Technische Universität Graz

An adaptive finite element method for distributed elliptic optimal control problems with variable energy regularization

U. Langer, R. Löscher, O. Steinbach, H. Yang

Berichte aus dem Institut für Angewandte Mathematik

Technische Universität Graz

An adaptive finite element method for distributed elliptic optimal control problems with variable energy regularization

U. Langer, R. Löscher, O. Steinbach, H. Yang

Berichte aus dem Institut für Angewandte Mathematik

Technische Universität Graz
Institut für Angewandte Mathematik
Steyrergasse 30
A 8010 Graz

WWW: `http://www.applied.math.tugraz.at`

An adaptive finite element method for distributed elliptic optimal control problems with variable energy regularization

Ulrich Langer[*], Richard Löscher[†], Olaf Steinbach[‡], Huidong Yang[§]

Abstract

We analyze the finite element discretization of distributed elliptic optimal control problems with variable energy regularization, where the usual $L^2(\Omega)$ norm regularization term with a constant regularization parameter ϱ is replaced by a suitable representation of the energy norm in $H^{-1}(\Omega)$ involving a variable, mesh-dependent regularization parameter $\varrho(x)$. It turns out that the error between the computed finite element state $\widetilde{u}_{\varrho h}$ and the desired state $\overline{u}$ (target) is optimal in the $L^2(\Omega)$ norm provided that $\varrho(x)$ behaves like the local mesh size squared. This is especially important when adaptive meshes are used in order to approximate discontinuous target functions. The adaptive scheme can be driven by the computable and localizable error norm $\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$ between the finite element state $\widetilde{u}_{\varrho h}$ and the target $\overline{u}$. The numerical results not only illustrate our theoretical findings, but also show that the iterative solvers for the discretized reduced optimality system are very efficient and robust.

Keywords: Distributed elliptic optimal control problem, variable energy regularization, finite element discretization, error estimates, adaptivity, solvers
2010 MSC: 49J20, 49M05, 35J05, 65M60, 65M15, 65N22

1 Introduction

Optimal control [2, 10, 17, 22] and inverse problems [8, 12, 20] subject to partial differential equations often involve some parameter-dependent cost or regularization terms, see also the recent special issue [7] on optimal control and inverse problems. While in optimal control problems the regularization parameter ϱ is often considered as a given constant, in inverse problems, the parameter-dependent convergence as $\varrho \to 0$ is well studied, see, e.g., [1]. For tracking type cost functionals subject to elliptic partial differential equations, the regularization error was analyzed in [18] which depends on the regularity of the given target. In [15], and in the case of energy regularization, we have considered a related finite element analysis which resulted in an optimal choice of the regularization parameter $\varrho = h^2$, or vice versa, where h

---

[*]Institute of Computational Mathematics, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria, Email: ulanger@numa.uni-linz.ac.at

[†]Institut für Angewandte Mathematik, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria, Email: loescher@math.tugraz.at

[‡]Institut für Angewandte Mathematik, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria, Email: o.steinbach@tugraz.at

[§]Faculty of Mathematics, University of Vienna, Oskar–Morgenstern–Platz 1, 1090 Wien, Austria, and Christian Doppler Laboratory for Mathematical Modeling and Simulation of Next Generations of Ultrasound Devices (MaMSi), Oskar–Morgenstern–Platz 1, 1090 Wien, Austria, Email: huidong.yang@univie.ac.at

is the mesh size of the globally quasi-uniform finite element mesh. While the latter can be used to approximate smooth target functions in the state space, adaptively refined finite element meshes should be used when considering less regular target functions, e.g., discontinuous or singular, or violating Dirichlet boundary conditions which are involved in the definition of the state space. This motivates the use of a variable regularization parameter function which can be defined by using the local finite element mesh sizes h_ℓ. We are not aware of any paper on optimal control problems dealing with such an approach. However, there are several papers in imaging considering a similar approach: In [11], a variable regularization parameter is used for an adaptive balancing of the data fidelity and the regularization term, see equation (3) in [11]. A variable $L^2(\Omega)$ (TV) regularization is considered in [6, equation (1.1)]. Finally, a spatially adapted total generalized variation model was already used in [4, equation (1.4)], see also the more recent work [9].

As model problem we consider the optimal control problem to minimize the cost functional

$$\mathcal{J}(u_\varrho, z_\varrho) = \frac{1}{2} \int_\Omega [u_\varrho(x) - \overline{u}(x)]^2 dx + \frac{1}{2} \varrho \, \|z_\varrho\|^2_{H^{-1}(\Omega)} \tag{1.1}$$

subject to the Dirichlet boundary value problem for the Poisson equation,

$$- \Delta u_\varrho = z_\varrho \quad \text{in } \Omega, \quad u_\varrho = 0 \quad \text{on } \partial\Omega. \tag{1.2}$$

We assume that $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$, is a bounded Lipschitz domain, and $\varrho \in \mathbb{R}_+$ is some, at this time constant, regularization parameter, on which the solution (u_ϱ, z_ϱ) depends on. Our particular interest is in the behavior of $\|u_\varrho - \overline{u}\|_{L^2(\Omega)}$ as $\varrho \to 0$, see [18]. Note that the energy norm as used in (1.1) is given by

$$\|z_\varrho\|^2_{H^{-1}(\Omega)} = \|\nabla u_\varrho\|^2_{L^2(\Omega)} = \langle z_\varrho, u_\varrho \rangle_\Omega, = \langle z_\varrho, \mathcal{S}z_\varrho \rangle_\Omega,$$

where $u_\varrho = \mathcal{S}z_\varrho \in H^1_0(\Omega)$ is the weak solution of the primal Dirichlet boundary value problem (1.2), and $\mathcal{S} : H^{-1}(\Omega) \to H^1_0(\Omega) \subset L^2(\Omega)$ is the associated solution operator. Hence we can write the reduced cost functional as

$$\widetilde{\mathcal{J}}(z_\varrho) = \frac{1}{2} \, \|\mathcal{S}z_\varrho - \overline{u}\|^2_{L^2(\Omega)} + \frac{1}{2} \varrho \, \langle \mathcal{S}z_\varrho, z_\varrho \rangle_\Omega, \tag{1.3}$$

and its minimizer is given as the unique solution of the gradient equation

$$\mathcal{S}^*(\mathcal{S}z_\varrho - \overline{u}) + \varrho \, \mathcal{S}z_\varrho = 0 \, .$$

In addition to $u_\varrho = \mathcal{S}z_\varrho$ we now introduce the adjoint state $p_\varrho = \mathcal{S}^*(u_\varrho - \overline{u})$ as unique solution of the Dirichlet boundary value problem

$$- \Delta p_\varrho = u_\varrho - \overline{u} \quad \text{in } \Omega, \quad p_\varrho = 0 \quad \text{on } \partial\Omega. \tag{1.4}$$

Hence we can rewrite the gradient equation as

$$p_\varrho + \varrho \, u_\varrho = 0 \quad \text{in } \Omega.$$

When eliminating the adjoint state p_ϱ we can determine the optimal state u_ϱ as the solution of the singularly perturbed Dirichlet boundary value problem

$$- \varrho \, \Delta u_\varrho + u_\varrho = \overline{u} \quad \text{in } \Omega, \quad u_\varrho = 0 \quad \text{on } \partial\Omega, \tag{1.5}$$

which is also known as differential filter in fluid mechanics [13]. The variational formulation of (1.5) is to find $u_\varrho \in H^1_0(\Omega)$ such that

$$\varrho \, \langle \nabla u_\varrho, \nabla v \rangle_{L^2(\Omega)} + \langle u_\varrho, v \rangle_{L^2(\Omega)} = \langle \overline{u}, v \rangle_{L^2(\Omega)} \quad \text{for all } v \in H^1_0(\Omega). \tag{1.6}$$

2

For a finite element discretization of the variational formulation (1.5) we may use the ansatz space $V_h := S_h^1(\Omega) \cap H_0^1(\Omega)$ of piecewise linear and continuous basis functions which are defined with respect to some admissible and globally quasi-uniform decomposition of Ω into simplicial finite elements of mesh size h. The Galerkin variational formulation of (1.6) is to find $u_{\varrho h} \in V_h$ such that

$$\varrho \langle \nabla u_{\varrho h}, \nabla v_h \rangle_{L^2(\Omega)} + \langle u_{\varrho h}, v_h \rangle_{L^2(\Omega)} = \langle \overline{u}, v_h \rangle_{L^2(\Omega)} \quad \text{for all } v_h \in V_h. \qquad (1.7)$$

When combining the regularization error estimates for $\|u_\varrho - \overline{u}\|_{L^2(\Omega)}$ as given in [18] with finite element error estimates for the approximate solution $u_{\varrho h}$, i.e., for $\|u_{\varrho h} - u_\varrho\|_{L^2(\Omega)}$, we were able to derive estimates for the error $\|u_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$, see [15]. In particular for the optimal choice $\varrho = h^2$ this gives

$$\|u_{\varrho h} - \overline{u}\|_{L^2(\Omega)} \le c\, h^s\, \|\overline{u}\|_{H^s(\Omega)}, \qquad (1.8)$$

when assuming $\overline{u} \in H_0^s(\Omega) := [L^2(\Omega), H_0^1(\Omega)]_s$ for $s \in [0, 1]$, or $\overline{u} \in H_0^1(\Omega) \cap H^s(\Omega)$ for $s \in (1, 2]$. This error estimate remains true when considering optimal control problems with energy regularization subject to time-dependent partial differential equations, see [16] in the case of the heat equation. However, when considering less regular target functions $\overline{u}$, e.g., singular or discontinuous targets, the use of adaptive finite elements seems to be mandatory in order to gain optimal computational complexity. In this case it is not obvious how to choose a constant regularization parameter ϱ, e.g., $\varrho = h_{\min}^2$. Instead, one may use a locally adapted regularization function $\varrho(x)$, $x \in \Omega$. The energy norm in $H^{-1}(\Omega)$ can be realized by duality when solving a Poisson equation with zero Dirichlet boundary conditions. When including the (constant) regularization parameter ϱ, we can generalize this approach by considering a diffusion equation with $\varrho(x)^{-1}$ as diffusion coefficient in order to realize the variable energy regularization within an adaptive finite element discretization.

The rest of this paper is organized as follows. In Section 2, we derive the analogon of the optimal control problem (1.1) when using a regularization function $\varrho(x)$ instead of a constant regularization parmeter ϱ, and the corresponding reduced optimality systems. Section 3 provides estimates of the derivation of the state u_ϱ from the desired state $\overline{u}$ with respect to the L^2-norm in terms of the regularization function $\varrho(x)$, and the regularity of the target $\overline{u}$. In Section 4 we analyze the L^2-norm $\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$ of the error between the computed finite element state $\widetilde{u}_{\varrho h}$ and the desired state $\overline{u}$ leading to an elementwise adaption of the regularization function $\varrho(x)$ to the local mesh size h_ℓ. The first part of Section 5 is devoted to numerical studies of the error $\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$ for benchmark problems with discontinuous targets $\overline{u}$, whereas the second part provides numerical studies of the proposed iterative solvers in the three-dimensional case $n = 3$. Finally, in Section 6, we draw some conclusions, and give some outlook on further research work.

2 Distributed optimal control problems with variable energy regularization

To give a motivation for the optimization problem to be solved, let us consider an alternative representation of the energy norm, still using a constant regularization parameter ϱ:

$$\varrho \|z_\varrho\|_{H^{-1}(\Omega)}^2 = \|\sqrt{\varrho}\, z_\varrho\|_{H^{-1}(\Omega)}^2 = \langle \sqrt{\varrho}\, z_\varrho, w_\varrho \rangle_\Omega,$$

where $w_\varrho \in H_0^1(\Omega)$ is the weak solution of the Dirichlet boundary value problem

$$-\Delta w_\varrho = \sqrt{\varrho}\, z_\varrho \quad \text{in } \Omega, \quad w_\varrho = 0 \quad \text{on } \partial\Omega. \qquad (2.1)$$

3

Now we may introduce $\widetilde{w}_\varrho = \sqrt{\varrho}\, w_\varrho$ to conclude the diffusion equation

$$-\operatorname{div}\left[\varrho^{-1}\, \nabla \widetilde{w}_\varrho\right] = z_\varrho \quad \text{in } \Omega, \quad \widetilde{w}_\varrho = 0 \quad \text{on } \partial\Omega,$$

and the norm representation

$$\varrho \, \|z_\varrho\|_{H^{-1}(\Omega)}^2 = \langle \sqrt{\varrho}\, z_\varrho, w_\varrho \rangle_\Omega = \langle z_\varrho, \widetilde{w}_\varrho \rangle_\Omega.$$

Instead of using a constant regularization parameter ϱ we now consider a diffusion equation with a variable diffusion coefficient $\varrho \in L^\infty(\Omega)$, that is uniformely bounded from above and below, i.e., there exists positive constants $\underline{\varrho}$ and $\overline{\varrho}$ such that $0 < \underline{\varrho} \le \varrho(x) \le \overline{\varrho} < \infty$ for almost every $x \in \Omega$. More precisely, we consider the Dirichlet boundary value problem

$$- \operatorname{div}\left[\frac{1}{\varrho(x)}\, \nabla \widetilde{w}_\varrho(x)\right] = z_\varrho(x) \quad \text{for } x \in \Omega, \quad \widetilde{w}_\varrho(x) = 0 \quad \text{for } x \in \partial\Omega, \qquad (2.2)$$

and its variational formulation to find $\widetilde{w}_\varrho \in H_0^1(\Omega)$ such that

$$\langle A_{1/\varrho}\widetilde{w}_\varrho, v\rangle_\Omega := \int_\Omega \frac{1}{\varrho(x)}\, \nabla \widetilde{w}_\varrho(x) \cdot \nabla v(x)\, dx = \int_\Omega z_\varrho(x)\, v(x)\, dx$$

is satisfied for all $v \in H_0^1(\Omega)$, i.e., we have $\widetilde{w}_\varrho = A_{1/\varrho}^{-1} z_\varrho$. Instead of (1.3) we now consider the reduced cost functional

$$\widetilde{\mathcal{J}}(z_\varrho) = \frac{1}{2}\, \|\mathcal{S}z_\varrho - \overline{u}\|_{L^2(\Omega)}^2 + \frac{1}{2}\, \langle A_{1/\varrho}^{-1} z_\varrho, z_\varrho\rangle_\Omega, \qquad (2.3)$$

whose minimizer is given as the unique solution of the gradient equation

$$\mathcal{S}^*(\mathcal{S}z_\varrho - \overline{u}) + A_{1/\varrho}^{-1} z_\varrho = 0,$$

i.e.,

$$p_\varrho + \widetilde{w}_\varrho = 0 \quad \text{in } \Omega. \qquad (2.4)$$

The optimality system to be solved is now given by the primal problem (1.2), the adjoint problem (1.4), the gradient equation (2.4), and (2.2). When eliminating $\widetilde{w}_\varrho$ and z_ϱ, we end up with a coupled system of the primal problem

$$- \operatorname{div}\left[\frac{1}{\varrho(x)}\, \nabla p_\varrho(x)\right] - \Delta u_\varrho(x) = 0 \quad \text{for } x \in \Omega, \quad u_\varrho(x) = 0 \quad \text{for } x \in \partial\Omega, \quad (2.5)$$

and the adjoint boundary value problem (1.4). The related variational formulation is to find $(u_\varrho, p_\varrho) \in H_0^1(\Omega) \times H_0^1(\Omega)$ such that

$$\int_\Omega \frac{1}{\varrho(x)}\, \nabla p_\varrho(x) \cdot \nabla v(x)\, dx + \int_\Omega \nabla u_\varrho(x) \cdot \nabla v(x)\, dx = 0 \qquad (2.6)$$

for all $v \in H_0^1(\Omega)$, and

$$\int_\Omega u_\varrho(x)\, q(x)\, dx - \int_\Omega \nabla p_\varrho(x) \cdot \nabla q(x)\, dx = \int_\Omega \overline{u}(x)\, q(x)\, dx \qquad (2.7)$$

for all $q \in H_0^1(\Omega)$. When introducing

$$\langle Bu, v\rangle_\Omega := \int_\Omega \nabla u(x) \cdot \nabla v(x)\, dx \quad \text{for all } u, v \in H_0^1(\Omega),$$

we can write the coupled variational formulation (2.6) and (2.7) in operator form as

$$A_{1/\varrho}p_\varrho + Bu_\varrho = 0, \quad u_\varrho - B^* p_\varrho = \overline{u},$$

4

and eliminating p_ϱ results in the Schur complement system to find $u_\varrho \in H_0^1(\Omega)$ such that

$$B^* A_{1/\varrho}^{-1} B u_\varrho + u_\varrho = \overline{u}. \tag{2.8}$$

Note that

$$S_\varrho := B^* A_\varrho^{-1} B : H_0^1(\Omega) \to H^{-1}(\Omega) \tag{2.9}$$

is bounded, self-adjoint, and $H_0^1(\Omega)$ elliptic. Note that for a constant regularization parameter $\varrho(x) = \varrho$, (2.8) coincides with (1.5).

3 Regularization error estimates

In this section, we consider estimates for the regularization error $\|u_\varrho - \overline{u}\|_{L^2(\Omega)}$ when $u_\varrho \in H_0^1(\Omega)$ is the weak solution of the operator equation

$$S_\varrho u_\varrho + u_\varrho = \overline{u}, \tag{3.1}$$

and where S_ϱ is as defined in (2.9). Note that S_ϱ induces a norm, satisfying

$$\|v\|_{S_\varrho} := \sqrt{\langle S_\varrho v, v \rangle_\Omega}, \quad \langle S_\varrho u, v \rangle_\Omega \le \|u\|_{S_\varrho} \|v\|_{S_\varrho} \quad \text{for all } u, v \in H_0^1(\Omega).$$

First we follow the general approach as given in [16, Section 2] in the case of a constant regularization parameter. The variational formulation of the operator equation (3.1) is to find $u_\varrho \in H_0^1(\Omega)$ such that

$$\langle S_\varrho u_\varrho, v \rangle_\Omega + \langle u_\varrho, v \rangle_{L^2(\Omega)} = \langle \overline{u}, v \rangle_{L^2(\Omega)} \tag{3.2}$$

is satisfied for all $v \in H_0^1(\Omega)$. Unique solvability of the variational formulation (3.2) follows from the boundedness and ellipticity of S_ϱ.

Theorem 1. *Let $u_\varrho \in H_0^1(\Omega)$ be the unique solution of the variational formulation (3.2). For $\overline{u} \in L^2(\Omega)$ there holds the estimate*

$$\|u_\varrho - \overline{u}\|_{L^2(\Omega)} \le \|\overline{u}\|_{L^2(\Omega)}. \tag{3.3}$$

For $\overline{u} \in H_0^1(\Omega)$ we have

$$\|u_\varrho - \overline{u}\|_{S_\varrho} \le \|\overline{u}\|_{S_\varrho}, \tag{3.4}$$

and

$$\|u_\varrho - \overline{u}\|_{L^2(\Omega)} \le \|\overline{u}\|_{S_\varrho}. \tag{3.5}$$

If $\overline{u} \in H_0^1(\Omega)$ is such that $S_\varrho \overline{u} \in L^2(\Omega)$ is satisfied, we also have

$$\|u_\varrho - \overline{u}\|_{L^2(\Omega)} \le \|S_\varrho \overline{u}\|_{L^2(\Omega)}, \tag{3.6}$$

and

$$\|u_\varrho - \overline{u}\|_{S_\varrho} \le \|S_\varrho \overline{u}\|_{L^2(\Omega)}. \tag{3.7}$$

Proof. When considering the variational formulation (3.2) for $v = u_\varrho \in H_0^1(\Omega)$, this gives

$$\langle S_\varrho u_\varrho, u_\varrho \rangle_\Omega + \langle u_\varrho, u_\varrho \rangle_{L^2(\Omega)} = \langle \overline{u}, u_\varrho \rangle_{L^2(\Omega)},$$

which can be written as

$$\langle S_\varrho u_\varrho, u_\varrho \rangle_\Omega + \langle u_\varrho - \overline{u}, u_\varrho - \overline{u} \rangle_{L^2(\Omega)} = \langle \overline{u} - u_\varrho, \overline{u} \rangle_{L^2(\Omega)},$$

i.e.,

$$\langle S_\varrho u_\varrho, u_\varrho \rangle_\Omega + \|u_\varrho - \overline{u}\|_{L^2(\Omega)}^2 = \langle \overline{u} - u_\varrho, u_\varrho \rangle_{L^2(\Omega)} \le \|u_\varrho - \overline{u}\|_{L^2(\Omega)} \|\overline{u}\|_{L^2(\Omega)}.$$

Hence we conclude (3.3). For $\overline{u} \in H_0^1(\Omega)$ we can consider the variational formulation (3.2) for $v = \overline{u} - u_\varrho \in H_0^1(\Omega)$ to obtain

$$
\begin{aligned}
\|\overline{u} - u_\varrho\|_{L^2(\Omega)}^2 &= \langle \overline{u} - u_\varrho, \overline{u} - u_\varrho \rangle_{L^2(\Omega)} = \langle S_\varrho u_\varrho, \overline{u} - u_\varrho \rangle_\Omega \\
&= -\langle S_\varrho \overline{u} - u_\varrho, \overline{u} - u_\varrho \rangle_\Omega + \langle S_\varrho \overline{u}, \overline{u} - u_\varrho \rangle_\Omega,
\end{aligned}
$$

i.e.,

$$
\|u_\varrho - \overline{u}\|_{L^2(\Omega)}^2 + \|u_\varrho - \overline{u}\|_{S_\varrho}^2 = \langle S_\varrho \overline{u}, \overline{u} - u_\varrho \rangle_\Omega \le \|\overline{u}\|_{S_\varrho} \|u_\varrho - \overline{u}\|_{S_\varrho}.
$$

From this we conclude

$$
\|u_\varrho - \overline{u}\|_{S_\varrho} \le \|\overline{u}\|_{S_\varrho},
$$

that is (3.4), and

$$
\|u_\varrho - \overline{u}\|_{L^2(\Omega)} \le \|\overline{u}\|_{S_\varrho},
$$

i.e., (3.5). If $\overline{u} \in H_0^1(\Omega)$ is such that $S_\varrho \overline{u} \in L^2(\Omega)$ is satisfied, we also have

$$
\|u_\varrho - \overline{u}\|_{L^2(\Omega)}^2 + \|u_\varrho - \overline{u}\|_{S_\varrho}^2 = \langle S_\varrho \overline{u}, \overline{u} - u_\varrho \rangle_\Omega \le \|S_\varrho \overline{u}\|_{L^2(\Omega)} \|u_\varrho - \overline{u}\|_{L^2(\Omega)}.
$$

Hence we obtain

$$
\|u_\varrho - \overline{u}\|_{L^2(\Omega)} \le \|S_\varrho \overline{u}\|_{L^2(\Omega)},
$$

that is (3.6), and

$$
\|u_\varrho - \overline{u}\|_{S_\varrho} \le \|S_\varrho \overline{u}\|_{L^2(\Omega)},
$$

i.e., (3.7). $\qquad\square$

Let us now consider the operator S_ϱ as defined in (2.9). For $u \in H_0^1(\Omega)$ let $p_u = A_{1/\varrho}^{-1} B u \in H_0^1(\Omega)$ be the unique solution of the variational formulation

$$
\langle A_{1/\varrho} p_u, v \rangle_\Omega = \int_\Omega \frac{1}{\varrho(x)} \nabla p_u(x) \cdot \nabla v(x)\, dx = \int_\Omega \nabla u(x) \cdot \nabla v(x)\, dx = \langle B u, v \rangle_\Omega
$$

for all $v \in H_0^1(\Omega)$. We first conclude

$$
\|u\|_{S_\varrho}^2 = \langle S_\varrho u, u \rangle_\Omega = \langle B^* A_{1/\varrho}^{-1} B u, u \rangle_\Omega = \langle p_u, B u \rangle_\Omega = \langle A_{1/\varrho} p_u, p_u \rangle_\Omega.
$$

Moreover, using a weighted Cauchy–Schwarz inequality, this gives

$$
\begin{aligned}
\langle A_{1/\varrho} p_u, p_u \rangle_\Omega &= \int_\Omega \frac{1}{\varrho(x)} \nabla p_u(x) \cdot \nabla p_u(x)\, dx = \int_\Omega \nabla u(x) \cdot \nabla p_u(x)\, dx \\
&\le \left(\int_\Omega \varrho(x)\, \nabla u(x) \cdot \nabla u(x)\, dx \right)^{1/2} \left(\int_\Omega \frac{1}{\varrho(x)} \nabla p_u(x) \cdot \nabla p_u(x)\, dx \right)^{1/2},
\end{aligned}
$$

i.e.,

$$
\|u\|_{S_\varrho}^2 \le \int_\Omega \varrho(x)\, |\nabla u(x)|^2\, dx \quad \text{for all } u \in H_0^1(\Omega). \tag{3.8}
$$

When combining this with the regularization error estimate (3.5) this gives, for $\overline{u} \in H_0^1(\Omega)$,

$$
\|u_\varrho - \overline{u}\|_{L^2(\Omega)}^2 \le \int_\Omega \varrho(x)\, |\nabla \overline{u}(x)|^2\, dx. \tag{3.9}
$$

It remains to consider $\|S_\varrho u\|_{L^2(\Omega)}$ when S_ϱ is given as in (2.9), i.e.,

$$
\|S_\varrho u\|_{L^2(\Omega)} = \| -\Delta p_u \|_{L^2(\Omega)}, \quad -\mathrm{div}\left[\frac{1}{\varrho(x)} \nabla p_u(x) \right] = -\Delta u(x).
$$

We first compute

$$\varrho(x)\Delta u(x) = \varrho(x)\operatorname{div}\left[\frac{1}{\varrho(x)}\nabla p_u(x)\right] = \varrho(x)\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left[\frac{1}{\varrho(x)}\frac{\partial}{\partial x_k}p_u(x)\right]$$

$$= \varrho(x)\nabla\left(\frac{1}{\varrho(x)}\right)\cdot\nabla p_u(x) + \Delta p_u(x).$$

For the first part we further have

$$\varrho(x)\frac{\partial}{\partial x_k}\frac{1}{\varrho(x)} = \varrho(x)\left[-\frac{1}{[\varrho(x)]^2}\frac{\partial}{\partial x_k}\varrho(x)\right] = -\frac{1}{\varrho(x)}\frac{\partial}{\partial x_k}\varrho(x),$$

and hence,

$$\varrho(x)\nabla\left(\frac{1}{\varrho(x)}\right)\cdot\nabla p_u(x) = -\frac{1}{\varrho(x)}\nabla\varrho(x)\cdot\nabla p_u(x),$$

i.e.,

$$\Delta p_u(x) = \varrho(x)\Delta u(x) + \frac{1}{\varrho(x)}\nabla\varrho(x)\cdot\nabla p_u(x).$$

When taking the square and applying Hölder's inequality we obtain

$$[\Delta p_u(x)]^2 \leq 2\,[\varrho(x)\Delta\overline{u}(x)]^2 + 2\,\frac{1}{[\varrho(x)]^2}\left[\nabla\varrho(x)\cdot\nabla p_u(x)\right]^2$$

$$\leq 2\,[\varrho(x)\Delta\overline{u}(x)]^2 + 2\,\frac{1}{[\varrho(x)]^2}\,|\nabla\varrho(x)|^2\,|\nabla p_u(x)|^2,$$

and therefore

$$\|\Delta p_u\|_{L^2(\Omega)}^2 \leq 2\,\|\varrho\Delta\overline{u}\|_{L^2(\Omega)}^2 + 2\int_\Omega\frac{|\nabla\varrho(x)|^2}{[\varrho(x)]^2}\,|\nabla p_u(x)|^2\,dx$$

$$\leq 2\,\|\varrho\Delta\overline{u}\|_{L^2(\Omega)}^2 + 2\,\sup_{x\in\Omega}\frac{|\nabla\varrho(x)|^2}{[\varrho(x)]}\int_\Omega\frac{1}{[\varrho(x)]}\,|\nabla p_u(x)|^2\,dx$$

follows. Using (3.8) we finally obtain

$$\|S_\varrho u\|_{L^2(\Omega)}^2 \leq 2\,\|\varrho\Delta u\|_{L^2(\Omega)}^2 + 2\,\sup_{x\in\Omega}\frac{|\nabla\varrho(x)|^2}{[\varrho(x)]}\int_\Omega\varrho(x)]\,|\nabla u(x)|^2\,dx\,.$$

When assuming

$$\sup_{x\in\Omega}\frac{|\nabla\varrho(x)|^2}{[\varrho(x)]}\leq c_\varrho, \tag{3.10}$$

and combining this with (3.6) we obtain

$$\|u_\varrho - \overline{u}\|_{L^2(\Omega)}^2 \leq 2\,\|\varrho\Delta\overline{u}\|_{L^2(\Omega)}^2 + 2\,c_\varrho\int_\Omega\varrho(x)\,|\nabla\overline{u}(x)|^2\,dx\,. \tag{3.11}$$

While for a constant regularization parameter $\varrho(x) = \varrho$ we obviously have $c_\varrho = 0$, in the more general situation of a, e.g., piecewise linear parameter function $\varrho(x)$ we finally obtain a similar error estimate as in (3.9) when assuming $\overline{u} \in H_0^1(\Omega)$ only. Hence we restrict our considerations to $\overline{u} \in H_0^1(\Omega)$, and to less regular target functions, where we can formulate the following result.

Theorem 2. *Let $u_\varrho \in H_0^1(\Omega)$ be the unique solution of the Schur complement system (2.8), where the regularization function $\varrho \in L^\infty(\Omega)$ is assumed to be bounded and uniform positive. For $\overline{u} \in H_0^1(\Omega)$ there holds the regularization error estimate*

$$\|u_\varrho - \overline{u}\|_{L^2(\Omega)}^2 \leq \int_\Omega\varrho(x)\,|\nabla\overline{u}(x)|^2\,dx. \tag{3.12}$$

7

Remark 1. *Due to the assumption $\varrho \in L^\infty(\Omega)$ we can write (3.12) as*

$$\|u_\varrho - \overline{u}\|^2_{L^2(\Omega)} \leq \|\varrho\|_{L^\infty(\Omega)}\, \|\overline{u}\|_{H^1(\Omega)},$$

and using (3.3) together with an interpolation argument we conclude

$$\|u_\varrho - \overline{u}\|^2_{L^2(\Omega)} \leq \|\varrho\|^s_{L^\infty(\Omega)}\, \|\overline{u}\|_{H^s(\Omega)}$$

when assuming $\overline{u} \in [L^2(\Omega), H^1_0(\Omega)]_s$ for some $s \in [0,1]$.
   Alternatively, we can consider the eigenvalue problem

$$-div\left[\frac{1}{\varrho(x)}\,\nabla v(x)\right] = \lambda\, v(x) \quad for\, x \in \Omega, \quad v(x) = 0 \quad for\, x \in \partial\Omega, \quad \|v\|_{L^2(\Omega)} = 1,$$

where the eigenfunctions $\{v_i\}_{i\in\mathbb{N}}$ form an orthonormal basis in $L^2(\Omega)$, and the eigenvalues $\lambda_i = \lambda_i(\varrho) \in \mathbb{R}_+$ are positive and tend to infinity as $i \to \infty$. Hence we can define, for $s \in [0,1]$, the weighted Sobolev norms

$$\|u\|^2_{H^s_\varrho(\Omega)} := \sum_{i=1}^\infty \lambda_i^s\, u_i^2, \quad u_i = \langle u, v_i\rangle_{L^2(\Omega)}.$$

With this we can write the regularization error estimates (3.3) and (3.12) as

$$\|u_\varrho - \overline{u}\|_{L^2(\Omega)} \leq \|\overline{u}\|_{H^s_\varrho(\Omega)}$$

for $s = 0$ and $s = 1$, respectively. Using an interpolation argument, this remains true for all $s \in (0,1)$ when assuming $\overline{u} \in H^s_\varrho(\Omega)$. Under additional assumptions on ϱ, one can give a more explicit representation of the norm for $H^s_\varrho(\Omega)$ showing the explicit dependency on the powers of ϱ, see [21, Theorem 3.4.3].

4 Finite element error estimates

Let $V_h = S^1_h(\Omega) \cap H^1_0(\Omega) = \mathrm{span}\{\varphi_k\}_{k=1}^M$ be the finite element space of piecewise linear and continuous basis functions φ_k, which are defined with respect to some admissible locally quasi-uniform decomposition of the computational domain Ω into simplicial shape-regular finite elements τ_ℓ of local mesh size $h_\ell = \Delta_\ell^{1/n}$, where Δ_ℓ is the volume of the finite element τ_ℓ. As regularization we consider the piecewise constant function

$$\varrho(x) = h_\ell^2 \quad \text{for } x \in \tau_\ell. \tag{4.1}$$

The Galerkin variational formulation of the abstract operator equation (3.1) is to find $u_{\varrho h} \in V_h$ such that

$$\langle S_\varrho u_{\varrho h}, v_h\rangle_\Omega + \langle u_{\varrho h}, v_h\rangle_{L^2(\Omega)} = \langle \overline{u}, v_h\rangle_{L^2(\Omega)} \quad \text{for all } v_h \in V_h. \tag{4.2}$$

Using standard arguments we conclude Cea's lemma,

$$\|u_\varrho - u_{\varrho h}\|^2_{S_\varrho} + \|u_\varrho - u_{\varrho h}\|^2_{L^2(\Omega)} \leq \inf_{v_h \in V_h}\left[\|u_\varrho - v_h\|^2_{S_\varrho} + \|u_\varrho - v_h\|^2_{L^2(\Omega)}\right], \tag{4.3}$$

from which we further obtain

$$\begin{aligned}
\|u_\varrho - u_{\varrho h}\|^2_{L^2(\Omega)} &\leq 2\Big[\|u_\varrho - \overline{u}\|^2_{S_\varrho} + \|u_\varrho - \overline{u}\|^2_{L^2(\Omega)} \\
&\qquad + \inf_{v_h \in V_h}\left(\|\overline{u} - v_h\|^2_{S_\varrho} + \|\overline{u} - v_h\|^2_{L^2(\Omega)}\right)\Big].
\end{aligned} \tag{4.4}$$

When assuming $\overline{u} \in H_0^1(\Omega)$, and using the regularization error estimates (3.4) and (3.5), this gives

$$\|u_\varrho - u_{\varrho h}\|_{L^2(\Omega)}^2 \leq 2 \left[2 \|\overline{u}\|_{S_\varrho}^2 + \inf_{v_h \in V_h} \left(\|\overline{u} - v_h\|_{S_\varrho}^2 + \|\overline{u} - v_h\|_{L^2(\Omega)}^2 \right) \right].$$

Let $\Pi_h \overline{u} \in V_h$ be a quasi-interpolation of $\overline{u} \in H_0^1(\Omega)$, e.g., using the Scott–Zhang operator Π_h, see, e.g., [5], and satisfying the error estimates

$$\|\overline{u} - \Pi_h \overline{u}\|_{L^2(\tau_\ell)} \leq c\, h_\ell \,\|\nabla \overline{u}\|_{L^2(\omega_\ell)}, \tag{4.5}$$

and

$$\|\nabla(\overline{u} - \Pi_h \overline{u})\|_{L^2(\tau_\ell)} \leq c \,\|\nabla \overline{u}\|_{L^2(\omega_\ell)}. \tag{4.6}$$

Here, $\omega_\ell := \{\tau_j : \overline{\tau}_\ell \cap \overline{\tau}_j \neq \emptyset\}$ is the simplex patch of τ_ℓ. Then we can estimate, using (3.8) and (4.6),

$$\begin{aligned}
\|\overline{u} - \Pi_h \overline{u}\|_{S_\varrho}^2 &\leq \int_\Omega \varrho(x) \,|\nabla(\overline{u} - \Pi_h \overline{u}(x))|^2 \,dx \\
&= \sum_{\ell=1}^N h_\ell^2 \int_{\tau_\ell} |\nabla(\overline{u}(x) - \Pi_h \overline{u}(x))|^2 \,dx \\
&= \sum_{\ell=1}^N h_\ell^2 \|\nabla(\overline{u} - \Pi_h \overline{u})\|_{L^2(\tau_\ell)}^2 \leq c \sum_{\ell=1}^N h_\ell^2 \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2.
\end{aligned}$$

Moreover, using (4.5), we also have

$$\|\overline{u} - \Pi_h \overline{u}\|_{L^2(\Omega)}^2 = \sum_{\ell=1}^N \|\overline{u} - \Pi_h \overline{u}\|_{L^2(\tau_\ell)}^2 \leq c \sum_{\ell=1}^N h_\ell^2 \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2.$$

Together with (3.8) we then obtain

$$\|u_\varrho - u_{\varrho h}\|_{L^2(\Omega)}^2 \leq c \sum_{\ell=1}^N h_\ell^2 \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2, \tag{4.7}$$

and with (3.9) this finally gives

$$\|u_{\varrho h} - \overline{u}\|_{L^2(\Omega)}^2 \leq c \sum_{\ell=1}^N h_\ell^2 \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2. \tag{4.8}$$

The variational formulation (4.2) requires, for any given $u \in H_0^1(\Omega)$, the evaluation of $S_\varrho u = B^* A_{1/\varrho}^{-1} B u = B^* p_u$, where $p_u = A_{1/\varrho}^{-1} B u \in H_0^1(\Omega)$ is the unique solution of the variational formulation

$$\int_\Omega \frac{1}{\varrho(x)} \nabla p_u(x) \cdot \nabla v(x) \,dx = \int_\Omega \nabla u(x) \cdot \nabla v(x) \,dx \quad \text{for all } v \in H_0^1(\Omega). \tag{4.9}$$

Hence we can define the approximate solution $p_{uh} \in V_h$ satisfying

$$\int_\Omega \frac{1}{\varrho(x)} \nabla p_{uh}(x) \cdot \nabla v_h(x) \,dx = \int_\Omega \nabla u(x) \cdot \nabla v_h(x) \,dx \quad \text{for all } v_h \in V_h, \tag{4.10}$$

and therefore we can introduce an approximation $\widetilde{S}_\varrho u = B^* p_{uh}$ of $S_\varrho u = B^* p_u$. Instead of (4.2) we now consider the perturbed variational formulation to find $\widetilde{u}_{\varrho h} \in V_h$ such that

$$\langle \widetilde{S}_\varrho \widetilde{u}_{\varrho h}, v_h \rangle_\Omega + \langle \widetilde{u}_{\varrho h}, v_h \rangle_{L^2(\Omega)} = \langle \overline{u}, v_h \rangle_{L^2(\Omega)} \quad \text{for all } v_h \in V_h. \tag{4.11}$$

Unique solvability of (4.11) follows since the stiffness matrix of $\widetilde{S}_\varrho$ is positive semi-definite, while the mass matrix related to the inner product in $L^2(\Omega)$ is positive definite.

9

Lemma 1. *Let $\widetilde{u}_{\varrho h} \in V_h$ be the unique solution of the perturbed variational formulation (4.11). Then there holds the error estimate*

$$\|\widetilde{u}_{\varrho h} - u_{\varrho h}\|_{L^2(\Omega)} \le c \sum_{\ell=1}^{N} h_\ell^2 \, \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2. \tag{4.12}$$

Proof. The difference of the variational formulations (4.2) and (4.11) first gives the Galerkin orthogonality

$$\langle S_\varrho u_{\varrho h} - \widetilde{S}_\varrho \widetilde{u}_{\varrho h}, v_h \rangle_\Omega + \langle u_{\varrho h} - \widetilde{u}_{\varrho h}, v_h \rangle_{L^2(\Omega)} = 0 \quad \text{for all } v_h \in V_h,$$

which can be written as

$$\langle \widetilde{S}_\varrho(\widetilde{u}_{\varrho h} - u_{\varrho h}), v_h \rangle_\Omega + \langle \widetilde{u}_{\varrho h} - u_{\varrho h}, v_h \rangle_{L^2(\Omega)} = \langle (S_\varrho - \widetilde{S}_\varrho) u_{\varrho h}, v_h \rangle_\Omega \quad \text{for all } v_h \in V_h.$$

When chosing $v_h = \widetilde{u}_{\varrho h} - u_{\varrho h} \in V_h$, and using $\langle \widetilde{S}_\varrho u, u \rangle_\Omega \ge 0$ for all $u \in H_0^1(\Omega)$, this gives

$$
\begin{aligned}
\|\widetilde{u}_{\varrho h} - u_{\varrho h}\|_{L^2(\Omega)}^2 \;\le\;& \langle (S_\varrho - \widetilde{S}_\varrho) u_{\varrho h}, \widetilde{u}_{\varrho h} - u_{\varrho h} \rangle_\Omega \\
=\;& \langle B^*(p_{u_{\varrho h}} - p_{u_{\varrho h} h}), \widetilde{u}_{\varrho h} - u_{\varrho h} \rangle_\Omega \\
=\;& \int_\Omega \nabla(p_{u_{\varrho h}} - p_{u_{\varrho h} h}) \cdot \nabla(\widetilde{u}_{\varrho h} - u_{\varrho h}) \, dx \\
\le\;& \left(\int_\Omega \frac{1}{\varrho} |\nabla(p_{u_{\varrho h}} - p_{u_{\varrho h} h})|^2 \, dx \right)^{1/2} \left(\int_\Omega \varrho \, |\nabla(\widetilde{u}_{\varrho h} - u_{\varrho h})|^2 \, dx \right)^{1/2}.
\end{aligned}
$$

Using (4.1) and an inverse inequality locally, we further have

$$
\begin{aligned}
\int_\Omega \varrho \, |\nabla(\widetilde{u}_{\varrho h} - u_{\varrho h})|^2 \, dx \;=\;& \sum_{\ell=1}^{N} h_\ell^2 \, \|\nabla(\widetilde{u}_{\varrho h} - u_{\varrho h})\|_{L^2(\tau_\ell)}^2 \\
\le\;& c \sum_{\ell=1}^{N} \|\widetilde{u}_{\varrho h} - u_{\varrho h}\|_{L^2(\tau_\ell)}^2 \;=\; c \, \|\widetilde{u}_{\varrho h} - u_{\varrho h}\|_{L^2(\Omega)}^2,
\end{aligned}
$$

and hence,

$$
\begin{aligned}
\|\widetilde{u}_{\varrho h} - u_{\varrho h}\|_{L^2(\Omega)}^2 \;\le\;& c \int_\Omega \frac{1}{\varrho} |\nabla(p_{u_{\varrho h}} - p_{u_{\varrho h} h})|^2 \, dx \\
=\;& c \, \langle A_{1/\varrho}(p_{u_{\varrho h}} - p_{u_{\varrho h} h}), p_{u_{\varrho h}} - p_{u_{\varrho h} h} \rangle_\Omega \;=\; c \, \|p_{u_{\varrho h}} - p_{u_{\varrho h} h}\|_{A_{1/\varrho}}^2,
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
\|\widetilde{u}_{\varrho h} - u_{\varrho h}\|_{L^2(\Omega)} \;\le\;& c \, \|p_{u_{\varrho h}} - p_{u_{\varrho h} h}\|_{A_{1/\varrho}} \\
\le\;& c \left[\|p_{u_{\varrho h}} - p_{\overline{u}}\|_{A_{1/\varrho}} + \|p_{\overline{u} h} - p_{u_{\varrho h} h}\|_{A_{1/\varrho}} + \|p_{\overline{u}} - p_{\overline{u} h}\|_{A_{1/\varrho}} \right].
\end{aligned}
$$

Note that we have

$$\int_\Omega \frac{1}{\varrho} \nabla(p_{u_{\varrho h}} - p_{\overline{u}}) \cdot \nabla v \, dx = \int_\Omega \nabla(u_{\varrho h} - \overline{u}) \cdot \nabla v \, dx \quad \text{for all } v \in H_0^1(\Omega)$$

and

$$\int_\Omega \frac{1}{\varrho} \nabla(p_{u_{\varrho h} h} - p_{\overline{u} h}) \cdot \nabla v_h \, dx = \int_\Omega \nabla(u_{\varrho h} - \overline{u}) \cdot \nabla v_h \, dx \quad \text{for all } v_h \in V_h.$$

10

Hence we conclude

$$\|p_{u_{\varrho h}} - p_{\overline{u}}\|_{A_{1/\varrho}}^2 \le \int_\Omega \varrho \, |\nabla(u_{\varrho h} - \overline{u})|^2 \, dx,$$

as well as

$$\|p_{u_{\varrho h}h} - p_{\overline{u}h}\|_{A_{1/\varrho}}^2 \le \int_\Omega \varrho \, |\nabla(u_{\varrho h} - \overline{u})|^2 \, dx.$$

We can further obtain, inserting the Scott-Zhang interpolation $\Pi_h \overline{u}$,

$$
\begin{aligned}
\int_\Omega \varrho \, |\nabla(u_{\varrho h} - \overline{u})|^2 \, dx \;\; &\le \;\; 2 \left[\int_\Omega \varrho \, |\nabla(u_{\varrho h} - \Pi_h \overline{u})|^2 \, dx + \int_\Omega \varrho \, |\nabla(\overline{u} - \Pi_h \overline{u})|^2 \, dx \right] \\
&\le \;\; 2 \left[\int_\Omega \varrho \, |\nabla(u_{\varrho h} - \Pi_h \overline{u})|^2 \, dx + c \sum_{\ell=1}^N h_\ell^2 \, \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2 \right].
\end{aligned}
$$

Using an inverse inequality locally, we further estimate the first term by

$$
\begin{aligned}
\int_\Omega \varrho \, |\nabla(u_{\varrho h} - \Pi_h \overline{u})|^2 \, dx \;\; &= \;\; \sum_{\ell=1}^N h_\ell^2 \, \|\nabla(u_{\varrho h} - \Pi_h \overline{u})\|_{L^2(\tau_\ell)}^2 \\
&\le \; c \sum_{\ell=1}^N \|u_{\varrho h} - \Pi_h \overline{u}\|_{L^2(\tau_\ell)}^2 \;\; = \;\; \|u_{\varrho h} - \Pi_h \overline{u}\|_{L^2(\Omega)}^2 \\
&\le \; 2 \left[\|u_{\varrho h} - \overline{u}\|_{L^2(\Omega)}^2 + \|\overline{u} - \Pi_h \overline{u}\|_{L^2(\Omega)}^2 \right] \le c \sum_{\ell=1}^N h_\ell^2 \, \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2.
\end{aligned}
$$

Recall that $p_{\overline{u}} \in H_0^1(\Omega)$ solves

$$\int_\Omega \frac{1}{\varrho(x)} \, \nabla p_{\overline{u}} \cdot \nabla v(x) \, dx = \int_\Omega \nabla \overline{u}(x) \cdot \nabla v(x) \, dx \quad \text{for all } v \in H_0^1(\Omega),$$

while $p_{\overline{u}h} \in V_h \subset H_0^1(\Omega)$ solves

$$\int_\Omega \frac{1}{\varrho(x)} \, \nabla p_{\overline{u}h} \cdot \nabla v_h(x) \, dx = \int_\Omega \nabla \overline{u}(x) \cdot \nabla v_h(x) \, dx \quad \text{for all } v_h \in V_h.$$

Hence we conclude the Galerkin orthogonality

$$\int_\Omega \frac{1}{\varrho(x)} \, \nabla(p_{\overline{u}}(x) - p_{\overline{u}h}(x)) \cdot \nabla v_h(x) \, dx = 0 \quad \text{for all } v_h \in V_h,$$

and

$$\|p_{\overline{u}} - p_{\overline{u}h}\|_{A_{1/\varrho}} \le \|p_{\overline{u}}\|_{A_{1/\varrho}}.$$

Now the assertion follows from

$$\|p_{\overline{u}}\|_{A_{1/\varrho}}^2 = \int_\Omega \frac{1}{\varrho(x)} \, |\nabla p_{\overline{u}}(x)|^2 \, dx \le \int_\Omega \varrho(x) \, |\nabla \overline{u}(x)|^2 \, dx \le \sum_{\ell=1}^N h_\ell^2 \, \|\nabla \overline{u}\|_{L^2(\omega_\ell)}^2.$$

$\square$

Now we are in the position to state the main result of this paper.

Theorem 3. *Let $\widetilde{u}_{\varrho h} \in V_h \subset H_0^1(\Omega)$ be the unique solution of the perturbed variational formulation (4.11) where the regularization function $\varrho(x)$ is given as in (4.1), and where the underlying finite element mesh is assumed to be locally quasi-uniform. Then there holds the error estimate*

$$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}^2 \le c \sum_{\ell=1}^N h_\ell^2 \, \|\nabla \overline{u}\|_{L^2(\tau_\ell)}^2 = c \int_\Omega \varrho(x) \, |\nabla \overline{u}(x)|^2 \, dx. \qquad (4.13)$$

Proof. The estimate (4.13) follows from the finite element error estimates (4.8) and (4.12), since the finite element mesh is assumed to be locally quasi-uniform. $\qquad\square$

Remark 2. *Similar as before we also have the error estimate*

$$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)} \leq c \, \|\overline{u}\|_{H^s_\varrho(\Omega)},$$

when assuming $\overline{u} \in H^s_\varrho(\Omega)$ *for some* $s \in [0,1]$.

The perturbed Galerkin finite element formulation (4.11) can be written as coupled system to find $(\widetilde{u}_{\varrho h}, \widetilde{p}_{\varrho h}) \in V_h \times V_h$ such that

$$\int_\Omega \frac{1}{\varrho(x)} \nabla \widetilde{p}_{\varrho h}(x) \cdot \nabla v_h(x)\, dx + \int_\Omega \nabla \widetilde{u}_{\varrho h}(x) \cdot \nabla v_h(x)\, dx = 0 \qquad (4.14)$$

for all $v_h \in V_h$, and

$$\int_\Omega \widetilde{u}_{\varrho h}(x)\, q_h(x)\, dx - \int_\Omega \nabla \widetilde{p}_{\varrho h}(x) \cdot \nabla q_h(x)\, dx = \int_\Omega \overline{u}(x)\, q_h(x)\, dx \qquad (4.15)$$

for all $q_h \in V_h$. Note that this system corresponds to the finite element discretization of the coupled variational formulation (2.6) and (2.7).

The finite element variational formulation (4.14) and (4.15) is equivalent to a coupled linear system of algebraic equations,

$$K_{\varrho h} \underline{p} + K_h \underline{u} = \underline{0}, \qquad M_h \underline{u} - K_h \underline{p} = \underline{f}, \qquad (4.16)$$

where we use the standard finite element stiffness and mass matrices defined as

$$
\begin{aligned}
K_h[j,k] &= \int_\Omega \nabla \varphi_k(x) \cdot \nabla \varphi_j(x)\, dx, \\
K_{\varrho h}[j,k] &= \int_\Omega \frac{1}{\varrho(x)} \nabla \varphi_k(x) \cdot \nabla \varphi_j(x)\, dx, \\
M_h[j,k] &= \int_\Omega \varphi_k(x)\, \varphi_j(x)\, dx
\end{aligned}
$$

for $j,k = 1, \ldots, M$, and the entries of the load vector

$$f_j = \int_\Omega \overline{u}(x)\, \varphi_j(x)\, dx \quad \text{for } j = 1, \ldots, M.$$

Since the finite element stiffness matrix $K_{\varrho h}$ is invertible, we can eliminate the adjoint $\underline{p}$ to end up with the Schur complement system

$$\left[M_h + K_h K_{\varrho h}^{-1} K_h \right] \underline{u} = \underline{f}. \qquad (4.17)$$

Since all involved stiffness and mass matrices are symmetric and positive definite, unique solvability of the Schur complement system and therefore of the Galerkin variational formulation (4.14) and (4.15) follows.

5 Numerical results

5.1 Convergence studies

As a first numerical example we consider the two-dimensional domain $\Omega = (0,1)^2$, and the discontinuous target function

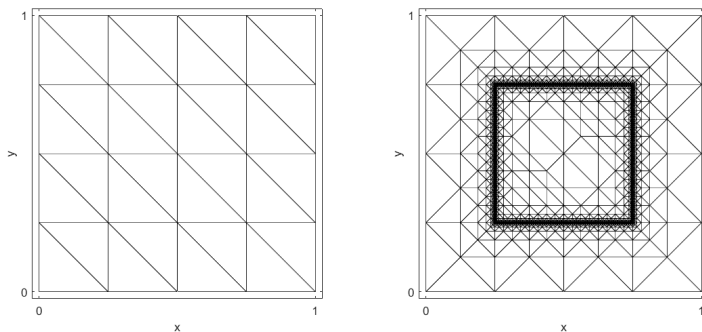$$\overline{u}_{2D}(x) = \begin{cases} 1 & \text{for } x \in (0.25, 0.75)^2, \\ 0 & \text{else.} \end{cases}$$

12

Figure 1: Initial mesh and adaptively refined mesh at level 14 for $\Omega = (0,1)^2$ and $\overline{u}_{2D}$.

The initial mesh consists of 32 triangular finite elements and 9 degrees of freedom, see Figure 1.

For a given mesh we compute the approximate solution $\widetilde{u}_{\varrho h}$, the global error

$$\eta := \|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)},$$

and the local error indicators

$$\eta_\ell := \|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\tau_\ell)}, \quad \eta^2 = \sum_{\ell=1}^{N} \eta_\ell^2.$$

We mark all elements τ_ℓ when

$$\eta_\ell > \theta \max_{\ell=1,\ldots,N} \eta_\ell$$

is satisfied, with $\theta = 0.5$. After 14 refinement steps we obtain the mesh as shown in Fig. 1 with 1.310.444 finite elements and 655.215 degrees of freedom. According to the final error estimate (4.13) we expect a linear order of convergence. The numerical results are shown in Fig. 2 where in addition to the present approach we also present the convergence results when considering energy regularization [16] with the optimal choice $\varrho = h^2$, and the regularization in $L^2(\Omega)$ with $\varrho = h^4$, see [14]. As expected, we observe a linear order of convergence when using the variable energy regularization in the adaptive version described above, while both the energy regularization and the regularization in $L^2(\Omega)$ almost coincide with half the order of convergence. For a comparison of the different approaches, see also the computed states as shown in Fig. 3.

Next we consider the three-dimensional domain $\Omega = (0,1)^3$, and the target function

$$\overline{u}_{3D}(x) = \begin{cases} 1 & \text{for } x \in (0.25, 0.75)^3, \\ 0 & \text{else.} \end{cases}$$

As shown in Fig. 4 we still observe a $h^{1/2}$ convergence for a uniform refinement in the case of both the $L^2(\Omega)$ and energy regularizations, but a $h^{3/4}$ convergence for the adaptive diffusion approach, where $h = N^{-1/3}$.

To explain the different convergence behaviour for the adaptive refinement in two and three space dimensions, we first consider the 2D case for the example $\overline{u}_{2D} \in H^{1/2-\varepsilon}(\Omega)$, $\varepsilon > 0$ for a uniform refinement of the triangulation with N triangles (see Fig. 1 (left)). Let further m denote approximately the number of elements in each row/column of the mesh grid, i.e., $N \sim m \cdot m = m^2$. Due to the
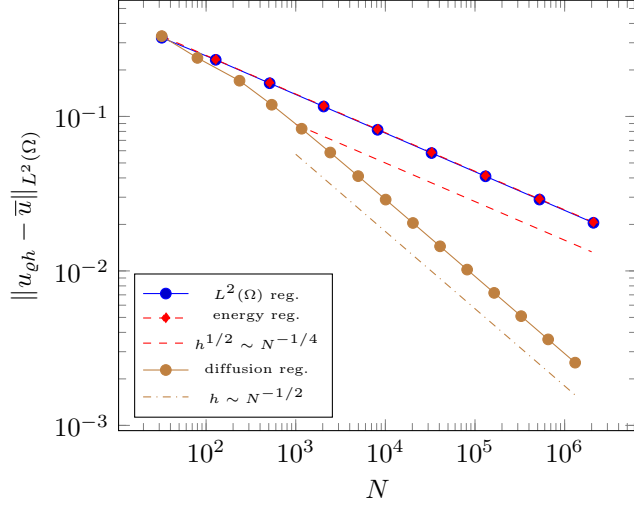
13

Figure 2: Convergence plots for $\overline{u}_{2D}$ choosing $\varrho = h^4$ for the $L^2(\Omega)$ regularization, $\varrho = h^2$ for the energy regularization, and $\varrho(x) = h_\ell^2$ for $x \in \tau_\ell$ for the diffusion regularization.

regularity of $\overline{u}_{2D}$, the optimal order of convergence is de facto $1/2$. Thus, refining all of the N elements uniformly, will lead to a an error reduction of order $h^{1/2}$, as observed. Now we aim for an adaptive refinement. Since for the particular test example the singularity of $\overline{u}_{2D}$ is only along the boundary of the square $(0.25, 0.75)^2$, it is sufficient (after an initializing phase), to refine only $\mathcal{O}(m) = \mathcal{O}(\sqrt{N})$ elements in the neighborhood of this boundary in order to have the optimal rate of $1/2$. Note, that for a uniform refinement, the number of elements would grow by a factor of 4 in each step, for the adaptive scheme, we only refine $\mathcal{O}(m) = \mathcal{O}(\sqrt{N})$ elements in the neighborhood of the discontinuity. Therefore, the number of elements only grows by a factor of 2. Hence, if we adaptively refine $\mathcal{O}(N)$ elements, we can expect an error reduction of order h, which is exactly what we see in the numerical example as given in Fig. 2.

Now let us look at the 3D case. Here again counting the elements along each edge, denoted by m, we get the relation $N \sim m^3$. For a uniform refinement we get a convergence rate $h^{1/2}$. In order to get the same rate with an adaptive scheme, we need to refine at least the elements in the neighborhood of the boundary of the cube $[0.25, 0.75]^3$, where $\overline{u}_{3D}$ jumps. Each side of the cube consists of approximately $\mathcal{O}(m^2)$ elements. So the whole boundary of the interior cube has approximately $\mathcal{O}(m^2) = \mathcal{O}(N^{2/3})$ neighboring elements. So, refining $O(N^{2/3})$ elements adaptively gives a rate of $1/2$. Hence, if we refine $\mathcal{O}(N)$ elements adaptively, we might expect an error reduction of order $h^{3/2 \cdot 1/2} = h^{3/4}$, which is exactly what we observe in Fig. 4.

Finally we consider the one-dimensional domain $\Omega = (0, 1)$ and the target

$$\overline{u}_{1D}(x) = \begin{cases} 1 & \text{for } x \in (0.25, 0.75), \\ 0 & \text{else.} \end{cases}$$

For $\overline{u}_{1D} \in H^{1/2-\varepsilon}(\Omega)$, $\varepsilon > 0$, and for uniformly refining N elements, we will get an error reduction of order $h^{1/2}$. Using an adaptive refinement though, it is enough to refine exactly $4 \sim \mathcal{O}(\log(N))$ elements in each step to get the optimal order of $1/2$. Thus we can expect exponential convergence, which is also what we observe in the numerical example in Fig. 5.
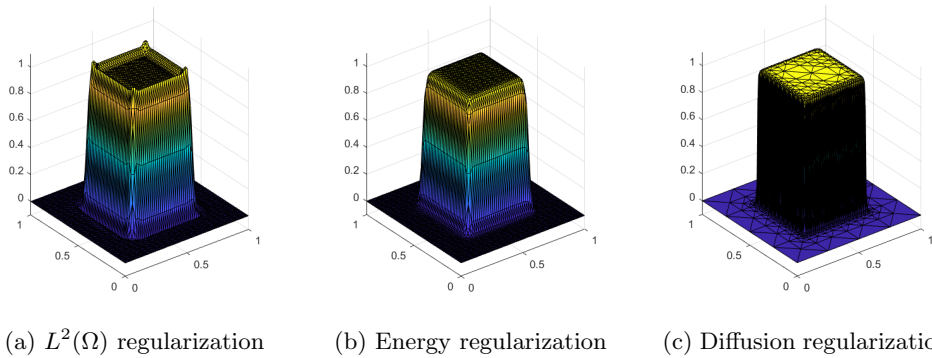
14

(a) $L^2(\Omega)$ regularization   (b) Energy regularization   (c) Diffusion regularization

Figure 3: Solution $\widetilde{u}_{\varrho h}$ for the three different regularization approaches, with 8192 finite elements (level 4) using a uniform refinement strategy for $L^2(\Omega)$ and energy regularization, and with 4972 finite elements (level 6) for an adaptive refinement for the diffusion regularization.

5.2 Solver studies

While all the numerical results as presented in the previous subsection were computed using Matlab with a sparse direct solver, we finally discuss the use of preconditioned iterative solution strategies which are robust with respect to the regularization parameter function $\varrho(x)$. Here we will restrict our considerations to the three-dimensional case with the target $\overline{u}_{3D}$.

We first consider the preconditioned conjugate gradient (PCG) solver applied to the Schur complement equation (4.17) with a proper preconditioner. When performing a uniform refinement, the diffusion coefficients (the inverse of the regularization parameters) are constant on all elements, and we may replace $\varrho(x)$ by h^2 with h being the global mesh size. Therefore, the Schur complement is simplified as $M_h + K_h K_{\varrho h}^{-1} K_h = M_h + \varrho K_h K_h^{-1} K_h = M_h + h^2 K_h$. Robust preconditioners for such a Schur complement have been studied in our previous work [15], and the spectral equivalence of this Schur complement to the mass matrix was analyzed in our recent work [14]. Since, in this case, the Schur complement $M_h + \varrho K_h K_h^{-1} K_h = M_h + h^2 K_h$ is spectrally equivalent to the mass matrix M_h, we can use a simple diagonal approximation to the mass matrix such as $\mathrm{diag}[M_h]$ or $\mathrm{lump}[M_h]$ as cheap preconditioners for the Schur complement. In the case of an adaptive refinement, both the local mesh refinement and the use of varying diffusion coefficients $\varrho(x)$, which are piecewise constant, play a role in developing robust Schur complement preconditioners with respect to the mesh size and diffusion coefficient jumps. At least, it is not straightforward to show the robustness of the preconditioner for the Schur complement $M_h + K_h K_{\varrho h}^{-1} K_h$ in this situation. However, thanks to the particular choice of the diffusion coefficient (4.1), the negative effect resulting from the varying mesh size seems to be compensated by the local diffusion coefficient (the inverse of the local regularization parameter) during the adaptive procedure. The spectral equivalence of the Schur complement to the mass matrix still seems to hold on the adaptive refinements. A rigorous analysis of the robust preconditioners needs to be further studied. The robustness with respect to the minimal and maximal local mesh size ($h_{\min}$ and $h_{\max}$) is numerically confirmed by the constant iteration numbers of the PCG method preconditioned by the lumped mass matrix (Its (PCG)) on the adaptive refinements in Table 1, in comparison with the increasing conjugate gradient (CG) iteration numbers without using the preconditioner (Its (CG)). Note that we solve the Schur complement equation (4.17) until the relative preconditioned residual error is reduced by a factor 10^6. For the inverse operation of $K_{\varrho h}^{-1}$ applied to
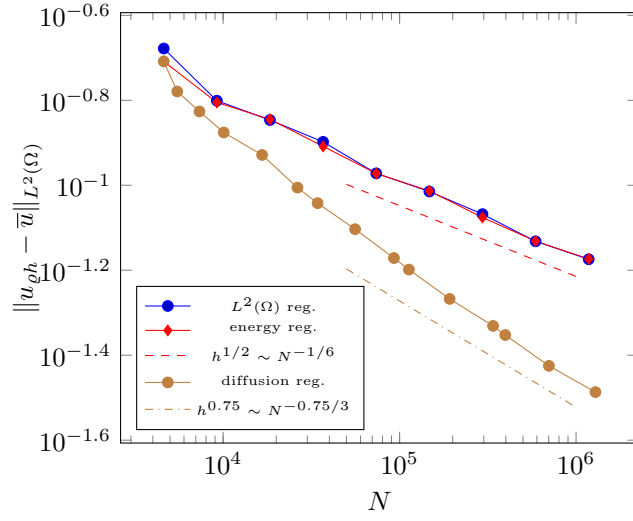
Figure 4: Convergence plots for $\overline{u}_{3D}$ choosing $\varrho = h^4$ for the $L^2(\Omega)$ regularization, $\varrho = h^2$ for the energy regularization, and $\varrho(x) = h_\ell^2$ for $x \in \tau_\ell$ for the diffusion regularization.
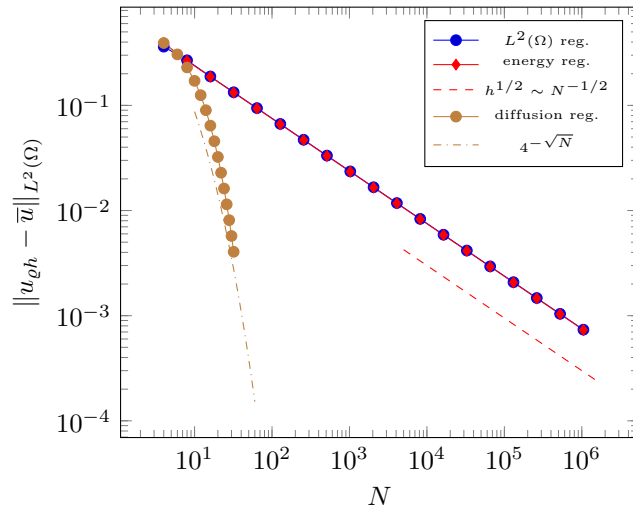

Figure 5: Convergence plots for $\overline{u}_{1D}$ choosing $\varrho = h^4$ for the $L^2(\Omega)$ regularization, $\varrho = h^2$ for the energy regularization, and $\varrho(x) = h_\ell^2$ for $x \in \tau_\ell$ for the diffusion regularization.

a vector $\underline{v}$ within the PCG iteration, we have used the classical Ruge–Stüben algebraic multigrid (AMG) [19] preconditioned CG method to solve $K_{\varrho h}\underline{w} = \underline{v}$ until the relative preconditioned residual error reaches 10^{-12} in order to perform a sufficiently accurate multiplication with the Schur complement. Alternatively, one can here use a sparse factorization in a preprocessing step.

Level	#Dofs	$h_{\min}$	$h_{\max}$	$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$	Its (PCG)	Its (CG)
L_1	125	2^{-2}	2^{-2}	3.01923e−1	7	8
L_2	223	2^{-3}	2^{-2}	2.55302e−1	18	22
L_3	1,059	2^{-4}	2^{-2}	1.79986e−1	23	52
L_4	4,728	2^{-5}	2^{-2}	1.26353e−1	26	127
L_5	18,827	2^{-6}	2^{-2}	8.84306e−2	25	317
L_6	75,603	2^{-7}	2^{-2}	6.21010e−2	24	829
L_7	303,782	2^{-8}	2^{-2}	4.37439e−2	23	2,103
L_8	1,218,846	2^{-9}	2^{-2}	3.08691e−2	23	5,219
L_9	4,884,317	2^{-10}	2^{-2}	2.18069e−2	19	> 10,000
L_{10}	19,553,202	2^{-11}	2^{-2}	1.54097e−2	17	> 20,000

Table 1: Comparison of the PCG (Its (PCG)) and CG iterations (Its (CG)) for the Schur complement equation (4.17) on the adaptive refinements.

Furthermore, we provide some numerical results concerning robust solvers for the coupled optimality system (4.16):

$$\begin{bmatrix} K_{\varrho h} & K_h \\ -K_h & M_h \end{bmatrix} \begin{bmatrix} \underline{p} \\ \underline{u} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \underline{f} \end{bmatrix}. \tag{5.1}$$

Since the system matrix is non-symmetric and positive definite, we apply the GMRES method with the following proposed block diagonal preconditioner:

$$\mathcal{P}_h = \begin{bmatrix} \widehat{K}_{\varrho h} & 0 \\ 0 & \mathrm{lump}[M_h] \end{bmatrix}.$$

The number of GMRES iterations (Its) using such a preconditioner are given in Table 2. We solve the system until the relative preconditioned residual error is reduced by a factor 10^6. In the preconditioner $\mathcal{P}_h$, we have utilized the Ruge–Stüben AMG preconditioner $\widehat{K}_{\varrho h}$ for $K_{\varrho h}$, whereas the lumped mass matrix $\mathrm{lump}[M_h]$ has been used as preconditioner for the Schur complement. We observe almost constant iteration numbers of the GMRES method preconditioned by $\mathcal{P}_h$ on the uniform refinement as well as on the adaptive refinements as given in Table 2. We only see slightly higher iteration numbers on the adaptive meshes than on the uniform ones. On the other hand, we may reformulate the coupled system (5.1) in the following equivalent form

$$\begin{bmatrix} K_{\varrho h} & K_h \\ K_h & -M_h \end{bmatrix} \begin{bmatrix} \underline{p} \\ \underline{u} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ -\underline{f} \end{bmatrix}. \tag{5.2}$$

When applying the Bramble–Pasciak transformation

$$\mathcal{T}_h = \begin{bmatrix} K_{\varrho h} C_h^{-1} - I_h & 0 \\ K_h C_h^{-1} & -I_h \end{bmatrix}$$

to the symmetric but indefinite system (5.2), this leads to the equivalent system

$$\mathcal{K}_h \begin{bmatrix} \underline{p} \\ \underline{u} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \underline{f} \end{bmatrix} \equiv \mathcal{T}_h \begin{bmatrix} \underline{0} \\ -\underline{f} \end{bmatrix},$$

Level	Adaptive			Uniform		
	#Dofs	$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$	Its	#Dofs	$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$	Its
L_1	250	3.01923e−1	14	250	3.01923e−1	14
L_2	446	2.55302e−1	33	1,458	2.26341e−1	28
L_3	2,118	1.79985e−1	40	9,826	1.61850e−1	36
L_4	9,456	1.26356e−1	46	71,874	1.14659e−1	36
L_5	37,656	8.84293e−2	48	549,250	8.10582e−2	34
L_6	151,212	6.20997e−2	47	4,293,378	5.72923e−2	32
L_7	607,586	4.37450e−2	46	33,949,186	4.04944e−2	30
L_8	2,437,880	3.08701e−2	44	270,011,394	2.86176e−2	28
L_9	9,769,496	2.18075e−2	44			
L_{10}	39,108,934	1.54100e−2	42			
L_{11}	156,568,020	1.08922e−2	42			

Table 2: Comparison of the preconditioned GMRES solver on both the adaptive and uniform refinements.

with the symmetric and positive definite system matrix

$$
\mathcal{K}_h = \begin{bmatrix} K_{\varrho h} C_h^{-1} - I_h & 0 \\ K_h C_h^{-1} & -I_h \end{bmatrix} \begin{bmatrix} K_{\varrho h} & K_h \\ K_h & -M_h \end{bmatrix}
$$
$$
= \begin{bmatrix} (K_{\varrho h} - C_h) C_h^{-1} K_{\varrho h} & (K_{\varrho h} - C_h) C_h^{-1} K_h \\ K_h C_h^{-1} (K_{\varrho h} - C_h) & K_h C_h^{-1} K_h + M_h \end{bmatrix}.
$$

Here, I_h denotes the identity, and C_h is some symmetric and positive definite (spd) matrix that is assumed to be spectrally equivalent to the matrix $K_{\varrho h}$, and less than $K_{\varrho h}$, i.e., $C_h < K_{\varrho h}$. In particular, we can again take the classical spd Ruge–Stüben AMG preconditioner $\widehat{K}_{\varrho h} = \delta K_{\varrho h}(I_h - AMG_{\varrho h})^{-1}$ with a proper scaling $\delta > 0$ as C_h, where $AMG_{\varrho h}$ denotes the AMG iteration matrix. Now using $C_h = \widehat{K}_{\varrho h}$ and the lumped mass matrix lump$[M_h]$ to the Schur complement $M_h + K_h K_{\varrho h}^{-1} K_h$, we arrive at the following (inexact) BP preconditioner:

$$
\mathcal{P}_h = \begin{bmatrix} K_{\varrho h} - C_h & 0 \\ 0 & \text{lump}[M_h] \end{bmatrix}.
$$

Details on the BP-CG can be found in the original paper [3]; see also [23] for improved convergence rate estimates. The number of BP-CG iterations (Its) using the preconditioner $\mathcal{P}_h$ are provided in Table 3. The solver stops the iteration when the relative preconditioned residual error is reduced by a factor 10^6. From the constant PB-CG iteration numbers on both the uniform and adaptive mesh refinements, we observe the robustness of the proposed preconditioner for the coupled optimality system.

6 Conclusion and outlook

We have studied finite element discretizations of the reduced optimality system for the standard distributed space-tracking elliptic optimal control problem, but using a new variable energy regularization technique. It has been shown that the choice of the local mesh-size squared as local regularization parameter $\varrho(x)$ leads to optimal rates of convergence of the computed finite element state $\widetilde{u}_{\varrho h}$ to the prescribed target $\overline{u}$ in the L^2 norm. In particular, this approach allows us to adapt the local regularization parameter to the local mesh-size when using an adaptive mesh refinement, where the adaptivity is driven by the localization of the L^2 norm of

Level	Adaptive			Uniform		
	#Dofs	$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$	Its	#Dofs	$\|\widetilde{u}_{\varrho h} - \overline{u}\|_{L^2(\Omega)}$	Its
L_1	250	3.01923e−1	13	250	3.01923e−1	13
L_2	446	2.55302e−1	32	1,458	2.26348e−1	25
L_3	2,118	1.79986e−1	34	9,826	1.61863e−1	30
L_4	9,456	1.26353e−1	38	71,874	1.14657e−1	31
L_5	37,654	8.84307e−2	39	549,250	8.10658e−2	25
L_6	151,206	6.21010e−2	38	4,293,378	5.73048e−2	24
L_7	607,558	4.37438e−2	36	33,949,186	4.05125e−2	23
L_8	2,437,674	3.08690e−2	35	270,011,394	2.86430e−2	23
L_9	9,768,526	2.18067e−2	33			
L_{10}	39,105,552	1.54096e−2	34			
L_{11}	156,550,890	1.089170−2	30			

Table 3: Comparison of the preconditioned PB-CG solver on both the adaptive and uniform refinements.

the error between $\overline{u}$ and $\widetilde{u}_{\varrho h}$ as computable local error indicator. Numerical studies made for discontinuous targets in one, two and three space dimensions illustrate that these simple adaptive schemes show a significantly better performance than the uniform refinement. We have also proposed iterative solvers for the finite element equations corresponding to the reduced optimality system. The numerical studies have shown that these solvers are robust and efficient at the same time.

Contrary to quasi-uniform meshes where we can choose $\varrho = h^2$, the numerical analysis of these solvers is still not complete in the case of variable regularization parameters. However, all numerical experiments have shown that the mass matrix M_h, and, therefore, also the lumped mass matrix lump$[M_h]$ are spectrally equivalent to the Schur complement $K_h K_{\varrho h}^{-1} K_h + M_h$, and can be used as robust preconditioners in the preconditioned BP-CG, MINRES, or GMRES solvers. Obviously, the classical plane Ruge-Stüben AMG preconditioner is doing this job for $K_{\varrho h}$. Here we may develop more efficient and robust preconditioners that are especially adapted to the diffusion coefficient $\varrho(x)$ that changes from element to element according to the mesh-sizes. Moreover, the parallelization of such iterative solution strategies should be implemented in order to solve really large-scale systems in three space dimensions. Finally, the variable energy regularization and the robust solvers for the corresponding linear algebraic equations can be extended to optimal control problems with state or control constraints, and subject to other state equations like elasticity, Maxwell, and Stokes equations.

Acknowledgments

---

[1] https://www3.risc.jku.at/projects/mach2/
[2] https://www.oeaw.ac.at/ricam/hpc

References

[1] V. Albani, A. De Cezaro, and J. P. Zubelli. On the choice of the Tikhonov regularization parameter and the discretization level: a discrepancy-based strategy. *Inverse Probl. Imaging*, 10(1):1–25, 2016.

[2] A. Borzì and V. Schulz. *Computational optimization of systems governed by partial differential equations*, volume 8 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012.

[3] J. H. Bramble and J. E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comp.*, 50(181):1–17, 1988.

[4] K. Bredies, Y. Dong, and M. Hintermüller. Spatially dependent regularization parameter selection in total generalized variation models for image restoration. *Int. J. Comput. Math.*, 90:109–123, 2013.

[5] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[6] C. Chung, J. De los Reyes, and C. Schönlieb. Learning optimal spatially-dependent regularization parameters in total variation image denoising. *Inverse Problems*, 33:074005, 2017.

[7] C. Clason and B. Kaltenbacher. Optimal control and inverse problems. *Inverse Problems*, 36:060301, 2020.

[8] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications*. Kluwer, Dordrecht, 1996.

[9] M. Hintermüller, K. Papafitsoros, C. N. Rautenberg, and H. Sun. Dualization and automatic distributed parameter selection of total generalized variation via bilevel optimization. *Numer. Funct. Anal. Optim.*, 43:887–932, 2022.

[10] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*, volume 23. Springer-Verlag, Berlin, 2009.

[11] B. Hong, J. Koo, and M. Burger. Adaptive regularization in convex composite optimization for variational imaging problems. In V. Roth and T. Vetter, editors, *Pattern Recognition. GCPR 2017*, volume 10496 of *Lecture Notes in Computer Science*, pages 268–280. Springer, Cham, 2017.

[12] V. Isakov. *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, Cham, third edition, 2017.

[13] V. John. *Finite Element Methods for Incompressible Flow Problems*, volume 51 of *Springer Series in Computational Mathematics*. Springer, 2016.

[14] U. Langer, R. Löscher, O. Steinbach, and H. Yang. Robust finite element discretization and solvers for distributed elliptic optimal control problems. Technical Report arXiv:2207.04664, arXiv.org, 2022.

[15] U. Langer, O. Steinbach, and H. Yang. Robust discretization and solvers for elliptic optimal control problems with energy regularization. *Comput. Meth. Appl. Math.*, 22:97–111, 2022.

[16] U. Langer, O. Steinbach, and H. Yang. Robust space-time finite element error estimates for parabolic distributed optimal control problems with energy regularization. Technical Report arXiv:2206.06455, arXiv.org, 2022.

[17] J.-L. Lions. *Optimal control of systems governed by partial differential equations.* Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin, 1971.

[18] M. Neumüller and O. Steinbach. Regularization error estimates for distributed control problems in energy spaces. *Math. Methods Appl. Sci.*, 44:4176–4191, 2021.

[19] J. W. Ruge and K. Stüben. Algebraic multigrid (AMG). In S. F. McCormick, editor, *Multigrid Methods*, pages 73–130. Society for Industrial and Applied Mathematics, Philadelphia, 1987.

[20] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. Kazimierski. *Regularization Methods in Banach Spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. de Gruyter, Berlin, 2012.

[21] H. Triebel. *Interpolation theory, function spaces, differential operators*, volume 18 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam-New York, 1978.

[22] F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[23] W. Zulehner. Analysis of iterative methods for saddle point problems: a unified approach. *Math. Comp.*, 71(238):479–505, 2002.