SIAM J. MATRIX ANAL. APPL.
Vol. 41, No. 4, pp. 1590–1615

ROBUST PRECONDITIONERS FOR MULTIPLE SADDLE POINT PROBLEMS AND APPLICATIONS TO OPTIMAL CONTROL PROBLEMS*

ALEXANDER BEIGL†, JARLE SOGN‡, AND WALTER ZULEHNER§

Abstract. In this paper we consider multiple saddle point problems with block tridiagonal Hessian in a Hilbert space setting. Well-posedness and the related issue of preconditioning are discussed. We give a characterization of all block structured norms which ensure well-posedness of multiple saddle point problems as a helpful tool for constructing block diagonal preconditioners. We subsequently apply our findings to a general class of PDE-constrained optimal control problems containing a regularization parameter α and derive α-robust preconditioners for the corresponding optimality systems. Finally, we demonstrate the generality of our approach with two optimal control problems related to the heat and the wave equation, respectively. Preliminary numerical experiments support the feasibility of our method.

Key words. saddle point problems, PDE-constrained optimization, optimal control, robust preconditioning

AMS subject classifications. 49J20, 49K20, 65F08, 65N22

DOI. 10.1137/19M1308426

1. Introduction.

In this paper we discuss the well-posedness of a particular class of saddle point problems in function spaces and the related topic of robust preconditioning. We consider linear operator equations

$$\mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \tag{1.1}$$

where $\mathcal{A} : \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ is a self-adjoint operator mapping from the product space $\boldsymbol{X} = X_1 \times X_2 \times \cdots \times X_n$ of Hilbert spaces X_i into its dual space $\boldsymbol{X}'$.

In particular, we are interested in the case where $\mathcal{A}\colon \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ is of n-by-n block tridiagonal form

$$\mathcal{A} = \begin{pmatrix} A_1 & B_1' & & \\ B_1 & -A_2 & \ddots & \\ & \ddots & \ddots & B_{n-1}' \\ & & B_{n-1} & (-1)^{n-1}A_n \end{pmatrix},$$

where $A_i\colon X_i \longrightarrow X_i'$, $B_i\colon X_i \longrightarrow X_{i+1}'$ are bounded linear operators, B_i' is the adjoint of B_i, and, additionally, A_i are self-adjoint and positive semidefinite. Under these

*Received by the editors December 23, 2019; accepted for publication (in revised form) by D. Orban July 9, 2020; published electronically October 14, 2020.
https://doi.org/10.1137/19M1308426
Funding: The work of the first author was supported by the Austrian Science Fund (FWF) project I3661-N27. The work of the second author was supported by the Austrian Science Fund (FWF) project P31048. The work of the third author was supported by the Austrian Science Fund (FWF) project S11702-N23.

†Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria (alexander.beigl@univie.ac.at).
‡Johann Radon Institute for Computational Mathematics (RICAM), Johannes Kepler University, 4040 Linz, Austria (jarle.sogn@ricam.oeaw.ac.at).
§Institute of Computational Mathematics, Johannes Kepler University, 4040 Linz, Austria (zulehner@numa.uni-linz.ac.at).

assumptions, the solutions to (1.1) are the multiple saddle points of the associated Lagrangian functional

$$\mathcal{L}(\boldsymbol{x}) = \frac{1}{2}\langle \mathcal{A}\boldsymbol{x}, \boldsymbol{x}\rangle - \langle \boldsymbol{b}, \boldsymbol{x}\rangle ,$$

where $\langle\cdot,\cdot\rangle$ denotes the duality product; see [19] for more details of this interpretation. The special case $n = 2$, where (1.1) is of the form

$$\begin{pmatrix} A & B' \\ B & -C \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1.2}$$

is usually referred to as a classical saddle point problem.

Saddle point problems in infinite-dimensional Hilbert spaces arise as the optimality systems of optimization problems in function spaces with a quadratic objective functional and constrained by a partial differential equation (PDE) or a system of PDEs. Other sources for such problems are mixed formulations of elliptic boundary value problems. For numerous applications of classical saddle point problems we refer to the seminal survey article [4] and for applications of multiple saddle point problem we refer to [19].

Classical saddle point problems ($n = 2$) are well-studied (see [6]). For $C = 0$, the well-known Brezzi conditions are sufficient and necessary conditions for well-posedness. This is generalized in [20], where sufficient and necessary conditions, including the case $C \neq 0$, are provided. The conditions in [20] also provided conditions for a robust preconditioner in the framework of operator preconditioning.

Multiple saddle point problems ($n > 2$) are less studied than classical saddle point problems. In [19] a block diagonal preconditioner was introduced whose diagonal blocks consist of a sequence of so-called Schur complements. The well-posedness of (1.1) could be shown with respect to the associated norm with robust estimates. However, Schur complements do not always exist. This already becomes apparent in the well-studied case (1.2), where A needs to be invertible only on the kernel of B. Then, of course, A might be not invertible and, consequently, the classical Schur complement $S = C + BA^{-1}B'$ would not exist. Therefore, a more general approach is undertaken here, where we consider general block diagonal preconditioners rather than the more restrictive class of preconditioners based on Schur complements.

An important field of applications are optimality systems of PDE-constrained optimization problems. In particular, optimal control problems are considered with objective functionals which contain a regularization term involving some regularization parameter α. Suitable Krylov subspace methods, e.g., the minimum residual method (MINRES), for solving the corresponding (discretized) optimality systems deteriorate for small α when using standard preconditioners. For most practical applications we have $0 < \alpha \ll 1$, and thus finding α-robust preconditioners is essential. For optimal control problems with an elliptic state equation α-robust preconditioners are provided by [18, 17, 15, 19]. Some time-depending problems are addressed in [16, 14]; however, the rigorous analysis of α-robust preconditioners always required full observation (observation throughout the whole domain). A special case with a hyperbolic state equation was studied in [2]. There a robust preconditioner was obtained also for a problem with partial observation. Based on a new abstract theory we will derive α-robust preconditioners without requiring full observation. The class of problems covered by the new approach include optimal control problems with elliptic, parabolic, or hyperbolic state equations. The work presented here can be seen as an extension of ideas presented in [15].

The paper is organized as follows. In section 2 the well-posedness of (1.1) is addressed in general Hilbert spaces. The main result is contained in Theorem 2.2, which provides a characterization of robust block diagonal preconditioners for (1.1). This result can be seen as an extension of corresponding results in [20] to multiple saddle point problems. In section 3 the application of the abstract results to optimal control problems are discussed in general. Section 4 contains particular examples of optimal control problems with parabolic, respectively, hyperbolic state equations. Preliminary numerical results are reported in section 5. Finally, an auxiliary result needed for the abstract analysis is provided in Appendix A.

2. Abstract theory. We introduce some notation which will be used throughout the paper.

Notation 1. For a real Hilbert space X with inner product $(\cdot,\cdot)_X$, the duality pairing in its dual space X' will be denoted by $\langle\cdot,\cdot\rangle_X$, where we omit the subscript when the space is clear from the context.

For a bounded linear operator $B : X \longrightarrow Y'$, where X and Y are Hilbert spaces, its adjoint $B' : Y \longrightarrow X'$ is given by

$$\langle B'y, x\rangle = \langle Bx, y\rangle \quad \text{for all } x \in X,\ y \in Y.$$

A bounded linear operator $A : X \longrightarrow X'$ is said to be self-adjoint, respectively, positive semidefinite, if

$$\langle Ay, x\rangle = \langle Ax, y\rangle, \quad \text{resp.} \quad \langle Ax, x\rangle \geq 0, \quad \text{for all } x, y \in X.$$

The operator A is positive definite (coercive) if

$$\langle Ax, x\rangle \geq \sigma \|x\|_X^2 \quad \text{for all } x \in X$$

for some positive constant σ.

Let $\boldsymbol{X} = X_1 \times X_2 \times \cdots \times X_n$ be the product space of Hilbert spaces X_i for $i = 1, 2, \ldots, n$, endowed with the canonical inner product

$$(\boldsymbol{x}, \boldsymbol{y})_{\boldsymbol{X}} = (x_1, y_1)_{X_1} + (x_2, y_2)_{X_2} + \cdots + (x_n, y_n)_{X_n},$$

and let the linear operator $\mathcal{A}\colon \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ be of n-by-n block tridiagonal form

$$\mathcal{A} = \begin{pmatrix} A_1 & B_1' & & \\ B_1 & -A_2 & \ddots & \\ & \ddots & \ddots & B_{n-1}' \\ & & B_{n-1} & (-1)^{n-1}A_n \end{pmatrix},$$

where $A_i\colon X_i \longrightarrow X_i'$, $B_i\colon X_i \longrightarrow X_{i+1}'$ are bounded linear operators, and, additionally, A_i are self-adjoint and positive semidefinite. Here, as usual, we identify the dual space $\boldsymbol{X}'$ with $X_1' \times X_2' \times \cdots \times X_n'$.

For a given right-hand side $\boldsymbol{b} \in \boldsymbol{X}'$, we consider the linear operator equation

$$\mathcal{A}\boldsymbol{x} = \boldsymbol{b}. \tag{2.1}$$

We introduce two linear operators associated to $\mathcal{A}$:

$$\mathcal{D} = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_n \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} 0 & B_1' & & \\ B_1 & 0 & \ddots & \\ & \ddots & \ddots & B_{n-1}' \\ & & B_{n-1} & 0 \end{pmatrix}.$$

Observe that $\mathcal{D}$ and $\mathcal{B}$ are self-adjoint, and, additionally, $\mathcal{D}$ is positive semidefinite. Furthermore, let

$$\widetilde{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ -x_2 \\ \vdots \\ (-1)^{n-1} x_n \end{pmatrix} \text{ for } \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \widetilde{\mathcal{D}} = \begin{pmatrix} A_1 & & & \\ & -A_2 & & \\ & & \ddots & \\ & & & (-1)^{n-1} A_n \end{pmatrix}.$$

This notation is used in the following analysis.

We start with the analysis of the uniqueness of a solution to (2.1).

LEMMA 2.1. $\ker \mathcal{A} = \ker \mathcal{D} \cap \ker \mathcal{B}$.

Proof. With the notation introduced above we have for all $\boldsymbol{x} \in \boldsymbol{X}$

$$\langle \mathcal{A}\boldsymbol{x}, \widetilde{\boldsymbol{x}} \rangle = \langle (\widetilde{\mathcal{D}} + \mathcal{B})\boldsymbol{x}, \widetilde{\boldsymbol{x}} \rangle = \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x} \rangle + \langle \mathcal{B}\boldsymbol{x}, \widetilde{\boldsymbol{x}} \rangle = \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x} \rangle.$$

Therefore, if $\boldsymbol{x} \in \ker \mathcal{A}$, then

$$\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \mathcal{A}\boldsymbol{x}, \widetilde{\boldsymbol{x}} \rangle = 0,$$

which implies that $\boldsymbol{x} \in \ker \mathcal{D}$, since $\mathcal{D}$ is self-adjoint and positive semidefinite. Furthermore, since $\ker \widetilde{\mathcal{D}} = \ker \mathcal{D}$, it follows that $\mathcal{B}\boldsymbol{x} = \mathcal{A}\boldsymbol{x} - \widetilde{\mathcal{D}}\boldsymbol{x} = 0$. This concludes the proof of $\ker \mathcal{A} \subset \ker \mathcal{D} \cap \ker \mathcal{B}$.

On the other hand, if $\boldsymbol{x} \in \ker \mathcal{D} \cap \ker \mathcal{B}$, then $\boldsymbol{x} \in \ker \widetilde{\mathcal{D}}$ and, consequently, $\mathcal{A}\boldsymbol{x} = \widetilde{\mathcal{D}}\boldsymbol{x} + \mathcal{B}\boldsymbol{x} = 0$, which shows that $\ker \mathcal{D} \cap \ker \mathcal{B} \subset \ker \mathcal{A}$. ∎

The next theorem deals with the well-posedness of (2.1).

THEOREM 2.2. *If there are positive constants $\underline{c}$ and $\overline{c}$ such that*

$$\underline{c}\,\|\boldsymbol{x}\|_{\boldsymbol{X}} \le \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \le \overline{c}\,\|\boldsymbol{x}\|_{\boldsymbol{X}} \quad \textit{for all } \boldsymbol{x} \in \boldsymbol{X}, \tag{2.2}$$

then

$$\underline{\gamma}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2 \le \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x} \rangle + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 \le \overline{\gamma}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2 \quad \textit{for all } \boldsymbol{x} \in \boldsymbol{X} \tag{2.3}$$

with positive constants $\underline{\gamma}$ and $\overline{\gamma}$ which depend only on $\underline{c}$ and $\overline{c}$. Vice versa, if there are positive constants $\underline{\gamma}$ and $\overline{\gamma}$ such that (2.3) *holds, then* (2.2) *holds with positive constants $\underline{c}$ and $\overline{c}$ which depend only on $\underline{\gamma}$ and $\overline{\gamma}$.*

Proof. First we show that (2.2) implies (2.3). For the estimate from above in (2.3) observe that

$$\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \mathcal{A}\boldsymbol{x}, \widetilde{\boldsymbol{x}} \rangle \le \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'}\, \|\widetilde{\boldsymbol{x}}\|_{\boldsymbol{X}} \le \overline{c}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}\, \|\widetilde{\boldsymbol{x}}\|_{\boldsymbol{X}} = \overline{c}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2.$$

In order to estimate $\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}$ we use

$$\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} = \|(\mathcal{A} - \widetilde{\mathcal{D}})\boldsymbol{x}\|_{\boldsymbol{X}'} \le \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} + \|\widetilde{\mathcal{D}}\boldsymbol{x}\|_{\boldsymbol{X}'} = \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} + \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}.$$

Since

$$\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{y}\rangle^2 \le \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \, \langle \mathcal{D}\boldsymbol{y}, \boldsymbol{y}\rangle \le \overline{c}^2 \, \|\boldsymbol{x}\|_{\boldsymbol{X}}^2 \, \|\boldsymbol{y}\|_{\boldsymbol{X}}^2,$$

it follows that

$$\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'} = \sup_{\boldsymbol{0}\neq\boldsymbol{y}\in\boldsymbol{X}} \frac{\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_{\boldsymbol{X}}} \le \overline{c} \, \|\boldsymbol{x}\|_{\boldsymbol{X}},$$

which allows one to complete the estimate of $\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}$:

$$\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} \le \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} + \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'} \le 2\,\overline{c} \, \|\boldsymbol{x}\|_{\boldsymbol{X}}.$$

The estimates of $\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle$ and $\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}$ lead directly to the estimate from above in (2.3) with $\overline{\gamma} = \overline{c} + 4\,\overline{c}^2$.

For showing the estimate from below in (2.3) we start with the following argument,

$$\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{y}\rangle^2 \le \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \, \langle \mathcal{D}\boldsymbol{y}, \boldsymbol{y}\rangle \le \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \, \|\mathcal{D}\boldsymbol{y}\|_{\boldsymbol{X}'} \|\boldsymbol{y}\|_{\boldsymbol{X}} \le \overline{c} \, \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \, \|\boldsymbol{y}\|_{\boldsymbol{X}}^2,$$

which implies

$$\|\mathcal{D}\boldsymbol{x}\|^2 = \sup_{\boldsymbol{0}\neq\boldsymbol{y}\in\boldsymbol{X}} \frac{\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{y}\rangle^2}{\|\boldsymbol{y}\|_{\boldsymbol{X}}^2} \le \overline{c} \, \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle.$$

Therefore,

$$\begin{aligned} \underline{c} \, \|\boldsymbol{x}\|_{\boldsymbol{X}} \le \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} = \|(\widetilde{\mathcal{D}} + \mathcal{B})\boldsymbol{x}\|_{\boldsymbol{X}'} \le \|\widetilde{\mathcal{D}}\boldsymbol{x}\|_{\boldsymbol{X}'} + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} = \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'} + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} \\ \le \overline{c}^{1/2} \, \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle^{1/2} + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} \le (\overline{c}+1)^{1/2} \, \left(\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2\right)^{1/2}, \end{aligned}$$

from which the estimate from below in (2.3) follows for $\underline{\gamma} = \underline{c}^2/(\overline{c}+1)$.

It remains to show that (2.3) implies (2.2). For the estimate from above in (2.2) we again use the triangle inequality and obtain

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \le \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'} + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}$$

(see above). Since

$$\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{y}\rangle^2 \le \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \, \langle \mathcal{D}\boldsymbol{y}, \boldsymbol{y}\rangle \le \overline{\gamma} \, \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \, \|\boldsymbol{y}\|_{\boldsymbol{X}}^2,$$

it follows that

$$\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 = \sup_{\boldsymbol{0}\neq\boldsymbol{y}\in\boldsymbol{X}} \frac{\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{y}\rangle^2}{\|\boldsymbol{y}\|_{\boldsymbol{X}}^2} \le \overline{\gamma} \, \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle, \tag{2.4}$$

which allows one to complete the estimate of $\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'}$:

$$\begin{aligned} \|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \le \overline{\gamma}^{1/2} \, \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle^{1/2} + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} \le (\overline{\gamma}+1)^{1/2} \, \left(\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2\right)^{1/2} \\ \le (\overline{\gamma}+1)^{1/2} \, \overline{\gamma}^{1/2} \, \|\boldsymbol{x}\|_{\boldsymbol{X}}. \end{aligned}$$

Then the estimate from above in (2.2) follows for $\overline{c} = [\overline{\gamma}\,(\overline{\gamma}+1)]^{1/2}$.

For the estimate from below in (2.2), we start with the following two estimates:

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \ge \left| \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} - \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'} \right| \tag{2.5}$$

and

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \ge \frac{\langle \mathcal{A}\boldsymbol{x}, \widetilde{\boldsymbol{x}}\rangle}{\|\widetilde{\boldsymbol{x}}\|_{\boldsymbol{X}}} = \frac{\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle}{\|\boldsymbol{x}\|_{\boldsymbol{X}}} \ge \frac{1}{\overline{\gamma}} \, \frac{\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2}{\|\boldsymbol{x}\|_{\boldsymbol{X}}} \quad \text{for } \boldsymbol{x} \neq \boldsymbol{0}. \tag{2.6}$$

The first estimate follows from the triangle inequality. For the second estimate we used (2.4). Next we need to estimate $\|\boldsymbol{x}\|_{\boldsymbol{X}}$ from above in terms of $\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}$ and $\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}$: From (2.3) and

$$\langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle \leq \frac{1}{2\,\varepsilon}\,\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \frac{\varepsilon}{2}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2$$

it follows that

$$\underline{\gamma}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2 \leq \langle \mathcal{D}\boldsymbol{x}, \boldsymbol{x}\rangle + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 \leq \frac{1}{2\,\varepsilon}\,\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \frac{\varepsilon}{2}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2.$$

For $\varepsilon = \underline{\gamma}$ we obtain

$$\frac{\underline{\gamma}}{2}\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2 \leq \frac{1}{2\underline{\gamma}}\,\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 \leq \max\left(\frac{1}{2\underline{\gamma}}, 1\right)\,\left(\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2\right),$$

which implies

$$\delta\,\|\boldsymbol{x}\|_{\boldsymbol{X}}^2 \leq \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 \quad \text{with} \quad \delta = \min\left(\underline{\gamma}^2, \underline{\gamma}/2\right).$$

With this estimate we obtain from the estimates (2.5) and (2.6) for $\boldsymbol{x} \neq 0$

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \geq \left|\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'} - \|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}\right| = |\eta - \xi|\,(\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2)^{1/2} \geq \delta^{1/2}\,|\eta - \xi|\,\|\boldsymbol{x}\|_{\boldsymbol{X}}$$

and

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \geq \frac{1}{\overline{\gamma}}\,\frac{\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2}{\|\boldsymbol{x}\|_{\boldsymbol{X}}} = \frac{1}{\overline{\gamma}}\,\xi^2\,\frac{\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2}{\|\boldsymbol{x}\|_{\boldsymbol{X}}} \geq (\delta/\overline{\gamma})\,\xi^2\,\|\boldsymbol{x}\|_{\boldsymbol{X}}$$

with

$$\xi = \frac{\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}}{(\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2)^{1/2}} \quad \text{and} \quad \eta = \frac{\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}}{(\|\mathcal{D}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 + \|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2)^{1/2}}.$$

Note that ξ and η are well-defined for $x \notin \ker \mathcal{A}$ by Lemma 2.1.

By combining these two estimates we obtain

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \geq (\delta/\overline{\gamma})\,\max\left(|\eta - \xi|, \xi^2\right)\,\|\boldsymbol{x}\|_{\boldsymbol{X}},$$

where we used that

$$\delta/\overline{\gamma} = \begin{cases} \underline{\gamma}^2/\overline{\gamma} \leq \underline{\gamma} = \sqrt{\delta} & \text{if } \underline{\gamma} \leq 1/2, \\ \underline{\gamma}/(2\overline{\gamma}) \leq 1/2 \leq \sqrt{\underline{\gamma}/2} = \sqrt{\delta} & \text{if } \underline{\gamma} \geq 1/2. \end{cases}$$

Observe that $\xi^2 + \eta^2 = 1$ and

$$\varphi(\xi, \eta) \geq \min\{\varphi(x, y)\colon x, y \geq 0,\ x^2 + y^2 = 1\} \geq 0.29$$

with $\varphi(x, y) = \max(|y - x|, x^2)$. Therefore

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'} \geq \underline{c}\,\|\boldsymbol{x}\|_{\boldsymbol{X}} \quad \text{with} \quad \underline{c} = 0.29\,(\delta/\overline{\gamma}) = 0.29\,\min\left(\underline{\gamma}^2, \underline{\gamma}/2\right)/\overline{\gamma},$$

which concludes the proof. ∎

Assume that we have self-adjoint and positive definite bounded linear operators $P_i\colon X_i \longrightarrow X_i'$ inducing inner products on X_i via

$$(x_i, y_i)_{P_i} = \langle P_i x_i, y_i \rangle \quad \text{for all } x_i, y_i \in X_i.$$

Then the block diagonal operator $\mathcal{P}\colon \boldsymbol{X} \longrightarrow \boldsymbol{X}'$, given by

$$\mathcal{P} = \begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_n \end{pmatrix},$$

defines an inner product on $\boldsymbol{X}$, called the $\mathcal{P}$-inner product, by virtue of

$$(\boldsymbol{x}, \boldsymbol{y})_{\mathcal{P}} = \langle \mathcal{P}\boldsymbol{x}, \boldsymbol{y} \rangle \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}.$$

The associated equivalent norm on $\boldsymbol{X}$, called the $\mathcal{P}$-norm, will be denoted by $\|\cdot\|_{\mathcal{P}}$.

With this notation, we want to express (2.3) in a more convenient form. For the Hilbert space $\boldsymbol{X}$ equipped with the $\mathcal{P}$-norm it follows by Lemma A.1 that

$$\|\mathcal{B}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 = \sup_{0 \neq \boldsymbol{y} \in \boldsymbol{X}} \frac{\langle \mathcal{B}\boldsymbol{x}, \boldsymbol{y} \rangle^2}{\langle \mathcal{P}\boldsymbol{y}, \boldsymbol{y} \rangle} = \left\langle \mathcal{B}\mathcal{P}^{-1}\mathcal{B}\boldsymbol{x}, \boldsymbol{x} \right\rangle \quad \text{for all } \boldsymbol{x} \in \boldsymbol{X}.$$

Therefore, the condition (2.3) of Theorem 2.2 can be written in the short form

$$\mathcal{P} \sim \mathcal{D} + \mathcal{B}\mathcal{P}^{-1}\mathcal{B}, \tag{2.7}$$

using the following notation.

Notation 2. Let $M, N : X \longrightarrow X'$ be two self-adjoint bounded linear operators. Then the following hold:

1. $M \leq N$ if and only if
$$\langle Mx, x \rangle \leq \langle Nx, x \rangle \quad \text{for all } x \in X.$$
2. $M \lesssim N$ if and only if there is a constant $c \geq 0$ such that $M \leq cN$.
3. $M \sim N$ if and only if $M \lesssim N$ and $N \lesssim M$. In this case we call M and N spectrally equivalent.

If the operators M and N depend on some parameters (like a regularization parameter α or a discretization parameter h), then we additionally assume that the involved constants are independent of those parameters.

With this notation, Theorem 2.2 offers a result on robust preconditioning of (2.1): For the Hilbert space $\boldsymbol{X}$ equipped with the $\mathcal{P}$-norm, given by a block diagonal operator $\mathcal{P} : \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ satisfying the relation (2.7), there exist parameter-independent constants $\underline{c}$, $\overline{c}$ such that (2.2) holds. Since

$$\|\mathcal{A}\boldsymbol{x}\|_{\boldsymbol{X}'}^2 = \left\langle \mathcal{A}\mathcal{P}^{-1}\mathcal{A}\boldsymbol{x}, \boldsymbol{x} \right\rangle = \left\langle \mathcal{A}\boldsymbol{x}, \mathcal{P}^{-1}\mathcal{A}\boldsymbol{x} \right\rangle = \left\| \mathcal{P}^{-1}\mathcal{A}\boldsymbol{x} \right\|_{\mathcal{P}}^2,$$

well-posedness (2.2) can be written as

$$\underline{c}\,\|\boldsymbol{x}\|_{\mathcal{P}} \leq \|\mathcal{P}^{-1}\mathcal{A}\boldsymbol{x}\|_{\mathcal{P}} \leq \overline{c}\,\|\boldsymbol{x}\|_{\mathcal{P}} \quad \text{for all } \boldsymbol{x} \in \boldsymbol{X}.$$

Consequently, it follows for the condition number

$$\kappa(\mathcal{P}^{-1}\mathcal{A}) = \|\mathcal{P}^{-1}\mathcal{A}\|_{\mathcal{P}}\|(\mathcal{P}^{-1}\mathcal{A})^{-1}\|_{\mathcal{P}} \leq \frac{\overline{c}}{\underline{c}},$$

where here $\|\cdot\|_{\mathcal{P}}$ indicates the operator norm associated to the $\mathcal{P}$-norm.

Therefore, the task of finding a good preconditioner $\mathcal{P} : \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ for the system (2.1) translates to choosing inner products $(\cdot,\cdot)_{P_i}$ on the Hilbert spaces X_i such that the condition (2.7) is satisfied.

We now illustrate (2.7) for the three interesting cases $n \in \{2, 3, 4\}$.

2.1. The case $n = 2$. Let

$$\mathcal{A} = \begin{pmatrix} A_1 & B_1' \\ B_1 & -A_2 \end{pmatrix}, \quad \mathcal{D} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 0 & B_1' \\ B_1 & 0 \end{pmatrix}, \quad \mathcal{P} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}.$$

Then

$$\mathcal{B}\mathcal{P}^{-1}\mathcal{B} = \begin{pmatrix} B_1' P_2^{-1} B_1 & 0 \\ 0 & B_1 P_1^{-1} B_1' \end{pmatrix}$$

and the spectral relation

$$\mathcal{P} \sim \mathcal{D} + \mathcal{B}\mathcal{P}^{-1}\mathcal{B}$$

is equivalent to

$$P_1 \sim A_1 + B_1' P_2^{-1} B_1 \quad \text{and} \quad P_2 \sim A_2 + B_1 P_1^{-1} B_1'.$$

Thus, we recover the result from [20].

2.2. The case $n = 3$. Let

$$\mathcal{A} = \begin{pmatrix} A_1 & B_1' & 0 \\ B_1 & -A_2 & B_2' \\ 0 & B_2 & A_3 \end{pmatrix},$$

$$\mathcal{D} = \begin{pmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 0 & B_1' & 0 \\ B_1 & 0 & B_2' \\ 0 & B_2 & 0 \end{pmatrix}, \quad \mathcal{P} = \begin{pmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & P_3 \end{pmatrix}.$$

Then

$$\mathcal{B}\mathcal{P}^{-1}\mathcal{B} = \begin{pmatrix} B_1' P_2^{-1} B_1 & 0 & B_1' P_2^{-1} B_2' \\ 0 & B_1 P_1^{-1} B_1' + B_2' P_3^{-1} B_2 & 0 \\ B_2 P_2^{-1} B_1 & 0 & B_2 P_2^{-1} B_2' \end{pmatrix}$$

and the spectral relation

$$\mathcal{P} \sim \mathcal{D} + \mathcal{B}\mathcal{P}^{-1}\mathcal{B}$$

is equivalent to

$$\begin{pmatrix} P_1 & 0 \\ 0 & P_3 \end{pmatrix} \sim \begin{pmatrix} A_1 + B_1' P_2^{-1} B_1 & B_1' P_2^{-1} B_2' \\ B_2 P_2^{-1} B_1 & A_3 + B_2 P_2^{-1} B_2' \end{pmatrix}$$

and

$$P_2 \sim A_2 + B_1 P_1^{-1} B_1' + B_2' P_3^{-1} B_2.$$

2.3. The case $n = 4$. Let

$$\mathcal{A} = \begin{pmatrix} A_1 & B_1' & 0 & 0 \\ B_1 & -A_2 & B_2' & 0 \\ 0 & B_2 & A_3 & B_3' \\ 0 & 0 & B_3 & -A_4 \end{pmatrix},$$

$$\mathcal{D} = \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & A_3 & 0 \\ 0 & 0 & 0 & A_4 \end{pmatrix}, \mathcal{B} = \begin{pmatrix} 0 & B_1' & 0 & 0 \\ B_1 & 0 & B_2' & 0 \\ 0 & B_2 & 0 & B_3' \\ 0 & 0 & B_3 & 0 \end{pmatrix}, \mathcal{P} = \begin{pmatrix} P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & P_3 & 0 \\ 0 & 0 & 0 & P_4 \end{pmatrix}.$$

Then

$$\mathcal{B}\mathcal{P}^{-1}\mathcal{B} = \begin{pmatrix} B_1'P_2^{-1}B_1 & 0 & B_1'P_2^{-1}B_2' & 0 \\ 0 & B_1P_1^{-1}B_1' + B_2'P_3^{-1}B_2 & 0 & B_2'P_3^{-1}B_3' \\ B_2P_2^{-1}B_1 & 0 & B_2P_2^{-1}B_2' + B_3'P_4^{-1}B_3 & 0 \\ 0 & B_3P_3^{-1}B_2 & 0 & B_3P_3^{-1}B_3' \end{pmatrix}$$

and the spectral relation

$$\mathcal{P} \sim \mathcal{D} + \mathcal{B}\mathcal{P}^{-1}\mathcal{B}$$

is equivalent to

$$\begin{pmatrix} P_1 & 0 \\ 0 & P_3 \end{pmatrix} \sim \begin{pmatrix} A_1 + B_1'P_2^{-1}B_1 & B_1'P_2^{-1}B_2' \\ B_2P_2^{-1}B_1 & A_3 + B_2P_2^{-1}B_2' + B_3'P_4^{-1}B_3 \end{pmatrix} \tag{2.8}$$

and

$$\begin{pmatrix} P_2 & 0 \\ 0 & P_4 \end{pmatrix} \sim \begin{pmatrix} A_2 + B_1P_1^{-1}B_1' + B_2'P_3^{-1}B_2 & B_2'P_3^{-1}B_3' \\ B_3P_3^{-1}B_2 & A_4 + B_3P_3^{-1}B_3' \end{pmatrix}. \tag{2.9}$$

3. Application to optimal control problems. We are now going to apply our theory to general abstract optimal control problems constrained by linear partial differential equations:

For given data d and fixed $\alpha > 0$, we consider the minimization problem of finding a state y and control u which minimize the functional

$$J : Y \times U \longrightarrow \mathbb{R}, \quad J(y,u) = \frac{1}{2}\|Ty - d\|_O^2 + \frac{\alpha}{2}\|u\|_U^2, \tag{3.1}$$

subject to the constraint

$$Ky + Cu = g. \tag{3.2}$$

Here, Y denotes the state space, U is the control space, and O is the observation space. The bounded linear observation operator $T : Y \longrightarrow O$ in (3.1) maps the state to the measurements.

The state equation (3.2) is given in terms of the bounded linear operators

$$K : Y \longrightarrow M' \text{ (state operator)} \quad \text{and} \quad C : U \longrightarrow M' \text{ (control operator)}.$$

Here, we assume that the test space M is a product space of Hilbert spaces where the first space is the same function space as used for the control,

$$M = U \times R.$$

The components of K will be denoted by the bounded linear operators

$$K_U : Y \longrightarrow U', \quad K_R : Y \longrightarrow R',$$

such that $Ky = (K_U y, K_R y)^\top$. Typically, the components of the state operator represent the differential expression and side conditions, such as boundary, and (or) initial conditions, respectively. For illustrative examples we refer to section 4.

The crucial assumption on the control operator C is that it is of the form

$$C = \begin{pmatrix} \mathcal{I}_U \\ 0 \end{pmatrix}.$$

Here, we used the following notation.

Notation 3. The inner product in a Hilbert space X induces a self-adjoint and positive definite bounded linear operator $\mathcal{I}_X \colon X \longrightarrow X'$, given by

$$\langle \mathcal{I}_X x, y \rangle = (x, y)_X \quad \text{for all } x, y \in X,$$

whose inverse is usually called the Riesz isomorphism associated to the Hilbert space X.

Note that the control operator C acts only on the first line of the state equation, which reads in detail

$$\begin{pmatrix} K_U \\ K_R \end{pmatrix} y + \begin{pmatrix} \mathcal{I}_U \\ 0 \end{pmatrix} u = \begin{pmatrix} g_U \\ g_R \end{pmatrix}. \tag{3.3}$$

Remark 3.1. We stress that the following treatment does not exclude the trivial case $R = \{0\}$, which corresponds to full control distributed on M.

The optimality system for the constrained optimization problem (3.1) and (3.3) reads as follows:

Find $(y, u, p_U, p_R) \in Y \times U \times U \times R$ such that

$$\begin{pmatrix} T'\mathcal{I}_O T & 0 & K_U' & K_R' \\ 0 & \alpha\mathcal{I}_U & \mathcal{I}_U & 0 \\ K_U & \mathcal{I}_U & 0 & 0 \\ K_R & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ u \\ p_U \\ p_R \end{pmatrix} = \begin{pmatrix} (d, T\cdot)_O \\ 0 \\ g_U \\ g_R \end{pmatrix}. \tag{3.4}$$

Remark 3.2. Note that $T'\mathcal{I}_O T : Y \longrightarrow Y'$ is given by $y \mapsto (Ty, T\cdot)_O$.

After a reordering, the optimality can equivalently be written in tridiagonal form:

Find $(u, p_U, y, p_R) \in U \times U \times Y \times R$ such that

$$\begin{pmatrix} \alpha\mathcal{I}_U & \mathcal{I}_U & 0 & 0 \\ \mathcal{I}_U & 0 & K_U & 0 \\ 0 & K_U' & T'\mathcal{I}_O T & K_R' \\ 0 & 0 & K_R & 0 \end{pmatrix} \begin{pmatrix} u \\ p_U \\ y \\ p_R \end{pmatrix} = \begin{pmatrix} 0 \\ g_U \\ (d, T\cdot)_O \\ g_R \end{pmatrix}. \tag{3.5}$$

It is obvious that the spectral relations (2.8) and (2.9) strongly depend on the properties of the involved operators K_U, K_R (and T). We are going to make the following assumptions:

(K1) The operator $K : Y \longrightarrow M'$, defined by $y \mapsto (K_U y, K_R y)^\top$, has closed range and is injective, or, equivalently, there exists a positive constant c_K such that

$$\|y\|_Y \le c_K \|Ky\|_{M'} = c_K \sqrt{\|K_U y\|^2_{U'} + \|K_R y\|^2_{R'}} \quad \text{for all } y \in Y. \tag{3.6}$$

(K2) The operator $K_R : Y \longrightarrow R'$ is surjective, or, equivalently, there exists a positive constant c_R such that

$$\sup_{0 \neq y \in Y} \frac{\langle K_R y, r \rangle}{\|y\|_Y} \ge c_R \|r\|_R \quad \text{for all } r \in R.$$

The assumption (K2) will also be considered in the stronger form:

(K2′) The operator $K_R|_{\ker K_U} : \ker K_U \longrightarrow R'$ is surjective, or, equivalently, there exists a positive constant c_R such that

$$\sup_{0 \neq y \in \ker K_U} \frac{\langle K_R y, r \rangle}{\|y\|_Y} \ge c_R \|r\|_R \quad \text{for all } r \in R.$$

Remark 3.3. Since we assumed that K_U, K_R are bounded linear operators, it follows from (3.6) that $\|K\cdot\|_{M'}$ induces an equivalent norm on Y.

Remark 3.4. Equation (3.6) can be seen as a natural a priori estimate for a linear partial differential equation of the form $Ky = g$, which states that *if* a unique solution to $Ky = g$ exists, then it needs to be bounded by the data $g \in M'$.

The next theorem deals with the well-posedness of (3.5) and offers a corresponding robust preconditioner. The derivation of the preconditioner is constructive in the following sense: Having a good guess for three out of four inner products on the Hilbert spaces $X_i \in \{U, U, Y, R\}$ leading to a robust preconditioner for the optimality system (3.5), the remaining fourth inner product follows almost as a gift from the spectral relation (2.7).

To be more precise, for the Hilbert spaces $X_1 = U$, $X_2 = U$, $X_4 = R$, we choose inner products corresponding to the operators

$$P_1 = \alpha \mathcal{I}_U, \quad P_2 = \alpha^{-1} \mathcal{I}_U, \quad P_4 = \mathcal{I}_R,$$

respectively.

Remark 3.5. The motivation for this choice is based on the relations

$$P_1 \sim A_1 + B_1' P_1^{-1} B_1 \quad \text{and} \quad P_2 \sim A_2 + B_1 P_1^{-1} B_1' + B_2' P_3^{-1} B_2,$$

which follow from (2.8) and (2.9). Hence $P_1 \gtrsim A_1$ and $P_2 \gtrsim A_2 + B_1 P_1^{-1} B_1'$. These necessary conditions inspired us to try out $P_1 = A_1 = \alpha \mathcal{I}_U$ and $P_2 = A_2 + B_1 P_1^{-1} B_1' = \alpha^{-1} \mathcal{I}_U$. The choice of P_4 is simply motivated by the standard inner product in R represented by $\mathcal{I}_R$. We would like to stress that this is, of course, not a derivation but just a motivation for the choice of P_1, P_2, P_4, which will turn out to be successful for the particular problem class discussed here.

With this choice the condition (2.8) reads

$$\begin{pmatrix} \alpha \mathcal{I}_U & 0 \\ 0 & P_3 \end{pmatrix} \sim \begin{pmatrix} 2\alpha \mathcal{I}_U & \alpha K_U \\ \alpha K_U' & T' \mathcal{I}_O T + \alpha K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R \end{pmatrix}.$$

Then, by Lemma 3.7, the only possible candidate for P_3 is given by (up to spectral equivalence)

$$P_3 = T'\mathcal{I}_O T + \alpha\, K_U'\mathcal{I}_U^{-1}K_U + K_R'\mathcal{I}_R^{-1}K_R. \tag{3.7}$$

The next theorem guarantees that (3.7) is not only necessary but also sufficient. The proof relies on two auxiliary results, frequently used in the literature, which are listed in the following. For the convenience of the reader we present the proof of the first lemma and refer to [1] for the second lemma.

LEMMA 3.6. *Let $A : V \longrightarrow V'$, $B : V \longrightarrow Q'$, and $C : Q \longrightarrow Q'$ be linear operators, where V and Q are Hilbert spaces with dual spaces V' and Q'. Additionally assume that A and C are self-adjoint and coercive. Then the condition*

$$\left\langle BA^{-1}B'q, q\right\rangle \leq \langle Cq, q\rangle \quad \textit{for all } q \in Q$$

is equivalent to the condition

$$\left\langle B'C^{-1}Bv, v\right\rangle \leq \langle Av, v\rangle \quad \textit{for all } v \in V.$$

Proof. Using Lemma A.1 we have

$$\left\langle BA^{-1}B'q, q\right\rangle = \sup_{0\neq v\in V} \frac{\langle Bv, q\rangle^2}{\langle Av, v\rangle} \quad \text{and} \quad \left\langle B'C^{-1}Bv, v\right\rangle = \sup_{0\neq q\in Q} \frac{\langle B'q, v\rangle^2}{\langle Cq, q\rangle}.$$

Then it immediately follows that the first condition is equivalent to the condition

$$\langle Bv, q\rangle^2 \leq \langle Av, v\rangle \langle Cq, q\rangle \quad \text{for all } v \in V,\, q \in Q$$

and the second condition is equivalent to the condition

$$\langle B'q, v\rangle^2 \leq \langle Av, v\rangle \langle Cq, q\rangle \quad \text{for all } v \in V,\, q \in Q.$$

These two new conditions are obviously equivalent. ∎

LEMMA 3.7. *Let V and Q be Hilbert spaces with dual spaces V' and Q'.*

Let $\mathcal{M} : V \times Q \longrightarrow V' \times Q'$ be a self-adjoint and positive definite linear operator of a 2-by-2 block form

$$\mathcal{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

and $\mathcal{D} : V \times Q \longrightarrow V' \times Q'$ be of 2-by-2 block diagonal form with self-adjoint and positive definite diagonal blocks

$$\mathcal{D} = \begin{pmatrix} D_{11} & 0 \\ 0 & D_{22} \end{pmatrix}.$$

Then $\mathcal{M} \sim \mathcal{D}$ is equivalent to

$$M_{11} \sim D_{11}, \quad M_{22} \sim D_{22} \quad \textit{and} \quad M_{11} \lesssim M_{11} - M_{12}M_{22}^{-1}M_{21} \tag{3.8}$$

as well as to

$$M_{11} \sim D_{11}, \quad M_{22} \sim D_{22} \quad \textit{and} \quad M_{22} \lesssim M_{22} - M_{21}M_{11}^{-1}M_{12}. \tag{3.9}$$

We will now prove the main result of this section.

THEOREM 3.8. *Let $\alpha > 0$ and assume that assumptions* (K1) *and* (K2) *are satisfied. Then the linear operator $\mathcal{A} : U \times U \times Y \times R = \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ defined in* (3.5) *is a self-adjoint isomorphism. Furthermore, for the Hilbert space $\boldsymbol{X}$ endowed with the inner product*

$$(\boldsymbol{x}, \boldsymbol{y})_{\mathcal{P}} = \langle \mathcal{P}\boldsymbol{x}, \boldsymbol{y} \rangle \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X},$$

where $\mathcal{P} : \boldsymbol{X} \longrightarrow \boldsymbol{X}'$ is given by

$$\mathcal{P} = \begin{pmatrix} \alpha \mathcal{I}_U & & & \\ & \alpha^{-1} \mathcal{I}_U & & \\ & & T' \mathcal{I}_O T + \alpha\, K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R & \\ & & & \mathcal{I}_R \end{pmatrix},$$

there exist positive constants $\underline{c}$ and $\overline{c}$, both independent of $\alpha \in (0, 1]$, such that

$$\underline{c}\, \|\boldsymbol{x}\|_{\mathcal{P}} \leq \|\mathcal{P}^{-1} \mathcal{A} \boldsymbol{x}\|_{\mathcal{P}} \leq \overline{c}\, \|\boldsymbol{x}\|_{\mathcal{P}} \quad \text{for all } \boldsymbol{x} \in \boldsymbol{X}. \tag{3.10}$$

Under the stronger assumption (K2′), *the constants in* (3.10) *are independent of all $\alpha > 0$.*

Proof. Denoting

$$P_3 = T' \mathcal{I}_O T + \alpha\, K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R,$$

by Theorem 2.2 and (2.8) and (2.9), it suffices to show

$$\begin{pmatrix} \alpha\, \mathcal{I}_U & 0 \\ 0 & P_3 \end{pmatrix} \sim \begin{pmatrix} 2\alpha\, \mathcal{I}_U & \alpha\, K_U \\ \alpha\, K_U' & T' \mathcal{I}_O T + \alpha\, K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R \end{pmatrix}$$

and

$$\begin{pmatrix} \alpha^{-1}\, \mathcal{I}_U & 0 \\ 0 & \mathcal{I}_R \end{pmatrix} \sim \begin{pmatrix} \alpha^{-1}\, \mathcal{I}_U + K_U P_3^{-1} K_U' & K_U P_3^{-1} K_R' \\ K_R P_3^{-1} K_U' & K_R P_3^{-1} K_R' \end{pmatrix}.$$

The first condition follows from its equivalence to (3.9) in Lemma 3.7, since $\alpha \mathcal{I}_U \sim 2\alpha \mathcal{I}_U$, $P_3 = T' \mathcal{I}_O T + \alpha\, K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R$ and

$$\begin{aligned} T' \mathcal{I}_O T + \alpha\, K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R - \frac{1}{2} \alpha\, K_U' \mathcal{I}_U^{-1} K_U & \\ = T' \mathcal{I}_O T + \frac{\alpha}{2}\, K_U' \mathcal{I}_U^{-1} K_U + K_R' \mathcal{I}_R^{-1} K_R \sim P_3. & \end{aligned}$$

Concerning the second condition, from $\alpha\, K_U' \mathcal{I}_U^{-1} K_U \leq P_3$ and Lemma 3.6 it follows that

$$K_U P_3^{-1} K_U' \leq \alpha^{-1}\, \mathcal{I}_U,$$

which implies

$$\alpha^{-1}\, \mathcal{I}_U \sim \alpha^{-1}\, \mathcal{I}_U + K_U P_3^{-1} K_U'. \tag{3.11}$$

From $K_R' \mathcal{I}_R^{-1} K_R \leq P_3$ and Lemma 3.6 it also follows that

$$K_R P_3^{-1} K_R' \leq \mathcal{I}_R.$$

So, in order to ensure

$$\mathcal{I}_R \sim K_R P_3^{-1} K_R', \tag{3.12}$$

it suffices to show $\mathcal{I}_R \lesssim K_R P_3^{-1} K_R'$, or, equivalently,

$$\sup_{0\neq y\in Y} \frac{\langle K_R y, r\rangle}{\langle P_3 y, y\rangle^{1/2}} \gtrsim \|r\|_R \quad \text{for all } r \in R.$$

This easily follows from assumption (K2), since

$$\langle P_3 y, y\rangle \lesssim \|y\|_Y^2 \quad \text{for all } y \in Y,$$

under the mild condition that α is uniformly bounded, e.g., $\alpha \leq 1$, and, therefore,

$$\sup_{0\neq y\in Y} \frac{\langle K_R y, r\rangle}{\langle P_3 y, y\rangle^{1/2}} \gtrsim \sup_{0\neq y\in Y} \frac{\langle K_R y, r\rangle}{\|y\|_Y} \gtrsim \|r\|_R \quad \text{for all } r \in R.$$

Therefore, (3.12) holds and $K_R P_3^{-1} K_R'$ is nonsingular. Then, by Lemma 3.6, it follows that

$$K_R'[K_R P_3^{-1} K_R']^{-1} K_R \leq P_3,$$

and as a consequence,

$$K_U P_3^{-1} K_R'[K_R P_3^{-1} K_R']^{-1} K_R P_3^{-1} K_U' \leq K_U P_3^{-1} K_U'.$$

Therefore,

$$\alpha^{-1}\mathcal{I}_U \leq \alpha^{-1}\mathcal{I}_U + K_U P_3^{-1} K_U' - K_U P_3^{-1} K_R'[K_R P_3^{-1} K_R']^{-1} K_R P_3^{-1} K_U'. \tag{3.13}$$

The assertion then follows from (3.11) (3.12), (3.13) using (3.8) in Lemma 3.7.

Under no restrictions on α we have

$$\langle P_3 y, y\rangle \lesssim \|y\|_Y^2 \quad \text{for all } y \in \ker K_U,$$

and, therefore,

$$\sup_{0\neq y\in Y} \frac{\langle K_R y, r\rangle}{\langle P_3 y, y\rangle^{1/2}} \gtrsim \sup_{0\neq y\in \ker K_U} \frac{\langle K_R y, r\rangle}{\|y\|_Y} \gtrsim \|r\|_R \quad \text{for all } r \in R$$

under the stronger assumption assumption (K2′). ∎

Remark 3.9. Theorem 3.8 also holds true under the relaxed condition that $T : Y \longrightarrow O$ is invertible on the kernel of $K : Y \longrightarrow M'$, if $\ker K \neq \{0\}$, as was done in [15].

Remark 3.10. The optimality system (3.4) can be phrased as a classical saddle point problem: Find $x = (y, u) \in Y \times U = X$ and $p = (p_U, p_R) \in U \times R = M$ such that

$$\begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{3.14}$$

where $A : X \longrightarrow X'$, $B : X \longrightarrow M'$ are given by

$$A = \begin{pmatrix} T'\mathcal{I}_O T & 0 \\ 0 & \alpha\mathcal{I}_U \end{pmatrix}, \quad B = \begin{pmatrix} K_U & \mathcal{I}_U \\ K_R & 0 \end{pmatrix},$$

and $f \in X'$, $g \in M'$ are given by

$$f(w) = (d, Tz)_O, \quad g(q) = g_U(q_U) + g_R(q_R)$$

for all $w = (z, v) \in X$, $q = (q_U, q_R) \in M$, respectively.

If, in accordance with Theorem 3.8, the Hilbert spaces X, M are endowed with the α-dependent norms

$$\|w\|_X^2 = \|Tz\|_O^2 + \alpha \|K_U z\|_{U'}^2 + \|K_R z\|_{R'}^2 + \alpha \|v\|_U^2, \qquad w = (z, v) \in X,$$
$$\|q\|_M^2 = \alpha^{-1} \|q_U\|_U^2 + \|q_R\|_R^2, \qquad q = (q_U, q_R) \in M,$$

then one can show the following conditions under assumptions (K1) and (K2′):

1. The linear operator A is bounded:
$$\langle Ax, w\rangle \le c_A \, \|x\|_X \, \|w\|_X \quad \text{for all } x, w \in X \quad \text{with} \quad c_A = 1.$$
2. The linear operator B is bounded:
$$\langle Bw, q\rangle \le c_B \, \|w\|_X \, \|q\|_M \quad \text{for all } w \in X,\ q \in M \quad \text{with} \quad c_B = \sqrt{2}.$$
3. The linear operator A is coercive on $\ker B$:
$$\langle Aw, w\rangle \ge \gamma_0 \, \|w\|_X^2 \quad \text{for all } w \in \ker B \quad \text{with} \quad \gamma_0 = \frac{1}{2}.$$
4. The linear operator B satisfies an inf-sup condition:
$$\sup_{0 \ne w \in X} \frac{\langle Bw, q\rangle}{\|w\|_X} \ge k_0 \, \|q\|_M \quad \text{for all } q \in M \quad \text{with} \quad k_0 = \frac{1}{\sqrt{\|T\|_{L(Y,O)}^2 \, c_K^2 + 1}},$$
where $\|T\|_{L(Y,O)}$ denotes the operator norm of $T\colon Y \longrightarrow O$ and c_K is the positive constant from (3.6).

So, as expected, the conditions of Brezzi's theorem, which are equivalent to the well-posedness of the saddle point problem (see [6]), are satisfied with α-independent constants c_A, c_B, γ_0, k_0. For details, see [3] (version v1).

4. Examples. Under the assumptions (K1) and (K2), respectively (K2′), Theorem 3.8 guarantees well-posedness of the optimality system (3.5) and proposes a robust preconditioner. The question arises for which particular applications, that is, linear PDEs, these conditions are fulfilled.

Usually, the operators K_U and K_R represent the differential operator and the side conditions of the PDE, respectively. Elliptic control problems of this form (with $R = \{0\}$) have been considered in [15], where the space for the control and the test space for the nonstandard variational formulation in the strong form of the state equation coincide. In this sense the setting in [15] fits into our framework and, therefore, we will focus here as an alternative on time-dependent problems in the following.

In the two examples to come we have a linear and bijective state operator $K = (K_U, K_R)^\top : Y \longrightarrow M'$, where M is a product space of Hilbert spaces U, R, and Y is a linear space. There is a natural way of introducing a canonical Hilbert space structure on Y such that the assumptions (K1) and (K2′) are satisfied.

LEMMA 4.1. *Under the assumptions for Y, M, and $K : Y \longrightarrow M'$ from above, Y is a Hilbert space endowed with the inner product*

$$\begin{aligned}(y, z)_Y &= \langle Ky, \mathcal{I}_M^{-1} Kz\rangle \\ &= \langle K_U y, \mathcal{I}_U^{-1} K_U z\rangle + \langle K_R y, \mathcal{I}_R^{-1} K_R z\rangle \quad \textit{for all } y, z \in Y,\end{aligned}$$

where $\mathcal{I}_M$, $\mathcal{I}_U$, and $\mathcal{I}_R$ represent the canonical inner products on M, U, and R, respectively (see Notation 3).

Moreover, for Y equipped with the inner product $(\cdot,\cdot)_Y$, the state operator K satisfies assumptions (K1) *and* (K2′).

Proof. Observe that $\langle \cdot, \mathcal{I}_M^{-1} \cdot \rangle$ defines an inner product on M' by the definition of $\mathcal{I}_M$. Since K is linear and injective, it follows that $(\cdot,\cdot)_Y = \langle K\cdot, \mathcal{I}_M^{-1} K\cdot\rangle$ defines an inner product on Y.

In order to show that Y is complete with respect to $\|\cdot\|_Y = \sqrt{(\cdot,\cdot)_Y} = \|K\cdot\|_{M'}$ let $(y_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in Y. Then $(Ky_k)_{k\in\mathbb{N}}$ is a Cauchy sequence in M' which possesses a limit in M' denoted by g. Since K is bijective, there exists a unique $y \in Y$ such that $Ky = g$. Consequently,

$$\|y_k - y\|_Y = \|K(y_k - y)\|_{M'} = \|Ky_k - g\|_{M'} \to 0 \quad \text{as} \quad k \to \infty.$$

Thus the Cauchy sequence is converging and thus the space Y is complete.

Note that assumption (K1) is trivially satisfied by the definition of $\|\cdot\|_Y$, that is,

$$\|y\|_Y = \|Ky\|_{M'} = \sqrt{\|K_U y\|_{U'}^2 + \|K_R y\|_{R'}^2} \quad \text{for all } y \in Y.$$

Assumption (K2′) is fulfilled as well since for any $g_R \in R'$ the system $Ky = (0, g_R)^\top$ is uniquely solvable. ∎

This lemma is needed for the following two examples of the heat equation and the wave equation considered over a bounded domain $\Omega \subset \mathbb{R}^N$, $N \geq 1$, with Lipschitz boundary $\partial\Omega$, and over a finite time interval $(0,T)$. Throughout the remainder of the paper we will introduce the space-time cylinder by $Q_T = \Omega \times (0,T)$ and its lateral surface by $\Gamma_T = \partial\Omega \times [0,T]$.

4.1. Heat equation. Consider the heat equation with homogeneous Dirichlet boundary conditions on Γ_T,

$$\begin{aligned} \partial_t y - \Delta y &= f \quad \text{in} \quad Q_T, \\ y &= 0 \quad \text{on} \quad \Gamma_T, \\ y(0) &= y_0 \quad \text{in} \quad \Omega \end{aligned}$$

for given data f, y_0. For this problem we introduce the following function spaces. $L^2(D)$ denotes the standard Lebesgue space of square-integrable functions on a domain D. $H_0^1(\Omega)$ is the subspace of the standard Sobolev space $H^1(\Omega)$ of functions on Ω with vanishing trace on $\partial\Omega$. We use the inner product $(v,w)_{H_0^1(\Omega)} = (\nabla v, \nabla w)_{L^2(\Omega)}$ for functions $v, w \in H_0^1(\Omega)$. Moreover, let

$$H(\Delta,\Omega) = \{v \in L^2(\Omega) : \Delta v \in L^2(\Omega)\}$$

with inner product $(v,w)_{H(\Delta,\Omega)} = (v,w)_{L^2(\Omega)} + (\Delta v, \Delta w)_{L^2(\Omega)}$. Finally, for a Hilbert space H, $L^2((0,T);H)$ denotes the Bochner space of square-integrable functions from $(0,T)$ to H with inner product

$$(f,g)_{L^2((0,T);H)} = \int_0^T (f(t), g(t))_H \, dt.$$

Then the well-known solution theory for this initial-boundary value problem can be summarized in the spirit of Lemma 4.1.

THEOREM 4.2 (see [12, Chapter 3, Theorem 2.1]). *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and $T > 0$ finite. Define*

$$\begin{aligned} Y &= \{y \in L^2((0,T); H_0^1(\Omega) \cap H(\Delta,\Omega)) : \partial_t u \in L^2(Q_T)\}, \\ M &= L^2(Q_T) \times H_0^1(\Omega). \end{aligned} \tag{4.1}$$

Then the operator

$$K : Y \longrightarrow M', \quad y \mapsto \begin{pmatrix} (\partial_t y - \Delta y, \cdot)_{L^2(Q_T)} \\ (y(0), \cdot)_{H_0^1(\Omega)} \end{pmatrix}$$

is linear and bijective.

Combining Theorem 3.8 and Lemma 4.1 we obtain the following result.

COROLLARY 4.3. *Let Y, M, and $K : Y \longrightarrow M'$ be as in Theorem* 4.2. *Then Y is a complete space with respect to the norm $\|\cdot\|_Y$, given by*

$$\|y\|_Y^2 = \|\partial_t y - \Delta y\|_{L^2(Q_T)}^2 + \|y(0)\|_{H_0^1(\Omega)}^2 \quad \text{for all } y \in Y.$$

Moreover, for any decomposition of M as a product space of Hilbert spaces U and R and any bounded linear observation operator $T : Y \longrightarrow O$, there exists a unique minimizer of the constrained optimization problem (3.1) *and* (3.3) *which is characterized by the solution of the optimality system* (3.4).

Optimal control problem for the heat equation. Let ω be a nonempty open subset of Ω and denote $q_T = \omega \times (0,T) \subset Q_T$. We consider an optimal control problem of minimizing a tracking-type quadratic cost functional with (possibly) limited observation plus a regularization term, where the constraint is the heat equation. More precisely, for Y given by (4.1), we want to minimize the functional

$$J : Y \times L^2(Q_T) \longrightarrow \mathbb{R}, \qquad J(y,u) = \frac{1}{2}\|y - d\|_{L^2(q_T)}^2 + \frac{\alpha}{2}\|u\|_{L^2(Q_T)}^2,$$

subject to

$$\begin{aligned} (\partial_t y - \Delta y, q_U)_{L^2(Q_T)} + (u, q_U)_{L^2(Q_T)} &= 0 && \text{for all } q_U \in L^2(Q_T), \\ (y(0), q_R)_{H_0^1(\Omega)} &= (y_0, q_R)_{H_0^1(\Omega)} && \text{for all } q_R \in H_0^1(\Omega) \end{aligned}$$

for given initial value $y_0 \in H_0^1(\Omega)$ and data $d \in L^2(q_T)$.

The optimality system then reads as follows:

Find $(y, u, p_U, p_R) \in Y \times U \times U \times R$ such that

$$\begin{pmatrix} T'\mathcal{I}_O T & 0 & K_U' & K_R' \\ 0 & \alpha\mathcal{I}_U & \mathcal{I}_U & 0 \\ K_U & \mathcal{I}_U & 0 & 0 \\ K_R & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ u \\ p_U \\ p_R \end{pmatrix} = \begin{pmatrix} (d, T\cdot)_{L^2(q_T)} \\ 0 \\ 0 \\ (y_0, \cdot)_{H_0^1(\Omega)} \end{pmatrix} \tag{4.2}$$

with the spaces Y given by (4.1),

$$U = L^2(Q_T), \quad R = H_0^1(\Omega), \quad O = L^2(q_T)$$

and the operators

$$\begin{aligned} T &: Y \longrightarrow O, && y \mapsto y|_{q_T}, \\ K_U &: Y \longrightarrow U', && y \mapsto (\partial_t y - \Delta y, \cdot)_{L^2(Q_T)}, \\ K_R &: Y \longrightarrow R', && y \mapsto (y(0), \cdot)_{H_0^1(\Omega)}. \end{aligned} \tag{4.3}$$

It follows from Corollary 4.3 that the system (4.2) is well-posed. Additionally, by Theorem 3.8, the $\mathcal{P}$-norm leading to an α-robust preconditioner is given by

$$\begin{aligned} \|(z, v, q_U, q_R)\|_{\mathcal{P}}^2 = \|z\|_{L^2(q_T)}^2 + \alpha\|\partial_t z - \Delta z\|_{L^2(Q_T)}^2 + \|\nabla z(0)\|_{L^2(\Omega)}^2 \\ + \alpha\|v\|_{L^2(Q_T)}^2 + \frac{1}{\alpha}\|q_U\|_{L^2(Q_T)}^2 + \|\nabla q_R\|_{L^2(\Omega)}^2. \end{aligned}$$

4.2. Wave equation. Consider the wave equation with homogeneous Dirichlet boundary conditions on Γ_T,

$$\begin{aligned} \partial_{tt} y - \Delta y &= f \quad \text{in} \quad Q_T, \\ y &= 0 \quad \text{on} \quad \Gamma_T, \\ y(0) &= y_0 \quad \text{in} \quad \Omega, \\ \partial_t y(0) &= y_1 \quad \text{in} \quad \Omega \end{aligned}$$

for given data f, y_0, y_1. For this problem we need one further function space. $C([0,T];H)$ denotes the space of continuous functions from $[0,T]$ to a Hilbert space H. Then the well-known solution theory for this initial-boundary value problem can be summarized in the spirit of Lemma 4.1.

THEOREM 4.4 (see [13, Chapter 3, Theorem 8.2]). *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and $T > 0$ finite. Define*

$$\begin{aligned} &Y = \{y \in C([0,T]; H_0^1(\Omega)) : \partial_t y \in C([0,T]; L^2(\Omega)),\ \partial_{tt} y - \Delta y \in L^2(Q_T)\}, \\ &M = L^2(Q_T) \times H_0^1(\Omega) \times L^2(\Omega). \end{aligned} \tag{4.4}$$

Then the operator

$$K : Y \longrightarrow M', \quad y \mapsto \begin{pmatrix} (\partial_{tt} y - \Delta y, \cdot)_{L^2(Q_T)} \\ (y(0), \cdot)_{H_0^1(\Omega)} \\ (\partial_t y(0), \cdot)_{L^2(\Omega)} \end{pmatrix} \tag{4.5}$$

is linear and bijective.

Combining Theorem 3.8 and Lemma 4.1 we obtain the following result.

COROLLARY 4.5. *Let Y, M, and $K : Y \longrightarrow M'$ be as in Theorem 4.4. Then Y is a complete space with respect to*

$$\|y\|_Y^2 = \|\partial_{tt} y - \Delta y\|_{L^2(Q_T)}^2 + \|y(0)\|_{H_0^1(\Omega)}^2 + \|\partial_t y(0)\|_{L^2(\Omega)}^2 \quad \text{for all } y \in Y.$$

Moreover, for any decomposition of M as a product space of Hilbert spaces U and R and any bounded linear observation operator $T : Y \longrightarrow O$ there exists a unique minimizer of the constrained optimization problem (3.1) *and* (3.3) *which is characterized by the solution of the optimality system* (3.4).

In a previous work we considered the problem of controlling the initial condition of the wave equation,

$$y(0) = u, \quad u \in U = H_0^1(\Omega);$$

see [2] for further details. This particular study was the starting point for our current work. Here, we want to demonstrate the flexibility of our approach and consider in the following a constrained optimization problem for the wave equation where we control the differential expression

$$\partial_{tt} y - \Delta y + u = 0, \quad u \in U = L^2(Q_T).$$

Optimal control problem for the wave equation. Let ω be a nonempty open subset of Ω and denote $q_T = \omega \times (0,T) \subset Q_T$. We consider an optimal control

problem of minimizing a tracking-type quadratic cost functional with (possibly) limited observation plus a regularization term where the constraint is the wave equation. More precisely, for Y given by (4.4), we want to minimize the functional

$$J : Y \times L^2(Q_T) \longrightarrow \mathbb{R}, \qquad J(y,u) = \frac{1}{2}\,\|y-d\|^2_{L^2(q_T)} + \frac{\alpha}{2}\,\|u\|^2_{L^2(Q_T)}\,,$$

subject to

$$\begin{aligned} (\partial_{tt} y - \Delta y, q_U)_{L^2(Q_T)} + (u, q_U)_{L^2(Q_T)} &= 0 && \text{for all } q_U \in L^2(Q_T),\\ (y(0), q_{R_1})_{H^1_0(\Omega)} &= (y_0, q_{R_1})_{H^1_0(\Omega)} && \text{for all } q_{R_1} \in H^1_0(\Omega),\\ (\partial_t y(0), q_{R_2})_{L^2(\Omega)} &= (y_1, q_{R_2})_{L^2(\Omega)} && \text{for all } q_{R_2} \in L^2(\Omega) \end{aligned}$$

for given initial values $(y_0, y_1) \in H^1_0(\Omega) \times L^2(\Omega)$ and data $d \in L^2(q_T)$.

The optimality system then reads as follows:

Find $(y, u, p_U, p_{R_1}, p_{R_2}) \in Y \times U \times U \times R_1 \times R_2$ such that

$$\begin{pmatrix} T'\mathcal{I}_O T & 0 & K'_U & K'_{R_1} & K'_{R_2} \\ 0 & \alpha\mathcal{I}_U & \mathcal{I}_U & 0 & 0 \\ K_U & \mathcal{I}_U & 0 & 0 & 0 \\ K_{R_1} & 0 & 0 & 0 & 0 \\ K_{R_2} & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ u \\ p_U \\ p_{R_1} \\ p_{R_2} \end{pmatrix} = \begin{pmatrix} (d, T\cdot)_{L^2(q_T)} \\ 0 \\ 0 \\ (y_0, \cdot)_{H^1_0(\Omega)} \\ (y_1, \cdot)_{L^2(\Omega)} \end{pmatrix} \tag{4.6}$$

with the spaces Y given by (4.4),

$$U = L^2(Q_T), \quad R_1 = H^1_0(\Omega), \quad R_2 = L^2(\Omega), \quad O = L^2(q_T)$$

and the operators

$$\begin{aligned} T &: Y \longrightarrow O, & y &\mapsto y|_{q_T},\\ K_U &: Y \longrightarrow U', & y &\mapsto (\partial_{tt} y - \Delta y, \cdot)_{L^2(Q_T)}\,,\\ K_{R_1} &: Y \longrightarrow R_1', & y &\mapsto (y(0), \cdot)_{H^1_0(\Omega)}\,,\\ K_{R_2} &: Y \longrightarrow R_2', & y &\mapsto (\partial_t y(0), \cdot)_{L^2(\Omega)}\,. \end{aligned} \tag{4.7}$$

It follows from Corollary 4.5 that the system (4.6) is well-posed. Additionally, by Theorem 3.8, the $\mathcal{P}$-norm leading to an α-robust preconditioner is given by

$$\begin{aligned} &\|(z, v, q_U, q_{R_1}, q_{R_2})\|^2_{\mathcal{P}} \\ &\quad = \|z\|^2_{L^2(q_T)} + \alpha\,\|\partial_{tt} z - \Delta z\|^2_{L^2(Q_T)} + \|\nabla z(0)\|^2_{L^2(\Omega)} + \|\partial_t z(0)\|^2_{L^2(\Omega)} \\ &\qquad + \alpha\,\|v\|^2_{L^2(Q_T)} + \frac{1}{\alpha}\,\|q_U\|^2_{L^2(Q_T)} + \|\nabla q_{R_1}\|^2_{L^2(\Omega)} + \|q_{R_2}\|^2_{L^2(\Omega)}\,. \end{aligned} \tag{4.8}$$

Remark 4.6. The examples in this section could be modified to either include final time observations or to have final time observations only instead of observations distributed in time, since final time observations (in appropriate norms) constitute linear bounded observation operators in the respective state spaces.

Remark 4.7. So far we have only discussed space-time problems, where the control u acted on the whole domain Q_T. Problems where the control acts only on a subdomain fit also in the general framework presented at the beginning of section 3. Consider, e.g., an elliptic optimal control problem with the state equation

$$\begin{aligned} -\Delta y + u &= f \qquad \text{in } \Omega, \\ y &= 0 \qquad \text{on } \Gamma, \end{aligned} \tag{4.9}$$

where the control u is supported on a subset Ω_c with $\overline{\Omega}_c \subset \Omega$. We assume that the boundaries of Ω and Ω_c are sufficiently smooth; more precisely, Ω and Ω_c are of class $\mathscr{C}^{1,1}$ (see [9]). Then, for $u \in L^2(\Omega_c)$, we have $y \in H^2(\Omega) \cap H_0^1(\Omega)$. In this case (4.9) can be equivalently reformulated as

$$\begin{aligned} -\Delta y + u &= f \qquad \text{in } \Omega_c, \\ -\Delta y &= f \qquad \text{in } \Omega \setminus \overline{\Omega}_c \\ [\![y]\!] = [\![\partial_n y]\!] &= 0 \qquad \text{on } \Gamma_c, \\ y &= 0 \qquad \text{on } \Gamma, \end{aligned}$$

where $[\![y]\!]$ and $[\![\partial_n y]\!]$ denote the jump of y and the normal derivative of y across Γ_c, respectively. The associated state operator $K : Y \longrightarrow M'$ with $M = U \times R$, given by its components

$$\begin{aligned} K_U y &= -(\Delta y, \cdot)_{L^2(\Omega_c)}, \\ K_R y &= \big(-(\Delta y, \cdot)_{L^2(\Omega \setminus \overline{\Omega}_c)}, ([\![y]\!], \cdot)_{H^{3/2}(\Gamma_c)}, ([\![\partial_n y]\!], \cdot)_{H^{1/2}(\Gamma_c)} \big)^\top \end{aligned}$$

and the spaces $Y = H^2(\Omega) \cap H_0^1(\Omega)$, $U = L^2(\Omega_c)$, $R = L^2(\Omega \setminus \overline{\Omega}_c) \times H^{3/2}(\Gamma_c) \times H^{1/2}(\Gamma_c)$, satisfies assumptions (K1) and (K2′), since K is linear, bounded, and bijective.

This approach extends naturally to the heat equation and the wave equation.

5. Discretization and numerical experiments. In order to illustrate the theoretical results we shortly discuss in this section (as one selected example) the discretization of the optimality system (4.6) of the optimal control problem for the wave equation and present some first numerical results.

5.1. Discretization. We consider conforming discretization spaces, that is,

$$Y_h \subset Y, \quad U_h \subset L^2(Q_T), \quad R_{1,h} \subset H_0^1(\Omega), \quad \text{and} \quad R_{2,h} \subset L^2(\Omega). \tag{5.1}$$

Let K_{U_h}, $K_{R_{1,h}}$, $K_{R_{2,h}}$ be the matrix representations of the linear operators K_U, K_{R_1}, K_{R_2} defined in (4.7), on Y_h, U_h, $R_{h,1}$, $R_{h,2}$ relative to the chosen bases in these spaces, respectively. Let $M_{Q_T,h}$ and $M_{q_T,h}$ be the matrix representations of the linear operators $\mathcal{I}_U$ and $T'\mathcal{I}_O T$ on U_h and Y_h, respectively. These matrices are mass matrices; they represent the inner products $(\cdot,\cdot)_{L^2(Q_T)}$ and $(\cdot,\cdot)_{L^2(q_T)}$ on U_h and Y_h, respectively.

Applying Galerkin's principle to (4.6) leads to the following linear problem:

Find $(y_h, u_h, p_{U_h}, p_{R_{1,h}}, p_{R_{2,h}}) \in Y_h \times U_h \times U_h \times R_{1,h} \times R_{2,h}$ such that

$$\underbrace{\begin{pmatrix} M_{q_T,h} & 0 & K_{U_h}^T & K_{R_{1,h}}^T & K_{R_{2,h}}^T \\ 0 & \alpha M_{Q_T,h} & M_{Q_T,h} & 0 & 0 \\ K_{U_h} & M_{Q_T,h} & 0 & 0 & 0 \\ K_{R_{1,h}} & 0 & 0 & 0 & 0 \\ K_{R_{2,h}} & 0 & 0 & 0 & 0 \end{pmatrix}}_{=\mathcal{A}_h} \begin{pmatrix} \underline{y}_h \\ \underline{u}_h \\ \underline{p}_{U_h} \\ \underline{p}_{R_{1,h}} \\ \underline{p}_{R_{2,h}} \end{pmatrix} = \begin{pmatrix} \underline{d}_h \\ 0 \\ 0 \\ \underline{y}_{1,h} \\ \underline{y}_{2,h} \end{pmatrix}, \tag{5.2}$$

where d_h, $y_{2,h}$ are L^2 projections of d, y_2 on Y_h, $R_{2,h}$, respectively, $y_{1,h}$ is the H_0^1 projection of y_1 on $R_{1,h}$, and underlined quantities denote the vector representations of the corresponding functions from Y_h, U_h, $R_{1,h}$, $R_{2,h}$ relative to the chosen bases in these spaces.

Motivated by the analysis of the continuous problem we propose the following block diagonal preconditioner, which is the matrix representation of the $\mathcal{P}$-inner product on the discretization spaces:

$$\mathcal{P}_h = \begin{pmatrix} P_{Y_h} & 0 & 0 & 0 & 0 \\ 0 & \alpha\, P_{U_h} & 0 & 0 & 0 \\ 0 & 0 & \alpha^{-1}\, P_{U_h} & 0 & 0 \\ 0 & 0 & 0 & P_{R_{1,h}} & 0 \\ 0 & 0 & 0 & 0 & P_{R_{2,h}} \end{pmatrix} \tag{5.3}$$

with

$$\begin{aligned} \left\langle P_{Y_h} \underline{y}_h, \underline{z}_h \right\rangle &= (y_h, z_h)_{L^2(q_T)} + \alpha \left(\partial_{tt} y_h - \Delta y_h, \partial_{tt} z_h - \Delta z_h\right)_{L^2(Q_T)} \\ &\quad + \left(\nabla y_h(0), \nabla z_h(0)\right)_{L^2(\Omega)} + \left(\partial_t y_h(0), \partial_t z_h(0)\right)_{L^2(\Omega)}, \\ \langle P_{U_h} \underline{u}_h, \underline{v}_h \rangle &= (u_h, v_h)_{L^2(Q_T)}, \\ \left\langle P_{R_{1,h}} \underline{p}_{R_{1,h}}, \underline{q}_{R_{1,h}} \right\rangle &= \left(\nabla p_{R_{1,h}}, \nabla q_{R_{1,h}}\right)_{L^2(\Omega)}, \\ \left\langle P_{R_{2,h}} \underline{p}_{R_{2,h}}, \underline{q}_{R_{2,h}} \right\rangle &= \left(p_{R_{2,h}}, q_{R_{2,h}}\right)_{L^2(\Omega)}. \end{aligned}$$

The preconditioner $\mathcal{P}_h$ is a symmetric and positive definite block diagonal matrix. This matrix is also sparse provided basis functions with local support are chosen.

Observe that the α-robust preconditioner which results from applying Theorem 3.8 directly to the discrete problem is similar to (5.3), but with P_{Y_h} replaced by

$$\begin{aligned} \left\langle \widetilde{P}_{Y_h} \underline{y}_h, \underline{z}_h \right\rangle &= (y_h, z_h)_{L^2(q_T)} + \alpha \left\langle K'_{U_h} \mathcal{I}_{U_h}^{-1} K_{U_h} \underline{y}_h, \underline{z}_h \right\rangle \\ &\quad + \langle K'_{R_{1,h}} \mathcal{I}_{R_{1,h}}^{-1} K_{R_{1,h}} \underline{y}_h, \underline{z}_h \rangle + \langle K'_{R_{2,h}} \mathcal{I}_{R_{2,h}}^{-1} K_{R_{2,h}} \underline{y}_h, \underline{z}_h \rangle. \end{aligned} \tag{5.4}$$

In general, $\widetilde{P}_{Y_h}$ is not sparse. Therefore, the application of the corresponding block diagonal preconditioner $\widetilde{\mathcal{P}}_h$ is rather costly. On the other hand, the choice $\widetilde{P}_{Y_h}$ would ensure α-robustness of $\widetilde{\mathcal{P}}_h$ by Theorem 3.8 provided assumptions (K1) and (K2) hold. In the next lemma we present sufficient conditions on the discretization spaces, Y_h, U_h, R_h, which ensure that $P_{Y_h} = \widetilde{P}_{Y_h}$ as well as assumption (K1).

LEMMA 5.1. *Assume the discretization space Y_h is of the following form:*

$$Y_h = Y_h^t \otimes Y_h^x, \tag{5.5}$$

where Y_h^t is the time discretization and Y_h^x space discretization. If the following conditions hold,

$$(\partial_{tt} - \Delta) Y_h \subset U_h, \quad Y_h^x \subset R_{1,h}, \quad Y_h^x \subset R_{2,h}, \tag{5.6}$$

then $\widetilde{P}_{Y_h} = P_{Y_h}$ and assumption (K1) *holds for the discretized state equation with the same constant c_K as for the continuous state equation.*

Proof. Let $y_h \in Y_h$ be arbitrary but fixed. Using Lemma A.1 we get

$$\left\langle K'_{U_h} \mathcal{I}_{U_h}^{-1} K_{U_h} \underline{y}_h, \underline{y}_h \right\rangle = \sup_{u_h \in U_h} \frac{\left\langle K_{U_h} \underline{y}_h, \underline{u}_h \right\rangle^2}{(u_h, u_h)_{L^2(Q_T)}} = \sup_{u_h \in U_h} \frac{((\partial_{tt} - \Delta) y_h, u_h)^2_{L^2(Q_T)}}{(u_h, u_h)_{L^2(Q_T)}}.$$

Since $(\partial_{tt} - \Delta) Y_h \subset U_h$, the supremum is attained for $u_h = (\partial_{tt} - \Delta) y_h$, and we have

$$\sup_{u_h \in U_h} \frac{((\partial_{tt} - \Delta) y_h, u_h)^2_{L^2(Q_T)}}{(u_h, u_h)_{L^2(Q_T)}} = \|(\partial_{tt} - \Delta) y_h\|^2_{L^2(Q_T)}.$$

Therefore,

$$\left\langle K'_{U_h} \mathcal{I}_{U_h}^{-1} K_{U_h} \underline{y}_h, \underline{y}_h \right\rangle = \|(\partial_{tt} - \Delta) y_h\|^2_{L^2(Q_T)}.$$

Similarly it follows that

$$\langle K'_{R_{1,h}} \mathcal{I}_{R_{1,h}}^{-1} K_{R_{1,h}} \underline{y}_h, \underline{y}_h \rangle = \|y_h(0)\|^2_{H_0^1(\Omega)}$$

and

$$\langle K'_{R_{2,h}} \mathcal{I}_{R_{2,h}}^{-1} K_{R_{2,h}} \underline{y}_h, \underline{y}_h \rangle = \|\partial_t y_h(0)\|^2_{L^2(\Omega)}.$$

This shows that the P_{Y_h}-norm and the $\widetilde{P}_{Y_h}$-norm coincide and, therefore, the associated inner products coincide. This implies $P_{Y_h} = \widetilde{P}_{Y_h}$.

The three identities from above can be rewritten as

$$\|K_{U_h} y_h\|_{U'_h} = \|K_U y_h\|_{U'}$$

and

$$\|K_{R_{1,h}} y_h\|_{R'_{1,h}} = \|K_{R_1} y_h\|_{R'_1}, \quad \|K_{R_{2,h}} y_h\|_{R'_{2,h}} = \|K_{R_2} y_h\|_{R'_2},$$

from which it immediately follows that assumption (K1) for the continuous state operator K implies assumption (K1) for the state operator $K_h = (K_{U_h}, K_{R_{1,h}}, K_{R_{2,h}})^\top$ of the discretized problem. ∎

So, under the assumptions of Lemma 5.1, the preconditioner $\mathcal{P}_h = \widetilde{\mathcal{P}}_h$ is sparse and assumption (K1) holds for the discretized problem with a constant independent of the discretization spaces.

In order to apply Theorem 3.8 and ensure α-robustness of the preconditioner, assumption (K2) is additionally required. It is easy to see that the condition

$$R_{1,h} + t\, R_{2,h} \subset Y_h \tag{5.7}$$

implies the surjectivity of $K_{R_h} = (K_{R_{1,h}}, K_{R_{2,h}})^\top$. Here $R_{1,h} + t\, R_{2,h}$ denotes the space of all functions q_{R_h} on Q_T of the form $q_{R_h}(x,t) = q_{R_{1,h}}(x) + t\, q_{R_{2,h}}(x)$ with $q_{R_{1,h}} \in R_{1,h}$, $q_{R_{2,h}} \in R_{2,h}$. Consequently, assumption (K2) holds and the preconditioner is robust for all positive values of α below some threshold, which was set to 1 in Theorem 3.8 for simplicity. But it is not clear how the constant c_R in assumption (K2) depends on this threshold and the discretization spaces. An analysis of this dependency remains an unresolved problem. Instead we present in the next section numerical experiments for a particular choice of the discretization spaces and report on promising preliminary numerical results.

5.2. Numerical results. We consider the optimal control problem from subsection 4.2 with $\Omega = (0,1)^2$, $\omega = (1/4, 3/4)^2$, $T = 1$, and homogeneous data. The following discretization spaces are used:

$$\begin{aligned} Y_h &= S_{p,\ell}(0,T) \otimes \left[S_{p,\ell}(\Omega) \cap H_0^1(\Omega)\right], \\ U_h &= S_{p,\ell,p-3}(0,T) \otimes S_{p,\ell,p-3}(\Omega), \\ R_{1,h} &= S_{p,\ell}(\Omega) \cap H_0^1(\Omega), \\ R_{2,h} &= S_{p,\ell}(\Omega) \cap H_0^1(\Omega). \end{aligned}$$

Here, $S_{p,\ell,k}(a,b)$ denotes the space of splines of degree p on an equidistant knot span of the interval (a,b) of mesh size $h = (b-a)/2^\ell$ which are k-times continuously differentiable. Spline spaces of maximal continuity, i.e., $k = p-1$, are denoted $S_{p,\ell}(a,b)$. Spline spaces on Ω are defined as tensor products of univariate splines spaces. It is easy to see that the chosen discretization spaces satisfy (5.1), (5.5), (5.6), and (5.7) for spline degree $p \geq 2$.

We use the sparse preconditioner $\mathcal{P}_h$ from (5.3). The application of the preconditioner $\mathcal{P}_h$ requires the multiplication of the inverses of its diagonal blocks with vectors. The action of the inverse of P_{U_h} and $P_{R_{1,h}}$ is efficiently computed by exploiting the tensor product structure and computing the inverse of univariate mass matrices. For P_{Y_h} and $P_{R_{2,h}}$ sparse direct solvers are used.

The system (5.2) is solved using the preconditioned MINRES method with random initial guess. The performance of MINRES depends subtly on the distribution of the eigenvalues of the preconditioned matrix. A well-known estimate for the residual r_k of the kth iterate with $k = 2l$ in terms of the condition number is given by

$$\|\mathcal{P}_h^{-1} r_{2l}\|_{\mathcal{P}_h} \leq 2\, q^l\, \|\mathcal{P}_h^{-1} r_0\|_{\mathcal{P}_h} \quad \text{with} \quad q = \frac{\kappa\left(\mathcal{P}_h^{-1}\mathcal{A}_h\right) - 1}{\kappa\left(\mathcal{P}_h^{-1}\mathcal{A}_h\right) + 1}$$

(see, e.g., [10]), which also motivates our stopping criteria. MINRES is terminated if $\|\mathcal{P}_h^{-1} r_k\|_{\mathcal{P}_h}$ is reduced by a factor of 10^8 from its initial value $\|\mathcal{P}_h^{-1} r_0\|_{\mathcal{P}_h}$.

Table 1 for $p = 2$ and Table 2 for $p = 3$ show the degrees of freedom (DoFs) of the systems for several levels ℓ of refinements and the iteration numbers of MINRES for different values of α.

Reasonably small iteration numbers were observed for $0 < \alpha \ll 1$. For $\alpha = 1$, the iteration numbers are significantly larger. As expected the performance of MINRES does not deteriorate for small values of α. The dependence on the mesh size h is moderate.

Remark 5.2. The iteration numbers in Tables 1 and 2 indicate that the constant c_R in assumption (K2) might depend moderately on the refinement level, while the stronger assumption (K2′) might not hold. Then the observed large iteration numbers

TABLE 1
Iteration numbers using $p = 2$.

$\ell\backslash\alpha$	10^0	10^{-3}	10^{-6}	10^{-9}	DoFs
2	92	39	42	13	3 584
3	174	45	43	22	28 416
4	250	46	43	34	226 304
5	367	52	41	41	1 806 336

TABLE 2
Iteration numbers using $p = 3$.

$\ell\backslash\alpha$	10^0	10^{-3}	10^{-6}	10^{-9}	DoFs
2	132	41	43	38	4 619
3	192	43	43	42	32 303
4	304	45	43	43	241 367
5	388	48	43	43	1 865 639

for $\alpha = 1$ are consistent with Theorem 3.8, which ensures bounded condition numbers for increasing values of α only under the stronger assumption (K2′). The precise influence of the combination of values for α of the order 1 and high refinement levels ℓ remains unclear.

However, a value of α of the order 1 does not pose any difficulty for the construction of an alternative preconditioner based on the standard norms of the spaces Y, U, and R.

Remark 5.3. For more complex domains, isogeometric analysis (cf. [7, 11]) can be used to obtain smooth conforming discretization subspaces, and multipatch domains can be dealt with by methods described in [5] and the references within. For large-scale problems sparse direct solvers eventually fail due to memory limitations. The methods described in [8] can then be considered.

6. Concluding remarks. The main theoretical result of this paper, Theorem 2.2 is the characterization of all robust block diagonal preconditioners for a class of multiple saddle point problems by the relation (2.7). This characterization does not provide an explicit formula for robust preconditioners but it can serve as a tool for constructing them, as was demonstrated in Theorem 3.8 for a quite general class of optimal control problems. Examples of elliptic, parabolic, and hyperbolic control problems were discussed. The hyperbolic optimal control problem was studied in more detail, for which a preconditioner was proposed which performs robustly with respect to the regularization parameter α, if α lies below some threshold, and which shows only a mild dependency on the mesh size h. The robustness with respect to α is shown; the dependency on the threshold for α and on h is not yet covered by the theoretical results.

Appendix A. An auxiliary result. Throughout the paper we have used the following well-known result, for which we present a short proof, for convenience of the reader.

LEMMA A.1. *Let $A : V \longrightarrow V'$ and $B : V \longrightarrow Q'$ be linear operators, where V and Q are Hilbert spaces with dual spaces V' and Q'. Additionally assume that A is*

self-adjoint and coercive. Then we have

$$\left\langle BA^{-1}B'q, q\right\rangle = \sup_{0\neq v\in V} \frac{\langle Bv, q\rangle^2}{\langle Av, v\rangle} \quad \textit{for all } q \in Q.$$

Proof. Observe that

$$\left\langle BA^{-1}B'q, q\right\rangle = \left\langle B'q, A^{-1}B'q\right\rangle = \|B'q\|^2_{A^{-1}}$$

with the norm $\|f\|_{A^{-1}} = (f,f)^{1/2}_{A^{-1}}$ on V', given by the inner product $(f,g)_{A^{-1}} = \langle f, A^{-1}g\rangle$ on V'. By Cauchy's inequality it easily follows for any inner product and associated norm

$$\|f\| = \sup_{0\neq g} \frac{(f,g)}{\|g\|}.$$

In particular, we have

$$\|B'q\|^2_{A^{-1}} = \sup_{0\neq g\in V'} \frac{(g, B'q)^2_{A^{-1}}}{\|g\|^2_{A^{-1}}} = \sup_{0\neq g\in V'} \frac{\left\langle g, A^{-1}B'q\right\rangle^2}{\langle g, A^{-1}g\rangle}.$$

By substituting g by Av it follows that

$$\sup_{0\neq g\in V'} \frac{\left\langle g, A^{-1}B'q\right\rangle^2}{\langle g, A^{-1}g\rangle} = \sup_{0\neq v\in V} \frac{\left\langle Av, A^{-1}B'q\right\rangle^2}{\langle Av, v\rangle} = \sup_{0\neq v\in V} \frac{\langle B'q, v\rangle^2}{\langle Av, v\rangle} = \sup_{0\neq v\in V} \frac{\langle Bv, q\rangle^2}{\langle Av, v\rangle}.$$

∎

Acknowledgment. The authors would like to thank the anonymous referees for their valuable comments and suggestions, which helped to improve this manuscript.

REFERENCES

[1] O. Axelsson and I. Gustafsson, *Preconditioning and two-level multigrid methods of arbitrary degree of approximation*, Math. Comp., 40 (1983), pp. 219–242.

[2] A. Beigl, O. Scherzer, J. Sogn, and W. Zulehner, *Preconditioning inverse problems for hyperbolic equations with applications to photoacoustic tomography*, Inverse Problems, 36 (2019), 014002.

[3] A. Beigl, J. Sogn, and W. Zulehner, *Robust Preconditioners for Multiple Saddle Point Problems and Applications to Optimal Control Problems*, https://arxiv.org/abs/1912.09995, 2019.

[4] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[5] K. Birner and M. Kapl, *The space of c1-smooth isogeometric spline functions on trilinearly parameterized volumetric two-patch domains*, Comput. Aided Geom. Design, 70 (2019), pp. 16–30.

[6] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer Ser. Comput. Math. 44, Springer, New York, 2013.

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

[8] I. Georgieva and C. Hofreither, *Greedy low-rank approximation in Tucker format of solutions of tensor linear systems*, J. Comput. Appl. Math., 358 (2019), pp. 206–220.

[9] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier–Stokes Equations: Theory and Algorithms*, Springer Ser. Comput. Math. 5, Springer, New York, 1986.

[10] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, Front. Appl. Math. 17, SIAM, Philadelphia, 1997.

[11] T. Hughes, J. Cottrell, and Y. Bazilevs, *Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 4135–4195.

[12] O. A. Ladyzhenskaya, *The Boundary Value Problems of Mathematical Physics*, Springer, New York, 1985.
[13] J.-L. Lions and E. Magenes, *Non-Homogeneous Boundary Value Problems and Applications* I, Grundlehren Math. Wiss. 181, Springer, New York, 1972.
[14] J. Liu and J. W. Pearson, *Parameter-robust preconditioning for the optimal control of the wave equation*, Numer. Algorithms, 83 (2020), pp. 1171–1203.
[15] K.-A. Mardal, B. F. Nielsen, and M. Nordaas, *Robust preconditioners for PDE-constrained optimization with limited observations*, BIT, 57 (2017), pp. 405–431.
[16] J. W. Pearson, M. Stoll, and A. J. Wathen, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1126–1152.
[17] J. W. Pearson and A. J. Wathen, *A new approximation of the Schur complement in preconditioners for PDE-constrained optimization*, Numer. Linear Algebra Appl., 19 (2012), pp. 816–829.
[18] 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 (2007), pp. 752–773.
[19] J. Sogn and W. Zulehner, *Schur complement preconditioners for multiple saddle point problems of block tridiagonal form with application to optimization problems*, IMA J. Numer. Anal., 39 (2018), pp. 1328–1359.
[20] W. Zulehner, *Nonstandard norms and robust estimates for saddle point problems*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 536–560.