Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Robust preconditioning and error estimates for optimal control of the convection-diffusion-reaction equation with limited observation in Isogeometric Analysis

K.-A. Mardal, J. Sogn, S. Takacs

RICAM-Report 2020-46

ROBUST PRECONDITIONING AND ERROR ESTIMATES FOR OPTIMAL CONTROL OF THE CONVECTION-DIFFUSION-REACTION EQUATION WITH LIMITED OBSERVATION IN ISOGEOMETRIC ANALYSIS[*]

KENT-ANDRE MARDAL[†], JARLE SOGN[‡], AND STEFAN TAKACS[§]

Abstract. In this paper we analyze an optimization problem with limited observation governed by a convection–diffusion–reaction equation. Motivated by a Schur complement approach, we arrive at continuous norms that enable analysis of well-posedness and subsequent derivation of error analysis and a preconditioner that is robust with respect to the parameters of the problem. We provide conditions for inf-sup stable discretizations and present one such discretization for box domains with constant convection. We also provide a priori error estimates for this discretization. The preconditioner requires a fourth order problem to be solved. For this reason, we use Isogeometric Analysis as a method of discretization. To efficiently realize the preconditioner, we consider geometric multigrid with a standard Gauss-Seidel smoother as well as a new macro Gauss-Seidel smoother. The latter smoother provides good results with respect to both the geometry mapping and the polynomial degree.

Key words. PDE-constrained optimization, optimal control, robust preconditioning, error estimates

AMS subject classifications. 49K20, 65F08, 65N22, 65N15

1. Introduction.

In this paper, we consider an optimal control problem involving a linear Convection–Diffusion–Reaction (CDR) problem, which reads as follows:

$$\text{Minimize} \quad J(u,q) := \frac{1}{2}\|u-u_d\|^2_{L^2(\mathcal{O})} + \frac{\alpha}{2}\|q\|^2_{L^2(\Omega)} \quad \text{for} \quad u \in U, q \in L^2(\Omega) \tag{1.1}$$

subject to

$$\begin{aligned} -\varepsilon\Delta u + \beta\cdot\nabla u + \sigma u &= f - q && \text{in} \quad \Omega,\\ u &= 0 && \text{on} \quad \partial\Omega. \end{aligned} \tag{1.2}$$

Here and in what follows, Ω is a bounded open subset of $\mathbb{R}^d$ $(d = 1, 2, 3)$ with Lipschitz boundary, $f \in L^2(\Omega)$, $\varepsilon, \sigma \in \mathbb{R}$ with $\varepsilon > 0, \sigma \geq 0$, $\beta \in L^\infty(\Omega)^d$ with $\nabla\cdot\beta = 0$ and $\mathcal{O} \subseteq \Omega$ is measurable in $\mathbb{R}^d$. For certain choices of the parameters, like $\varepsilon \ll \beta, \sigma$, convection–diffusion–reaction problems are singular perturbation problems exhibiting sharp gradients and a potential for loss of regularity. To overcome the problems associated to the loss of regularity, significant effort has been put in the development of methods with low regularity, such as discontinuous Galerkin methods [2, 9]. We take the opposite approach and investigate to what extent higher regularity may be used in the setting of optimal control problems.

[*] **Funding:** The first author acknowledges support from the Research Council of Norway, grant 300305 and 301013. The second and the third author are supported by the Austrian Science Fund (FWF): P31048.

[†] Department of Mathematics, University of Oslo, Oslo, Norway. Center for Biomedical Computing, Simula Research Laboratory, Lysaker, Norway (kent-and@simula.no).

[‡] Johann Radon Institute for Computational Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria (jarle.sogn@ricam.oeaw.ac.at).

[§] Johann Radon Institute for Computational Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria (stefan.takacs@ricam.oeaw.ac.at).

There are two main problems with (1.1)–(1.2), namely: 1) potential sharp gradients leading to non-physical oscillations in the numerical solution and 2) ill-posedness due to limited observations, this is, when $\mathcal{O}$ is a subset of Ω. Motivated by the fact that higher regularity has been exploited in the cases with limited observations [17, 25, 4], we derive order optimal preconditioners via stability analysis in non-standard Sobolev spaces.

When solving the CDR problem, it is common to consider some stabilization method (e.g. the streamline upwind Petrov Galerkin (SUPG) method) or adaptive grids (e.g. Shishkin grids) to reduce the oscillatory behavior, cf. [10]. This is also the case in optimal control settings, see, e.g., [21, 3, 13, 8]. We do not use any such stabilization techniques, but we remark that the trial and test functions involved in the state equation differ. That means, the state equation, if considered isolated, is discretized by a Petrov-Galerkin method, although the complete optimality system is discretized by a standard Galerkin method, this is, trial and test functions agree. In particular, in the continuous setting, the trial space and test spaces are $H^2(\Omega)\cap H^1_0(\Omega)$ and $L^2(\Omega)$, respectively, with properly weighted norms.

By considering a Schur complement of the optimal control problem, we derive non-standard norms in which well-posedness is obtained, assuming extra regularity. From the well-posedness of the continuous system we subsequently analyse corresponding discrete systems to arrive at both error estimates and preconditioners that are robust with respect to the problem parameters $\alpha, \varepsilon, \beta$ and σ. In detail, we provide a condition for the discretization which ensures that the preconditioner is sparse and that the preconditioned system is stable. Further, we give an example of such a discretization based on Isogeometric Analysis (IgA) [14, 5]. For our approach, IgA provides useful discretization methods since the extra regularity leads $H^2(\Omega)$–conforming approximation spaces. Using these discretization methods, a priori error estimates are derived, where we detail the dependencies of the problem parameters. We remark that the error estimates required extending some approximation error estimates for tensor-product B-splines, which is done in Appendix B.

Similar Schur complement preconditioners were used, on the linear algebra level, in [19, 20] for optimal control problems of the CDR equation. [20] also considers mixed constraints. The preconditioners perform well for different values of the problem parameters. The Schur complement preconditioners were replaced with approximations based on the factorization approach by [19]. However, this approach does not work well for problems with limited observation, i.e., when $\mathcal{O} \subsetneq \Omega$.

To solve the resulting linear system we use preconditioned Krylov subspace methods. We consider two approaches to realize our preconditioner: sparse direct methods and multigrid methods. For mid-sized problems, sparse direct solvers work well since each component of the preconditioner is symmetric and positive definite. For large-sized problems, we use a multigrid method to realize the fourth-order operator. Combining the results from this paper and from [23, 24], it follows that the multigrid method we consider is robust in the grid-size, however, not necessarily in any of the other problem parameters. Finding a multigrid method which is robust in the grid-size, the chosen spline degrees, $\alpha, \varepsilon, \beta, \sigma$ and $\mathcal{O}$, remains an open problem.

The outline of the paper is as follows: in the next section we perform the analysis of the continuous problem. In Section 3, we analyse the discrete problem and provide a condition for a stable discretization. IgA is then introduced in Section 4 along with the proposed discretization. In Section 5, a priori error estimates are derived. Section 6 contains a discussion of the solution of an one-dimensional problem and in Section 7 we perform numerical experiments on two-dimensional problems and see how the preconditioner behaves.

2. Analysis of the continuous problem. To obtain a standard (weak) variational formulation of the state equation (1.2), one would choose the state variable u and test function $\tilde{w}$ to be in $H_0^1(\Omega)$. Instead, we consider the strong variational formulation, this is, find $u \in H^2(\Omega) \cap H_0^1(\Omega)$ such that

$$(q, \tilde{w})_{L^2(\Omega)} + (-\varepsilon \Delta u + \beta \cdot \nabla u + \sigma u, \tilde{w})_{L^2(\Omega)} = (f, \tilde{w})_{L^2(\Omega)} \qquad \forall\, \tilde{w} \in L^2(\Omega).$$

The Lagrangian functional associated to (1.1)–(1.2) is

$$\begin{aligned}\mathcal{L}(q, w, u) := &\frac{1}{2}\|u - u_d\|^2_{L^2(\mathcal{O})} + \frac{\alpha}{2}\|q\|^2_{L^2(\Omega)} \\ &+ (q, w)_{L^2(\Omega)} + (-\varepsilon \Delta u + \beta \cdot \nabla u + \sigma u, w)_{L^2(\Omega)} - (f, w)_{L^2(\Omega)},\end{aligned}$$

where $u \in H^2(\Omega) \cap H_0^1(\Omega)$, $q \in L^2(\Omega)$ and the Lagrangian multiplier $w \in L^2(\Omega)$. From the first order necessary optimality conditions

$$\frac{\partial \mathcal{L}}{\partial q}(q, w, u) = 0, \quad \frac{\partial \mathcal{L}}{\partial w}(q, w, u) = 0, \quad \frac{\partial \mathcal{L}}{\partial u}(q, w, u) = 0,$$

which are also sufficient here, we obtain the optimality system:

PROBLEM 2.1. *Find $(q, w, u) \in L^2(\Omega) \times L^2(\Omega) \times H^2(\Omega) \cap H_0^1(\Omega)$ such that*

$$\begin{aligned} \alpha(q, \tilde{q})_{L^2(\Omega)} + (w, \tilde{q})_{L^2(\Omega)} &= 0 && \forall\, \tilde{q} \in L^2(\Omega), \\ (q, \tilde{w})_{L^2(\Omega)} + (-\varepsilon \Delta u + \beta \cdot \nabla u + \sigma u, \tilde{w})_{L^2(\Omega)} &= (f, \tilde{w})_{L^2(\Omega)} && \forall\, \tilde{w} \in L^2(\Omega), \\ (w, -\varepsilon \Delta \tilde{u} + \beta \cdot \nabla \tilde{u} + \sigma \tilde{u})_{L^2(\Omega)} + (u, \tilde{u})_{L^2(\mathcal{O})} &= (u_d, \tilde{u})_{L^2(\mathcal{O})} && \forall\, \tilde{u} \in H^2(\Omega) \cap H_0^1(\Omega). \end{aligned}$$

Problem 2.1 can be written as

$$\mathcal{A} \begin{pmatrix} q \\ w \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ Mf \\ \tilde{M}_{\mathcal{O}} u_d \end{pmatrix} \quad \text{with} \quad \mathcal{A} := \begin{pmatrix} \alpha M & M & 0 \\ M & 0 & K \\ 0 & K' & M_{\mathcal{O}} \end{pmatrix}. \tag{2.1}$$

Here, $M : L^2(\Omega) \to (L^2(\Omega))'$ represents the $L^2(\Omega)$-inner product, that is, we have

$$\langle Mq, w \rangle = (q, w)_{L^2(\Omega)},$$

where $\langle \cdot, \cdot \rangle$ denotes the duality product. The notation " $'$" is used to denote both dual spaces and dual operators. $K : H^2(\Omega) \cap H_0^1(\Omega) \to (L^2(\Omega))'$ is the state operator:

$$\langle Ku, w \rangle = (-\varepsilon \Delta u + \beta \cdot \nabla u + \sigma u, w)_{L^2(\Omega)}.$$

Finally, $M_{\mathcal{O}} : H^2(\Omega) \cap H_0^1(\Omega) \to (H^2(\Omega) \cap H_0^1(\Omega))'$ and $\tilde{M}_{\mathcal{O}} : L^2(\mathcal{O}) \to (H^2(\Omega) \cap H_0^1(\Omega))'$, and both represent the $L^2(\mathcal{O})$-inner product on the subdomain $\mathcal{O}$.

We observe that the block operator $\mathcal{A}$ has a block tridiagonal form. Such tridiagonal operators are studied in [25, 4]. We use the Schur complement preconditioner proposed in [25]:

$$\mathcal{S}(\mathcal{A}) := \begin{pmatrix} S_q & 0 & 0 \\ 0 & S_w & 0 \\ 0 & 0 & S_u \end{pmatrix}, \tag{2.2}$$

where the components are

$$S_q := \alpha M, \quad S_w := \frac{1}{\alpha} M, \quad S_u := M_{\mathcal{O}} + \alpha K' M^{-1} K. \tag{2.3}$$

These Schur complements define weighted norms as follows:

$$\begin{aligned} \|q\|_{S_q}^2 &:= \langle S_q q, q\rangle = \alpha \|q\|_{L^2(\Omega)}^2, \\ \|w\|_{S_w}^2 &:= \langle S_w w, w\rangle = \frac{1}{\alpha}\|w\|_{L^2(\Omega)}^2, \\ \|u\|_{S_u}^2 &:= \langle S_u u, u\rangle = \|u\|_{L^2(\mathcal{O})}^2 + \alpha\| - \varepsilon\Delta u + \beta\cdot\nabla u + \sigma u\|_{L^2(\Omega)}^2. \end{aligned} \tag{2.4}$$

The last norm follows from

$$\begin{aligned} \langle K' M^{-1} K u, u\rangle &= \sup_{0\neq w\in L^2(\Omega)} \frac{\langle Ku, w\rangle^2}{\langle Mw, w\rangle} = \sup_{0\neq w\in L^2(\Omega)} \frac{(-\varepsilon\Delta u + \beta\cdot\nabla u + \sigma u, w)_{L^2(\Omega)}^2}{\|w\|_{L^2(\Omega)}^2} \\ &= \frac{\| - \varepsilon\Delta u + \beta\cdot\nabla u + \sigma u\|_{L^2(\Omega)}^4}{\| - \varepsilon\Delta u + \beta\cdot\nabla u + \sigma u\|_{L^2(\Omega)}^2} = \| - \varepsilon\Delta u + \beta\cdot\nabla u + \sigma u\|_{L^2(\Omega)}^2. \end{aligned}$$

We show well-posedness with respect to the norms (2.4) by showing that the operator

$$\mathcal{A} : L^2(\Omega)\times L^2(\Omega)\times H^2(\Omega)\cap H_0^1(\Omega) \to L^2(\Omega)'\times L^2(\Omega)'\times (H^2(\Omega)\cap H_0^1(\Omega))' \tag{2.5}$$

is an isomorphism with respect to the norms (2.4). This is done by using the main result in [25], which for our problem reads as follows.

Theorem 2.2. *Assume the Schur complements in (2.3) are well defined and positive definite, that is,*

$$\langle S_q q, q\rangle \geq \sigma_q \|q\|_{L^2(\Omega)}^2, \quad \langle S_w w, w\rangle \geq \sigma_w \|w\|_{L^2(\Omega)}^2, \quad \langle S_u u, u\rangle \geq \sigma_u \|u\|_{H^2(\Omega)}^2 \tag{2.6}$$

for some positive constants σ_q, σ_w and σ_u, which can depend on the given parameters. Then, $\mathcal{A}$ in (2.5) is an isomorphism, moreover, the condition number of the preconditioned operator $\mathcal{S}^{-1}\mathcal{A}$ is bounded:

$$\kappa\left(\mathcal{S}^{-1}\mathcal{A}\right) \leq \frac{\cos(\pi/7)}{\sin(\pi/14)} \approx 4.05. \tag{2.7}$$

The conditions in (2.6) ensure that the spaces $L^2(\Omega)$, $L^2(\Omega)$ and $H^2(\Omega)\cap H_0^1(\Omega)$, equipped with norms $\|\cdot\|_{S_q}$, $\|\cdot\|_{S_w}$ and $\|\cdot\|_{S_u}$, are complete. Before proving Condition (2.6), we provide a useful lemma which bounds the H^2-norm. The proof of this lemma is presented in Appendix A.

Lemma 2.3. *If the domain Ω has a Lipschitz boundary and*

- *the boundary is a polygon (polyhedron) or*
- *the domain is the image of a geometric mapping $\mathbf{G} : \widehat{\Omega} := (0,1)^d \to \Omega$, where both $\|\nabla^r \mathbf{G}\|_{L^\infty(\widehat{\Omega})}$ and $\|(\nabla^r \mathbf{G})^{-1}\|_{L^\infty(\widehat{\Omega})}$ are bounded for $r\in\{1,2,3\}$,*

then the H^2-norm is bounded by the L^2-norm of the Laplacian, i.e.,

$$\|u\|_{H^2(\Omega)} \leq C_\Omega \|\Delta u\|_{L^2(\Omega)} \quad \forall\, u\in H^2(\Omega)\cap H_0^1(\Omega), \tag{2.8}$$

for a constant C_Ω depending only on Ω.

THEOREM 2.4. *If Ω is a domain such that the conditions of Lemma 2.3 hold, then the assumptions of Theorem 2.2 are satisfied for Problem 2.1.*

Proof. For simplicity, we prove this lemma only for $\sigma = 0$. An extension to the case $\sigma > 0$ is straight-forward.

The first two conditions are trivial since $\langle S_q q, q\rangle = \alpha \|q\|^2_{L^2(\Omega)}$ and $\langle S_w w, w\rangle = \frac{1}{\alpha}\|w\|^2_{L^2(\Omega)}$. For the third condition, let

$$\delta := \frac{\|\beta\|}{\frac{\varepsilon}{c_P} + \|\beta\|},$$

where $\|\beta\| = \|\beta\|_{L^\infty(\Omega)}$ and c_P is constant from the Poincaré inequality, that is, we have

$$\|u\|_{L^2(\Omega)} \leq c_p \|\nabla u\|_{L^2(\Omega)}.$$

Note that $\delta \in [0, 1)$. Let $u \in H^2(\Omega) \cap H^1_0(\Omega)$ be arbitrary but fixed. We consider the two cases:

$$\|\beta\|\|\nabla u\|_{L^2(\Omega)} < \delta\varepsilon\|\Delta u\|_{L^2(\Omega)} \quad \text{and} \quad \|\beta\|\|\nabla u\|_{L^2(\Omega)} \geq \delta\varepsilon\|\Delta u\|_{L^2(\Omega)}. \tag{2.9}$$

First case. From the definition, we have

$$\left\langle K' M^{-1} K u, u \right\rangle = \sup_{0 \neq w \in L^2(\Omega)} \frac{(\beta \cdot \nabla u - \varepsilon \Delta u, w)^2_{L^2(\Omega)}}{\|w\|^2_{L^2(\Omega)}}.$$

By setting $w = -\varepsilon\Delta u$, we get using the Cauchy–Schwarz inequality that

$$\begin{aligned}\left\langle K' M^{-1} K u, u \right\rangle &\geq \frac{((\beta \cdot \nabla u, -\varepsilon \Delta u)_{L^2(\Omega)} + \varepsilon^2 \|\Delta u\|^2_{L^2(\Omega)})^2}{\varepsilon^2 \|\Delta u\|^2_{L^2(\Omega)}} \\ &\geq \frac{(-\|\beta\|\|\nabla u\|_{L^2(\Omega)}\varepsilon\|\Delta u\|_{L^2(\Omega)} + \varepsilon^2\|\Delta u\|^2_{L^2(\Omega)})^2}{\varepsilon^2 \|\Delta u\|^2_{L^2(\Omega)}} \\ &= (-\|\beta\|\|\nabla u\|_{L^2(\Omega)} + \varepsilon\|\Delta u\|_{L^2(\Omega)})^2.\end{aligned}$$

Using the first inequality in (2.9), we obtain

$$\begin{aligned}\left\langle K' M^{-1} K u, u \right\rangle \geq (-\|\beta\|\|\nabla u\|_{L^2(\Omega)} + \varepsilon\|\Delta u\|_{L^2(\Omega)})^2 &\geq ((1-\delta)\varepsilon\|\Delta u\|_{L^2(\Omega)})^2 \\ &= \frac{\varepsilon^4}{(\varepsilon + c_P\|\beta\|)^2}\|\Delta u\|^2_{L^2(\Omega)}.\end{aligned}$$

Second case. By setting $w = u$, we get

$$\left\langle K' M^{-1} K u, u \right\rangle \geq \frac{((\beta \cdot \nabla u, u)_{L^2(\Omega)} + \varepsilon\|\nabla u\|^2_{L^2(\Omega)})^2}{\|u\|^2_{L^2(\Omega)}} = \frac{\varepsilon^2\|\nabla u\|^4_{L^2(\Omega)}}{\|u\|^2_{L^2(\Omega)}}$$

using integration by parts. Due to the homogeneous Dirichlet boundary conditions and $\nabla \cdot \beta = 0$, the term $(\beta \cdot \nabla u, w)_{L^2(\Omega)}$ is skew symmetric and vanishes for $w = u$. Finally, we use $\|u\|_{L^2(\Omega)} \leq c_p\|\nabla u\|_{L^2(\Omega)}$ and the second inequality in (2.9), which gives

$$\begin{aligned}\left\langle K' M^{-1} K u, u \right\rangle &\geq \frac{\varepsilon^2\|\nabla u\|^4_{L^2(\Omega)}}{\|u\|^2_{L^2(\Omega)}} \geq \frac{\varepsilon^2}{c_P^2}\|\nabla u\|^2_{L^2(\Omega)} \\ &\geq \frac{\delta^2\varepsilon^4}{\|\beta\|^2 c_P^2}\|\Delta u\|^2_{L^2(\Omega)} = \frac{\varepsilon^4}{(\varepsilon + c_P\|\beta\|)^2}\|\Delta u\|^2_{L^2(\Omega)}.\end{aligned}$$

To summarize, in both cases we get

$$\begin{aligned}\langle S_u u, u\rangle &= \|u\|^2_{L^2(\mathcal{O})} + \alpha \left\langle K' M^{-1} K u, u\right\rangle \geq \alpha \left\langle K' M^{-1} K u, u\right\rangle \\ &\geq \alpha \frac{\varepsilon^4}{(\varepsilon + c_P \|\beta\|)^2} \|\Delta u\|^2_{L^2(\Omega)} \geq \alpha \frac{C_\Omega \varepsilon^4}{(\varepsilon + c_P \|\beta\|)^2} \|u\|^2_{H^2(\Omega)}.\end{aligned}$$

The last inequality follows from Lemma 2.3. □

Theorem 2.2 and Theorem 2.4 show that Problem 2.1 is well-posed with respect to the norms in (2.4). The boundedness and coercivity constants are bounded independent from the regularization parameter α as well as the problem parameters ε, β and σ. Consequently, the operator preconditioner (2.2) is a robust preconditioner for the optimality system, that is, the condition number is uniformly bounded independently of the above mentioned parameters. So far, we have only analyzed the problem on the continuous level. In the next section, we carry this analysis over to the discrete case and provide a computationally feasible preconditioner.

3. Analysis of the discrete problem. We consider conforming discretizations, that is, we choose the finite-dimensional spaces Q_h and U_h such that they satisfy

$$Q_h \subset L^2(\Omega) \quad \text{and} \quad U_h \subset H^2(\Omega) \cap H^1_0(\Omega).$$

Applying Galerkins principle to (2.1) leads to the discrete variational problem for the functions $(q_h, w_h, u_h) \in Q_h \times Q_h \times U_h$, which we immediately write in matrix-vector notation. We denote the vector representation of functions in these spaces by underlined versions of the corresponding symbols, i.e., for $q_h \in Q_h$ the corresponding coefficient vector is $\underline{q}_h \in \mathbb{R}^{\dim Q_h}$. With a slight abuse of notation, we use the same notation also for the right-hand-side vectors $\underline{f}_h$ and $\underline{u}_{d,h}$, which are obtained by testing the corresponding linear functionals with the basis functions in Q_h and U_h, respectively, see, e.g., [18, Section 6] for further details. Furthermore, operators with subscript h denote matrix representations of the operators.

Using this notation, the discrete problem reads as follows.

Problem 3.1. *Find* $(\underline{q}_h, \underline{w}_h, \underline{u}_h) \in \mathbb{R}^{\dim Q_h} \times \mathbb{R}^{\dim Q_h} \times \mathbb{R}^{\dim U_h}$ *such that*

$$\mathcal{A}_h \begin{pmatrix} \underline{q}_h \\ \underline{w}_h \\ \underline{u}_h \end{pmatrix} = \begin{pmatrix} 0 \\ \underline{f}_h \\ \underline{u}_{d,h} \end{pmatrix} \quad \textit{with} \quad \mathcal{A}_h := \begin{pmatrix} \alpha M_h & M_h & 0 \\ M_h & 0 & K_h \\ 0 & K_h^T & M_{\mathcal{O},h} \end{pmatrix}. \tag{3.1}$$

The exact Schur complement preconditioner (2.2) of the discretized system is

$$\mathcal{S}(\mathcal{A}_h) = \begin{pmatrix} \alpha M_h & 0 & 0 \\ 0 & \frac{1}{\alpha} M_h & 0 \\ 0 & 0 & M_{\mathcal{O},h} + \alpha K_h^T M_h^{-1} K_h \end{pmatrix}. \tag{3.2}$$

Under the mild condition $U_h \subseteq Q_h$ this preconditioner is symmetric positive definite. This is a straight forward extension of [25, Lemma 4.4] by using the fact that $(\beta \cdot \nabla u_h, u_h)_{L^2(\Omega)} = 0$. In this case, Theorem 2.2 yields the following condition number bound:

$$\kappa\left((\mathcal{S}(\mathcal{A}_h))^{-1} \mathcal{A}_h\right) \leq \frac{\cos(\pi/7)}{\sin(\pi/14)} \approx 4.05\,.$$

This preconditioner cannot be efficiently realized since the matrix M_h^{-1} is dense. So, we use the following preconditioner motivated by the norms (2.4) on the discretization spaces instead:

$$\mathcal{S}_h := \begin{pmatrix} \alpha M_h & 0 & 0 \\ 0 & \frac{1}{\alpha} M_h & 0 \\ 0 & 0 & M_{\mathcal{O},h} + \alpha B_h \end{pmatrix}, \tag{3.3}$$

where B_h is the matrix representation of the linear operator $B : H^2(\Omega) \cap H_0^1(\Omega) \to (H^2(\Omega) \cap H_0^1(\Omega))'$,

$$\langle Bu, \tilde{u} \rangle = (-\varepsilon\Delta u + \beta \cdot \nabla u + \sigma u, -\varepsilon\Delta\tilde{u} + \beta \cdot \nabla \tilde{u} + \sigma \tilde{u})_{L^2(\Omega)} \tag{3.4}$$

on U_h. On the continuous level, the operators B and $K'M^{-1}K$ coincide. In general, this does not carry over to the discrete case. The following lemma gives sufficient conditions that guarantee that B_h and $K_h^T M_h^{-1} K_h$ coincide.

Lemma 3.2. *If*

$$(-\varepsilon\Delta + \beta \cdot \nabla + \sigma) U_h \subset Q_h, \tag{3.5}$$

then $K_h^T M_h^{-1} K_h = B_h$ *and thus* $\mathcal{S}(\mathcal{A}_h) = \mathcal{S}_h$.

Proof. Let $u_h \in U_h$ be arbitrary but fixed with coefficient vector $\underline{u}_h$. The definitions yield

$$\begin{aligned} \left\langle K_h^T M_h^{-1} K_h \underline{u}_h, \underline{u}_h \right\rangle &= \sup_{\underline{w}_h \in \mathbb{R}^{\dim Q_h}} \frac{\langle K_h \underline{u}_h, \underline{w}_h \rangle^2}{\langle M_h \underline{w}_h, \underline{w}_h \rangle} \\ &= \sup_{w_h \in Q_h} \frac{(-\varepsilon\Delta u_h + \beta \cdot \nabla u_h + \sigma u_h, w_h)^2_{L^2(\Omega)}}{\|w_h\|^2_{L^2(\Omega)}}. \end{aligned}$$

Since $(-\varepsilon\Delta + \beta \cdot \nabla + \sigma) U_h \subset Q_h$, the supremum is attained for $w_h = -\varepsilon\Delta u_h + \beta \cdot \nabla u_h + \sigma u_h$, and we have

$$\begin{aligned} \left\langle K_h^T M_h^{-1} K_h \underline{u}_h, \underline{u}_h \right\rangle &= \sup_{w_h \in Q_h} \frac{(-\varepsilon\Delta u_h + \beta \cdot \nabla u_h + \sigma u_h, w_h)^2_{L^2(\Omega)}}{\|w_h\|^2_{L^2(\Omega)}} \\ &= \| - \varepsilon\Delta u_h + \beta \cdot \nabla u_h + \sigma u_h\|^2_{L^2(\Omega)} = \langle B_h \underline{u}_h, \underline{u}_h \rangle . \end{aligned}$$

Therefore, $K_h^T M_h^{-1} K_h = B_h$ and thus $\mathcal{S}(\mathcal{A}_h) = \mathcal{S}_h$. □

4. Isogeometric analysis.

Due to requirement $U_h \subset H^2(\Omega) \cap H_0^1(\Omega)$, we need a smooth discretization space. We achieve this by using IgA. We give a brief introduction to the approximation spaces in use. Let $S_{p,k,\ell}(0,1)$ be the space of B-spline functions on the unit interval $(0,1)$ which are k-times continuously differentiable and piecewise polynomials of degree p on a uniform grid with grid size $2^{-\ell}$. For the space of B-splines with maximum continuity, that is, with $k = p - 1$, we only write $S_{p,\ell}$.

On the parameter domain $\widehat{\Omega} := (0,1)^d$, we use a tensor-product B-spline space, denoted by

$$S^d_{p,k,\ell} := \bigotimes_{i=1}^{d} S_{p,k,\ell}(0,1).$$

For ease of notation, we assume to have the same spline degree p, the same continuity k and the same number of uniform refinement steps ℓ, for each spatial dimension. We assume that the domain Ω can be parametrized by a geometry mapping $\mathbf{G} : \widehat{\Omega} \to \Omega = G(\widehat{\Omega})$ with the property

$$\|\nabla^r \mathbf{G}\|_{L^\infty(\widehat{\Omega})} \le c_1 \quad \text{and} \quad \| \left(\nabla^r \mathbf{G}\right)^{-1} \|_{L^\infty(\widehat{\Omega})} \le c_2, \quad \text{for} \quad r = 1, 2, 3, \tag{4.1}$$

for some constants c_1 and c_2. The discretization space $S_{p,k,\ell}$ on the domain Ω is defined using the pull-back principle as

$$S_{p,k,\ell}(\Omega) := \left\{ f \circ \mathbf{G}^{-1} : f \in S^d_{p,k,\ell} \right\}.$$

For more information on IgA, see the survey article [5] and the references therein. We use spline spaces with maximum smoothness as the discrete state space and reduce the smoothness for Q_h accordingly. More precisely, we use

$$Q_h := S_{p,p-3,\ell}(\Omega) \quad \text{and} \quad U_h := S_{p,\ell}(\Omega) \cap H^1_0(\Omega) \quad \text{with} \quad p \ge 2. \tag{4.2}$$

The following lemma shows that, if we consider the special case of box domains (if trivially parametrized) and constant convection, the condition of Lemma 3.2 holds.

THEOREM 4.1. *If $\Omega = (0,1)^d$, $Q_h = S^d_{p,p-3,\ell}$ and $U_h = S^d_{p,\ell} \cap H^1_0(\Omega)$ and if the convection β is constant, then*

$$\mathcal{S}(\mathcal{A}_h) = \mathcal{S}_h \quad \textit{and} \quad \kappa\left(\mathcal{S}_h^{-1}\mathcal{A}_h\right) \le \frac{\cos(\pi/7)}{\sin(\pi/14)} \approx 4.05.$$

Proof. For sake of simplicity, we restrict the proof to the two-dimensional case. Clearly,

$$f' \in S_{p-1,k-1,\ell}(0,1) \quad \forall\, f \in S_{p,k,\ell}(0,1)$$

together with

$$S_{p-1,k-1,\ell}(0,1) \subset S_{p,k-1,\ell}(0,1) \quad \text{and} \quad S_{p,k,\ell}(0,1) \subset S_{p,k-2,\ell}(0,1),$$

see [5]. Let $u \in S^2_{p,k,\ell} = S_{p,k,\ell}(0,1) \otimes S_{p,k,\ell}(0,1)$. Then,

$$\frac{\partial^2 u}{\partial x_1^2} \in S_{p-2,k-2,\ell}(0,1) \otimes S_{p,k,\ell}(0,1) \subset S^2_{p,k-2,\ell},$$
$$\frac{\partial^2 u}{\partial x_2^2} \in S_{p,k,\ell}(0,1) \otimes S_{p-2,k-2,\ell}(0,1) \subset S^2_{p,k-2,\ell},$$

and, by combining these results, also

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} \in S^2_{p,k-2,\ell}.$$

Since $\beta = (\beta_1, \beta_2)$ is constant we also have

$$\beta \cdot \nabla u = \beta_1 \frac{\partial u}{\partial x_1} + \beta_2 \frac{\partial u}{\partial x_2} \in S^2_{p,k-2,\ell}.$$

To summarize, for every $u \in S^2_{p,k,\ell}$, we have

$$-\varepsilon \Delta u + \beta \cdot \nabla u + \sigma u \in S^2_{p,k-2,\ell},$$

hence Condition (3.5) in Lemma 3.2 holds and we have $\mathcal{S}(\mathcal{A}_h) = \mathcal{S}_h$. The condition number bound follows from Theorem 2.2. ∎

If the domain is not a box domain or if the convection is not constant, then Theorem 4.1 cannot be applied. Nevertheless, the numerical results presented in Section 7 indicate that the spaces proposed in (4.2) work well even for more complex domains and variable convection.

5. Error estimates. In this section, we derive discretization error estimates for Problem 3.1. Let $A(\mathbf{x}, \tilde{\mathbf{x}})$ denote the bilinear form in Problem 2.1, where $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbf{X}$ denotes the triplet $(q, w, u) \in L^2(\Omega) \times L^2(\Omega) \times H^2(\Omega) \cap H_0^1(\Omega)$, with the norm

$$\|\mathbf{x}\|_{\mathcal{S}}^2 = \|q\|_{S_q}^2 + \|w\|_{S_w}^2 + \|u\|_{S_u}^2.$$

The discrete triplet $(q_h, w_h, u_h) \in Q_h \times Q_h \times U_h$ is denoted by $\mathbf{x}_h \in \mathbf{X}_h$. Galerkin orthogonality reads as follows:

$$A(\mathbf{x} - \mathbf{x}_h, \mathbf{y}_h) = 0 \quad \forall\, \mathbf{y}_h \in \mathbf{X}_h. \tag{5.1}$$

Since Problem 2.1 is well-posed (Theorem 2.2), we have boundedness

$$A(\mathbf{x}, \mathbf{y}) \le \overline{c} \|\mathbf{x}\|_{\mathcal{S}} \|\mathbf{y}\|_{\mathcal{S}} \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbf{X} \tag{5.2}$$

and inf-sup stability

$$\sup_{0 \ne \mathbf{y} \in \mathbf{X}} \frac{A(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|_{\mathcal{S}}} \ge \underline{c} \|\mathbf{x}\|_{\mathcal{S}} \quad \forall\, \mathbf{x} \in \mathbf{X}, \tag{5.3}$$

where $\overline{c}/\underline{c} = \kappa(\mathcal{S}^{-1}\mathcal{A})$. The boundedness also holds for a conforming discretization space $\mathbf{X}_h \subset \mathbf{X}$ with the same constant $\overline{c}$. If the condition (3.5) in Lemma 3.2 holds, then also the inf-sup holds with the same constant $\underline{c}$. Using this and the ideas of [1], we derive the following discretization error estimate.

Lemma 5.1. *If $\mathbf{x}_h \in \mathbf{X}_h$ is the solution to (3.1) for a discretization space satisfying condition (3.5) and if $\mathbf{x} \in \mathbf{X}$ is the solution of (2.1), then we have the estimate*

$$\|\mathbf{x} - \mathbf{x}_h\|_{\mathcal{S}} \le (1 + \kappa(\mathcal{S}^{-1}\mathcal{A})) \inf_{\mathbf{y}_h \in \mathbf{X}_h} \|\mathbf{x} - \mathbf{y}_h\|_{\mathcal{S}}. \tag{5.4}$$

Proof. Let $\mathbf{y}_h$ and $\mathbf{z}_h$ be arbitrary functions in $\mathbf{X}_h$. Due to Galerkin orthogonality (5.1), we have

$$A(\mathbf{x}_h - \mathbf{y}_h, \mathbf{z}_h) = A(\mathbf{x}_h - \mathbf{x} + \mathbf{x} - \mathbf{y}_h, \mathbf{z}_h) = A(\mathbf{x} - \mathbf{y}_h, \mathbf{z}_h).$$

Combining this with the boundedness condition (5.2) gives

$$A(\mathbf{x}_h - \mathbf{y}_h, \mathbf{z}_h) \le \overline{c} \|\mathbf{x} - \mathbf{y}_h\|_{\mathcal{S}} \|\mathbf{z}_h\|_{\mathcal{S}} \quad \forall\, \mathbf{y}_h, \mathbf{z}_h \in \mathbf{X}_h.$$

Using the discrete inf-sup gives

$$\underline{c} \|\mathbf{x}_h - \mathbf{y}_h\|_{\mathcal{S}} \le \sup_{0 \ne \mathbf{z}_h \in \mathbf{X}_h} \frac{A(\mathbf{x}_h - \mathbf{y}_h, \mathbf{z}_h)}{\|\mathbf{z}_h\|_{\mathcal{S}}} \le \overline{c} \|\mathbf{x} - \mathbf{y}_h\|_{\mathcal{S}} \quad \forall\, \mathbf{y}_h \in \mathbf{X}_h.$$

Finally, we use the triangle inequality and $\overline{c}/\underline{c} = \kappa(\mathcal{S}^{-1}\mathcal{A})$ to obtain the desired result

$$\|\mathbf{x} - \mathbf{x}_h\|_{\mathcal{S}} \le \|\mathbf{x} - \mathbf{y}_h\|_{\mathcal{S}} + \|\mathbf{x}_h - \mathbf{y}_h\|_{\mathcal{S}} \le (1 + \kappa(\mathcal{S}^{-1}\mathcal{A})) \|\mathbf{x} - \mathbf{y}_h\|_{\mathcal{S}} \quad \forall\, \mathbf{y}_h \in \mathbf{X}_h. \qquad \square$$

Let $\Omega = \widehat{\Omega} := (0,1)^d$ and let the convection β be constant. Since we assume a trivial parametrization, we have the discretization spaces $Q_h = S^d_{p,p-3,\ell}$ and $U_h = S^d_{p,\ell} \cap H^1_0(\widehat{\Omega})$.

To derive the error estimates, we assume that the solution of (2.1) satisfies the regularity assumption $(q,w,u) \in H^1(\widehat{\Omega}) \times H^1(\widehat{\Omega}) \times H^3(\widehat{\Omega}) \cap H^1_0(\widehat{\Omega})$.

Now, we estimate the approximation error term in (5.4) from above. First, we observe that

$$\inf_{\mathbf{y}_h \in \mathbf{X}_h} \|\mathbf{x} - \mathbf{y}_h\|_{\mathcal{S}} \leq$$

$$\inf_{q_h \in S^d_{p,p-3,\ell}} \alpha \|q - q_h\|^2_{L^2(\widehat{\Omega})} + \inf_{w_h \in S^d_{p,p-3,\ell}} \frac{1}{\alpha} \|w - w_h\|^2_{L^2(\widehat{\Omega})} + \inf_{u_h \in S^d_{p,\ell} \cap H^1_0(\widehat{\Omega})} \|u - u_h\|^2_{S_u}.$$

The two first terms can be bounded by using the following approximation error estimate

$$\inf_{q_h \in S^d_{p,p-3,\ell}} \|q - q_h\|_{L^2(\widehat{\Omega})} \leq \frac{h}{4\sqrt{3}} \|\nabla q\|_{L^2(\widehat{\Omega})}, \tag{5.5}$$

see [22, Corollary 1]. The estimate for the last term

$$\inf_{u_h \in S^d_{p,\ell} \cap H^1_0(\widehat{\Omega})} \|u - u_h\|^2_{S_u}$$

is more involved. Before handling this term, we need a convenient notation and some auxiliary approximation error estimates.

Notation 5.2. In what follows, c is a generic positive constant independent of α, σ, β, ε, h and p, but may depend on the spatial dimension d and the observation domain $\mathcal{O}$.

In Appendix B, we extend some of the results of [22, 23, 24]. The main result is summarized in the following theorem.

THEOREM 5.3. *Let $\mathbf{\Pi}_p : H^3(\widehat{\Omega}) \cap H^1_0(\widehat{\Omega}) \to S^d_{p,\ell} \cap H^1_0(\widehat{\Omega})$ be the H^2-orthogonal projector, where $p \geq 3$ and $\ell \geq 1$. Then,*

$$\|\nabla^2 (I - \mathbf{\Pi}_p) u\|_{L^2(\widehat{\Omega})} \leq ch \ \|\nabla^3 u\|_{L^2(\widehat{\Omega})}, \tag{5.6}$$

$$\|\nabla (I - \mathbf{\Pi}_p) u\|_{L^2(\widehat{\Omega})} \leq ch^2 \|\nabla^3 u\|_{L^2(\widehat{\Omega})}, \tag{5.7}$$

$$\|(I - \mathbf{\Pi}_p) u\|_{L^2(\widehat{\Omega})} \leq ch^3 \|\nabla^3 u\|_{L^2(\widehat{\Omega})} \quad \forall\, u \in H^3(\widehat{\Omega}) \cap H^1_0(\widehat{\Omega}). \tag{5.8}$$

Remark 5.4. In the statement of Theorem 5.3, the parameter domain $\widehat{\Omega}$ is considered. This result can be extended to physical domains if the corresponding geometry function $\mathbf{G}$ is sufficiently smooth, cf. [24].

With Theorem 5.3, we can derive an error estimate for our problem.

THEOREM 5.5. *Let $\Omega := \widehat{\Omega} := (0,1)^d$ with $d \in \mathbb{N}$, let $\beta \in \mathbb{R}^d$ be constant and let $(q_h, w_h, u_h) \in S^d_{p,p-3,\ell} \times S^d_{p,p-3,\ell} \times S^d_{p,\ell} \cap H^1_0(\widehat{\Omega})$ with $p \geq 3$ and $\ell \geq 1$, be the solution to (3.1). If $(q,w,u) \in H^1(\widehat{\Omega}) \times H^1(\widehat{\Omega}) \times H^3(\widehat{\Omega}) \cap H^1_0(\widehat{\Omega})$ is the solution of (2.1), then we have the following estimates:*

$$\|q - q_h\|_{S_q} + \|w - w_h\|_{S_w} + \|u - u_h\|_{S_u} \leq$$

$$ch\left(\sqrt{\alpha}\|\nabla q\|_{L^2(\widehat{\Omega})} + \frac{1}{\sqrt{\alpha}}\|\nabla w\|_{L^2(\widehat{\Omega})} + \sqrt{\alpha}\max\left\{\varepsilon, \|\beta\|h, \left(\sigma + \frac{1}{\sqrt{\alpha}}\right)h^2\right\} \|\nabla^3 u\|_{L^2(\widehat{\Omega})}\right)$$

Proof. Let $\tilde{u}_h := \mathbf{\Pi}_p u$. Then, we have for all $u \in H^3(\widehat{\Omega}) \cap H^1_0(\widehat{\Omega})$

$$\|u - \tilde{u}_h\|^2_{S_u} = \|u - \tilde{u}_h\|^2_{L^2(\mathcal{O})} + \alpha\|(-\varepsilon\Delta + \beta\cdot\nabla + \sigma)(u - \tilde{u}_h)\|^2_{L^2(\widehat{\Omega})}.$$

For the first term, we simply extend the norm from $\mathcal{O}$ to $\widehat{\Omega}$ and obtain

$$\|u - \tilde{u}_h\|^2_{L^2(\mathcal{O})} \leq \|u - \tilde{u}_h\|^2_{L^2(\widehat{\Omega})} \leq c\, h^6 \|\nabla^3 u\|^2_{L^2(\widehat{\Omega})}.$$

For the second term, we use the triangle inequality and Theorem 5.3, and we get

$$\begin{aligned}
&\alpha\|(-\varepsilon\Delta + \beta\cdot\nabla + \sigma)(u - \tilde{u}_h)\|^2_{L^2(\widehat{\Omega})} \\
&\quad\leq 3\alpha\left(\varepsilon^2\|\Delta(u - \tilde{u}_h)\|^2_{L^2(\widehat{\Omega})} + \|\beta\|^2\|\nabla(u - \tilde{u}_h)\|^2_{L^2(\widehat{\Omega})} + \sigma^2\|u - \tilde{u}_h\|^2_{L^2(\widehat{\Omega})}\right) \\
&\quad\leq c\,\alpha\left(\varepsilon^2 h^2 + \|\beta\|^2 h^4 + \sigma^2 h^6\right)\|\nabla^3 u\|^2_{L^2(\widehat{\Omega})}.
\end{aligned}$$

Combining these results gives

$$\inf_{\tilde{u}_h \in S^d_{p,\ell} \cap H^1_0(\widehat{\Omega})} \|u - \tilde{u}_h\|_{S_u} \leq c\sqrt{\alpha}\, h \,\max\left\{\varepsilon, \|\beta\| h, (\sigma + 1/\sqrt{\alpha})h^2\right\}\|\nabla^3 u\|_{L^2(\widehat{\Omega})}. \tag{5.9}$$

The theorem follows from combining Lemma 5.1 with Equation (5.5) and Equation (5.9). ∎

Remark 5.6. Theorem 5.5 is somewhat restrictive since it requires the domain to be a box domain and the convection to be constant. These requirements are needed for the proof of the discrete inf-sup stability. The numerical results presented in Section 7 indicate that the restrictions are not needed.

6. Numerical experiments: Accuracy of the solution. In this section we compare the solution of the forward problem to the (state) solution of the optimal control problem to investigate the fact that the optimal control problem naturally introduces a non-standard Petrov-Galerkin method for the state equation. We consider a well-known one-dimensional problem [7]:

$$-\partial_x u(x) - \varepsilon\,\partial_{xx} u(x) = 0 \quad \text{in} \quad (0,1), \qquad u(0) = 0, \qquad u(1) = 1,$$

whose exact solution is

$$u(x) = \frac{e^{-x/\varepsilon} - 1}{e^{-1/\varepsilon} - 1}.$$

This problem is used as state equation in our optimal control problem (Problem 1.1), where we choose $\beta = -1$, $\sigma = 0$, $f = 0$, u_d to be the exact solution of the forward problem, and the boundary conditions on u to be as for the forward problem. The analytical solution of the optimal control problem is

$$q(x) = 0, \quad w(x) = 0, \quad u(x) = \frac{e^{-x/\varepsilon} - 1}{e^{-1/\varepsilon} - 1}.$$

We use the discretization spaces

$$Q_h = S_{p,p-3,\ell}(0,1) \quad \text{and} \quad U_h = \{u_h \in S_{p,\ell}(0,1) \,:\, u(0) = 0,\, u(1) = 1\}$$

for the optimal control problem, which satisfy the condition (3.5). We compare the numerical solution for the state with the numerical solution of the forward problem, where we use U_h as trial and test space. No stabilization techniques are used. The diffusion is set to $\varepsilon = 0.01$ and $\alpha = 0.001$ for the optimal control problem. Three observation domains are considered: Full observation, that is, $\mathcal{O} = \Omega = (0,1)$, and partial observation on $\mathcal{O} = (0, \frac{1}{4})$ and on $\mathcal{O} = (\frac{3}{4}, 1)$. The numerical solutions are displayed in Figures 1 to 4. The plots indicate that the forward solution is unstable for coarse discretizations. The non-physical oscillations start in the boundary layer and propagate into the remainder of the computational domain. These kinds of instabilities are often remedied by upwind and/or Petrov-Galerkin schemes [7]. We remark though that our Petrov-Galerkin like approach for the state equation does not resemble any of the common Petrov-Galerkin schemes for this equation, as far as we know. The state solution (of the optimal control problem) does not have these instabilities. In fact, the state operator K_h is discretized with a PetrovGalerkin method as the trial space is U_h and the test space is Q_h.

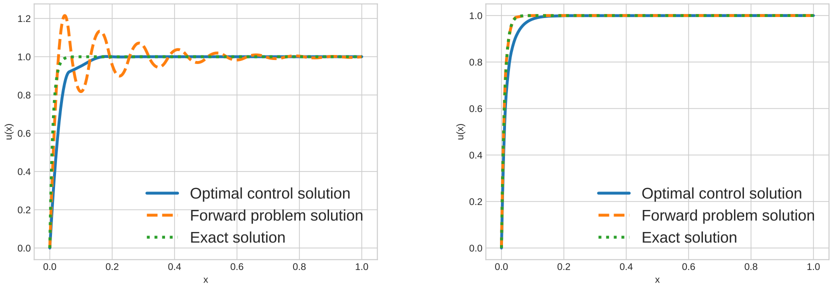

Fig. 1. *Full observation on $\mathcal{O} = (0,1)$ with $p = 2$ (both), $\ell = 4$ (left), $\ell = 6$ (right).*

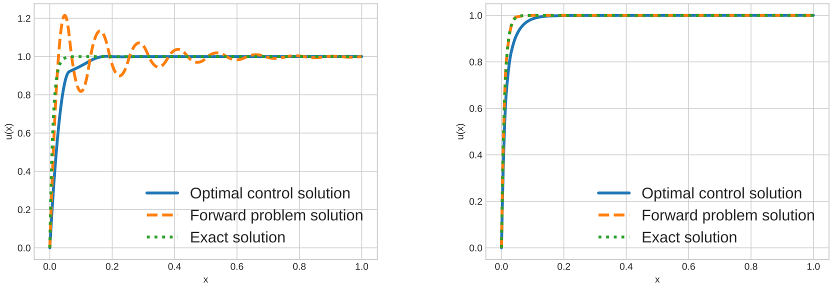

Fig. 2. *Partial observation on $\mathcal{O} = (0, \frac{1}{4})$ with $p = 2$ (both), $\ell = 4$ (left), $\ell = 6$ (right).*

In Figure 2, we consider the optimal control problem with observation on $(0, \frac{1}{4})$. This is only a quarter of the whole domain, but it is located at the boundary layer. The solutions for the state variable are almost identical to those obtained for the case of full observation.

Next we look at the solution where the observation domain is $(\frac{3}{4}, 1)$. Here the solution is almost constant ($u(x) \approx 1$). From Figures 3 and 4, we see that the approximation is not good. In the left plot of Figure 3, we see that the boundary layer is not captured. However, the error does not propagate into the observation domain. For h-refinement, see Figure 3 (right) and Figure 4 (left), we observe that the approximation

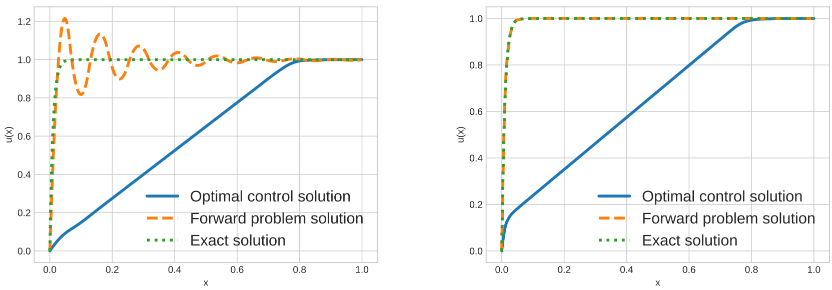

FIG. 3. *Partial observation on* $\mathcal{O} = (\frac{3}{4}, 1)$ *with* $p = 2$ *(both),* $\ell = 4$ *(left),* $\ell = 6$ *(right).*

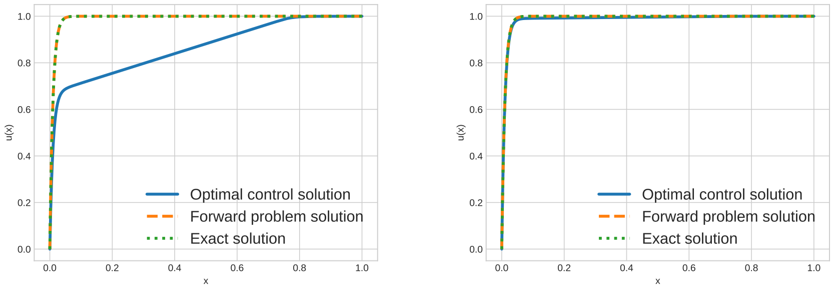

FIG. 4. *Partial observation on* $\mathcal{O} = (\frac{3}{4}, 1)$ *with* $p = 2$, $\ell = 8$ *(left) and* $p = 4$, $\ell = 6$ *(right).*

improves slowly. For p-refinement, see Figure 4 (right), the approximation improves significantly. Since we use splines, increasing the spline degree by one means that the number of degrees of freedom is only increased by one, while each h-refinement step doubles the number of degrees of freedom.

Remark 6.1. The effect of increasing the spline degree compared to h-refinement as shown in Figure 4 is somewhat surprising. We are not completely sure why larger spline degrees are so effective. Unfortunately, the error estimate in Theorem 5.5 does not provide any explanation for this behavior. Further analysis is needed to explain this properly.

7. Numerical experiments for exact and inexact preconditioners. In this section, we analyze the convergence of Krylov space solvers when the proposed preconditioner is used. In the first subsection, we consider an exact realization of the preconditioner. A multigrid approximation is then considered in the second subsection.

We have done the numerical experiments for two model domains, in both cases for $d = 2$. The first domain is a box-domain, more precisely, Ω is the unit square, see Figure 5 (left), which is parameterized with the identity function. For this domain, the conditions of Theorem 4.1 are satisfied. In Model problem 7.1, we have full observation and in Model problem 7.2, the observation is restricted to the subdomain represented by the smaller area in Figure 5 (left). In all model problems, the desired state u_d is a step function with value $u_d = 1$ inside a circle and with value $u_d = 0$ outside the circle. The support of u_d is shown as the dashed lines in Figure 5.

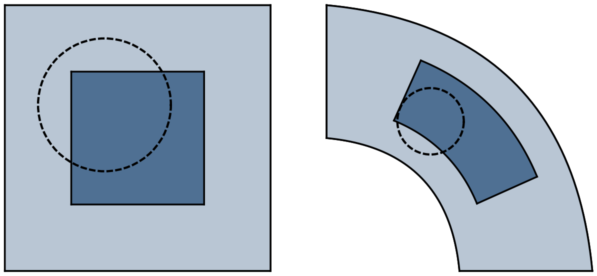

FIG. 5. *Computational domains Ω, partial observation domains $\mathcal{O}$ (dark blue), and support of desired state (inside of dashed circles).*

MODEL PROBLEM 7.1 (Unit square and constant convection with full observation). *Let $\Omega := \mathcal{O} := (0,1)^2$ be the computational domain, which is also the observation domain. The convection is $\beta = (-2,1)$ and there is no reaction $\sigma = 0$ or source term $f = 0$. The desired state is*

$$u_d(x,y) = \begin{cases} 1 & \textit{if} \quad (x-\frac{3}{8})^2 + (y-\frac{5}{8})^2 \le \frac{1}{16} \\ 0 & \textit{otherwise.} \end{cases}$$

The diffusion ε and regularization parameter α will vary.

MODEL PROBLEM 7.2 (Unit square and constant convection with limited observation). *Let $\Omega = (0,1)^2$ be the computational domain and $\mathcal{O} = (\frac{1}{4},\frac{3}{4})^2$ be the observation domain. The remainder of this problem is the same as for Model problem 7.1.*

Furthermore, we consider a non-trivial geometry Ω, which is a approximation of a quarter annulus by means of a B-spline parameterization, see Figure 5 (right). Again, Model problem 7.3 is a problem with full observation and the observation domain in Model problem 7.4 is the smaller area in Figure 5 (right). We observe that for this domain, the conditions of Theorem 4.1 are not satisfied.

MODEL PROBLEM 7.3 (Quarter annulus and varying convection with full observation). *Let $\Omega = \mathcal{O} = \mathbf{G}((0,1)^2)$ with $\mathbf{G} : (0,1)^2 \to \mathbb{R}^2$ and*

$$\mathbf{G}(\widehat{x}) = \begin{pmatrix} (1+\widehat{x}_1)(1-\widehat{x}_2)(1+2(\sqrt{2}-1)\widehat{x}_2) \\ (1+\widehat{x}_1)\widehat{x}_2(2\sqrt{2}-1-2(\sqrt{2}-1)\widehat{x}_2) \end{pmatrix} \tag{7.1}$$

be the computational domain, which is also the observation domain. The convection is $\beta = (y, 1+x^2)$ and there is no reaction $\sigma = 0$ or source term $f = 0$. The desired state is

$$u_d(x,y) = \begin{cases} 1, & \textit{if} \quad (x-x_0)^2 + (y-y_0)^2 \le \frac{1}{16} \\ 0, & \textit{otherwise,} \end{cases}$$

where $(x_0,y_0) = \mathbf{G}(\frac{3}{8},\frac{5}{8})$. The diffusion ε and regularization parameter α will vary.

MODEL PROBLEM 7.4 (Quarter annulus and varying convection with limited observation). *Let $\Omega = \mathbf{G}((0,1)^2)$ be the computational domain and $\mathcal{O} = \mathbf{G}((\frac{1}{4},\frac{3}{4})^2)$ be the observation domain, where $\boldsymbol{G}$ is as in (7.1). The remainder of this problem is the same as for Model problem 7.3.*

For all model problems, we consider a discretization of the optimality system using the spaces given in (4.2) as outlined in Section 4. The resulting linear system of equations

$$\mathcal{A}_h \underline{\mathbf{x}}_h = \underline{\mathbf{b}}_h$$

is solved using the MINRES method, preconditioned with the proposed Schur complement preconditioner (3.3). We use a random initial guess $\underline{\mathbf{x}}_{h,0}$. The stopping criterion is

$$\|\underline{\mathbf{r}}_k\| \leq 10^{-8} \|\underline{\mathbf{r}}_0\|,$$

where $\underline{\mathbf{r}}_k := \underline{\mathbf{b}}_h - \mathcal{A}_h \underline{\mathbf{x}}_{h,k}$ denotes the residual and $\| \cdot \|$ is the Euclidean norm.

7.1. Results with exact preconditioner. In this section, we present the results for the Schur complement preconditioner (3.3) when realized using a sparse Cholesky decomposition.

Table 1 shows the iteration numbers needed to reach the stopping criteria for full and partial observation (Model problems 7.1 and 7.2). In these tables, α and ε are varied, while $p = 2$ and $\ell = 6$ are fixed. In Table 2, we set $\varepsilon = 10^{-3}$ and vary the refinement level ℓ and α. From the tables, we observe that for the partial observation problem, we need a few more iterations for small values of α. This is probably because $M_{\mathcal{O},h}$ is singular in case of partial observation. The iteration numbers are relatively small for all considered values of α, ε and ℓ. This is predicted by the theory as Model problems 7.1 and 7.2 satisfy the conditions of Theorem 4.1.

Table 1

Iteration numbers: Model problem 7.1 (left) and 7.2 (right), $p = 2$, $\ell = 6$.

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	12	26	60	72
10^{-3}	15	47	26	11
10^{-6}	15	46	26	11
10^{-9}	15	46	26	11

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	12	20	57	78
10^{-3}	15	41	54	19
10^{-6}	14	41	53	19
10^{-9}	14	41	53	19

Table 2

Iteration numbers: Model problem 7.1 (left) and 7.2 (right), $p = 2$, $\varepsilon = 10^{-3}$.

$\ell \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
4	15	46	14	7
5	15	47	19	8
6	15	47	26	11
7	15	46	38	11

$\ell \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
4	15	46	35	15
5	15	43	44	17
6	15	41	54	19
7	15	39	55	22

Next, we consider Model problems 7.3 and 7.4, which are the problems where the computational domain is a quarter annulus. The iteration numbers are shown in Tables 3 and 4. Note that the conditions of Theorem 4.1 are not satisfied. Nevertheless, the iteration numbers are comparable with those of Tables 1 and 2.

Remark 7.5. For $\varepsilon = 1$, the iteration numbers in Table 1 (and Table 3) are growing as α becomes smaller. This may appear strange since we have proven that the condition number (for Table 1) is less then 4.05. Describing convergence estimates for Krylov subspace methods in term of only the condition number can be misleading and/or insufficient, cf. [16]. The different iteration numbers for various values of ε and α can be explained by the distribution of the eigenvalues. For small iteration numbers the eigenvalues are more clustered. For $\varepsilon = 1$, the iteration numbers starts decreasing when $\alpha < 10^{-9}$.

Table 3

Iteration numbers: Model problem 7.3 (left) and 7.4 (right), $p = 2$, $\ell = 6$.

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	17	41	62	64
10^{-3}	18	48	29	11
10^{-6}	18	48	29	11
10^{-9}	18	48	29	11

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	17	32	60	76
10^{-3}	17	46	54	25
10^{-6}	17	46	54	25
10^{-9}	17	46	54	25

Table 4

Iteration numbers: Model problem 7.3 (left) and 7.4 (right), $p = 2$, $\varepsilon = 0.001$.

$\ell \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
4	18	47	16	10
5	18	48	21	11
6	18	48	29	11
7	16	48	42	12

$\ell \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
4	16	48	40	25
5	16	46	46	25
6	17	46	54	25
7	16	46	55	28

7.2. Results with inexact preconditioner. So far, we have realized the proposed preconditioners using sparse direct solvers. This approach works well for mid-sized problems. For large-sized problems, alternatives are of interest since they might be faster or have a smaller memory footprint. We replace $\mathcal{S}_h$ by a spectrally equivalent approximation $\widetilde{\mathcal{S}}_h$, where the action of $\widetilde{\mathcal{S}}_h^{-1}$ can be calculated efficiently. The spectral equivalence should be robust with respect to the parameters of interest.

For the approximation of the mass matrix M_h, which is found in the first and the second block of the overall preconditioner, we exploit the fact that the mass matrix on the parameter domain is the Kronecker product of two mass matrices that correspond to the discretization of a univariate problem, i.e., we have

$$M_h = M^{(1)} \otimes M^{(2)},$$

where $\otimes$ denotes the Kronecker product and $M^{(1)}$ and $M^{(2)}$ denote the univariate mass matrices. For the Model problems 7.3 and 7.4, we use a similar preconditioner that incorporates a tensor-rank-1 approximation of the geometry, which is derived as follows. As common in IgA, the bilinear forms are computed by transformation to the parameter domain, i.e., we have

$$(u, v)_{L_2(\Omega)} = \int_0^1 \int_0^1 J(x_1, x_2)\, u(x_1, x_2)\, v(x_1, x_2)\, \mathrm{d}x_2\, \mathrm{d}x_1,$$

where $J(x) = |\det \nabla \mathbf{G}(x)|$, which we approximate by

$$(u, v)_{\widetilde{M}} := \frac{1}{J(\frac{1}{2}, \frac{1}{2})} \int_0^1 \int_0^1 J(x_1, \tfrac{1}{2})\, J(\tfrac{1}{2}, x_2)\, u(x_1, x_2)\, v(x_1, x_2)\, \mathrm{d}x_2\, \mathrm{d}x_1.$$

The corresponding mass matrix $\widetilde{M}_h$ has tensor-product structure:

$$\widetilde{M}_h = \widetilde{M}^{(1)} \otimes \widetilde{M}^{(2)},$$

where $\widetilde{M}^{(1)}$ and $\widetilde{M}^{(2)}$ denote the univariate mass matrices, which are lumped with $J(x_1, \frac{1}{2})$ and $J(\frac{1}{2}, x_2)$, respectively. Straight-forward computations show that the relative condition number of the exact mass matrix and its approximation can be bounded uniformly by a term that only depends on $\mathbf{G}$. For realizing the inverse of $\widetilde{M}_h$

efficiently, we make use of the fact that the application of the inverse of a Kronecker product to some vector can be efficiently realized using sparse direct solvers that realize the application of $(\widetilde{M}^{(1)})^{-1}$ and $(\widetilde{M}^{(2)})^{-1}$.

For the approximation of the inverse of the matrix $M_{\mathcal{O},h} + \alpha B_h$, representing a fourth-order PDE, we use a geometric multigrid solver. Following the standard approach, we assume to have a hierarchy of quasi-uniform grids, where the grid sizes of two consecutive grids differ by a factor of two. Since we have tensor-product grids in Isogeometric Analysis, such a grid hierarchy can be easily constructed by coarsening. The coarsest grid level is chosen such that there are no inner knots. On each of these grid levels $\ell = 0, 1, \dots, L$, we introduce a discretization space $U_{h_\ell} = S_{p,\ell}(\Omega) \cap H_0^1(\Omega)$. One iterate of the multigrid solver consists of the following steps:

- Apply $\nu = 2$ forward Gauss-Seidel sweeps as pre-smoother.
- Apply coarse-grid correction. Since we have nested grids ($U_{h_\ell} \subset U_{h_{\ell+1}}$), the coarse-grid correction is realized based on canonical embedding. For $\ell > 1$, the problem on the next coarser level $\ell - 1$ is solved by applying 1 step of the multigrid method recursively (V-cycle). Only on the coarsest grid level $\ell = 0$, the problem is solved using a direct solver.
- Apply $\nu = 2$ backward Gauss-Seidel sweeps as post-smoother.

The robustness of that multigrid method in the grid size is a straight-forward extension of the known results for the biharmonic problem, cf. [23]. It is worth mentioning that this argument does not cover the robustness in any of the other parameters that affect the multigrid solver.

We again use a MINRES solver, now preconditioned with the presented tensor-rank-one approximation of the mass matrices and with one step of the multigrid solver. The corresponding numerical results are presented in Tables 5 and 6. In Table 5, we observe that the iteration counts are uniformly bounded for all choices of ϵ and α, however with much larger values than for the exact preconditioner. This is related to the well-known fact that standard Gauss-Seidel smoothers do not perform well in the framework of Isogeometric Analysis. The convergence deteriorates particularly if the spline degree is increased, which can also be seen in Table 6. Furthermore, in Table 6, we can also study the dependence of the convergence on the grid size. Although, the convergence theory predicts a robust convergence behavior, this is not observed in practice for the grid levels considered. Apparently, this is the case since the Gauss-Seidel smoother does not work well for spline bases, even for moderate values of p, cf., e.g., [24].

Table 5
Iteration numbers: Model problem 7.3 (left) and 7.4 (right), $p = 2$, $\ell = 6$.

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	128	108	104	67
10^{-3}	169	81	42	26
10^{-6}	179	81	42	26
10^{-9}	179	81	42	26

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	128	104	132	175
10^{-3}	171	112	139	172
10^{-6}	178	112	142	174
10^{-9}	178	112	142	174

To obtain a better convergence behavior, we consider a second approach for the smoother: a macro Gauss-Seidel approach. This approach makes use of the tensor-product structure of the discretization. For two dimensions, the degrees of freedom can be represented as a grid in the plane, see Figure 6 (left). Each dot represents one degree of freedom or basis function. We start by introducing a macro grid that groups $a \times a$ degrees of freedom. (If the number of rows or columns is not divisible by

Table 6
Iteration numbers: Model problem 7.3 (left) and 7.4 (right), $\alpha = 0.001$, $\varepsilon = 0.001$.

$\ell \setminus p$	2	3	5	7
4	49	64	191	730
5	55	61	150	510
6	81	86	137	440
7	118	134	179	380

$\ell \setminus p$	2	3	5	7
4	57	63	152	567
5	77	85	135	437
6	112	123	170	414
7	152	162	204	364

a, the last macro elements in each direction are correspondingly smaller.) The macro grid is depicted in Figure 6 (left).

Each of the macro elements consists of the degrees of freedom that belong to the element of the macro grid and of degrees of freedom of the neighboring elements of the macro grid. Here, we use b additional rows and columns each on each of the sides, see Figure 6 (right).

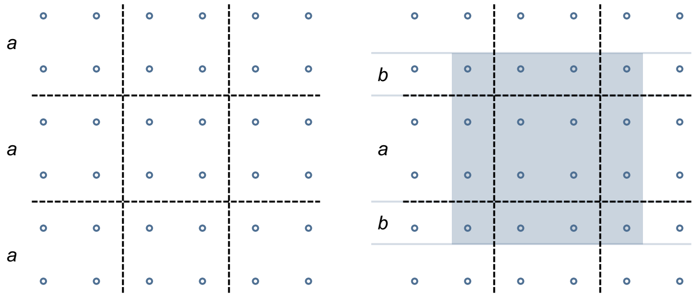

Fig. 6. *The construction of the macro Gauss-Seidel approach.*

Then, a macro Gauss-Seidel sweep is a standard multiplicative Schwarz method, where the subspaces are the degrees of freedom that belong to each of the macro elements. So, the choice $a := 1$ and $b := 0$ corresponds to a standard Gauss-Seidel sweep.

In the following, we use the patch size $a := p$ and the overlap size $b := p - 1$. As for the standard Gauss-Seidel case, we apply a forward sweep for pre-smoothing and a backward sweep, i.e., with the reverse ordering of the macro elements, for post-smoothing. The problem within the (relatively small) subspaces is solved by means of a direct solver. The number of smoothing steps ν is set to 1.

The corresponding iteration counts are presented in Tables 7 and 8. In all cases, we obtain significantly better convergence rates than for a standard Gauss-Seidel smoother. Table 8 shows that the resulting method is robust in the spline degree, and that the method is quite robust in the grid size. Table 7 shows that the overall method is also robust in the parameter ε and well-bounded for α.

Appendix A. Proof of Lemma 2.3. The inequality in Lemma 2.3 is sometimes referred to as *the second fundamental inequality*, cf. [15]. For domains with polygonal (polyhedral) Lipschitz boundary the result is known, but for domains which are images of geometry mappings we were unable to find any result. We therefore provide a proof in this appendix. We start with providing a density result.

Lemma A.1. *Let the domain Ω have a Lipschitz boundary and be the image of a*

TABLE 7
Iteration numbers: Model problem 7.3 (left) and 7.4 (right), $p = 2$, $\ell = 6$.

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	50	51	62	64
10^{-3}	91	49	29	13
10^{-6}	96	49	29	13
10^{-9}	96	49	29	13

$\varepsilon \setminus \alpha$	10^0	10^{-3}	10^{-6}	10^{-9}
10^0	50	46	72	98
10^{-3}	94	77	99	103
10^{-6}	96	77	99	103
10^{-9}	96	77	99	103

TABLE 8
Iteration numbers: Model problem 7.3 (left) and 7.4 (right), $\alpha = 0.001$, $\varepsilon = 0.001$.

$\ell \setminus p$	2	3	5	7
4	47	48	48	48
5	48	48	48	48
6	49	48	48	48
7	64	49	48	48

$\ell \setminus p$	2	3	5	7
4	49	48	46	48
5	57	52	46	46
6	77	64	53	49
7	96	80	58	52

geometric mapping $\mathbf{G} : \widehat{\Omega} := (0,1)^d \to \Omega$, where both $\|\nabla^r \mathbf{G}\|_{L^\infty}$ and $\|(\nabla^r \mathbf{G})^{-1}\|_{L^\infty}$ are bounded for $r \in \{1,2,3\}$, then $H^3(\Omega) \cap H_0^1(\Omega)$ is dense in $H^2(\Omega) \cap H_0^1(\Omega)$.

Proof. Let $V := H^3(\Omega) \cap H_0^1(\Omega)$ and $U := H^2(\Omega) \cap H_0^1(\Omega)$, we want that for any $\epsilon > 0$ and $u \in U$, there exist a $v \in V$ such that

$$\|u - v\|_{H^2(\Omega)} \le \epsilon. \tag{A.1}$$

From [11, Theorem 1.6.2] we know that $\widehat{V} := H^3(\widehat{\Omega}) \cap H_0^1(\widehat{\Omega})$ is dense in $\widehat{U} := H^2(\widehat{\Omega}) \cap H_0^1(\widehat{\Omega})$, i.e., for any $\epsilon > 0$ and $u \in \widehat{U}$, there exist a $\widehat{v} \in \widehat{V}$ such that

$$\|\widehat{u} - \widehat{v}\|_{H^2(\widehat{\Omega})} \le \epsilon. \tag{A.2}$$

Now, we know using the standard IgA-results that

$$\|w\|_{H^2(\Omega)} \le c_g \|w \circ \mathbf{G}\|_{H^2(\widehat{\Omega})} \tag{A.3}$$

holds for all $w \in H^2(\Omega)$, where c_g only depends on the geometry. Moreover, we have

$$v \in V \Leftrightarrow v \circ \mathbf{G} \in \widehat{V} \qquad \text{and} \qquad u \in U \Leftrightarrow u \circ \mathbf{G} \in \widehat{U}.$$

Now we prove (A.1). Let $u \in U$ and $\epsilon > 0$ be given. Let $\widehat{u} := u \circ \mathbf{G}$. Using (A.2), we know that there exist a $\widehat{v} \in V_g$ such that

$$\|\widehat{u} - \widehat{v}\|_{H^2(\widehat{\Omega})} \le \epsilon / c_g.$$

By choosing $v := \widehat{v} \circ \mathbf{G}^{-1}$, we have

$$\|(u - v) \circ \mathbf{G}\|_{H^2(\widehat{\Omega})} \le \epsilon / c_g.$$

and using (A.3) consequently

$$\|u - v\|_{H^2(\Omega)} \le \epsilon.$$

This means that we have found a proper $v \in V$ such that (A.1) holds. □

Next, we state a weighted trace theorem [12, Theorem 1.5.1.10].

THEOREM A.2. *Let Ω be a bounded open subset of $\mathbb{R}^d$ with Lipschitz boundary and let T be the trace operator. Then there a exist a constant c which only depend on Ω such that*

$$\int_{\partial\Omega} |Tv(x)|^2 \; ds \leq c \left[\sqrt{\delta} \int_{\Omega} |\nabla v(x)|^2 \; dx + \frac{1}{\sqrt{\delta}} \int_{\Omega} |v(x)|^2 \; dx \right], \tag{A.4}$$

hold for all $v(x) \in H^1(\Omega)$ and all $\delta \in (0, 1)$.

We can now prove Lemma 2.3.

Proof. When Ω has a polygonal (polyhedral) Lipschitz boundary the result follows from [11, 12]. A detailed proof of this case can be found in [25, Lemma 3.3]. We consider the case where Ω is the image of a geometric mapping and has Lipschitz boundary. According to [12, Theorem 3.1.1.2] we have

$$\int_{\Omega} |\nabla \cdot \psi(x)|^2 \; dx = \int_{\Omega} \nabla\psi(x) : (\nabla\psi(x))^T \; dx - \int_{\partial\Omega} g(x)(\psi_n(x))^2 \; ds,$$

for all $\psi \in H^2(\Omega)^d$ with $\psi_n := \psi \cdot n$ and $\psi_T := \psi - \psi_n \, n = 0$. Here, $g(x)$ is a function which depends on the curvature of boundary $\partial\Omega$. This can be bounded from above by a constant c_g depending only on Ω:

$$\int_{\Omega} |\nabla \cdot \psi(x)|^2 \; dx \geq \int_{\Omega} \nabla\psi(x) : (\nabla\psi(x))^T \; dx - c_g \int_{\partial\Omega} (\psi_n(x))^2 \; ds. \tag{A.5}$$

Applying this inequality to $\psi = \nabla u$ with $u \in H^3(\Omega) \cap H_0^1(\Omega)$, we now bound the last term by using Theorem A.2

$$\begin{aligned} -c_g \int_{\partial\Omega} (\psi_n(x))^2 \; ds = -c_g \int_{\partial\Omega} (\nabla u(x) \cdot n)^2 \; ds &\geq -c_g d \int_{\partial\Omega} |\nabla u(x)|^2 \; ds \\ &\geq -c \left[\sqrt{\delta} \int_{\Omega} |\nabla^2 u(x)|^2 \; dx + \frac{1}{\sqrt{\delta}} \int_{\Omega} |\nabla u(x)|^2 \; dx \right]. \end{aligned}$$

By using integration by parts, the Cauchy–Schwarz inequality and the Poincaré inequality, we can bound the last term by

$$\int_{\Omega} |\nabla u(x)|^2 \; dx \leq c_P^2 \|\Delta u\|_{L^2(\Omega)}^2,$$

where c_P is the Poincaré constant. Combining the last two inequalities gives

$$-c_g \int_{\partial\Omega} (\psi_n(x))^2 \; ds \geq -c \left[\sqrt{\delta} \|\nabla^2 u\|_{L^2(\Omega)}^2 + \frac{c_P^2}{\sqrt{\delta}} \|\Delta u\|_{L^2(\Omega)}^2 \right].$$

Inserting the inequality above and $\psi = \nabla u$ into (A.5) gives

$$\|\Delta u\|_{L^2(\Omega)}^2 \geq \|\nabla^2 u\|_{L^2(\Omega)}^2 - c \left[\sqrt{\delta} \|\nabla^2 u\|_{L^2(\Omega)}^2 + \frac{c_P^2}{\sqrt{\delta}} \|\Delta u\|_{L^2(\Omega)}^2 \right].$$

Note that this holds for any $\delta \in (0, 1)$. We now choose δ such that $1 - c\sqrt{\delta}$ is positive and we get

$$\|\nabla^2 u\|_{L^2(\Omega)} \leq c \, \|\Delta u\|_{L^2(\Omega)}^2 \quad \forall \, u \in H^3(\Omega) \cap H_0^1(\Omega).$$

We that note due to the boundary condition and the Poincaré inequality it follows that $\|\nabla^2 u\|_{L^2(\Omega)}$ is equivalent to the H^2-norm. So, we have now shown inequality (2.8) for $u \in H^3(\Omega) \cap H_0^1(\Omega)$. Since $H^3(\Omega) \cap H_0^1(\Omega)$ is dense in $H^2(\Omega) \cap H_0^1(\Omega)$ (Lemma A.1) the result also holds for all $u \in H^2(\Omega) \cap H_0^1(\Omega)$. □

Appendix B. Approximation error estimates for B-splines. In this Appendix, we prove Theorem 5.5 and some auxiliary results required for that proof. We consider B-splines with maximum smoothness on the parameter domain $\widehat{\Omega} := (0,1)^d$, that is, we consider the space $S^d_{p,\ell}$. We point out that for functions in $H^2((0,1)^d) \cap H^1((0,1)^d)$ the H^2-semi-norm and L^2-norm of the Laplacian coincide, that is,

$$\|\Delta u\|_{L^2(0,1)^d} = \|\nabla^2 u\|_{L^2(0,1)^d} \quad \forall\, H^2((0,1)^d) \cap H^1((0,1)^d).$$

For any $d \in \mathbb{N}$ and $p \in \mathbb{N}$ with $p \geq 3$, let $\mathbf{\Pi}_p : H^2((0,1)^d) \cap H^1_0((0,1)^d) \to S^d_{p,\ell} \cap H^1_0((0,1)^d)$ be the H^2-orthogonal projector, defined by

$$(\Delta \mathbf{\Pi}_p u, \Delta \tilde{u})_{L^2((0,1)^d)} = (\Delta u, \Delta \tilde{u})_{L^2((0,1)^d)} \quad \forall\, \tilde{u} \in S^d_{p,\ell} \cap H^1_0.$$

To better distinguish the univariate case ($d = 1$), we write $\Pi_p := \mathbf{\Pi}_p$ for that case.

Theorem B.1. *Let $d \in \mathbb{N}$ and $p \in \mathbb{N}$ with $p \geq 3$. Then there exits a constant $c > 0$ such that*

$$\|u - \mathbf{\Pi}_p u\|_{L^2((0,1)^d)} \leq c\, h^2 \|\Delta u\|_{L^2((0,1)^d)} \quad \forall\, u \in H^2((0,1)^d) \cap H^1_0((0,1)^d).$$

Proof. Let $u \in H^2((0,1)^d) \cap H^1_0((0,1)^d)$ be arbitrary but fixed. [23, Theorem 9.3] states that

$$\|u - \widetilde{\mathbf{\Pi}}_p u\|_{L^2((0,1)^d)} \leq c\, h^2 \|\Delta u\|_{L^2((0,1)^d)},$$

where $\widetilde{\mathbf{\Pi}}_p : H^2((0,1)^d) \cap H^1_0((0,1)^d) \to \widetilde{S}$ is the H^2-orthogonal projector into some space $\widetilde{S} \subset S^d_{p,\ell}$. Using $\widetilde{\mathbf{\Pi}}_p \mathbf{\Pi}_p = \widetilde{\mathbf{\Pi}}_p$, the triangle inequality and the stability statement $\|\Delta \mathbf{\Pi}_p u\|_{L^2((0,1)^d)} \leq \|\Delta u\|_{L^2((0,1)^d)}$, we immediately obtain the desired result. □

Next, we provide an H^2–H^4 error estimate for the univariate case.

Theorem B.2. *Let $p \in \mathbb{N}$ with $p \geq 3$. Then,*

$$\|\partial^2 (u - \Pi_p u)\|_{L^2(0,1)} \leq 2h^2 \|\partial^4 u\|_{L^2(0,1)} \quad \forall\, H^4(0,1) \cap H^1_0(0,1).$$

Proof. See [24, Theorem 3]. □

We define projectors $\Pi_p^{x_k}$ on $C^\infty((0,1)^d)$ as follows:

$$(\Pi_p^{x_k}) u(x_1, \ldots, x_{k-1}, \cdot, x_{k+1}, \ldots, x_d) := \Pi_p u(x_1, \ldots, x_{k-1}, \cdot, x_{k+1}, \ldots, x_d)$$
$$\forall\, (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_d) \in (0,1)^{d-1} \quad \text{for} \quad k = 1, \ldots, d.$$

These projectors act on one variable. We also introduce projectors $\mathbf{\Pi}_p^{x_k}$ that act on every variable except one, which are given by

$$(\mathbf{\Pi}_p^{x_k}) u(\cdot, \ldots, \cdot, x_k, \cdot, \ldots, \cdot) := \mathbf{\Pi}_p u(\cdot, \ldots, \cdot, x_k, \cdot, \ldots, \cdot)$$
$$\forall\, x_k \in (0,1) \quad \text{for} \quad k = 1, \ldots, d.$$

Similarly we define a Laplace operator on the form

$$\Delta^{x_k} = \sum_{i \in \{1, \ldots, d\} \setminus \{k\}} \partial_{x_i x_i}, \quad \text{where} \quad \partial_{x_i x_i} := \frac{\partial^2}{\partial^2_{x_i}}.$$

We note that all projectors are commutative, cf. [26]. Using this notation, we can extend Theorem B.2 to an arbitrary number of dimensions.

Theorem B.3. *Let $d \in \mathbb{N}$ and $p \in \mathbb{N}$ with $p \geq 3$. Then, there exits a constant $c > 0$ such that*

$$|u - \mathbf{\Pi}_p u|_{H^2((0,1)^d)} \leq c\,h^2 |u|_{H^4((0,1)^d)} \quad \forall\, u \in H^4((0,1)^d) \cap H^1_0((0,1)^d).$$

Proof. Assume first that $u \in C^\infty((0,1)^d) \cap H^1_0((0,1)^d)$. We prove the statement of the theorem using induction with respect to d. Assume that it holds for $d-1$, that is,

$$\|\Delta^{x_k}(u - \mathbf{\Pi}_p^{x_k} u)\| \leq c\,h^2 \|\Delta^{x_k}\Delta^{x_k} u\| \quad \text{for} \quad k = 1, \dots, d. \tag{B.1}$$

Here and in what follows, all norms are $L^2((0,1)^d)$-norms unless stated otherwise. Now, we show that the statement holds also for d. By using the fact that $\mathbf{\Pi}_p$ minimizes the H^2-semi-norm (Laplace norm) and $\|\Delta u\|^2 \leq d \sum_{j=1}^d \|\partial_{x_j x_j} u\|^2$, we get

$$\|\Delta(u - \mathbf{\Pi}_p u)\|^2 \leq \frac{1}{d} \sum_{k=1}^d \|\Delta(u - \mathbf{\Pi}_p^{x_k} \Pi_p^{x_k} u)\|^2 \leq \sum_{k=1}^d \sum_{j=1}^d \|\partial_{x_j x_j}(u - \mathbf{\Pi}_p^{x_k} \Pi_p^{x_k} u)\|^2.$$

We separate this into two groups: $j = k$ and $j \neq k$. We start with $j = k$. Using the triangle inequality, the commutativity of the two projectors, Theorem B.2, the H^2-stability of $\Pi_p^{x_k}$, and the fact that that $\partial_{x_k x_k}$ and $\mathbf{\Pi}_p^{x_k}$ are commutative, we obtain

$$\begin{aligned}\sum_{k=1}^d \|\partial_{x_k x_k}(u - \mathbf{\Pi}_p^{x_k} \Pi_p^{x_k} u)\|^2 &\leq 2 \sum_{k=1}^d \left(\|\partial_{x_k x_k}(u - \Pi_p^{x_k} u)\|^2 + \|\partial_{x_k x_k} \Pi_p^{x_k}(u - \mathbf{\Pi}_p^{x_k} u)\|^2 \right) \\ &\leq 2 \sum_{k=1}^d \left(4h^4 \|\partial_{x_k x_k x_k x_k} u\|^2 + \|(I - \mathbf{\Pi}_p^{x_k})\partial_{x_k x_k} u\|^2 \right).\end{aligned}$$

Now, we use Theorem B.1 to obtain

$$\sum_{k=1}^d \|\partial_{x_k x_k}(u - \mathbf{\Pi}_p^{x_k} \Pi_p^{x_k} u)\|^2 \leq c\,h^4 \sum_{k=1}^d \left(\|\partial_{x_k x_k x_k x_k} u\|^2 + \|\Delta^{x_k} \partial_{x_k x_k} u\|^2 \right). \tag{B.2}$$

For the second group ($j \neq k$), we use the triangle inequality, the induction hypothesis (B.1) and the H^2-stability of $\mathbf{\Pi}^{x_k}$ to obtain

$$\begin{aligned}\sum_{j \neq k} \|\partial_{x_j x_j}(u - \mathbf{\Pi}_p^{x_k} \Pi_p^{x_k} u)\|^2 &\leq 2 \sum_{j \neq k} \left(\|\partial_{x_j x_j}(u - \mathbf{\Pi}_p^{x_k} u)\|^2 + \|\partial_{x_j x_j} \mathbf{\Pi}_p^{x_k}(u - \Pi_p^{x_k} u)\|^2 \right) \\ &\leq 2 \left(\|\Delta^{x_k}(u - \mathbf{\Pi}_p^{x_k} u)\|^2 + \|\Delta^{x_k} \mathbf{\Pi}_p^{x_k}(u - \Pi_p^{x_k} u)\|^2 \right) \\ &\leq c \left(h^4 \|\Delta^{x_k}\Delta^{x_k} u)\|^2 + \|(I - \Pi_p^{x_k})\Delta^{x_k} u\|^2 \right).\end{aligned}$$

Again, we use Theorem B.2 and obtain

$$\sum_{k=1}^d \sum_{j \neq k} \|\partial_{x_j x_j}(u - \mathbf{\Pi}_p^{x_k} \Pi_p^{x_k} u)\|^2 \leq c\,h^4 \sum_{k=1}^d \left(\|\Delta^{x_k}\Delta^{x_k} u)\|^2 + \|\partial_{x_k x_k} \Delta^{x_k} u\|^2 \right).$$

Combining this with (B.2), we finally get

$$\begin{aligned}\|\Delta(u - \mathbf{\Pi}_p u)\|^2 &\leq c\,h^4 \sum_{k=1}^d \left(\|\Delta^{x_k}\Delta^{x_k} u)\|^2 + 2\|\partial_{x_k x_k} \Delta^{x_k} u\|^2 + \|\partial_{x_k x_k x_k x_k} u\|^2 \right) \\ &\leq c\,h^4 \|\Delta\Delta u\|^2 \quad \forall\, u \in C^\infty((0,1)^d) \cap H^1_0((0,1)^d).\end{aligned}$$

Note that $\|\Delta\Delta u\| \leq \sqrt{d}\,|u|_{H^4((0,1)^d)}$. Using a standard density argument, we obtain the result also for $u \in H^4((0,1)^d) \cap H_0^1((0,1)^d)$. □

Using interpolation theory (cf. [6]) and the $H^2 - H^4$ result above, we obtain a $H^2 - H^3$ result, cf. [24, Theorem 6].

Theorem B.4. *Let $d \in \mathbb{N}$ and $p \in \mathbb{N}$ with $p \geq 3$. Then there exits a constant c such that*

$$|u - \mathbf{\Pi}_p u|_{H^2((0,1)^d)} \leq c\,h|u|_{H^3((0,1)^d)} \quad \forall\, u \in H^3((0,1)^d) \cap H_0^1((0,1)^d).$$

We also use interpolation theory and the $L^2 - H^2$ result (Theorem B.1) to obtain a $H^1 - H^2$ result.

Theorem B.5. *Let $d \in \mathbb{N}$ and $p \in \mathbb{N}$ with $p \geq 3$. Then, there exits a constant c such that*

$$|u - \mathbf{\Pi}_p u|_{H^1((0,1)^d)} \leq c\,h|u|_{H^2((0,1)^d)} \quad \forall\, u \in H^2((0,1)^d) \cap H_0^1((0,1)^d).$$

By combining these auxiliary results, we can prove Theorem 5.5:

Proof. Inequality (5.6) is proven in Theorem B.4. For the inequality (5.7), we combine Theorem B.5 and Theorem B.4 as follows:

$$\|\nabla(I - \mathbf{\Pi}_p)u\| = \|\nabla(I - \mathbf{\Pi}_p)(I - \mathbf{\Pi}_p)u\| \leq c\,h\|\nabla^2(I - \mathbf{\Pi}_p)u\| \leq c\,h^2\|\nabla^3 u\|,$$

where $\|\cdot\|$ again denotes the L^2-norm. Finally, the inequality (5.8) is proven by combining Theorem B.1 and Theorem B.4 as follows:

$$\|(I - \mathbf{\Pi}_p)u\| = \|(I - \mathbf{\Pi}_p)(I - \mathbf{\Pi}_p)u\| \leq c\,h^2\|\nabla^2(I - \mathbf{\Pi}_p)u\| \leq c\,h^3\|\nabla^3 u\|.$$

This concludes the proof. □

REFERENCES

[1] I. Babuška, *Error-bounds for finite element method*, Numerische Mathematik, 16 (1971), pp. 322–333.

[2] C. E. Baumann and J. T. Oden, *A discontinuous hp finite element method for convection-diffusion problems*, Computer Methods in Applied Mechanics and Engineering, 175 (1999), pp. 311–341.

[3] R. Becker and B. Vexler, *Optimal control of the convection-diffusion equation using stabilized finite element methods*, Numerische Mathematik, 106 (2007), pp. 349–367.

[4] A. Beigl, J. Sogn, and W. Zulehner, *Robust preconditioners for multiple saddle point problems and applications to optimal control problems*, SIAM Journal on Matrix Analysis and Applications, 41 (2020), pp. 1590–1615.

[5] L. Beirão da Veiga, A. Buffa, G. Sangalli, and R. Vázquez, *Mathematical analysis of variational isogeometric methods*, Acta Numerica, 23 (2014), pp. 157–287.

[6] J. Bergh and J. Löfström, *Interpolation spaces: an introduction*, vol. 223, Springer Science & Business Media, 2012.

[7] A. N. Brooks and T. J. Hughes, *Streamline upwind/petrov-galerkin formulations for convection dominated flows with particular emphasis on the incompressible navier-stokes equations*, Computer methods in applied mechanics and engineering, 32 (1982), pp. 199–259.

[8] G. Chen, W. Hu, J. Shen, J. R. Singler, Y. Zhang, and X. Zheng, *An HDG method for distributed control of convection diffusion PDEs*, Journal of Computational and Applied Mathematics, 343 (2018), pp. 643–661.

[9] B. Cockburn and C.-W. Shu, *The local discontinuous galerkin method for time-dependent convection-diffusion systems*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 2440–2463.

[10] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Oxford University Press, USA, 2014.

[11] P. Grisvard, *Singularities in boundary value problems*, Springer, Berlin, 1992.

[12] P. Grisvard, *Elliptic Problems in Nonsmooth Domains. Reprint of the 1985 hardback ed.*, Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2011.

[13] M. Hinze, N. Yan, and Z. Zhou, *Variational discretization for optimal control governed by convection dominated diffusion equations*, Journal of Computational Mathematics, (2009), pp. 237–253.

[14] T. Hughes, J. Cottrell, and Y. Bazilevs, *Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement*, Computer Methods in Applied Mechanics and Engineering, 194 (2005), pp. 4135–4195.

[15] O. A. Ladyzhenskaya, *The boundary value problems of mathematical physics*, vol. 49, Springer Science & Business Media, 2013.

[16] J. Málek and Z. Strakoš, *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs*, vol. 1, SIAM, 2014.

[17] K.-A. Mardal, B. F. Nielsen, and M. Nordaas, *Robust preconditioners for PDE-constrained optimization with limited observations*, BIT, 57 (2017), pp. 405–431.

[18] K.-A. Mardal and R. Winther, *Preconditioning discretizations of systems of partial differential equations*, Numerical Linear Algebra with Applications, 18 (2011), pp. 1–40.

[19] J. W. Pearson and A. J. Wathen, *A new approximation of the Schur complement in preconditioners for PDE-constrained optimization*, Numerical Linear Algebra with Applications, 19 (2012), pp. 816–829.

[20] M. Porcelli, V. Simoncini, and M. Tani, *Preconditioning of active-set Newton methods for PDE-constrained optimal control problems*, SIAM Journal on Scientific Computing, 37 (2015), pp. S472–S502.

[21] A. Quarteroni et al., *Optimal control and numerical adaptivity for advection–diffusion equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 39 (2005), pp. 1019–1040.

[22] E. Sande, C. Manni, and H. Speleers, *Explicit error estimates for spline approximation of arbitrary smoothness in isogeometric analysis*, Numerische Mathematik, (2020), pp. 1–41.

[23] J. Sogn, *Schur complement preconditioners for multiple saddle point problems and applications*, PhD thesis, Johannes Kepler University Linz, 2018.

[24] J. Sogn and S. Takacs, *Robust multigrid solvers for the biharmonic problem in isogeometric analysis*, Computers & Mathematics with Applications, 77 (2019), pp. 105–124.

[25] J. Sogn and W. Zulehner, *Schur complement preconditioners for multiple saddle point problems of block tridiagonal form with application to optimization problems*, IMA Journal of Numerical Analysis, 39 (2018), pp. 1328–1359.

[26] S. Takacs, *Robust approximation error estimates and multigrid solvers for isogeometric multi-patch discretizations*, Mathematical Models and Methods in Applied Sciences, 28 (2018), pp. 1899–1928.