in the case of linear Finite Elements and as h^2 in the case of quadratic Finite Elements. All the other terms are negligible. As expected, the norm $\|\cdot\|_n$ of $\hat{f}_h - f_0$ and the L^2-norm of $\hat{g}_h - g_0$ decay as h^2 in the case of linear Finite Elements and at least as h^3 in the case of quadratic Finite Elements.

Case B1 (bias with exact b.c., no noise) We solve the estimation problem with linear Finite Elements and we study the convergence for different values of the roughness parameter λ. Recall that, in this case, $g_0 = \Delta f_0 - u \neq 0$ satisfies the homogeneous Dirichlet boundary conditions. Figure 3 shows the rate of convergence

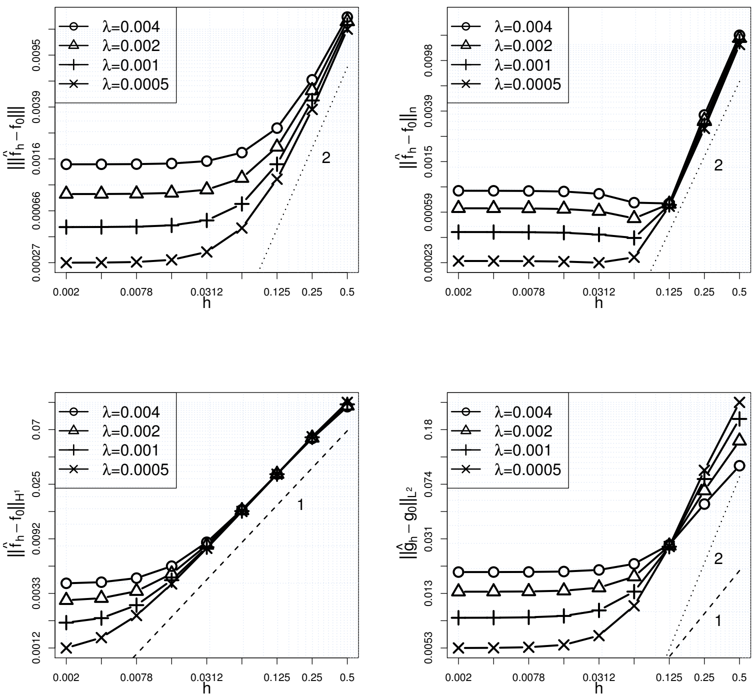

Figure 3: Test 1, case B1: convergence rates of the bias of the estimator obtained with linear Finite Elements, when $\lambda = 0.05, 0.1, 0.2, 0.4$. Top left: $\|\hat{f}_h - f_0\|$, top right: $\|\hat{f}_h - f_0\|_n$, bottom left: $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, bottom right: $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$.

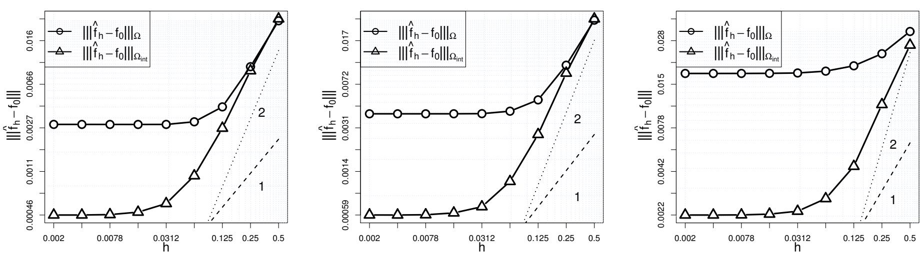

Figure 4: Test 1, case B2: convergence rates of $\|\hat{f}_h - f_0\|_\Omega$ and $\|\hat{f}_h - f_0\|_{\Omega_{\mathrm{int}}}$ using linear Finite Elements with $\lambda = 10^{-5}$. Left: case a), center: case b), right: c).

of the error in different norms, when $\lambda = 0.05, 0.1, 0.2, 0.4$. As in case A, since the observations are without noise, $\mathbb{E}[\hat{f}_h - f_0] = \hat{f}_h - f_0$ and $\mathbb{E}[\hat{g}_h - g_0] = \hat{g}_h - g_0$. Notice that when the mesh is fine the approximation error in the norm $\|\!|\cdot|\!\|$ asymptotically approaches a value proportional to $\sqrt{\lambda}$, as expected from Theorem 3; this behavior is caused by the presence of the bias term in the error bound (32). The dominant term is in this case the L^2-norm of $\hat{g}_h - g_0$. This term has a different behavior for different values of λ: if λ is sufficiently small, it decays as h^2 before approaching the asymptote, otherwise it decays as h. It is thus necessary to use small values of λ in order to recover the expected convergence rate h^2 but even when using a large value of λ the rate of convergence of $\|\!|\hat{f}_h - f_0|\!\|$ is at least linear before reaching the saturation level caused by the bias term. The other two terms instead decay with the expected convergence rate for all the values of λ, before approaching the asymptote.

Case B2 (bias with wrong b.c., no noise) We consider three different forcing terms u such that $g_0 = \Delta f_0 - u \neq 0$ does not satisfy the homogeneous Dirichlet boundary conditions. In this case we study the error in the norm $\|\!|\cdot|\!\|$ over the whole domain Ω, which will be denoted by $\|\!|\hat{f}_h - f_0|\!\|_\Omega$, as well as over the subdomain $\Omega_{\text{int}} = [0.1, 0.9] \times [0.1, 0.9]$, denoted by $\|\!|\hat{f}_h - f_0|\!\|_{\Omega_{\text{int}}}$. As highlighted in Remark 2, the former error should be affected by a "boundary term" decaying as $h^{1/2}$, while the latter does not include the error at the boundary. As in case A and B1, since the observations are without noise, $\mathbb{E}[\hat{f}_h - f_0] = \hat{f}_h - f_0$ and $\mathbb{E}[\hat{g}_h - g_0] = \hat{g}_h - g_0$. The results obtained with the three forcing terms are represented respectively in the left, center and right panels of Figure 4. In theory we would expect a different rate of convergence for the two errors, which should be more clearly visible when the mesh is fine. On the other hand the numerical simulations do not display any significant difference between the convergence rates of the two errors in all the three cases; this is due to the presence of the bias, which is asymptotically approached by both the error terms, that hides the expected convergence rate. Thus, using a forcing term such that g_0 does not satisfy the homogeneous Dirichlet boundary conditions, does not affect too much the surface estimation.

Case C (no bias, with noise) We add some noise to the pointwise evaluations $f_0(\mathbf{p}_i)$ of the surface: for each location point we sample independent errors, $\epsilon_1, \ldots, \epsilon_n$, from a zero mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$, with different standard deviations $\sigma = 0.005, 0.01, 0.02$. The first value of σ corresponds to a rather high signal to noise ratio, since the value of the true surface varies from 0 to 0.062, while the last corresponds to a very low signal to noise ratio with errors of the same order of magnitude as the variation of f_0. The values z_i, obtained from model (1), are shown in Figure 5. We can notice that the observations with small additive noise, represented in the left panel of Figure 5, are similar to the evaluation of f_0 in the sampling points, while the observations with large errors, represented in the right panel of the same figure, are far from the true underlying surface. The results obtained solving the estimation problem with linear Finite Elements, fixed roughness parameter $\lambda = 0.01$ and the exact PDE penalized are shown in Figure 6. The represented results concern a single replicate of the experiment and show a typical behavior of the error convergence in different norms. Notice that in Figure 6 the represented errors include both the approximation error and the error associated to the noisy observations. Due to the presence of noise, both $\|\!|\hat{f}_h - f_0|\!\|$ and $\|\hat{f}_h - f_0\|_n$

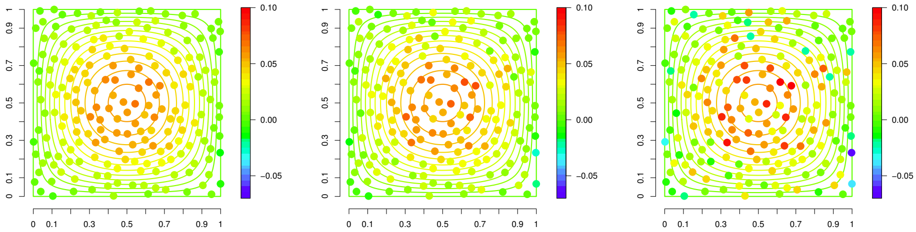

Figure 5: Test 1, case C: value of the observations $z_1, \ldots, z_n$ obtained from model (1) adding noise with different values of standard deviation σ, superimposed to the true underlying surface f_0 (the image displays the isolines $(0, 0.005, 0.01, \ldots, 0.06)$). Left: $\sigma = 0.005$, center: $\sigma = 0.01$, right: $\sigma = 0.02$.

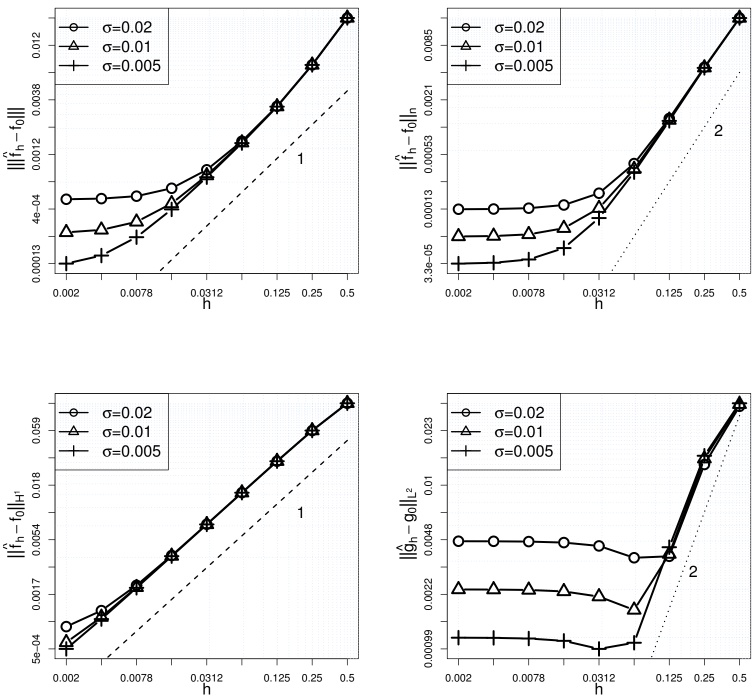

Figure 6: Test 1, case C: convergence rates of the bias of the estimator obtained with linear Finite Elements and $\lambda = 1$, when the error in the observations is generated form a Gaussian distribution with different standard deviations $\sigma = 0.005, 0.01, 0.02$. Top left: $\|\hat{f}_h - f_0\|$, top right: $\|\hat{f}_h - f_0\|_n$, bottom left: $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, bottom right: $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$.

reach quite soon a saturation limit proportional to the standard deviation of the noise σ. Refining further the mesh still provides better approximation of the first derivatives as shown by the convergence of $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$.

7.2 Test 2

We test the convergence also in a different simulation study concerning a diffusion-transport-reaction (DTR) PDE. We consider the domain $\Omega = [0,1] \times [0,1]$ and we assume that the true underlying field f_0 satisfies the following PDE

$$\begin{cases} Lf_0 = -1 & \text{in } \Omega \\ f_0 = 0 & \text{on } \partial\Omega \end{cases} \tag{39}$$

where the operator L is the diffusion-transport-reaction operator defined in (3) with

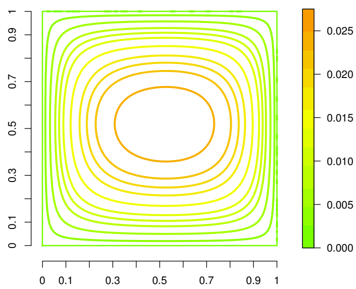

Figure 7: True surface f_0 used for the simulation study of Test 2; the image displays the isolines $(0, 0.0025, 0.005, \ldots, 0.03)$.

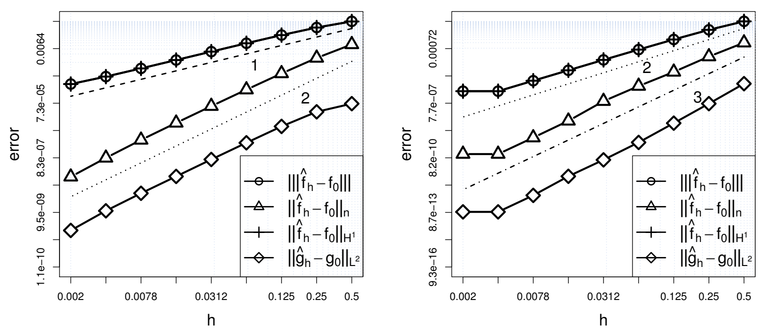

Figure 8: Test 2: convergence rates of the bias of the estimator in the norms $\|\hat{f}_h - f_0\|$, $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ with $\lambda = 1$. Left: linear mixed Finite Element approximation. Right: quadratic mixed Finite Element approximation.

parameters $\mathbf{K}_{11} = 4$, $\mathbf{K}_{22} = 1$, $\mathbf{K}_{12} = \mathbf{K}_{21} = 0$, $\mathbf{b}_1 = 2$, $\mathbf{b}_2 = 1$ and $c = 1$, $\mathbf{K}_{ij}$ and $\mathbf{b}_i$ being respectively the element (i, j) of the diffusion tensor $\mathbf{K}$ and the i-th element of the transport vector $\mathbf{b}$. The solution of the PDE (39) is represented in Figure 7. We consider the $n = 200$ observation points $\mathbf{p}_1, \ldots, \mathbf{p}_n$ represented in Figure 1, Right. In this case we only show the convergence of $\|\mathbb{E}[\hat{f}_h - f_0]\| = \|\hat{f}_h - f_0\|$ when the observations are without noise and the functional $J(f)$ penalizes the misfit of the governing PDE (39). The results obtained solving the estimation problem with linear and quadratic Finite Elements and $\lambda = 1$ on different uniform structured meshes with size $h = 1/2, 1/4, \ldots, 1/2^9$ are shown in Figure 8. We can notice that also in the DTR case we obtain a rate of convergence equal to or higher than the expected rate for all the error terms both with the linear and the quadratic approximation. The H^1-norm is still the dominating term while all the other terms are negligible. As in the Laplacian case, the error terms $\|\hat{f}_h - f_0\|_n$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ decay as h^2 for linear Finite Elements and at least as h^3 for quadratic Finite Elements.

7.3 Test 3

In the last simulation study we test the error convergence in a setting similar to the applied motivating problem and the simulation studies presented in [1]. We consider in particular a circular domain centered in zero, with unitary radius, $\Omega = B_1$, which is similar to the almost circular artery cross-section used in the blood velocity field estimation problem in [1]. We assume that the true underlying field f_0 satisfies the following PDE

$$\begin{cases} \Delta f_0 = y & \text{in } \Omega \\ f_0 = 0 & \text{on } \partial\Omega \end{cases} \tag{40}$$

whose solution $f_0 = y/8(1 - x^2 + y^2)$ is represented in Figure 9, Left. We consider the $n = 200$ observation points $\mathbf{p}_1, \ldots, \mathbf{p}_n$, represented in Figure 9, Right, on the circular domain. As in the previous simulation study, in this case we only show the

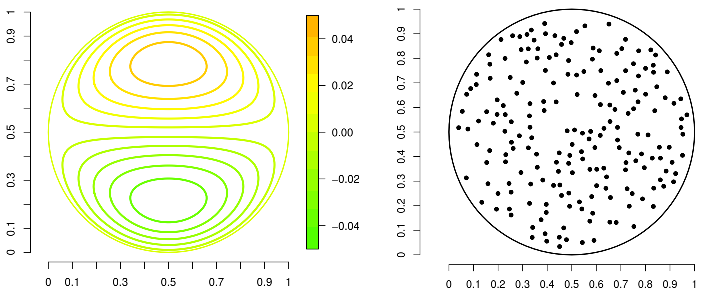

Figure 9: Left: true surface f_0 used for the simulation study of Test 3; the image displays the isolines $(-0.055, -0.045, -0.035, \ldots, 0.055)$. Right: location points sampled uniformly on the domain for Test 3.

convergence when the observations are without noise and the penalized PDE is the exact PDE (40). The results obtained solving the estimation problem with linear Finite Elements and $\lambda = 1$ on 6 different uniform unstructured meshes are shown

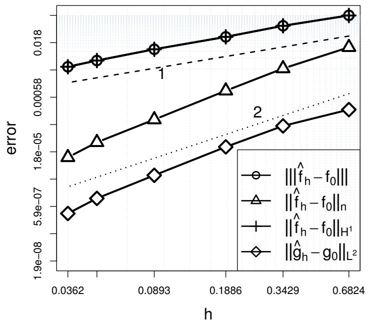

Figure 10: Test 3: convergence rates of the bias of the estimator in the norms $\|\hat{f}_h - f_0\|$, $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ with $\lambda = 1$ and linear mixed Finite Element approximation.

in Figure 10. As expected, the rate of convergence of the error $\|\mathbb{E}[\hat{f}_h - f_0]\| = \|\hat{f}_h - f_0\|$ is h, since $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$ is the dominating term, while $\|\hat{f}_h - f_0\|_n$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ decay as h^2. We thus obtain an optimal convergence rate of the error also in the case of a circular domain.

8 Surface estimator for areal data

The field smoothing method presented in Section 2 can be extended to the case of areal data that represent quantities computed on some subregions. This is useful in many applications of interest and it is for instance the case of the estimation of blood velocity field from Echo-Doppler data presented in [1]; the Echo-Doppler data represent in fact the mean velocity of the blood cells on a subdomain within an artery and cannot be approximated with pointwise observations.
Let $D_i \subset \Omega$, for $i = 1, \ldots, N$, be some subdomains and $\bar{z}_i$, for $i = 1, \ldots, N$, the mean value of a quantity of interest on the subdomains. We consider the following model for the observations $\bar{z}_i$:

$$\bar{z}_i = \frac{1}{|D_i|} \int_{D_i} f_0 + \eta_i. \tag{41}$$

The error terms η_i have zero mean and variances $\bar{\sigma}_i^2$. The variances $\bar{\sigma}_i^2$ depend inversely on the dimension of the beams D_i under the assumption that the number of observations in a subdomain is proportional to its dimension. This model can be derived starting from the model for pointwise observations; see [1] for more details.
In order to estimate the field we propose to minimize the penalized sum-of-square-error functional

$$\bar{J}(f) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{|D_i|} \left(\int_{D_i} (f - \bar{z}_i)\, d\mathbf{p} \right)^2 + \lambda \int_{\Omega} (Lf - u)^2 \tag{42}$$

over the space V, defined in Section 2. The first term is a weighted least-square-error functional for areal data over subdomains D_i, weighted with the inverse of

the variances $\bar{\sigma}_i^2$, under the assumption that $\bar{\sigma}_i^2 \propto 1/|D_i|$.
Existence and uniqueness of the estimator $\hat{f} = \text{argmin}_{f \in V} \bar{J}(f)$ is provided by the following theorem.

Theorem 4. *The estimator $\hat{f}$ exists, is unique and is obtained solving the system of PDEs:*

$$\begin{cases} L\hat{f} = u + \hat{g} & \text{in } \Omega \\ \mathcal{B}_c \hat{f} = h & \text{on } \partial\Omega \end{cases} \qquad \begin{cases} L^* \hat{g} = -\frac{1}{N\lambda} \sum_{i=1}^{N} \frac{1}{|D_i|} \mathbb{I}_{D_i} \int_{D_i} (\hat{f} - \bar{z}_i) & \text{in } \Omega \\ \mathcal{B}_c{}^* \hat{g} = 0 & \text{on } \partial\Omega \end{cases} \tag{43}$$

where $\hat{g} \in \mathcal{G}$ represents the misfit of the penalized PDE, L^* is the adjoint operator of L, described by equation (13), and $\mathcal{B}_c{}^*$ is the operator that defines the boundary conditions of the adjoint problem, summarized in (14).

The proof is analogous to the proof of Theorem 2. The existence and uniqueness of the estimator is in fact obtained, thanks to Theorem 1, writing the functional $\bar{J}(f)$ as the quadratic form (8). The proof of the well posedness of the problem in the areal case is easier than the one presented in Section 3 and it is similar to classical results in control theory. Data are in fact distributed and it's not necessary to require more regularity as in the case of punctual observations.

The estimator is then discretized by means of the mixed Finite Element method described in Section 5. The Finite Element estimator $\hat{f}_h$ can be written as $\hat{f}_h = \hat{\mathbf{f}}^T \boldsymbol{\psi} + \hat{\mathbf{f}}_D^T \boldsymbol{\psi}^D$ where $\hat{\mathbf{f}}$ is the solution of the linear system

$$\begin{bmatrix} \bar{\boldsymbol{\Psi}}^T \mathbf{W} \bar{\boldsymbol{\Psi}}/(N\lambda) & \mathbf{A}^T \\ \mathbf{A} & -\mathbf{R} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\Psi}}^T \mathbf{W} \bar{\mathbf{z}}/(N\lambda) - \bar{\boldsymbol{\Psi}}^T \bar{\boldsymbol{\Psi}}^D \mathbf{f}_D/(N\lambda) \\ \mathbf{u} + \mathbf{h}_N + \mathbf{h}_R - \mathbf{A}^D \mathbf{f}_D \end{bmatrix}. \tag{44}$$

where $\bar{\boldsymbol{\Psi}}_{ik} = 1/|D_i| \int_{D_i} \psi_k$ and $\bar{\boldsymbol{\Psi}}_{ik}^D = 1/|D_i| \int_{D_i} \psi_k^D$ represents the spatial average of the basis functions on the subdomains D_i, $\mathbf{W} = \text{diag}(|D_1|, \ldots, |D_N|)$ is the weight matrix and $\bar{\mathbf{z}} = (\bar{z}_1, \ldots, \bar{z}_N)^T$ is the vector of mean values on subdomains. For more details on the properties of the estimator see [1].

As in the case of pointwise observations we can obtain a bound for the bias of the estimator $\hat{f}$ that corresponds exactly to the bound (22). We can use the results obtained in the pointwise case also for the study of the bias of the Finite Element estimator. In the areal data case we can also relax the hypothesis on f_0 and g_0 in Theorem 3.

Theorem 5. *Using Finite Elements of degree r, if $f_0 \in H^{r+1}(\Omega)$ with $f_0|_{\Gamma_D} = h_D$ and $g_0 \in H^r(\Omega)$ with $g_0|_{\Gamma_D} = 0$, we obtain, under Assumption 1,*

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^{N} \frac{1}{|D_i|} \int_{D_i} (f_0 - \mathbb{E}[\hat{f}^*])^2 + \lambda \left[\left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_{H^1(\Omega)}^2 + \| g_0 - \mathbb{E}[\hat{g}_h] \|_{L^2(\Omega)}^2 \right] \\ &\leq C \left[h^{2r} \left(|f_0|_{H^{r+1}(\Omega)}^2 + |g_0|_{H^r(\Omega)}^2 \right) + \lambda \| g_0 \|_{L^2(\Omega)}^2 \right]. \end{aligned} \tag{45}$$

Proof. If we define $\hat{f}_h^* = \mathbb{E}[\hat{f}_h]$, $\hat{g}_h^* = \mathbb{E}[\hat{g}_h]$ and the bilinear form $l(\cdot, \cdot)$ as

$$l(f, \psi) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{|D_i|} \int_{D_i} f \psi$$

we can easily obtain the bound (30) for the norm of the bias defined as:

$$\left|\!\left|\!\left| \hat{f}_h^* - f_0 \right|\!\right|\!\right| = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|D_i|}\int_{D_i}(\hat{f}_h^* - f_0)^2 + \lambda\left[\left\|\hat{f}_h^* - f_0\right\|_{H^1(\Omega)}^2 + \|\hat{g}_h^* - g_0\|_{L^2(\Omega)}^2\right].$$

Since the norm associated to the bilinear form $l(\cdot,\cdot)$ is bounded by the H^1-norm we have that

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|D_i|}\int_{D_i}(f_0 - \hat{f}_h^*)^2 + \lambda\left[\left\|f_0 - \hat{f}_h^*\right\|_{H^1(\Omega)}^2 + \|g_0 - \hat{g}_h^*\|_{L^2(\Omega)}^2\right]\\ &\le C\left\{\inf_{(\varphi_h,p_h)\in\mathcal{W}_h}\left[\|f_0 - \varphi_h\|_{H^1(\Omega)}^2 + \lambda\|g_0 - p_h\|_{L^2(\Omega)}^2\right] + \lambda\|g_0\|_{L^2(\Omega)}^2\right\}.\end{aligned}$$

The inequality (33) still holds for $(\varphi_h, p_h) \in \mathcal{W}_h$ and $z_h \in V_h^r$ and we obtain

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|D_i|}\int_{D_i}(f_0 - \hat{f}_h^*)^2 + \lambda\left[\left\|f_0 - \hat{f}_h^*\right\|_{H^1(\Omega)}^2 + \|g_0 - \hat{g}_h^*\|_{L^2(\Omega)}^2\right] \le\\ &C\left\{\inf_{\substack{z_h\in V_h^r\\ z_h|_{\Gamma_D}=h_{D,h}}}\|f_0 - z_h\|_{H^1(\Omega)}^2 + \lambda\inf_{p_h\in V_{h,\Gamma_D}^r}\|g_0 - p_h\|_{L^2(\Omega)}^2 + \lambda\|g_0\|_{L^2(\Omega)}^2\right\}\end{aligned}$$

Using the classic error bound (37) we obtain the desired result. □

9 Conclusion and future work

In this work we have studied the properties of the SR-PDE smoothing technique. This smoothing method has a very broad applicability, since PDEs are commonly used to model physical phenomena. The method is actually not applicable to PDEs with discontinuous parameters, pointwise forcing term or defined on irregular domains, due to the extra regularity required to the parameters of the penalized PDE. This request however is not restrictive in spatial statistics and in the smoothing framework since the field is normally assumed to be very regular. The proposed mixed Finite Element method requires moreover g_0 to satisfy the Dirichlet boundary conditions on Γ_D. This hypothesis could be sometimes restrictive since it means that the second derivatives of the field at the boundary are clumped to zero; other discretization methods for fourth order problems could be considered in the future. However, we have observed numerically that whenever g_0 does not satisfy the homogeneous Dirichlet boundary conditions on Γ_D, the extra consistency error is of the same order as the bias contribution and therefore does not really compromise the optimal convergence rate of the method.

The convergence studied in this work concerns the bias of the estimator when the characteristic mesh size h goes to zero and neglects instead the error induced by the presence of noise in the observations. Classical results concerning smoothing splines and thin-plate splines (see, e.g., [6, 7, 8, 13]) show the consistency of these estimators when the smoothing parameter λ goes to zero, as $n \to +\infty$, with a proper rate. Unfortunately these results cannot be directly extended to SR-PDE and a different approach needs to be developed to show the consistency of these models.

We are currently studying the (infill) asymptotic properties of the estimator when the number of observations n goes to infinity. In particular we are studying the convergence of the variance term $\hat{w}$, both in the continuous and the discrete setting, when n goes to infinity and we are looking for a proper rate of λ that makes both the bias and variance vanish.

It will be also interesting to balance the discretization error induced by the Finite Element approximation with the bias of the estimator and the variance term related to the noise of the observations. A possible way to solve the problem is the development of a proper mesh adaptation technique, based on a posteriori estimates of noise, variance and bias. This technique should locally refine the mesh in order to obtain a local discretization error of the same order or smaller than the bias and the noise standard deviation σ.

Acknowledgments This work is supported by MIUR Ministero dell'Istruzione dell'Università e della Ricerca, *FIRB Futuro in Ricerca* research project "Advanced statistical and numerical methods for the analysis of high dimensional functional data in life sciences and engineering"
(http://mox.polimi.it/users/sangalli/firbSNAPLE.html) and by the program Dote Ricercatore Politecnico di Milano - Regione Lombardia, research project "Functional data analysis for life sciences".

References

[1] L. Azzimonti, L.M. Sangalli, P. Secchi, M. Domanin, and F. Nobile. Blood flow velocity field estimation via spatial regression with PDE penalization. Technical Report 19/2013, MOX - Dipartimento di Matematica, Politecnico di Milano, 2013. available at http://mox.polimi.it/it/progetti/pubblicazioni/.

[2] S.C. Brenner and L.R. Scott. *The mathematical theory of Finite Element methods*. Springer, 2008.

[3] S.C. Brenner and L.Y. Sung. C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains. *J. Sci. Comput*, 22/23:83–118, 2005.

[4] F. Brezzi and M. Fortin. *Mixed and hybrid finite elements methods*. Springer, 1991.

[5] P.G. Ciarlet. *The Finite Element method for elliptic problems*. Society for Industrial and Applied Mathematics, 2002.

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

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

[8] F Cucker and D.X. Zhou. *Learning theory: an approximation theory viewpoint*. Cambridge Monographs on Applied and Computational Mathematics, 2007.

[9] A. Ern and J.L. Guermond. *Theory and Practice of Finite Elements*. Springer, 2004.

[10] B. Ettinger, S. Perotto, and L.M. Sangalli. Spatial regression models over two-dimensional manifolds. Technical Report 54/2012, MOX - Dipartimento di Matematica, Politecnico di Milano, 2012. available at http://mox.polimi.it/it/progetti/pubblicazioni/.

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

[12] E.H. Georgoulis, P. Houston, and Virtanen J. An a posteriori error indicator for discontinuous Galerkin approximations of fourth-order elliptic problems. *IMA J Numer. Anal.*, (31):281–298, 2011.

[13] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 1990.

[14] T.J. Hastie, R.J. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[15] J.L. Lions. *Optimal control of systems governed by partial differential equations*. Springer, 1971.

[16] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, 2005.

[17] T. Ramsay. Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, (64):307–319, 2002.

[18] L.M. Sangalli, J.O. Ramsay, and T. Ramsay. Spatial spline regression models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, (75,4):1–23, 2013.

[19] G. Stone. *Bivariate splines*. PhD thesis, University of Bath, Bath, 1988.

[20] E. Süli and Mozolevski I. hp-version interior penalty DGFEMs for the biharmonic equation. *Comput. Methods Appl. Mech. Engrg.*, (196):1851 –1863, 2007.

[21] G. Wahba. *Spline models for observational data*. Society for Industrial and Applied Mathematics, 1990.

[22] S.N. Wood, M.V. Bravington, and S.L. Hedley. Soap film smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, (70):931–955, 2008.