Technische Universität Graz

Mass-lumping discretization and solvers for distributed elliptic optimal control problems

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

Berichte aus dem Institut für Angewandte Mathematik

Bericht 2023/3

Technische Universität Graz

Mass-lumping discretization and solvers for distributed elliptic optimal control problems

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

Berichte aus dem Institut für Angewandte Mathematik

Bericht 2023/3

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

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

Mass-lumping discretization and solvers for distributed elliptic optimal control problems

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

Abstract

The purpose of this paper is to investigate the effects of the use of mass-lumping in the finite element discretization of the reduced first-order optimality system arising from a standard tracking-type, distributed elliptic optimal control problem with L_2 regularization. We show that mass-lumping will not affect the L_2 error between the desired state and the computed state, but will lead to a Schur-complement system that allows for a fast matrix-by-vector multiplication. We show that the use of the Schur-Complement Preconditioned Conjugate Gradient method in a nested iteration setting leads to an asymptotically optimal solver with respect to the complexity.

Keywords: Elliptic optimal control problems, L_2 regularization, finite element discretization, mass lumping, preconditioned conjugate gradient method, nested iteration.

1 Introduction

We consider the following tracking-type, distributed elliptic optimal control problem with standard L_2 regularization: find the state $y_\varrho \in Y = H^1_0(\Omega)$ and the control $u_\varrho \in U = L_2(\Omega)$ minimizing the cost functional

$$J(y_\varrho, u_\varrho) := \frac{1}{2} \|y_\varrho - y_d\|^2_{L_2(\Omega)} + \frac{\varrho}{2} \|u_\varrho\|^2_{L_2(\Omega)}, \tag{1}$$

subject to (s.t.) the elliptic boundary value model problem

$$-\Delta y_\varrho = u_\varrho \text{ in } \Omega, \quad y_\varrho = 0 \text{ on } \partial\Omega, \tag{2}$$

for some given desired state (target) $y_d \in L_2(\Omega)$, and some regularization parameter $\varrho > 0$, where $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, is a bounded Lipschitz domain with the boundary $\partial\Omega$. We here use the standard notations for Lebesgue and Sobolev spaces. Since the state equation (2) has a unique solution $y_\varrho \in Y$ for every given control $u_\varrho \in U$, the optimal control problem (1)-(2) has a unique solution $(y_\varrho, u_\varrho) \in Y \times U$ too; see, e.g., [29], [19], or [41]. Moreover, the state y_ϱ obviously belongs to $H^\Delta(\Omega) = \{y \in$

[*]Institute of Numerical Mathematics, Johannes Kepler University Linz, and Johann Radon Institute for Computational and Applied Mathematics of the Austrian Academy of Sciences, 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

$H_0^1(\Omega) : \Delta y \in L_2(\Omega)\}$, and the solution operator S mapping u_ϱ to y_ϱ (control-to-state map) is an isomorphism between $L_2(\Omega)$ and $H^\Delta(\Omega)$.

The finite element (fe) discretization of the reduced (after elimination of the control u_ϱ) optimality system, which defines the solution to the optimal control problem (1)-(2), leads to the solution of a large-scale symmetric, but indefinite linear system of algebraic equations for defining the fe nodal adjoint state vector $\mathbf{p}_h \in \mathbb{R}^{n_h}$ and the fe nodal state vector $\mathbf{y}_h \in \mathbb{R}^{n_h}$ such that

$$\begin{pmatrix} \varrho^{-1} M_h & K_h \\ K_h & -M_h \end{pmatrix} \begin{pmatrix} \mathbf{p}_h \\ \mathbf{y}_h \end{pmatrix} = \begin{pmatrix} \mathbf{0}_h \\ -\mathbf{y}_{dh} \end{pmatrix}, \tag{3}$$

where the stiffness matrix K_h and the mass matrix M_h are symmetric and positive definite (spd), $\mathbf{y}_{dh} \in \mathbb{R}^{n_h}$ is nothing but the fe load vector representing the desired state y_d, and h denotes a suitable discretization parameter. For fixed ϱ, discretization error estimates can be found, e.g., in [19]. There is a huge number of publications on efficient preconditioned iterative solvers for symmetric, but indefinite systems in general; see, e.g., the unified approach proposed in [42], the survey paper [8], the review article [31], the books [15] and [6], the more recent papers [1, 3, 4, 33], and the literature cited therein. Special iterative solvers for discrete optimality systems such as (3) should be not only robust with respect to (wrt) the mesh refinement quantified by the discretization parameter h but also wrt the regularization parameter ϱ that can be quite small depending on the cost that we are willing to pay. Such kind of h and ϱ robust preconditioned iterative methods have been proposed and investigated in [1, 5, 35, 37, 43]; see also [2, 14, 34, 36, 40], for handling control and state constraints, and the references therein. Alternatively, we can use all-at-once multigrid methods to solve saddle-point problems such as (3) efficiently; see, e.g., [38] and the review paper [10].

In this paper, we are interested in the case $\varrho = h^4$ leading to asymptotically optimal balanced estimates of the L_2-error between the desired state y_d and the computed finite element state $y_{\varrho h}$ that is related to $\mathbf{y}_h$ by the fe isomorphism; see [23]. Asymptotically optimal preconditioned iterative solvers for the saddle-point system (3) were proposed in [23] and [22] for constant and variable L_2 regularizations, respectively. More precisely, it turns out that very cheap preconditioners for the MINRES and BP-CG can be constructed on the basis of simple diagonal approximations of the mass matrix M_h. Of course, we can further reduce the saddle-point system (3) to the Schur-Complement (SC) system

$$(\varrho K_h M_h^{-1} K_h + M_h)\mathbf{y}_h = \mathbf{y}_{dh} \tag{4}$$

by eliminating the adjoint state $\mathbf{p}_h$. The system matrix is spd, and we would like to solve this system by means of the Preconditioned Conjugate Gradient (PCG) method. Although we can use very cheap diagonal matrices D_h such as $\mathrm{diag}(M_h)$ or the lumped mass matrix $\mathrm{lump}(M_h)$ as preconditioners that are spectrally equivalent to the Schur complement $\varrho K_h M_h^{-1} K_h + M_h$ for $\varrho = h^4$ [22, 23], we cannot simply replace the mass matrix M_h^{-1} by the lumped mass matrix $(\mathrm{lump}(M_h))^{-1}$ in the Schur complement without a precise analysis of the impact of this replacement to the discretization error. In Section 2, we just provide this analysis, and show that, in the case of continuous, piecewise linear (Courant's) finite element spaces, the discretization error is not affected at all. This theoretical result is supported by our numerical results presented in Section 4. Now we have to solve the mass-lumped SC system

$$(\varrho K_h (\mathrm{lump}(M_h))^{-1} K_h + M_h)\hat{\mathbf{y}}_h = \mathbf{y}_{dh} \tag{5}$$

instead of the original SC system (4). Using the diagonal preconditioner $D_h = \mathrm{lump}(M_h)$, we can now solve (5) in asymptotically optimal complexity for some

fixed relative accuracy. In Section 3, we show how we can use this SC-PCG in a nested iteration setting in order to compute a fe approximation to the desired state y_d, which differs from y_d in the L_2-norm in the order of the discretization error, with asymptotically optimal complexity $O(n_h)$. These theoretical results are again quantitatively illustrated by numerical experiments in Section 4.

2 Mass-Lumping and Error Analysis

The first-order optimality system, derived from (1)–(2), is given by the equations

$$-\Delta y_\varrho = u_\varrho,\ -\Delta p_\varrho = y_\varrho - y_d, \text{ and } p_\varrho + \varrho u_\varrho = 0 \quad \text{in } \Omega, \tag{6}$$

with the boundary conditions

$$y_\varrho = 0 \text{ and } p_\varrho = 0 \quad \text{on } \partial\Omega. \tag{7}$$

Eliminating the control u_ϱ, we arrive at the reduced first-order optimality system, the variational form of which reads as follows: find $(y_\varrho, p_\varrho) \in H_0^1(\Omega) \times H_0^1(\Omega)$ such that

$$\begin{aligned} \frac{1}{\varrho}\left\langle p_\varrho, q\right\rangle_{L_2(\Omega)} + \left\langle \nabla y_\varrho, \nabla q\right\rangle_{L_2(\Omega)} &= 0,\ \ \forall q \in H_0^1(\Omega), & (8)\\ -\left\langle \nabla p_\varrho, \nabla v\right\rangle_{L_2(\Omega)} + \left\langle y_\varrho, v\right\rangle_{L_2(\Omega)} &= \left\langle y_d, v\right\rangle_{L_2(\Omega)},\ \ \forall v \in H_0^1(\Omega). & (9)\end{aligned}$$

Introducing the variable $\tilde{p}_\varrho = \frac{1}{\sqrt{\varrho}} p_\varrho$, we further derive the scaled system

$$\begin{aligned} \frac{1}{\sqrt{\varrho}}\left\langle \tilde{p}_\varrho, q\right\rangle_{L_2(\Omega)} + \left\langle \nabla y_\varrho, \nabla q\right\rangle_{L_2(\Omega)} &= 0,\ \ \forall q \in H_0^1(\Omega), & (10)\\ -\left\langle \nabla \tilde{p}_\varrho, \nabla v\right\rangle_{L_2(\Omega)} + \frac{1}{\sqrt{\varrho}}\left\langle y_\varrho, v\right\rangle_{L_2(\Omega)} &= \frac{1}{\sqrt{\varrho}}\left\langle y_d, v\right\rangle_{L_2(\Omega)},\ \ \forall v \in H_0^1(\Omega). & (11)\end{aligned}$$

For simplicity, we assume from now on that $\Omega \subset \mathbb{R}^d$ is polygonally ($d = 2$) or polyhedrally ($d = 3$) bounded. Let $\mathcal{T}_h = \{\tau_e\}_{e=1}^{N_h}$ be an admissible, globally quasi-uniform and shape-regular decomposition of Ω into simplicies τ_e, with the mesh-size $h_e = |\tau_e|^{1/d}$, such that $\overline{\Omega} = \bigcup_{e=1}^{N_h} \overline{\tau}_e$. Let $S_h^1(\mathcal{T}_h) = \text{span}\{\varphi_j^h\}_{j=1}^{\overline{n}_h}$ denote the space of piecewise linear, globally continuous functions spanned by the Lagrange basis functions φ_j^h (hat functions), which fulfil the equations

$$\sum_{j=1}^{\overline{n}_h} \varphi_j^h(x) = 1\ \ \forall x \in \Omega, \quad \text{and} \quad \varphi_j^h(x_i) = \delta_{i,j} \text{ for each node } x_i,\ i = 1, \ldots, \overline{n}_h. \tag{12}$$

Further, we define $V_h := S_h^1(\mathcal{T}_h) \cap H_0^1(\Omega) = \text{span}\{\varphi_j^h\}_{j=1}^{n_h}$, where we assume that the ordering of the basis functions is such that the indices $j = 1, \ldots, n_h$ correspond to vertices $x_j \in \Omega$ and $j = n_h + 1, \ldots, \overline{n}_h$ corresponds to the vertices on the boundary, $x_j \in \partial\Omega$. We refer to the books [11, 16, 39] for more details on standard finite element discretizations of elliptic PDEs.

A conforming discretization of (10)-(11) is then to find $(y_{\varrho h}, \tilde{p}_{\varrho h}) \in V_h \times V_h$ such that

$$\begin{aligned} \frac{1}{\sqrt{\varrho}}\left\langle \tilde{p}_{\varrho h}, q_h\right\rangle_{L_2(\Omega)} + \left\langle \nabla y_{\varrho h}, \nabla q_h\right\rangle_{L_2(\Omega)} &= 0,\ \ \forall q_h \in V_h, & (13)\\ -\left\langle \nabla \tilde{p}_{\varrho h}, \nabla v_h\right\rangle_{L_2(\Omega)} + \frac{1}{\sqrt{\varrho}}\left\langle y_{\varrho h}, v_h\right\rangle_{L_2(\Omega)} &= \frac{1}{\sqrt{\varrho}}\left\langle y_d, v_h\right\rangle_{L_2(\Omega)},\ \ \forall v_h \in V_h. & (14)\end{aligned}$$

In [23], we were able to show the following result for the L_2 error between the desired state y_d and the computed finite element state $y_{\varrho h}$.

Theorem 1 ([23, Corollary 1]). *Let $(y_{\varrho h}, \tilde{p}_{\varrho h}) \in V_h \times V_h$ be the unique solution of the coupled finite element variational formulation* (13)-(14). *Let $y_d \in H_0^s(\Omega)$ for $s \in [0,1]$ or $y_d \in H^s(\Omega) \cap H_0^1(\Omega)$ for $s \in (1,2]$. Then*

$$\|y_{\varrho h} - y_d\|_{L_2(\Omega)} \leq c\, h^s \, \|y_d\|_{H^s(\Omega)},$$

provided that $\varrho = h^4$.

We recall that $v_h \in V_h$ can be represented in the form

$$v_h(x) = \sum_{i=1}^{n_h} v_i \varphi_i^h(x), \tag{15}$$

where $v_i = v_h(x_i)$. Thus, we can associate each finite element function with its coefficient vector via the finite element isomorphism $v_h \leftrightarrow \mathbf{v}$, where $\mathbf{v}[i] = v_i$. With this, the matrix system corresponding to the fe scheme (13)-(14) can be written in the form: find $(\mathbf{y}_h, \tilde{\mathbf{p}}_h) \in \mathbb{R}^{n_h} \times \mathbb{R}^{n_h}$ such that

$$\begin{pmatrix} \frac{1}{\sqrt{\varrho}} M_h & K_h \\ -K_h^\top & \frac{1}{\sqrt{\varrho}} M_h \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{p}}_h \\ \mathbf{y}_h \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \frac{1}{\sqrt{\varrho}} \mathbf{y}_{dh} \end{pmatrix}, \tag{16}$$

where M_h and K_h denote the mass resp. stiffness matrix with the entries

$$M_h[i,j] = \int_\Omega \varphi_j^h(x) \varphi_i^h(x)\, dx \quad \text{and} \quad K_h[i,j] = \int_\Omega \nabla \varphi_j^h(x) \cdot \nabla \varphi_i^h(x)\, dx,$$

and the load vector

$$\mathbf{y}_{dh}[i] = \int_\Omega y_d(x) \varphi_i^h(x)\, dx.$$

We note that the system (16) is equivalent to (4). Moreover, when eliminating $\tilde{\mathbf{p}}_h$ resp. $\mathbf{p}_h$, we arrive at the same Schur-complement system (4).

As already mentioned in the introductionary Section 1, we would like to replace the inverse of the mass matrix M_h in the spd Schur complement $\varrho K_h M_h^{-1} K_h + M_h$ by the inverse of the lumped mass matrix $\mathrm{lump}(M_h)$ that is diagonal. The entries of the lumped mass matrix $\mathrm{lump}(M_h)$ are given as

$$\mathrm{lump}(M_h)[i,j] = \delta_{i,j} \sum_{k=1}^{\overline{n}_h} \widetilde{M}_h[i,k], \quad i,j = 1, \ldots, n_h, \tag{17}$$

where $\widetilde{M}_h \in \mathbb{R}^{\overline{n}_h} \times \mathbb{R}^{\overline{n}_h}$ denotes the mass matrix on $S_h^1(\mathcal{T}_h)$ with entries

$$\widetilde{M}_h[i,j] := \int_\Omega \varphi_j^h(x) \varphi_i^h(x)\, dx, \quad i,j = 1, \ldots, \overline{n}_h.$$

The Schur complement system is then given by (5) with the Schur complement $S_h = \varrho K_h (\mathrm{lump}(M_h))^{-1} K_h + M_h$ as system matrix. Unique solvability of the discrete system follows immediately, since $M_h = M_h^\top > 0$ is symmetric and positive definite (spd) and $K_h (\mathrm{lump}(M_h))^{-1} K_h > 0$ is spd, as $\mathrm{lump}(M_h)$ and K_h are spd.

Now the aim is to show an equivalent result to Theorem 1, when using the lumped mass matrix. We will exploit ideas from [7]. The discrete variational formulation for the lumped case reads as follows: find $(\hat{y}_{\varrho h}, \hat{p}_{\varrho h}) \in V_h \times V_h$ such that

$$\frac{1}{\sqrt{\varrho}} \langle \hat{p}_{\varrho h}, q_h \rangle_h + \langle \nabla \hat{y}_{\varrho h}, \nabla q_h \rangle_{L_2(\Omega)} = 0, \ \ \forall q_h \in V_h, \tag{18}$$

$$-\langle \nabla \hat{p}_{\varrho h}, \nabla v_h \rangle_{L_2(\Omega)} + \frac{1}{\sqrt{\varrho}} \langle \hat{y}_{\varrho h}, v_h \rangle_{L_2(\Omega)} = \frac{1}{\sqrt{\varrho}} \langle y_d, v_h \rangle_{L_2(\Omega)}, \ \ \forall v_h \in V_h, \tag{19}$$

where $\langle p_h, q_h \rangle_h = \mathbf{q}_h^\top \mathrm{lump}(M_h) \mathbf{p}_h$ denotes the underintegrated inner product on $L_2(\Omega)$ that is nothing but the realization of the mass lumping.

Lemma 1. *For $p_h, q_h \in V_h$ the realization of the lumped mass matrix admits the representation*

$$\langle p_h, q_h \rangle_h = \int_\Omega I_h^1(p_h q_h)\,dx,$$

where $I_h^1 : \mathcal{C}(\overline{\Omega}) \to V_h$ denotes the interpolation operator, given as

$$I_h^1 v(x) = \sum_{i=1}^{n_i} v_i \varphi_i^h(x),\ x \in \overline{\Omega},$$

where $v_i = v(x_i)$, $v \in \mathcal{C}(\overline{\Omega})$. Furthermore, it holds that

$$\frac{1}{d+2}\,\|p_h\|_h^2 \le \|p_h\|_{L_2(\Omega)}^2 \le \|p_h\|_h^2 := \langle p_h, p_h \rangle_h \ \text{ for all } p_h \in V_h. \tag{20}$$

Proof. With the representation (15), we have the coefficient vectors $\mathbf{p}_h \leftrightarrow p_h$ and $\mathbf{q}_h \leftrightarrow q_h$. Then we compute with (17), using (12),

$$\begin{aligned}
\langle p_h, q_h \rangle_h &= \mathbf{q}_h^\top \operatorname{lump}(M_h)\mathbf{p}_h = \sum_{i=1}^{n_h}\sum_{j=1}^{n_h} p_i q_j \delta_{i,j} \sum_{k=1}^{\overline{n}_h} \widetilde{M}_h[i,k] \\
&= \sum_{i=1}^{n_h} p_i q_i \sum_{k=1}^{\overline{n}_h} \widetilde{M}_h[i,k] = \sum_{i=1}^{n_h} p_i q_i \sum_{k=1}^{\overline{n}_h} \int_\Omega \varphi_i^h(x)\varphi_k^h(x)\,dx \\
&= \int_\Omega \underbrace{\sum_{i=1}^{n_h} p_i q_i \varphi_i^h(x)}_{I_h^1(p_h q_h)} \underbrace{\sum_{k=1}^{\overline{n}_h} \varphi_k^h(x)}_{=1}\,dx.
\end{aligned}$$

The estimate (20) follows from, e.g., [39, Lemma 9.4],

$$\frac{|\tau_e|}{(d+1)(d+2)} \sum_{x_i \in \overline{\tau}_e} p_i^2 \le \|p_h\|_{L_2(\tau_e)}^2 \le \frac{|\tau_e|}{d+1} \sum_{x_i \in \overline{\tau}_e} p_i^2,$$

and

$$\int_{\tau_e} I_h^1(p_h^2)(x)\,dx = \sum_{x_i \in \overline{\tau}_e} p_i^2 \int_{\tau_e} \varphi_i^h(x)\,dx = \frac{|\tau_e|}{d+1} \sum_{x_i \in \overline{\tau}_e} p_i^2,$$

when summing up over all elements τ_e.

□

With this representation, we can compute the consistency error.

Lemma 2. *Let $h = \max_{e=1,\dots,N_h} h_e$. Then, for $p_h, q_h \in V_h$, it holds*

$$\begin{aligned}
\Big|\langle p_h, q_h \rangle_{L_2(\Omega)} - \langle p_h, q_h \rangle_h\Big| &= \left|\int_\Omega \Big[p_h(x)q_h(x) - I_h^1(p_h q_h)(x)\Big]\,dx\right| \\
&\le c\,h^2\Big(\varepsilon^2\,\|\nabla p_h\|_{L_2(\Omega)}^2 + \frac{1}{\varepsilon^2}\|\nabla q_h\|_{L_2(\Omega)}^2\Big),
\end{aligned}$$

for any $\varepsilon > 0$.

Proof. The first representation follows from Lemma 1. Let τ_e be a simplicial finite element with the nodes x_{e_i}, $i = 1, \dots, d+1$. The associated nodal values of a

piecewise linear finite element function p_h are the coefficients p_{e_i}, $i = 1, \dots, d+1$. In particular, for $d = 1$ and $x \in \tau_e$, we then compute

$$\begin{aligned}
&\int_{\tau_e} \Big[p_h(x) q_h(x) - I_h^1(p_h q_h)(x) \Big] \, dx \\
&\quad = \int_{x_{e_1}}^{x_{e_2}} \bigg(\Big[p_{e_1} + \frac{x - x_{e_1}}{h_e} (p_{e_2} - p_{e_1}) \Big] \Big[q_{e_1} + \frac{x - x_{e_1}}{h_e} (q_{e_2} - q_{e_1}) \Big] \\
&\qquad\qquad - \Big[p_{e_1} q_{e_1} + \frac{x - x_{e_1}}{h_e} (p_{e_2} q_{e_2} - p_{e_1} q_{e_1}) \Big] \bigg) \, dx \\
&\quad = \frac{1}{6} \, h_e \, (p_{e_2} - p_{e_1}) \, (q_{e_2} - q_{e_1}) \, = \, \frac{1}{6} \, h_e^2 \int_{x_{e_1}}^{x_{e_2}} \frac{p_{e_2} - p_{e_1}}{h_e} \, \frac{q_{e_2} - q_{e_1}}{h_e} \, dx \\
&\quad = \frac{1}{6} \, h_e^2 \int_{x_{e_1}}^{x_{e_2}} p_h'(x) \, q_h'(x) \, dx \, \le \, \frac{1}{6} \, h_e^2 \, \|\nabla_x p_h\|_{L^2(\tau_e)} \|\nabla_x q_h\|_{L^2(\tau_e)} \, .
\end{aligned}$$

For $d = 2$ and $x \in \tau_e$, we introduce the representation $x = x_{e_1} + J_e \eta$ with respect to the reference element $\tau = \{\eta \in \mathbb{R}^2 : \eta_1 \in (0,1), \eta_2 \in (0, 1 - \eta_1)\}$ and we write $p_h(x) = p_h(x_{e_1} + J_e \eta) = \widetilde{p}_h(\eta)$, $\eta \in \tau$. Similar as in the case $d = 1$ we then compute, using $\det J_e = 2 \, |\tau_e|$,

$$\begin{aligned}
&\int_{\tau_e} \Big[p_h(x) q_h(x) - I_h^1(p_h q_h)(x) \Big] \, dx \, = \, \int_{\tau} \Big[\widetilde{p}_h(\eta) \widetilde{q}_h(\eta) - I_h^1(\widetilde{p}_h \widetilde{q}_h)(\eta) \Big] \, \det J_e \, d\eta \\
&= \frac{|\tau_e|}{12} \Big[(p_0 - p_2)(q_2 - q_0) + (p_1 - p_0)(q_0 - q_1) + (p_1 - p_2)(q_2 - q_1) \Big] \\
&\le \frac{|\tau_e|}{12} \Big[(p_0 - p_2)^2 + (p_1 - p_0)^2 + (p_1 - p_2)^2 \Big]^{1/2} \\
&\qquad\qquad \cdot \Big[(q_2 - q_0)^2 + (q_0 - q_1)^2 + (q_2 - q_1)^2 \Big]^{1/2} .
\end{aligned}$$

With

$$(p_1 - p_2)^2 = (p_1 - p_0 + p_0 - p_2)^2 \le 2 \, (p_1 - p_0)^2 + 2 \, (p_0 - p_2)^2,$$

we further have, e.g., [39, Lemma 9.1],

$$\begin{aligned}
&\int_{\tau_e} \Big[p_h(x) q_h(x) - I_h^1(p_h q_h)(x) \Big] \, dx \\
&\qquad \le \, \frac{|\tau_e|}{4} \Big[(p_0 - p_2)^2 + (p_1 - p_0)^2 \Big]^{1/2} \Big[(q_2 - q_0)^2 + (q_0 - q_1)^2 \Big]^{1/2} \\
&\qquad = \, \frac{|\tau_e|}{4} \left[2 \int_{\tau} |\nabla_\eta \widetilde{p}_h|^2 d\eta \right]^{1/2} \left[2 \int_{\tau} |\nabla_\eta \widetilde{q}_h|^2 d\eta \right]^{1/2} \\
&\qquad = \, \frac{|\tau_e|}{2} \, \|\nabla_\eta \widetilde{p}_h\|_{L^2(\tau)} \|\nabla_\eta \widetilde{q}_h\|_{L^2(\tau)} \, \le \, c \, h_e^2 \, \|\nabla_x p_h\|_{L^2(\tau_e)} \|\nabla_x q_h\|_{L^2(\tau_e)} \, .
\end{aligned}$$

For $d = 3$, we proceed in the same way. Now the reference element is given by $\tau = \{\eta \in \mathbb{R}^3 : \eta_1 \in (0,1), \eta_2 \in (0, 1 - \eta_1), \eta_3 \in (0, 1 - \eta_1 - \eta_2)\}$, and $\det J_e = 6 \, |\tau_e|$.

Then,

$$
\begin{aligned}
&\int_{\tau_e} \Big[p_h(x)q_h(x) - I_h^1(p_hq_h)(x)\Big]\,dx \\
&= 6\,|\tau_e| \int_0^1 \int_0^{1-\eta_1} \int_0^{1-\eta_1-\eta_2} \Big[\widetilde{p}_h(\eta)\widetilde{q}_h(\eta) - I_h^1(\widetilde{p}_h\widetilde{q}_h)(\eta)\Big]d\eta_3 d\eta_2 d\eta_1 \\
&= \frac{|\tau_e|}{20}\Big[(p_0-p_1)(q_1-q_0) + (p_0-p_2)(q_2-q_0) + (p_0-p_3)(q_3-q_0) \\
&\qquad\qquad +(p_1-p_2)(q_2-q_1) + (p_1-p_3)(q_3-q_1) + (p_2-p_3)(q_3-q_2)\Big] \\
&\le \frac{|\tau_e|}{20}\Big[(p_0-p_1)^2 + (p_0-p_2)^2 + (p_0-p_3)^2 + (p_1-p_2)^2 + (p_1-p_3)^2 + (p_2-p_3)^2\Big]^{1/2} \\
&\qquad \cdot\Big[(q_1-q_0)^2 + (q_2-q_0)^2 + (q_3-q_0)^2 + (q_2-q_1)^2 + (q_3-q_1)^2 + (q_3-q_2)^2\Big]^{1/2} \\
&\le \frac{|\tau_e|}{4}\Big[(p_0-p_1)^2 + (p_0-p_2)^2 + (p_0-p_3)^2\Big]^{1/2}\Big[(q_1-q_0)^2 + (q_2-q_0)^2 + (q_3-q_0)^2\Big]^{1/2} \\
&= \frac{|\tau_e|}{4}\left[6\int_\tau |\nabla_\eta \widetilde{p}_h|^2 d\eta\right]^{1/2}\left[6\int_\tau |\nabla_\eta|\widetilde{q}_h|^2 d\eta\right]^{1/2} \\
&= \frac{3}{2}\,|\tau_e|\,\|\nabla_\eta \widetilde{p}_h\|_{L^2(\tau)}\|\nabla_\eta \widetilde{q}_h\|_{L^2(\tau)} \;\le\; c\,h_e^2\,\|\nabla_x p_h\|_{L^2(\tau_e)}\|\nabla_x q_h\|_{L^2(\tau_e)}\,.
\end{aligned}
$$

Hence, using Young's inequality,

$$
\|\nabla_x p_h\|_{L_2(\tau_e)}\|\nabla_x q_h\|_{L_2(\tau_e)} \le \frac{1}{2}\left(\varepsilon^2\,\|\nabla_x p_h\|^2_{L_2(\tau_e)} + \frac{1}{\varepsilon^2}\,\|\nabla_x q_h\|^2_{L_2(\tau_e)}\right),
$$

and summing up over all elements τ_e, this gives the desired estimate. □

We need one more preliminary result, before we can state the main theorem.

Lemma 3. *Let $(y_\varrho, \tilde{p}_\varrho) \in H_0^1(\Omega) \times H_0^1(\Omega)$ be the unique solution of the reduced optimality system* (10) *and* (11)*. Then there holds the regularization error estimate*

$$
\|y_\varrho - y_d\|_{H^{-1}(\Omega)} \le c\,\sqrt{\varrho}\,|y_d|_{H^1(\Omega)} \quad \text{for } y_d \in H_0^1(\Omega).
$$

Additionally, for $y_d \in H^\Delta(\Omega) \cap H_0^1(\Omega)$, there holds

$$
\|y_\varrho - y_d\|_{L_2(\Omega)} \le \sqrt{\varrho}\,\|\Delta y_d\|_{L_2(\Omega)} \quad \text{and} \quad \|\Delta y_\varrho\|_{L_2(\Omega)} \le \|\Delta y_d\|_{L_2(\Omega)}.
$$

Proof. The first estimate is given in [32, Theorem 4.1, (4.7)]. The second and third estimate can be found in [23, Lemma 1, (2.5)] and in the proof of this lemma. But for clarity, we will recall the proof. We note that, by the optimality system, we have the equations $-\Delta y_\varrho = u_\varrho$, and $p_\varrho = -\varrho u_\varrho$. First, assuming the regularity $y_\varrho, y_d \in H^\Delta(\Omega) \cap H_0^1(\Omega)$, using (8) and (9), and integration by parts, we obtain

$$
\begin{aligned}
\|y_\varrho - y_d\|^2_{L_2(\Omega)} &= \langle y_\varrho - y_d, y_\varrho - y_d\rangle_{L_2(\Omega)} = \langle \nabla p_\varrho, \nabla(y_\varrho - y_d)\rangle_{L_2(\Omega)} \\
&= \langle p_\varrho, -\Delta y_\varrho\rangle_{L_2(\Omega)} + \langle p_\varrho, \Delta y_d\rangle_{L_2(\Omega)} \\
&= -\varrho\,\langle u_\varrho, -\Delta y_\varrho\rangle_{L_2(\Omega)} - \varrho\,\langle u_\varrho, \Delta y_d\rangle_{L_2(\Omega)} \\
&= -\varrho\,\|\Delta y_\varrho\|^2_{L_2(\Omega)} + \varrho\,\langle \Delta y_\varrho, \Delta y_d\rangle_{L_2(\Omega)}\,.
\end{aligned}
$$

From this we conclude

$$
\|y_\varrho - y_d\|^2_{L_2(\Omega)} + \varrho\,\|\Delta y_\varrho\|^2_{L_2(\Omega)} \le \varrho\,\|\Delta y_\varrho\|_{L_2(\Omega)}\|\Delta y_d\|_{L_2(\Omega)},
$$

and further

$$\|\Delta y_\varrho\|_{L_2(\Omega)} \le \|\Delta y_d\|_{L_2(\Omega)} \quad \text{and} \quad \|y_\varrho - y_d\|_{L_2(\Omega)} \le \sqrt{\varrho}\,\|\Delta y_d\|_{L_2(\Omega)}.$$

From the first estimate we conclude that $y_d \in H^\Delta(\Omega) \cap H_0^1(\Omega)$ implies $y_\varrho \in H^\Delta(\Omega) \cap H_0^1(\Omega)$. This confirms that it is sufficient to require the regularity on y_d only. □

The main statement of this paper is formulated in the following theorem.

Theorem 2. *Let $(\hat{y}_{\varrho h}, \hat{p}_{\varrho h}) \in V_h \times V_h$ be the unique solution of the variational formulation* (18) *and* (19)*. Assume that $\mathcal{T}_h$ is globally quasi-uniform such that a global inverse inequality holds true. Further, choose $\varrho = h^4$. Then,*

$$\|\hat{y}_{\varrho h} - y_d\|_{L_2(\Omega)} \le \begin{cases} c\,h\,\|y_d\|_{H_0^1(\Omega)}, & \text{if } y_d \in H_0^1(\Omega), \\ c\,h^2\,\|y_d\|_{H^2(\Omega)}, & \text{if } y_d \in H^2(\Omega) \cap H_0^1(\Omega) \text{ and } \Omega \text{ is convex.} \end{cases}$$

Proof. Let $(y_{\varrho h}, \tilde{p}_{\varrho h}) \in V_h \times V_h$ be the unique solution of (13) and (14). By the triangle inequality, we get that

$$\|\hat{y}_{\varrho h} - y_d\|_{L_2(\Omega)} \le \|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)} + \|y_{\varrho h} - y_d\|_{L_2(\Omega)}.$$

By Theorem 1, the second term fulfils the estimate. Thus, it is sufficient to bound the first term. Therefore, subtracting the variational formulation (13) and (14) from (18) and (19) with $v_h = \hat{y}_{\varrho h} - y_{\varrho h}$ and $q_h = \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}$, we obtain the equalities

$$\begin{aligned} \frac{1}{\sqrt{\varrho}}\,\|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)}^2 &= \frac{1}{\sqrt{\varrho}}\,\langle \hat{y}_{\varrho h} - y_{\varrho h}, \hat{y}_{\varrho h} - y_{\varrho h}\rangle_{L_2(\Omega)} \\ &= \langle \nabla(\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}), \nabla(\hat{y}_{\varrho h} - y_{\varrho h})\rangle_{L_2(\Omega)} \\ &= \frac{1}{\sqrt{\varrho}}\left(\langle \tilde{p}_{\varrho h}, \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\rangle_{L_2(\Omega)} - \langle \hat{p}_{\varrho h}, \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\rangle_h\right) \\ &= \frac{1}{\sqrt{\varrho}}\left(\langle \tilde{p}_{\varrho h}, \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\rangle_{L_2(\Omega)} - \|\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\|_h^2 - \langle \tilde{p}_{\varrho h}, \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\rangle_h\right). \end{aligned}$$

Multiplying by $\sqrt{\varrho}$ and using Lemma 1, we further get

$$\begin{aligned} \|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)}^2 + \|\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\|_{L_2(\Omega)}^2 &\le \|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)}^2 + \|\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\|_h^2 \\ &= \langle \tilde{p}_{\varrho h}, \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\rangle_{L_2(\Omega)} - \langle \tilde{p}_{\varrho h}, \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\rangle_h \\ &= \int_\Omega \left[\tilde{p}_{\varrho h}(x)(\hat{p}_{\varrho h}(x) - \tilde{p}_{\varrho h}(x)) - I_h^1(\tilde{p}_{\varrho h}(\hat{p}_{\varrho h} - \tilde{p}_{\varrho h})(x)\right] dx. \end{aligned}$$

With Lemma 2, choosing $p_h = \tilde{p}_{\varrho h}$ and $q_h = \hat{p}_{\varrho h} - \tilde{p}_{\varrho h}$, we estimate, for some $\varepsilon > 0$ to be specified,

$$\begin{aligned} &\|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)}^2 + \|\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\|_{L_2(\Omega)}^2 \\ &\qquad \le c\,h^2\left(\varepsilon^2\,\|\nabla \tilde{p}_{\varrho h}\|_{L_2(\Omega)}^2 + \frac{1}{\varepsilon^2}\,\|\nabla(\hat{p}_{\varrho h} - \tilde{p}_{\varrho h})\|_{L_2(\Omega)}^2\right). \end{aligned}$$

Using an inverse inequality, we estimate the second term by

$$\frac{ch^2}{\varepsilon^2}\,\|\nabla(\hat{p}_{\varrho h} - \tilde{p}_{\varrho h})\|_{L_2(\Omega)}^2 \le \frac{cc_I}{\varepsilon^2}\,\|\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\|_{L_2(\Omega)}^2 = \|\hat{p}_{\varrho h} - \tilde{p}_{\varrho h}\|_{L_2(\Omega)}^2,$$

when choosing $\varepsilon = \sqrt{cc_I}$. Thus, it holds

$$\|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)}^2 \le \tilde{c}\,h^2\,\|\nabla \tilde{p}_{\varrho h}\|_{L_2(\Omega)}^2. \tag{21}$$

Now it is sufficient to bound $\|\nabla \tilde{p}_{\varrho h}\|_{L_2(\Omega)}$ suitably. Let $(y_\varrho, \tilde{p}_\varrho) \in H_0^1(\Omega) \times H_0^1(\Omega)$ be the unique solution of the coupled variational formulation (8) and (9). Using the triangle inequality and the trivial inequality $(a+b)^2 \leq 2(a^2+b^2)$, we get

$$\|\nabla \tilde{p}_{\varrho h}\|^2_{L_2(\Omega)} \leq 2\left(\|\nabla \tilde{p}_\varrho\|^2_{L_2(\Omega)} + \|\nabla(\tilde{p}_{\varrho h} - \tilde{p}_\varrho)\|^2_{L_2(\Omega)}\right). \tag{22}$$

For $(y_\varrho, \widetilde{p}_\varrho) \in H_0^1(\Omega) \times H_0^1(\Omega)$ and $(y_{\varrho h}, \widetilde{p}_{\varrho h}) \in V_h \times V_h$ as solutions of (8)-(9) and (13)-(14), respectively, we can show, as in the proof of [23, Theorem 1], using an inverse inequality and $\varrho = h^4$, that Cea's Lemma

$$\begin{aligned} h^{-2}\,\|y_\varrho - y_{\varrho h}\|^2_{L_2(\Omega)} + \|\nabla(y_\varrho - y_{\varrho h})\|^2_{L_2(\Omega)} \\ + h^{-2}\,\|\tilde{p}_\varrho - \tilde{p}_{\varrho h}\|^2_{L_2(\Omega)} + \|\nabla(\tilde{p}_\varrho - \tilde{p}_{\varrho h})\|^2_{L_2(\Omega)} \\ \leq\; c\Big[h^{-2}\,\|y_\varrho - v_h\|^2_{L_2(\Omega)} + \|\nabla(y_\varrho - v_h)\|^2_{L_2(\Omega)} \\ + h^{-2}\,\|\tilde{p}_\varrho - q_h\|^2_{L_2(\Omega)} + \|\nabla(\tilde{p}_\varrho - q_h)\|^2_{L_2(\Omega)}\Big] \end{aligned}$$

holds true for all $(v_h, q_h) \in V_h \times V_h$. Further, using best approximation results, we get, for $y_d \in H_0^1(\Omega)$, that

$$\begin{aligned} h^{-2}\,\|y_\varrho - y_{\varrho h}\|^2_{L_2(\Omega)} + \|\nabla(y_\varrho - y_{\varrho h})\|^2_{L_2(\Omega)} \\ + h^{-2}\,\|\tilde{p}_\varrho - \tilde{p}_{\varrho h}\|^2_{L_2(\Omega)} + \|\nabla(\tilde{p}_\varrho - \tilde{p}_{\varrho h})\|^2_{L_2(\Omega)} \;\leq\; c\,|y_d|^2_{H^1(\Omega)}, \end{aligned}$$

and, for $y_d \in H_0^1(\Omega) \cap H^2(\Omega)$, that

$$\begin{aligned} h^{-2}\,\|y_\varrho - y_{\varrho h}\|^2_{L_2(\Omega)} + \|\nabla(y_\varrho - y_{\varrho h})\|^2_{L_2(\Omega)} \\ + h^{-2}\,\|\tilde{p}_\varrho - \tilde{p}_{\varrho h}\|^2_{L_2(\Omega)} + \|\nabla(\tilde{p}_\varrho - \tilde{p}_{\varrho h})\|^2_{L_2(\Omega)} \;\leq\; c\,h^2\,|y_d|^2_{H^2(\Omega)}. \end{aligned}$$

From this we immediately conclude that

$$\|\nabla(\tilde{p}_{\varrho h} - \tilde{p}_\varrho)\|^2_{L_2(\Omega)} \leq \begin{cases} c\,|y_d|^2_{H^1(\Omega)}, & \text{for } y_d \in H_0^1(\Omega), \\ c\,h^2\,|y_d|^2_{H^2(\Omega)}, & \text{for } y_d \in H_0^1(\Omega) \cap H^2(\Omega). \end{cases} \tag{23}$$

For the first term, we use the first inequality of Lemma 3 for $y_d \in H_0^1(\Omega)$ to estimate

$$\begin{aligned} \|\nabla \tilde{p}_\varrho\|^2_{L_2(\Omega)} = \langle \nabla \tilde{p}_\varrho, \nabla \tilde{p}_\varrho \rangle_{L_2(\Omega)} = \frac{1}{\sqrt{\varrho}}\,\langle y_\varrho - y_d, \tilde{p}_\varrho \rangle_{L_2(\Omega)} \\ \leq \frac{1}{\sqrt{\varrho}}\,\|y_\varrho - y_d\|_{H^{-1}(\Omega)} \|\nabla \tilde{p}_\varrho\|_{L_2(\Omega)} \;\leq\; c\,|y_d|_{H^1(\Omega)} \|\nabla \tilde{p}_\varrho\|_{L_2(\Omega)}, \end{aligned}$$

from which we conclude

$$\|\nabla \tilde{p}_\varrho\|_{L_2(\Omega)} \leq c\,|y_d|_{H^1(\Omega)}. \tag{24}$$

For a convex domain Ω, we have $H^\Delta(\Omega) = H^2(\Omega)$. Thus $y_d \in H^2(\Omega) \cap H_0^1(\Omega)$, and we get with the same reasoning, using the second inequality of Lemma 3,

$$\begin{aligned} \|\nabla \tilde{p}_\varrho\|^2_{L_2(\Omega)} = \frac{1}{\sqrt{\varrho}}\,\langle y_\varrho - y_d, \tilde{p}_\varrho \rangle_{L_2(\Omega)} \\ \leq \frac{1}{\sqrt{\varrho}}\,\|y_\varrho - y_d\|_{L_2(\Omega)} \|\tilde{p}_\varrho\|_{L_2(\Omega)} \;\leq\; |y_d|_{H^2(\Omega)} \|\tilde{p}_\varrho\|_{L_2(\Omega)}. \end{aligned}$$

Now, we recall that, from the optimality system (6)-(9), we have

$$\tilde{p}_\varrho = \frac{1}{\sqrt{\varrho}}\,p_\varrho = -\sqrt{\varrho}\,u_\varrho = \sqrt{\varrho}\,\Delta y_\varrho \quad \text{in } \Omega,$$

and thus, with Lemma 3, we obtain

$$\|\tilde{p}_\varrho\|_{L_2(\Omega)} = \sqrt{\varrho}\,\|\Delta y_\varrho\|_{L_2(\Omega)} \le \sqrt{\varrho}\,\|\Delta y_d\|_{L_2(\Omega)} \le \sqrt{\varrho}\,|y_d|_{H^2(\Omega)}.$$

Thus, for $\varrho = h^4$ and $y_d \in H^2(\Omega) \cap H_0^1(\Omega)$, we arrive at the estimate

$$\|\nabla \tilde{p}_\varrho\|_{L_2(\Omega)} \le h\,|y_d|_{H^2(\Omega)}. \tag{25}$$

Now, combining (21) with (22), (23),(24), and (25), this gives

$$\|\hat{y}_{\varrho h} - y_{\varrho h}\|_{L_2(\Omega)} \le h\,\|\nabla \tilde{p}_{\varrho h}\|_{L_2(\Omega)} \le \begin{cases} c\,h\,|y_d|_{H^1(\Omega)}, & \text{for } y_d \in H_0^1(\Omega), \\ c\,h^2\,|y_d|_{H^2(\Omega)}, & \text{for } y_d \in H^2(\Omega) \cap H_0^1(\Omega). \end{cases}$$

□

Lemma 4. *Let $(\hat{y}_{\varrho h}, \hat{p}_{\varrho h}) \in V_h \times V_h$ be the unique solution of the coupled variational formulation* (18) *and* (19)*. Then, for $y_d \in L_2(\Omega)$, we get the error estimate*

$$\|\hat{y}_{\varrho h} - y_d\|_{L_2(\Omega)} \le \|y_d\|_{L_2(\Omega)}.$$

Proof. Choosing $q_h = \hat{p}_{\varrho h}$ and $v_h = \hat{y}_{\varrho h}$ in (18) and (19), summing up the equations, and multiplying with $\sqrt{\varrho}$, this gives

$$\langle \hat{p}_{\varrho h}, \hat{p}_\varrho \rangle_h + \langle \hat{y}_{\varrho h}, \hat{y}_{\varrho h} \rangle_{L_2(\Omega)} = \langle y_d, \hat{y}_{\varrho h} \rangle_{L_2(\Omega)}.$$

Rewriting this equality gives

$$\langle \hat{p}_{\varrho h}, \hat{p}_{\varrho h} \rangle_h + \langle \hat{y}_{\varrho h} - y_d, \hat{y}_{\varrho h} - y_d \rangle_{L_2(\Omega)} = \langle y_d - \hat{y}_{\varrho h}, y_d \rangle_{L_2(\Omega)},$$

which yields the desired estimate. □

Theorem 3. *Let $(\hat{y}_{\varrho h}, \hat{p}_{\varrho h}) \in V_h \times V_h$ be the unique solution of the coupled variational formulation* (18)-(19)*. For $y_d \in H_0^s(\Omega)$, $s \in [0,1]$, and $y_d \in H^s(\Omega) \cap H_0^1(\Omega)$, $s \in (1,2]$, there holds the error estimate*

$$\|\hat{y}_{\varrho h} - y_d\|_{L_2(\Omega)} \le c\,h^s\,\|y_d\|_{H^s(\Omega)}. \tag{26}$$

Proof. This is a direct consequence of Theorem 2 and Lemma 4, together with a space interpolation argument. □

3 Nested PCG Iteration

Finally, we have to solve the spd mass-lumped Schur-complement system (5) that we now write in the compact from: find $\mathbf{y}_h \in \mathbb{R}^{n_h}$ such that

$$S_h \mathbf{y}_h = \mathbf{y}_{dh}, \tag{27}$$

where $S_h = \varrho K_h D_h^{-1} K_h + M_h$, and D_h is the lumped mass matrix $\mathrm{lump}(M_h)$. For simplicity, we omit the hat over $\mathbf{y}_h$ in (27) and throughout this section. The fast solution of the symmetric and indefinite system (3) with the original mass matrix M_h instead of D_h was studied in [23]. Since the matrix D_h is diagonal, the matrix-by-vector multiplication $S_h * \mathbf{y}_h$ can now be performed efficiently. Therefore, we can use the PCG method for solving (27). Moreover, it turns out that D_h can also serve as preconditioner in the case $\varrho = h^4$ that leads to the optimally balanced estimate of $\|\hat{y}_{\varrho h} - y_d\|_{L_2(\Omega)}$ as was shown in Section 2; see Theorem 2 and Theorem 3. First of all, we can easily show that D_h is spectrally equivalent to M_h, i.e., there exist positive, h-independent constants $\underline{c}_{\mathrm{MD}}$ and $\overline{c}_{\mathrm{MD}}$ such that

$$\underline{c}_{\rm MD} D_h \le M_h \le \overline{c}_{\rm MD} D_h, \tag{28}$$

where $\underline{c}_{\rm MD} = \lambda_{\min} = \lambda_{\min}(D_\tau^{-1} M_\tau) = 1/(d+2)$ and $\overline{c}_{\rm MD} = \lambda_{\max} = \lambda_{\max}(D_\tau^{-1} M_\tau) = 1$ are the minimal eigenvalue and maximal eigenvalue of the small generalized eigenvalue problem $M_\tau \mathbf{v}_\tau = \lambda D_\tau \mathbf{v}_\tau$ in $\mathbb{R}^{d+1}$, respectively. M_τ and D_τ denote resp. the mass matrix and the lumped mass matrix corresponding to the reference element (unit simplex) τ to which every element τ_e from $\mathcal{T}_h$ is mapped by an affine-linear mapping $x = x_{e_1} + J_e \eta$. We note that the spectral equivalence inequalities (28) are nothing but the algebraic version of the inequalities (20) where the constants were already explicitly computed.

In order to estimate the Schur complement $S_h = \varrho K_h D_h^{-1} K_h + M_h$ by the mass matrix M_h from below and above in the spectral sense, it is obviously enough to estimate $\varrho K_h D_h^{-1} K_h$ from above by M_h. Using the spectral equivalence inequalities (28), Cauchy's inequality, local inverse inequalities, and $\varrho = h^4$, we get

$$\begin{aligned}
(\varrho K_h D_h^{-1} K_h \mathbf{v}_h, \mathbf{v}_h) &= \varrho\,(D_h^{-1} K_h \mathbf{v}_h, K_h \mathbf{v}_h) \le \overline{c}_{\rm MD}\,\varrho\,(M_h^{-1} K_h \mathbf{v}_h, K_h \mathbf{v}_h) \\
&= \overline{c}_{\rm MD} (K_h (\varrho^{-1} M_h)^{-1} K_h \mathbf{v}_h, \mathbf{v}_h) \\
&= \overline{c}_{\rm MD} \sup_{\mathbf{q}_h \in \mathbb{R}^{n_h}} \frac{(K_h \mathbf{v}_h, \mathbf{q}_h)^2}{(\varrho^{-1} M_h \mathbf{q}_h, \mathbf{q}_h)} \\
&= \overline{c}_{\rm MD} \sup_{q_h \in V_h} \frac{\left[\int_\Omega \varrho^{1/4} \nabla v_h \cdot \varrho^{-1/4} \nabla q_h dx\right]^2}{\int_\Omega \varrho^{-1} [q_h(x)]^2 dx} \\
&\le \overline{c}_{\rm MD} \sup_{q_h \in V_h} \frac{\|\varrho^{1/4} \nabla v_h\|^2_{L_2(\Omega)} \|\varrho^{-1/4} \nabla q_h\|^2_{L_2(\Omega)}}{\int_\Omega \varrho^{-1} [q_h(x)]^2 dx} \\
&= \overline{c}_{\rm MD}\, \|\varrho^{1/4} \nabla v_h\|^2_{L_2(\Omega)} \sup_{q_h \in V_h} \frac{\sum\limits_{\tau_e \in \mathcal{T}_h} h^{-2} \int_{\tau_e} |\nabla q_h|^2 dx}{\int_\Omega \varrho^{-1} [q_h(x)]^2 dx} \\
&\le \overline{c}_{\rm MD}\, \|\varrho^{1/4} \nabla v_h\|^2_{L_2(\Omega)} \sup_{q_h \in V_h} \frac{\sum\limits_{\tau_e \in \mathcal{T}_h} h^{-4}\, c_{\rm inv}^2 \int_{\tau_e} (q_h)^2\, dx}{\sum\limits_{\tau_e \in \mathcal{T}_h} h^{-4} \int_{\tau_e} (q_h)^2 dx} \\
&= c_{\rm inv}^2\, \overline{c}_{\rm MD}\, \|\varrho^{1/4} \nabla v_h\|^2_{L_2(\Omega)} \\
&= c_{\rm inv}^2\, \overline{c}_{\rm MD} \sum_{\tau_e \in \mathcal{T}_h} h^2 \int_{\tau_e} |\nabla v_h|^2 dx \\
&\le c_{\rm inv}^4\, \overline{c}_{\rm MD} \sum_{\tau_e \in \mathcal{T}_h} \int_{\tau_e} (v_h)^2 dx \\
&= c_{\rm inv}^4\, \overline{c}_{\rm MD}\, (M_h \mathbf{v}_h, \mathbf{v}_h), \quad \forall \mathbf{v}_h \in V_h,
\end{aligned} \tag{29}$$

where $c_{\rm inv}$ is the universal positive constant in the local inverse inequalities

$$\|\nabla w_h\|_{L_2(\tau_e)} \le c_{\rm inv}\, h_e^{-1}\, \|w_h\|_{L_2(\tau_e)} \quad \forall w_h \in V_h,\ \forall \tau_e \in \mathcal{T}_h. \tag{30}$$

Here the local mesh size h_e can be replaced by the global mesh size h since we assumed quasi-uniform and shape-regular mesh $\mathcal{T}_h$. The local inverse inequalities

(30) can again be proved by mapping τ_e to the unit simplex τ. In this way the constant $c_{\rm inv}$ can even be computed explicitly in dependence of the mesh characteristics [12]. Therefore, we have just proved the spectral equivalence theorem that is fundamental for the efficient solution of the spd mass-lumped Schur-complement system (27) by means of PCG iteration.

Theorem 4. *Let us assume that the mesh $\mathcal{T}_h$ is globally quasi-uniform with the global mesh-size h, and $\varrho = h^4$. Then the spectral equivalence inequalities*

$$\underline{c}_{SD} D_h \leq \underline{c}_{SM} M_h \leq S_h = \varrho K_h D_h^{-1} K_h + M_h \leq \overline{c}_{SM} M_h \leq \overline{c}_{SD} D_h, \tag{31}$$

hold with the spectral equivalence constants

$$\underline{c}_{SM} = 1, \quad \overline{c}_{SM} = c_{inv}^4 \overline{c}_{MD} + 1, \quad \underline{c}_{SD} = \underline{c}_{MD} = \lambda_{min} = \lambda_{min}(D_T^{-1} M_T) = \frac{1}{d+2},$$

and $\overline{c}_{SD} = \overline{c}_{MD}^2 c_{inv}^4 + \overline{c}_{MD}$, where $\overline{c}_{MD} = \lambda_{max} = \lambda_{max}(D_T^{-1} M_T) = 1$.

Proof. The spectral equivalence inequalities (31) immediately follow from the inequalities (28), and (29). □

Remark 1. *The spectral estimate* (29) *can also be proved by Fourier analysis when one expands the vectors $\mathbf{v}_h$ into the orthonormal eigenvector basis corresponding to the eigenvalue problem $K_h \mathbf{e}_h = \lambda D_h \mathbf{e}_h$ as it was done in* [23] *for $D_h = M_h$. In* [22]*, we provide a rigorous analysis of the variable L_2 regularization with a technique that is different from the technique used for proving* (29)*. We note that the latter technique can be used to analyse the case of constant and variable energy regularizations for state equations leading to non-symmetric fe stiffness matrices K_h such as convection-diffusion problems as well as parabolic and hyperbolic problems when using space-time finite element discretizations; see* [27, 30].

Now we can efficiently solve the mass-lumped Schur-complement system (5) respectively (27) by means of the PCG methods because, thanks to mass lumping, the matrix-vector multiplication $S_h * \mathbf{y}_h^k$ can be performed in asymptotically optimal complexity $O(n_h)$, and, at the same time, the lumped mass matrix $D_h = \mathrm{lump}(M_h)$ is a perfect preconditioner. More precisely, let $\mathbf{y}_h^k \in \mathbb{R}^{n_h}$ be the kth PCG iterate. Due to the spectral equivalence inequalities (28) and (31), and the well-known convergence rate estimate for the PCG method (see, e.g., [39, Chapter 13]), we can estimate the L_2 error $\|\hat{y}_{\varrho h} - y_{\varrho h}^k\|_{L_2(\Omega)}$ between the fe functions $\hat{y}_{\varrho h}(x) = \sum_{i=1}^{n_h} y_i \varphi_i^h(x) \in V_h$ and $y_{\varrho h}^k(x) = \sum_{i=1}^{n_h} y_i^k \varphi_i^h(x) \in V_h$ corresponding to the solution $\mathbf{y}_h = (y_i)_{i=1,\ldots,n_h} \in \mathbb{R}^{n_h}$ of the Schur complement system (27) and the k-th PCG iterate $\mathbf{y}_h^k = (y_i^k)_{i=1,\ldots,n_h} \in \mathbb{R}^{n_h}$, respectively, as follows:

$$\begin{aligned}
\|\hat{y}_{\varrho h} - y_{\varrho h}^k\|_{L_2(\Omega)} &= \|\mathbf{y}_h - \mathbf{y}_h^k\|_{\mathbf{M}_h} := (\mathbf{M}_h(\mathbf{y}_h - \mathbf{y}_h^k), \mathbf{y}_h - \mathbf{y}_h^k)^{1/2} \\
&\leq (\mathbf{S}_h(\mathbf{y}_h - \mathbf{y}_h^k), \mathbf{y}_h - \mathbf{y}_h^k)^{1/2} \\
&= \|\mathbf{y}_h - \mathbf{y}_h^k\|_{\mathbf{S}_h} \leq 2\, q^k \, \|\mathbf{y}_h - \mathbf{y}_h^0\|_{\mathbf{S}_h} \\
&\leq 2\, \overline{c}_{\rm SM}^{1/2}\, q^k \, \|\mathbf{y}_h - \mathbf{y}_h^0\|_{\mathbf{M}_h} = 2\, \overline{c}_{\rm SM}^{1/2}\, q^k \, \|\hat{y}_{\varrho h} - y_{\varrho h}^0\|_{L_2(\Omega)},
\end{aligned} \tag{32}$$

where $q = (\sqrt{\mathrm{cond}_2(\mathbf{D}_h^{-1}\mathbf{S}_h)} - 1)/(\sqrt{\mathrm{cond}_2(\mathbf{D}_h^{-1}\mathbf{S}_h)} + 1) < 1$, and $\mathrm{cond}_2(\mathbf{D}_h^{-1}\mathbf{S}_h) = \lambda_{\max}(\mathbf{D}_h^{-1}\mathbf{S}_h)/\lambda_{\min}(\mathbf{D}_h^{-1}\mathbf{S}_h)$ denotes the spectral condition number of $\mathbf{D}_h^{-1}\mathbf{S}_h$ that can be bounded by the constant

$$\frac{\overline{c}_{\rm SD}}{\underline{c}_{\rm SD}} = \frac{\overline{c}_{\rm MD}^2 c_{\rm inv}^4 + \overline{c}_{\rm MD}}{\underline{c}_{\rm MD}} = \frac{\lambda_{\max}(D_T^{-1} M_T)^2 c_{\rm inv}^4 + \lambda_{\max}(D_T^{-1} M_T)}{\lambda_{\min}(D_T^{-1} M_T)} = (d+2)(c_{\rm inv}^4 + 1)$$

that is independent of h. Using the triangle inequality, the L_2-norm discretization error estimate (26) from Theorem 3, the L_2-norm iteration error estimate (32), the inequality

$$\|\hat{y}_{\varrho h}\|_{L_2(\Omega)} \leq \|y_d\|_{L_2(\Omega)} \tag{33}$$

that follow from (18)-(19) when we choose the test functions $q_h = \hat{p}_{\varrho h}$ and $v_h = \hat{y}_{\varrho h}$, we finally arrive at L_2-norm estimate between desired state y_d and the kth PCG iterate $y^k_{\varrho h}$ computed by the PCG method:

$$\begin{aligned}
\|y_d - y^k_{\varrho h}\|_{L_2(\Omega)} &\leq \|y_d - \hat{y}_{\varrho h}\|_{L_2(\Omega)} + \|\hat{y}_{\varrho h} - y^k_{\varrho h}\|_{L_2(\Omega)} \\
&\leq c\, h^s \, \|y_d\|_{H^s(\Omega)} + 2\, \overline{c}_{\rm SM}^{1/2}\, q^k \, \|\hat{y}_{\varrho h} - y^0_{\varrho h}\|_{L_2(\Omega)} \\
&\leq h^s \Big(c\, \|y_d\|_{H^s(\Omega)} + 2\, \overline{c}_{\rm SM}^{1/2} \, \|\hat{y}_{\varrho h} - y^0_{\varrho h}\|_{L_2(\Omega)} \Big) \\
&\leq h^s \Big(c\, \|y_d\|_{H^s(\Omega)} + 2\, \overline{c}_{\rm SM}^{1/2} \, \|y_d\|_{L_2(\Omega)} \Big) = h^s c(y_d)
\end{aligned} \tag{34}$$

provided that $q^k \leq h^s$ and that the initial guess $y^0_{\varrho h}$ is chosen to be zero. Therefore, $k \geq \ln h^{-s} / \ln q^{-1}$ ensures that the PCG computes an approximation $y^k_{\varrho h}$ to the desired state y_d that differs from y_d in the same order $O(h^s)$ as the discretization error $\|y_d - \hat{y}_{\varrho h}\|_{L_2(\Omega)}$ in the L_2 norm. Moreover, this can be done with $O(n_h \ln h^{-1}) = O(h^{-d} \ln h^{-1})$ arithmetical operations, i.e. the complexity is asymptotically optimal up to the logarithmical factor $\ln h^{-1}$.

This logarithmical factor can be avoided in a nested iteration setting on a sequence of refined (nested) meshes. Indeed, let us consider a sequence of uniformly refined meshes $\mathcal{T}_\ell = \mathcal{T}_{h_\ell}$ with the mesh size h_ℓ and the optimally balanced regularization parameter $\varrho_\ell = h_\ell^4$, $\ell = 1, \ldots, L$, where $h_\ell = h_{\ell-1}/2$, $\ell = 2, \ldots, L$. Thus, the coarsest mesh corresponds to the subindex 1, whereas the finest mesh is related to L. On every mesh $\mathcal{T}_\ell$, $\ell = 1, \ldots, L$, we have to solve the mass-lumped Schur-complement system (27) that we now write in form: find $\mathbf{y}_\ell = \mathbf{y}_{h_\ell} \in \mathbb{R}^{n_\ell} = \mathbb{R}^{n_{h_\ell}}$ such that

$$S_\ell \mathbf{y}_\ell = \mathbf{y}_{d\ell} \tag{35}$$

where $S_\ell = \varrho_\ell K_\ell D_\ell^{-1} K_\ell + M_\ell$, $K_\ell = K_{h_\ell}$. $D_\ell = D_{h_\ell}$, $M_\ell = M_{h_\ell}$, $\mathbf{y}_{d\ell} = \mathbf{y}_{dh_\ell}$, and $\varrho_\ell = h_\ell^4$.

Now the *nested iteration algorithm* works as follows. First we solve the coarse-mesh problem (35), $\ell = 1$, sufficiently accurate. More precisely, we compute an iterate $\mathbf{y}_1^{k_1} \in \mathbb{R}^{n_1}$ corresponding to the fe function $y_1^{k_1} = y^{k_1}_{\varrho_1 h_1} \in V_1 = V_{h_1}$ (short: $\mathbf{y}_1^{k_1} \leftrightarrow y_1^{k_1}$) such that

$$\|y_d - y_1^{k_1}\|_{L_2(\Omega)} \leq h_1^s c(y_d). \tag{36}$$

Due to (34), this can be done with $k_1 \geq \ln h_1^{-s} / \ln q^{-1}$ PCG iterations starting with $y_1^0 = 0$. Now, let us assume that, on level $\ell - 1 \in \{1, \ldots, L-1\}$, the iterate $y_{\ell-1}^{k_{\ell-1}} \in V_{\ell-1}$ fulfills the estimate

$$\|y_d - y_{\ell-1}^{k_{\ell-1}}\|_{L_2(\Omega)} \leq h_{\ell-1}^s c(y_d). \tag{37}$$

Let $y_\ell^0 = I_{\ell-1}^\ell y_{\ell-1}^{k_{\ell-1}}$ be the affine-linear interpolate of $y_{\ell-1}^{k_{\ell-1}}$. We note that $y_\ell^0 = I_{\ell-1}^\ell y_{\ell-1}^{k_{\ell-1}} = y_{\ell-1}^{k_{\ell-1}} \in V_{\ell-1} \subset V_\ell$ since the meshes are nested. Then we get the estimate

$$\begin{aligned}
\|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)} &\leq \|y_d - \hat{y}_\ell\|_{L_2(\Omega)} + \|\hat{y}_\ell - y_\ell^{k_\ell}\|_{L_2(\Omega)} \\
&\leq c h_\ell^s \|y_d\|_{H^s(\Omega)} + 2\, \overline{c}_{\rm SM}^{1/2}\, q^{k_\ell} \, \|\hat{y}_\ell - y_\ell^0\|_{L_2(\Omega)},
\end{aligned} \tag{38}$$

where $\hat{y}_\ell = \hat{y}_{\varrho_\ell h_\ell} \in V_\ell$ is the exact state solution of the finite element scheme (18)-(19) corresponding to the solution $\mathbf{y}_\ell \in \mathbb{R}^{n_\ell}$ of the mass-lumped Schur-complement

system (35). Now using $y_\ell^0 = I_{\ell-1}^\ell y_{\ell-1}^{k_{\ell-1}}$, the triangle inequality, Theorem 3, and estimate (37), we can continue to estimate the last term in (38) as follows:

$$\begin{aligned}\|\hat{y}_\ell - y_\ell^0\|_{L_2(\Omega)} &\leq \|\hat{y}_\ell - y_d\|_{L_2(\Omega)} + \|y_d - I_{\ell-1}^\ell y_{\ell-1}^{k_{l-1}}\|_{L_2(\Omega)} \\ &\leq \|\hat{y}_\ell - y_d\|_{L_2(\Omega)} + \|y_d - y_{\ell-1}^{k_{l-1}}\|_{L_2(\Omega)} \\ &\leq ch_\ell^s \|y_d\|_{H^s(\Omega)} + h_{\ell-1}^s c(y_d) \\ &\leq h_\ell^s (c\|y_d\|_{H^s(\Omega)} + 2^s c(y_d)) \end{aligned} \tag{39}$$

Inserting (39) into (38), we get

$$\begin{aligned}\|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)} &\leq h_\ell^s \left[c\|y_d\|_{H^s(\Omega)} + 2\,\overline{c}_{\mathrm{SM}}^{1/2}\, q^{k_\ell}(c\|y_d\|_{H^s(\Omega)} + 2^s c(y_d))\right] \\ &\leq h_\ell^s c(y_d) \end{aligned} \tag{40}$$

provided that $q^{k_\ell}(c\|y_d\|_{H^s(\Omega)} + 2^s c(y_d)) \leq \|y_d\|_{L_2(\Omega)}$. The latter inequality is ensured if we performed not more than

$$k_\ell = k_* \geq \ln(q(y_d)^{-1})/\ln q^{-1} \tag{41}$$

nested iterations, where $q(y_d) = \|y_d\|_{L_2(\Omega)}/((1+2^s)c\|y_d\|_{H^s(\Omega)} + 2^{1+s}\|y_d\|_{L_2(\Omega)}) < 1$. Here we exclude the trivial case that $y_d = 0$. Therefore, we have proved the following nested iteration theorem by induction.

Theorem 5. *If the coarse mesh problem on level $l = 1$ is solved by k_1 PCG iterations with the initial guess $\mathbf{y}_1^0 = \mathbf{0}_1$ such that (38) holds, and if k_* nested PCG iterations are used on all nested levels $\ell = 2, \ldots, L$, i.e. $k_2 = \ldots = k_L = k_*$ defined by (41), then the last iterate $y_L^{k_L} \leftrightarrow \mathbf{y}_L^{k_L}$ on the finest level $\ell = L$ differs from given desired state y_d in the order of the discretization error $O(h_L^s)$ with respect to the $L_2(\Omega)$ norm. More precisely, we get the estimate*

$$\|y_d - y_L^{k_L}\|_{L_2(\Omega)} \leq h_L^s c(y_d). \tag{42}$$

The computation of $y_L^{k_L} \leftrightarrow \mathbf{y}_L^{k_L}$ requires not more than $O(n_L) = O(h_L^{-d})$ arithmetical operations and memory, i.e. the nested iteration procedure proposed is asymptotically optimal.

Proof. The proof of estimate (42) follows from above by induction. The complexity analysis is based on simple use of the geometric series. □

We stop the nested iteration process as soon as we arrive at some desired relative accuracy $\varepsilon \in (0, 1)$ such that

$$\|y_d - y_L^{k_L}\|_{L_2(\Omega)} \leq \varepsilon \|y_d\|_{L_2(\Omega)}. \tag{43}$$

The apriori estimate (42) immediately yields that estimate (43) is guaranteed when $ch_L^s\|y_d\|_{H^s(\Omega)} \leq \varepsilon \|y_d\|_{L_2(\Omega)}$, but in practice we directly check (43) because all quantities are computable.

We will also stop the nested iteration if the cost for the control $u_\ell^{k_\ell}$ becomes too large, where $u_\ell^{k_\ell} \leftrightarrow \mathbf{u}_\ell^{k_\ell}$ is computed from the fe state equation

$$\mathbf{u}_\ell^{k_\ell} = -\varrho_\ell^{-1}\mathbf{p}_\ell^{k_\ell} = D_\ell^{-1} K_\ell \mathbf{y}_\ell^{k_\ell}$$

More precisely, let $c_u > 0$ be a given threshold for the control cost that we are willing to pay. Then we stop the nested iteration if

$$\|u_\ell^{k_\ell}\|_{L_2(\Omega)} = (M_\ell \mathbf{u}_\ell^{k_\ell}, \mathbf{u}_\ell^{k_\ell}) \leq \overline{c}_{\mathrm{MD}}(D_\ell \mathbf{u}_\ell^{k_\ell}, \mathbf{u}_\ell^{k_\ell}) = \overline{c}_{\mathrm{MD}}(K_\ell \mathbf{y}_\ell^{k_\ell}, \mathbf{y}_\ell^{k_\ell}) \leq c_u,$$

but $\|u_{\ell+1}^{k_{\ell+1}}\|_{L_2(\Omega)} > c_u$, where $\bar{c}_{\rm MD} = 1$. Then we set $L = \ell$.

Now we may proceed with *cascadic nested iteration* freezing the cost (regularization) parameter $\varrho_L = h_L^4$ and refining the mesh only , i.e.

$$\varrho_{\ell+1} = \varrho_L = h_L^4 = \text{const. and } h_{\ell+1} = h_\ell/2 \text{ for } \ell = L, \ldots, L+J-1, \qquad (44)$$

in order to improve the approximation of the control. We note that this further mesh refinement will not improve the approximation to the desired state y_d since this error is defined by the frozen cost parameter ϱ_L. Since we only use a few additional levels for the improvement of the control, we can proceed with the PCG preconditioned by D_ℓ as before as nested iteration, but replacing $\bar{c}_{\rm SD}$ by $\bar{c}_{{\rm SD},\ell} = \bar{c}_{\rm SD} 2^{4(\ell-L)}$ for $\ell = L+1, \ldots, L+J$. If we want to add many levels, i.e. $J >> 1$, then we may use some cascadic full multigrid Schur complement iteration using level L as coarse mesh and $y_{L+1}^0 = I_L^{L+1} y_L^{k_L} = y_L^{k_L} \in V_L \subset V_{L+1}$ as initial guess; see [18, 11] for L_2 convergent multigrid methods.

4 Numerical Results

In our numerical experiments, we consider the discontinuous desired state

$$y_d = \begin{cases} 1 & \text{in } (0.25, 0.75)^3, \\ 0 & \text{in } \overline{\Omega} \setminus (0.25, 0.75)^3, \end{cases}$$

in the computational domain $\Omega = (0,1)^3 \subset \mathbb{R}^{d=3}$. This discontinuous desired state y_d does not belong to $Y = H_0^1(\Omega)$, and has a rather low Sobolev regularity. More precisely, $y_d \in H^{1/2-\varepsilon}(\Omega)$ for any $\varepsilon > 0$. This discontinuous target has been utilized in the work [23, 26, 32] in both the cases of L_2 and energy (H^{-1}) regularization for distributed elliptic optimal control problems. So, we can easily compare the numerical results presented below for the mass-lumping discretization of the control term in the reduced optimality system with those of the L_2 regularization without mass lumping and the H^{-1} regularization.

We decompose the domain $\Omega = (0,1)^3$ into uniformly refined tetrahedral elements τ_e, and start with an initial mesh that contains 384 tetrahedral elements and 125 vertices, leading to the mesh size $h = 2^{-2}$. From such a mesh, we make successive refinements on the levels $\ell = 1, ..., 8$. On the finest level $\ell = L = 8$, we have $135,005,697$ vertices, $h = 2^{-9} = 1.9531\text{e}{-3}$, and $\varrho = h^4 = 2^{-36} = 1.4552\text{e}{-11}$. Further, we run tests on the adaptively refined meshes, in which we have employed the standard red-green refinement of tetrahedral elements, and we have chosen the locally varying regularization parameter $\varrho_\tau = h_e^4$ on each tetrahedral element τ_e. The adaptive procedure is simply based on the localization of the error $\|y_d - \tilde{y}_\ell\|_{L_2(\Omega)}$ that is explicitly computable for any known fe approximation $\tilde{y}_\ell$ to the given desired state y_d; see [21] for a detailed description.

As described in Section 3, thanks to the replacement of the mass matrix M_h by its diagonal approximation $D_h = \text{lump}(M_h)$, we can efficiently solve the spd mass-lumped Schur-complement system (27) by means of the PCG preconditioned by D_h. We first use the initial guess $\mathbf{y}_\ell^0 = \mathbf{0}$, and terminate the iteration as soon as the preconditioned residual is reduced by a factor 10^6. The number of PCG iterations (Its)and the computational time (Time) in seconds (s) are provided in Table 1 for both uniform and adaptive refinements.

Therein, we observe the robustness of our proposed preconditioner for (27) with respect to both the mesh size and local adaptivity under the choice of $\varrho_\tau = h_e^4$. We only see slightly more iterations for the adaptive refinements in comparison to the uniform refinements.

ℓ	Adaptive			Uniform		
	#Dofs	error	Its (Time)	#Dofs	error	Its (Time)
1	125	3.26e−1	10 (6.3e−4)	125	3.26e−1	10 (6.4e−4)
2	223	2.35e−1	62 (6.5e−3)	729	2.25e−1	55 (1.8e−2)
3	1,044	1.86e−1	106 (5.5e−2)	4,913	1.59e−1	79 (1.9e−1)
4	4,548	1.32e−1	123 (2.8e−1)	35,937	1.12e−1	85 (1.7e−0)
5	10,524	1.05e−1	116 (6.3e−1)	274,625	7.96e−2	81 (2.3e+1)
6	25,807	8.35e−2	113 (1.6e−0)	2,146,689	5.62e−2	74 (1.8e+2)
7	91,520	6.03e−2	100 (5.7e−0)	16,974,593	3.97e−2	68 (1.4e+3)
8	118,334	5.62e−2	102 (7.7e−0)	135,005,697	2.81e−2	66 (1.3e+4)
9	432,195	4.08e−2	93 (3.5e+1)			
10	473,638	3.97e−2	95 (6.3e+1)			
11	1,843,740	2.84e−2	91 (2.5e+2)			
12	1,937,983	2.79e−2	92 (3.2e+2)			
13	7,681,306	1.99e−2	91 (6.3e+2)			
14	7,922,574	1.96e−2	93 (1.0e+3)			
15	31,496,575	1.39e−2	83 (3.6e+3)			
16	32,000,845	1.38e−2	84 (5.4e+3)			
17	127,607,911	9.84e−3	68 (1.6e+4)			

Table 1: Comparison of the PCG iterations (Its) and computational time (Time) in seconds (indicated in the parentheses) for solving (27) for both adaptive and uniform refinements using the non-nested iterations, where error $= \|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)}$.

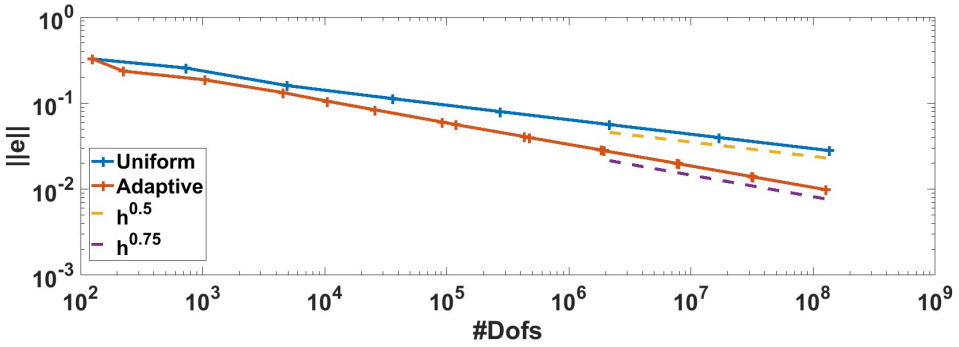

Figure 1: Comparison of the convergence history obtained from non-nested iterations for uniform and adaptive refinements, where $\|e\| = \|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)}$.

As shown in the theoretical part, solving the Schur complement equation with the lumped mass does not deteriorate the convergence of our finite element approximation. This is confirmed in our numerical experiments. The comparison of convergence on both uniform and adaptive refinements is given in Figure 1. We observe the convergence rate $h^{0.5}$ for the uniform refinement as predicted by Theorem3, and a much better convergence rate $h^{0.75}$ for the adaptive refinements; see [21] for the case of variable energy regularization. There one can also find an explanation of the convergence rate that can be achieved via this adaptive procedure.

In order to further reduce the computational cost, we utilize nested PCG iterations as described in Section 3. Here, on the coarsest level $\ell = 1$, we run the PCG iterations until the relative preconditioned residual reaches 10^{-6}. On the refined levels $\ell = 2, 3, ...$, we have utilized an adaptive tolerance

$$\alpha \left[n_\ell / n_{\ell-1}\right]^{-\frac{\beta}{3}}, \ \ell = 2, 3, \ldots, \tag{45}$$

for the relative preconditioned residual, with α being a scaling factor, $\beta = 0.5$ and 0.75 for the uniform and adaptive refinement, respectively, and n_ℓ the number of

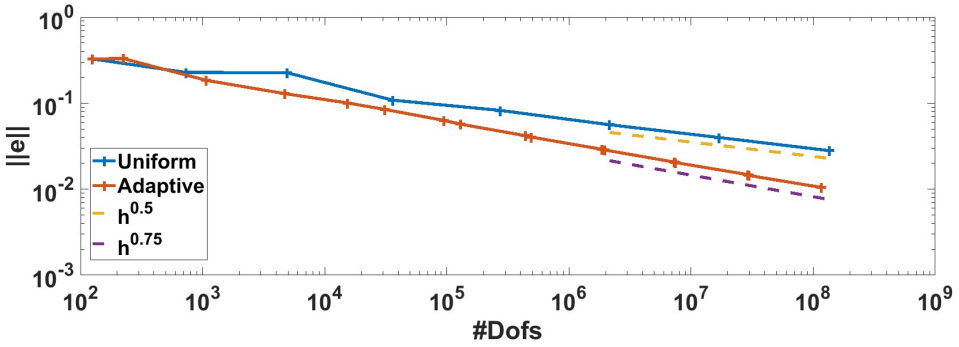

Figure 2: Comparison of the convergence history obtained from nested iterations for uniform and adaptive refinements, where $\|e\| = \|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)}$.

degrees of freedom (#Dofs) on the mesh level $\ell = 1, 2, \ldots$. The solution on the level $\ell - 1$ is used as an initial guess for the PCG iteration on the next finer level ℓ. The reduced number of nested iterations (Its) on both uniform and adaptive refinements is given in Table 2, where we have chosen $\alpha = 0.5$ and $\alpha = 1$ for the adaptive and uniform refinement, respectively. From this, we easily see much fewer iteration numbers and significantly less computational time in seconds in comparison with the case of non-nested iterations as shown in Table 1, without loss of accuracy of the numerical approximations; see Figure 2 for a comparison of convergence history for both uniform and adaptive refinements using the nested iterations.

ℓ	Adaptive			Uniform		
	#Dofs	error	Its (Time)	#Dofs	error	Its (Time)
1	125	3.26e−1	10 (6.3e−4)	125	3.26e−1	10 (6.3e−4)
2	223	3.30e−1	1 (2.6e−4)	729	2.27e−1	8 (2.9e−3)
3	1,067	1.84e−1	19 (1.1e−2)	4,913	2.25e−1	1 (4.6e−3)
4	4,705	1,28e−1	13 (3.3e−2)	35,937	1.08e−1	9 (1.9e−1)
5	15,368	1.00e−1	17 (1.4e−1)	274,625	8.22e−2	8 (1.5e−0)
6	30,996	8.45e−2	17 (4.0e−1)	2,146,689	5.60e−2	9 (1.4e+1)
7	94,176	6.30e−2	19 (1.3e−0)	16,974,593	3.98e−2	9 (2.1e+2)
8	129,760	5.68e−2	18 (1.7e−0)	135,005,697	2.81e−2	9 (2.2e+3)
9	440,572	4.18e−2	17 (1.2e+1)			
10	488,124	4.03e−2	17 (1.3e+1)			
11	1,860,339	2.90e−2	18 (6.1e+1)			
12	1,958,388	2.85e−2	16 (5.9e+1)			
13	7,254,384	2.06e−2	18 (2.6e+2)			
14	7,408,106	2.04e−2	16 (2.1e+2)			
15	29,094,073	1.47e−2	17 (6.9e+2)			
16	29,682,531	1.44e−2	16 (7.6e+2)			
17	116,229,104	1.04e−2	16 (3.7e+3)			

Table 2: Comparison of the PCG iterations (Its) and computational time (Time) in seconds (indicated in the parentheses) for solving (27) for both adaptive ($\alpha = 0.5$, $\beta = 0.75$) and uniform ($\alpha = 1$, $\beta = 0.5$) refinements using the nested iteration approach, where error $= \|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)}$

Another approach to reduce the computational time, especially, in the case of uniform refinement is the parallelization of the PCG solver. The parallelization of the conjugate gradient algorithm is now a standard procedure [13]. The crucial point is always the preconditioner. It is clear that the parallelization of a diagonal preconditioner is much easier than the parallelization of a multigrid preconditioner.

ℓ	#Cores					
	16	32	64	128	256	512
4	85 (4.0e−2)	-	-	-	-	-
5	84 (3.4e−1)	84 (1.6e−1)	84 (6.4e−2)	-	-	-
6	82 (2.9e−0)	82 (1.4e−0)	82 (7.3e−1)	82 (3.7e−1)	82 (1.8e−1)	82 (8.0e−2)
7	80 (2.5e+1)	80 (1.2e+1)	80 (6.3e−0)	80 (3.0e−0)	80 (1.5e−0)	80 (8.3e−1)
8	-	-	77 (5.1e+1)	77 (2.5e+1)	77 (1.3e+1)	77 (6.4e−0)

Table 3: Parallel performance on a distributed computer system for uniform refinement and non-nested iterations.

ℓ	#Cores					
	16	32	64	128	256	512
4	9 (5.8e−3)	-	-	-	-	-
5	9 (5.0e−2)	9 (2.7e−2)	9 (1.1e−2)	-	-	-
6	8 (3.5e−1)	8 (1.8e−1)	8 (9.4e−2)	8 (5.2e−2)	8 (2.8e−2)	8 (1.2e−2)
7	9 (3.1e−0)	9 (1.6e−0)	9 (8.1e−1)	9 (4.2e−1)	9 (2.2e−1)	9 (1.2e−1)
8	-	-	11 (7.9e−0)	11 (4.0e−0)	11 (2.0e−0)	11 (1.0e−0)

Table 4: Parallel performance on a distributed computer system for uniform refinement and nested iterations.

More precisely, the parallelization of a diagonal preconditioner such as D_h is trivial. For parallel performance studies, we have utilized the open source MFEM[1]. We observe from the diagonals of Table 3, e.g. from level 7 with 16 cores to level 8 with 512 cores (always factor 8), almost constant time, i.e. a good weak scaling behavior, whereas the horizontal lines show an almost perfect strong scaling. The latter one is also illustrated in Figure 3 for $\ell = 7$ and $\ell = 8$ corresponding to $16,974,593$ and $135,005,697$ Dofs, respectively. The largest problem with $135,005,697$ Dofs can be solved in 6.4 seconds using 512 cores. Similar scaling behaviors are also observed for the nested iterations approach; see Table 4 and Figure 4. The computational time in seconds (s) using nested iterations is further reduced by a factor of about 7 in comparison with the non-nested iterations. Using 512 cores, the largest problem with $135,005,697$ Dofs is solved in 1 second. Finally, we made some performance tests for the adaptive refinement using the nested iteration setting. The results are given in Table 5. We observe relatively good scaling in this case as well. Here, we have used the non-conforming simplicial complex and load balance from the open source MFEM.

We note that we used different computers and different codes for the single-core and parallel computations. More precisely, we used the shared-memory computer MACH2[2], that provides a big memory, and the distributed-memory computer RADON1[3] for the single-core and parallel computations, respectively.

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

#Dofs	#Cores				
	16	32	64	128	256
2.76154e+6	16 (1.0e−0)	16 (5.3e−1)	16 (2.8e−1)	16 (1.6e−1)	16 (1.0e−1)
1.06728e+7	-	17 (2.3e−0)	17 (1.2e−0)	16 (6.2e−1)	17 (3.3e−1)

Table 5: Parallel performance on a distributed computer system for adaptive refinement and nested iterations.

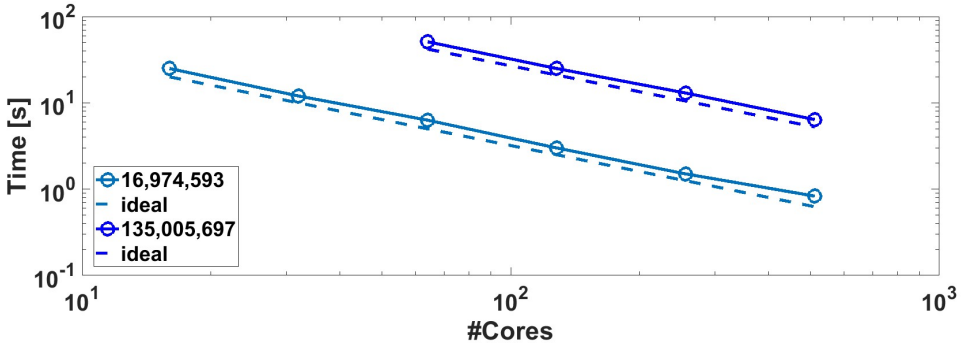

Figure 3: Strong scalability and computational time in seconds (s) with respect to the number of cores for uniform refinement and non-nested iterations

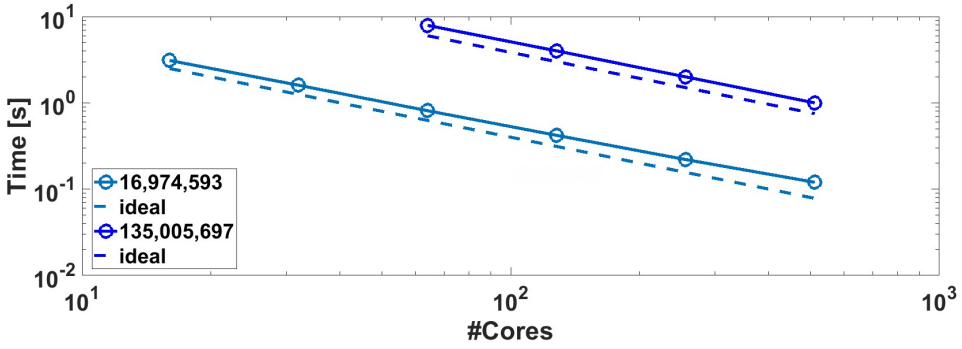

Figure 4: Strong scalability and computational time in seconds (s) with respect to the number of cores for uniform refinement and nested iterations

5 Conclusions and Outlook

We provide a rigorous analysis of the discretization error $\|y_d - \hat{y}_{\varrho h}\|_{L_2(\Omega)}$ when replacing the mass matrix M_h arising from the regularization term in the reduced optimality system by its lumped version $D_h = \mathrm{lump}(M_h)$. It turns out that the asymptotic behavior of the error is not affected by mass lumping when using affine-linear finite elements. More precisely, we again get the upper bound $ch^s\|y_d\|_{H^s(\Omega)}$, $s \in [0,2]$, for the choice $\varrho = h^4$ that provides the optimal balance between the regularization parameter ϱ and the mesh-size h. Moreover, this replacement of M_h by D_h opens the way to reduce the discrete reduced optimal optimality system further to a spd Schur complement problem that can efficiently be solved by PCG since now the matrix-by-vector multiplication is cheap and, surprisingly, D_h is a diagonal preconditioner that is spectrally equivalent to the Schur complement S_h. This PCG can efficiently be parallelized as the numerical results show. These findings provide the perfect ingredients for a nested PCG iteration producing iterates $y_\ell^{k_\ell}$ that differ from the desired state y_d in the order $O(h_\ell^s)$ of the discretization error in asymptotically optimal complexity $O(h_\ell^{-d})$. The nested iteration process will be stopped when some relative accuracy $\varepsilon \in (0,1)$ of the error is reached, or the cost we are willing to pay in terms of the control energy density $\|u_L\|^2_{L_2(\Omega)}$ becomes too large. In this case, we can freeze the regularization (cost) parameter $\varrho_L = h_L^4$, and continue the nested iteration process with mesh refinement only in order to improve to the approximation of the control.

We provide not only numerical results for the case of uniform refinement that nicely demonstrated the theoretical predictions but also for adaptive refinement when using variable regularization. The numerical results show that this adaptive approach works well, but a rigorous numerical analysis is still missing. Further investigation comprises this analysis, and the generalization to larger classes of PDEs like elliptic diffusion-convection-reaction, parabolic and hyperbolic state equations. We refer to [25, 24, 27] and [30] when using space-time fe discretization for parabolic and hyperbolic initial-boundary value problems, respectively. Another future research topic are the consideration of control and state (box) constraints in the framework discussed here; see [17] for first results. Finally, we mention that singular-perturbed problems as discussed here also appear in fluid mechanics where they are known as (discrete) differential filter that provide approximate deconvolution models of turbulence [9, 28, 20].

References

[1] O. Axelsson and J. Karátson. Superior properties of the PRESB preconditioner for operators on two-by-two block form with square blocks. *Numer. Math*, 146(2):335–368, 2020.

[2] O. Axelsson, M. Neytcheva, and A. Ström. An efficient preconditioning method for state box-constrained optimal control problems. *J. Numer. Math.*, 26(4):185–207, 2018.

[3] Z.-Z. Bai. Regularized HSS iteration methods for stabilized saddle-point problems. *IMA J. Numer. Anal.*, 39(4):1888–1923, 2019.

[4] Z.-Z. Bai and M. Benzi. Regularized HSS iteration methods for saddle-point linear systems. *BIT Numer. Math.*, 57(2):287–311, 2017.

[5] Z.-Z. Bai, M. Benzi, F. Chen, and Z.-Q. Wang. Preconditioned MHSS iteration methods for a class of block two-by-two linear systems with applications to distributed control problems. *IMA J. Numer. Anal.*, 33(1):343–369, 2013.

[6] Z.-Z. Bai and J.-Y. Pan. *Matrix Analysis and Computations*. SIAM, 2021.

[7] R. Becker and P. Hansbo. A simple pressure stabilization method for the Stokes equation. *Comm. Numer. Methods Engrg.*, 24(11):1421–1430, 2008.

[8] M. Benzi, G. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.

[9] L. Berselli, T. Iliescu, and W. Layton. *Mathematics of large eddy simulation of turbulent flows*. Scientific Computation. Springer-Verlag, Berlin, 2006.

[10] A. Borzi and V. Schulz. Multigrid methods for PDE optimization. *SIAM Review*, 51(2):361–395, 2009.

[11] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge, 2007.

[12] S. Chen and J. Zhao. Estimations of the constants in inverse inequalities for finite element functions. *J. Comput. Math.*, 31(5):522–531, 2013.

[13] C. Douglas, G. Haase, and U. Langer. *A Tutorial on Elliptic PDE Solvers and Their Parallelization*. Software, Environments, and Tools,. SIAM, Philadelphia, 2003.

[14] I. Dravins and M. Neytcheva. *On the Numerical Solution of State- and Control-constrained Optimal Control Problems*. Department of Information Technology, Uppsala Universitet, 2021.

[15] H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite elements and fast iterative solvers: With applications in incompressible fluid dynamics*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2005.

[16] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. Springer-Verlag, New York, 2004.

[17] P. Gangl, R. Löscher, and O. Steinbach. Regularization and finite element error estimates for distributed optimal control problems with energy regularization and state or control constraints. Technical report, TU Graz, 2023. in preparation.

[18] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer Verlag, 1985.

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

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

[21] U. Langer, R. Löscher, O. Steinbach, and H. Yang. An adaptive finite element method for distributed elliptic optimal control problems with variable energy regularization. Technical Report arXiv:2209.08811, arXiv.org, 2022.

[22] U. Langer, R. Löscher, O. Steinbach, and H. Yang. Robust iterative solvers for algebraic systems arising from elliptic optimal control problems. Berichte aus dem Institut für Angewandte Mathematik 2023/2, Technische Universität Graz, Institut für Angewandte Mathematik, February 2023. submitted to LSSC 2023 Proceedings.

[23] U. Langer, R. Löscher, O. Steinbach, and H. Yang. Robust finite element discretization and solvers for distributed elliptic optimal control problems. *Comput. Meth. Appl. Math.*, 2023.

[24] U. Langer, O. Steinbach, F. Tröltzsch, and H. Yang. Space-time finite element discretization of parabolic optimal control problems with energy regularization. *SIAM J. Numer. Anal.*, 59:675–695, 2021.

[25] U. Langer, O. Steinbach, F. Tröltzsch, and H. Yang. Unstructured space-time finite element methods for optimal control of parabolic equations. *SIAM J. Sci. Comput.*, 43:A744–A771, 2021.

[26] 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.

[27] 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.

[28] W. Layton and L. Rebholz. *Approximate deconvolution models of turbulence: Analysis, phenomenology and numerical analysis*, volume 2042 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2012.

[29] J. L. Lions. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod Gauthier-Villars, Paris, 1968.

[30] R. Löscher and O. Steinbach. Space-time finite element methods for distributed optimal control of the wave equation. Technical Report arXiv:2211.02562, arXiv.org, 2022.

[31] K.-A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, 18(1):1–40, 2011.

[32] M. Neumüller and O. Steinbach. Regularization error estimates for distributed control problems in energy spaces. *Math. Methods Appl. Sci.*, 44(5):4176–4191, 2021.

[33] Y. Notay. Convergence of some iterative methods for symmetric saddle point linear systems. *SIAM J. Matrix Anal. Appl.*, 40(1):122–146, 2019.

[34] J. Pearson, M. Stoll, and A. Wathen. Preconditioners for state-constrained optimal control problems with moreau-yosida penalty function. *Numer. Linear Algebra Appl.*, 21(1):81–97, 2014.

[35] J. Pearson and A. Wathen. A new approximation of the Schur complement in preconditioners for PDE-constrained optimization. *Numer. Linear Algebra Appl.*, 12(5):816–829, 2012.

[36] A. Schiela and S. Ulbrich. Operator preconditioning for a class of inequality constrained optimal control problems. *SIAM J. Optim.*, 24(1):435–466, 2014.

[37] J. Schöberl and W. Zulehner. Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems. *SIAM J. Matrix Anal. Appl.*, 29:752–773, 2007.

[38] V. Schulz and G. Wittum. Transforming smoothers for pde constrained optimization problems. *Comput. Visual. Sci.*, 11:207–219, 2008.

[39] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements*. Springer, New York, 2008.

[40] M. Stoll and A. Wathen. Preconditioning for partial differential equation constrained optimization with control constraints. *Numer. Linear Algebra Appl.*, 19:53–71, 2012.

[41] F. Tröltzsch. *Optimal control of partial differential equations: Theory, methods and applications*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2010.

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

[43] W. Zulehner. Nonstandard norms and robust estimates for saddle point problems. *SIAM J. Matrix Anal. Appl.*, 32(2):536–560, 2011.