JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Mathematics

Non-standard Norms and Robust Estimates for Saddle Point Problems

Walter Zulehner

Institute of Computational Mathematics, Johannes Kepler University
Altenberger Str. 69, 4040 Linz, Austria

NuMa-Report No. 2010-07 November 2010
Rev. 1 April 2011

Technical Reports before 1998:

1995

95-1	Hedwig Brandstetter *Was ist neu in Fortran 90?*	March 1995
95-2	G. Haase, B. Heise, M. Kuhn, U. Langer *Adaptive Domain Decomposition Methods for Finite and Boundary Element Equations.*	August 1995
95-3	Joachim Schöberl *An Automatic Mesh Generator Using Geometric Rules for Two and Three Space Dimensions.*	August 1995

1996

96-1	Ferdinand Kickinger *Automatic Mesh Generation for 3D Objects.*	February 1996
96-2	Mario Goppold, Gundolf Haase, Bodo Heise und Michael Kuhn *Preprocessing in BE/FE Domain Decomposition Methods.*	February 1996
96-3	Bodo Heise *A Mixed Variational Formulation for 3D Magnetostatics and its Finite Element Discretisation.*	February 1996
96-4	Bodo Heise und Michael Jung *Robust Parallel Newton-Multilevel Methods.*	February 1996
96-5	Ferdinand Kickinger *Algebraic Multigrid for Discrete Elliptic Second Order Problems.*	February 1996
96-6	Bodo Heise *A Mixed Variational Formulation for 3D Magnetostatics and its Finite Element Discretisation.*	May 1996
96-7	Michael Kuhn *Benchmarking for Boundary Element Methods.*	June 1996

1997

97-1	Bodo Heise, Michael Kuhn and Ulrich Langer *A Mixed Variational Formulation for 3D Magnetostatics in the Space $H(rot) \cap H(div)$*	February 1997
97-2	Joachim Schöberl *Robust Multigrid Preconditioning for Parameter Dependent Problems I: The Stokes-type Case.*	June 1997
97-3	Ferdinand Kickinger, Sergei V. Nepomnyaschikh, Ralf Pfau, Joachim Schöberl *Numerical Estimates of Inequalities in $H^{\frac{1}{2}}$.*	August 1997
97-4	Joachim Schöberl *Programmbeschreibung NAOMI 2D und Algebraic Multigrid.*	September 1997

From 1998 to 2008 technical reports were published by SFB013. Please see
`http://www.sfb013.uni-linz.ac.at/index.php?id=reports`
From 2004 on reports were also published by RICAM. Please see
`http://www.ricam.oeaw.ac.at/publications/list/`
For a complete list of NuMa reports see
`http://www.numa.uni-linz.ac.at/Publications/List/`

NON-STANDARD NORMS AND ROBUST ESTIMATES FOR SADDLE POINT PROBLEMS

WALTER ZULEHNER*

Abstract. In this paper we discuss how to find norms for parameter-dependent saddle point problems which lead to robust (i.e.: parameter-independent) estimates of the solution in terms of the data. In a first step a characterization of such norms is given for a general class of symmetric saddle point problems. Then, for special cases, explicit formulas for these norms are derived. Finally, we will apply these results to distributed optimal control problems for elliptic equations and for the Stokes equations. The norms which lead to robust estimates turn out to differ from the standard norms typically used for these problems. This will lead to block diagonal preconditioners for the corresponding discretized problems with mesh-independent and robust convergence rates if used in preconditioned Krylov subspace methods.

Key words. saddle point problems, PDE-constrained optimization, optimal control, robust estimates, block diagonal preconditioners

AMS subject classifications. 65F08, 65N22, 65K10, 49K40

1. Introduction. In this paper we consider mixed variational problems of the following form: Find $u \in V$ and $p \in Q$ such that

$$\begin{aligned} a(u,v) + b(v,p) &= f(v) \quad \text{for all } v \in V, \\ b(u,q) - c(p,q) &= g(q) \quad \text{for all } q \in Q, \end{aligned} \tag{1.1}$$

where V and Q are Hilbert spaces, a, b, and c are bounded bilinear forms on $V \times V$, $V \times Q$, and $Q \times Q$, respectively, and f, g are bounded linear functionals. Additionally, we will assume throughout the paper that a and c are symmetric, i.e.:

$$a(w,v) = a(v,w) \quad \text{for all } v,w \in V, \quad c(r,q) = c(q,r) \quad \text{for all } q,r \in Q, \tag{1.2}$$

and a and c are non-negative, i.e.:

$$a(v,v) \geq 0 \quad \text{for all } v \in V, \quad c(q,q) \geq 0 \quad \text{for all } q \in Q, \tag{1.3}$$

by which (1.1) becomes a symmetric and indefinite problem.

Examples of such problems are the Stokes problem in fluid mechanics, mixed formulations of elliptic boundary value problems and the optimality system of optimal control problems, where (in all these examples) V and Q are infinite-dimensional function spaces, as well as discretized versions of theses problems, where V and Q are finite-dimensional finite element spaces. In particular, we consider problems which involve some critical model parameter, say α, like the time step size in the time-discretized Stokes problem, the diffusion coefficient in a diffusion-reaction equation or a regularization parameter occurring in optimal control problems. In the discretized version an additional parameter is always involved: the mesh size, say h, of the underlying subdivision of the computational domain.

A fundamental issue is the question whether (1.1) is well-posed: Does there exist a unique solution (u,p) in $X = V \times Q$ for any data (f,g) from X^*, the dual space of X, and does the solution depend continuously on the data, or equivalently, is the

*Institute of Computational Mathematics, Johannes Kepler University, 4040 Linz, Austria (zulehner@numa.uni-linz.ac.at). The work was supported in part by the Austrian Science Foundation (FWF) under the grant W1214/DK12.

norm of the solution (u,p) in X bounded from above by the norm of the data (f,g) in X^*? Of course, the answer depends on the choice of the norm or better the inner product in X. We will concentrate on inner products in X of the particular form

$$\big((v,q),(w,r)\big)_X = (v,w)_V + (q,r)_Q, \tag{1.4}$$

where $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_Q$ are inner products of the Hilbert spaces V and Q, respectively.

Since we have assumed that the bilinear forms a, b, and c are bounded, the norm of the solution (u,p) can always be estimated from below by the norm of the data (f,g). The focus of this paper is the more specific question whether these estimates from above and from below are independent of the model parameter α and, for the discretized version, of the discretization parameter h. We will address this issue by characterizing those inner products in V and Q which lead to robust estimates. Additionally, for some classes of saddle point problems, we will derive explicit representations of such inner products.

Knowledge of robust estimates does not only contribute to the question of well-posedness but also to discretization error estimates and the construction of efficient solvers for the discretized problem. In the discretized case, having robust estimates for an inner product of the form (1.4) translates to having a block diagonal preconditioner for the linear operator describing the left-hand side of (1.1) with robust estimates for the condition number. This would immediately imply that Krylov subspace methods like the minimal residual method, see [22], converge with convergence rates independent of α and h.

Of course, for a wide range of problems robust estimates have been developed already, see the survey article [18] and the many references contained there. Most of the norms involved in these estimates were found on a case-by-case basis. Following the spirit of the survey article [18] the present paper is to be understood as a further contribution to a more systematic search for the "right" norms.

Remark 1. *Observe that Condition* (1.3) *characterizes exactly the case that the functional* $\mathscr{L}(v,q)$*, given by*

$$\mathscr{L}(v,q) = \frac{1}{2}\,a(v,v) + b(v,q) - \frac{1}{2}\,c(q,q) - f(v) - g(q),$$

is a convex function of $v \in V$ *and a concave function of* $q \in Q$*. Such a functional* $\mathscr{L}$ *is called a saddle function, see [25]. If, additionally, Condition* (1.2) *holds, it is easy to see that* (u,p) *solves* (1.1) *if and only if* (u,p) *is a saddle point of* $\mathscr{L}$*, i.e.:*

$$\mathscr{L}(u,q) \le \mathscr{L}(u,p) \le \mathscr{L}(v,p) \quad \text{for all } v \in V,\ q \in Q.$$

The paper is organized as follows: Section 2 contains the abstract framework. The main abstract result is presented in Theorem 2.6 describing necessary and sufficient conditions on the involved inner products for obtaining robust estimates. In Section 3 several special cases are discussed for which inner products leading to robust estimates are explicitly known. Section 4 deals with the application to optimal control problems. The paper ends with a few concluding remarks in Section 5.

Remark 2. *Note that Theorem 2.6 holds for quite general saddle point problems, it requires no further restrictions as those mentioned above, while, for the problems from optimal control discussed in Section 4, we heavily relied on a special common feature of these problems: It was possible to reformulate the problems such that the bilinear forms* a *and* c *only differ by a multiplicative factor.*

2. The abstract theory. Throughout the paper we will use the following notational convention:

Notation 1. *Let H be a Hilbert space with inner product $(\cdot,\cdot)_H$ and associated norm $\|\cdot\|_H$, given by*

$$\|x\|_H = \sqrt{(x,x)_H}.$$

The dual space of H is denoted by H^* with norm $\|\cdot\|_{H^*}$, given by

$$\|\ell\|_{H^*} = \sup_{0\neq x\in H} \frac{\ell(x)}{\|x\|_H}.$$

The duality pairing $\langle\cdot,\cdot\rangle_H$ on $H^*\times H$ is given by

$$\langle \ell, x\rangle_H = \ell(x) \quad \textit{for all } \ell\in H^*,\ x\in H.$$

We will usually drop the subscript H and write instead $\langle\cdot,\cdot\rangle$ for a duality pairing.

Let $\mathcal{I}_H\colon H\longrightarrow H^*$ be given by

$$\langle \mathcal{I}_H x, y\rangle = (x,y)_H.$$

It is well-known that $\mathcal{I}_H$ is an isometric isomorphism between H and its dual space H^*. The inverse $\mathcal{R}_H=\mathcal{I}_H^{-1}$ is called the Riesz-isomorphism, by which functionals in H^* can be identified with elements in H and we have:

$$\langle \ell, x\rangle = (\mathcal{R}_H\ell, x)_H.$$

The set of all linear and bounded operators from H_1 to H_2, where H_1 and H_2 are normed spaces, is denoted by $L(H_1,H_2)$.

The mixed variational problem (1.1) in V and Q can also be written as a variational problem on the product space $X=V\times Q$: Find $x=(u,p)\in X$ such that

$$\mathcal{B}(x,y) = \mathcal{F}(y) \quad \text{for all } y\in X \tag{2.1}$$

with

$$\mathcal{B}(z,y) \ = a(w,v)+b(v,r)+b(w,q)-c(r,q), \quad \mathcal{F}(y)=f(v)+g(q)$$

for $y=(v,q)$, $z=(w,r)$.

Since we have assumed that the bilinear forms a, b, and c are bounded, there is a constant, say $\overline{c}_x$, such that

$$\sup_{0\neq z\in X}\sup_{0\neq y\in X}\frac{\mathcal{B}(z,y)}{\|z\|_X\,\|y\|_X}\le \overline{c}_x<\infty. \tag{2.2}$$

A classical result due to Babuška, see [2], [3], reads in our symmetric situation: Problem (2.1) is well-posed if and only if there is a constant, say $\underline{c}_x$, such that

$$\inf_{0\neq z\in X}\sup_{0\neq y\in X}\frac{\mathcal{B}(z,y)}{\|z\|_X\,\|y\|_X}\ge \underline{c}_x>0 \tag{2.3}$$

and we have the following estimates from above and from below:

$$\frac{1}{\overline{c}_x}\,\|\mathcal{F}\|_{X^*}\le \|x\|_X\le \frac{1}{\underline{c}_x}\,\|\mathcal{F}\|_{X^*}.$$

Condition (2.3) is usually referred to as the inf-sup condition or the Babuška-Brezzi condition.

To each of the three bilinear forms a, b, and c we associate a corresponding linear operator $A \in L(V, V^*)$, $B \in L(V, Q^*)$, and $C \in L(Q, Q^*)$, respectively, given by

$$\langle Aw, v\rangle = a(w, v), \quad \langle Bw, q\rangle = b(w, q), \quad \langle Cr, q\rangle = c(r, q).$$

Additionally, $B^* \in L(Q, V^*)$ denotes the adjoint of $B \in L(V, Q^*)$, given by

$$\langle B^* r, v\rangle = \langle Bv, r\rangle.$$

The problem (1.1) now reads in operator notation:

$$\begin{array}{rcl} Au + B^* p &=& f, \\ Bu - Cp &=& g. \end{array} \tag{2.4}$$

In a similar way, we associate a linear operator $\mathcal{A} \in L(X, X^*)$ to the bilinear form $\mathcal{B}$, given by

$$\langle \mathcal{A}x, y\rangle = \mathcal{B}(x, y).$$

Then the problem (2.1), which is equivalent to (1.1), reads

$$\mathcal{A}x = \mathcal{F}. \tag{2.5}$$

In operator notation the important conditions (2.2) and (2.3) can be written in the following form:

$$\underline{c}_x \|z\|_X \le \|\mathcal{A}z\|_{X^*} \le \overline{c}_x \|z\|_X \quad \text{for all } z \in X. \tag{2.6}$$

The aim in this paper is to find inner products in V and Q which lead to such estimates with coefficients $\underline{c}_x$ and $\overline{c}_x$ independent of the model parameter α and, for the discretized version, also independent of the mesh size h. An immediate consequence of (2.6) is an estimation of the condition number $\kappa(\mathcal{A})$:

$$\kappa(\mathcal{A}) = \|\mathcal{A}\|_{L(X,X^*)} \|\mathcal{A}^{-1}\|_{L(X^*,X)} \le \frac{\overline{c}_x}{\underline{c}_x}.$$

So, robust estimates of the form (2.6) imply a robust estimate for the condition number, an important property in connection with convergence rates of iterative methods for solving (2.5).

We start the analysis of (1.1) by a very simple and helpful observation:

LEMMA 2.1. *Let $\ell_1 \in V^*$ and $\ell_2 \in Q^*$. Then*

$$\|\ell\|_{X^*}^2 = \|\ell_1\|_{V^*}^2 + \|\ell_2\|_{Q^*}^2$$

for $\ell \in X^*$, given by $\ell(v, q) = \ell_1(v) + \ell_2(q)$.

Proof. By using the Cauchy inequality we obtain

$$\begin{aligned} \|\ell\|_{X^*}^2 &= \sup_{0 \ne (v,q) \in X} \frac{\left(\langle \ell_1, v\rangle + \langle \ell_2, q\rangle\right)^2}{\|(v, q)\|_X^2} \\ &\le \sup_{0 \ne (v,q) \in X} \frac{\left(\|\ell_1\|_{V^*}\|v\|_V + \|\ell_2\|_{Q^*}\|q\|_Q\right)^2}{\|v\|_V^2 + \|q\|_Q^2} \le \|\ell_1\|_{V^*}^2 + \|\ell_2\|_{Q^*}^2. \end{aligned}$$

Equality follows for the choice $v = \mathcal{R}_V \ell_1$ and $q = \mathcal{R}_Q \ell_2$. ∎

As a consequence of an estimate of the form (2.6) in X we obtain two simple estimates, one in V and one in Q:

THEOREM 2.2. *If* (2.6) *holds for constants* $\underline{c}_x, \overline{c}_x > 0$, *then*

$$\underline{c}_x^2 \|w\|_V^2 \le \|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2 \le \overline{c}_x^2 \|w\|_V^2 \quad \textit{for all } w \in V \tag{2.7}$$

and

$$\underline{c}_x^2 \|r\|_Q^2 \le \|Cr\|_{Q^*}^2 + \|B^* r\|_{V^*}^2 \le \overline{c}_x^2 \|r\|_Q^2 \quad \textit{for all } r \in Q. \tag{2.8}$$

Proof. For $z = (w, r)$ we have:

$$\begin{aligned} \|\mathcal{A} z\|_{X^*} &= \sup_{0 \neq (v,q) \in X} \frac{\mathcal{B}((w,r),(v,q))}{\|(v,q)\|_X} \\ &= \sup_{0 \neq (v,q) \in X} \frac{a(w,v) + b(v,r) + b(w,q) - c(r,q)}{\|(v,q)\|_X} \quad = \sup_{0 \neq (v,q) \in X} \frac{\ell_1(v) + \ell_2(q)}{\|(v,q)\|_X} \end{aligned}$$

with

$$\begin{aligned} \ell_1(v) &= a(w,v) + b(v,r) = \langle Aw + B^* r, v \rangle, \\ \ell_2(q) &= b(w,q) - c(r,q) = \langle Bw - Cr, q \rangle. \end{aligned}$$

Therefore, by Lemma 2.1, we obtain

$$\|\mathcal{A} z\|_{X^*} = \left(\|Aw + B^* r\|_{V^*}^2 + \|Bw - Cr\|_{Q^*}^2 \right)^{1/2}.$$

Then the estimates (2.7) and (2.8) immediately follow from (2.6) for $r = 0$ and for $w = 0$, respectively. ∎

So, (2.7) and (2.8) are necessary conditions for (2.6). Next we will show that (2.7) and (2.8), not necessarily with the same constants, are also sufficient:

THEOREM 2.3. *If there are constants* $\underline{c}_v, \overline{c}_v, \underline{c}_q, \overline{c}_q > 0$ *such that*

$$\underline{c}_v^2 \|w\|_V^2 \le \|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2 \le \overline{c}_v^2 \|w\|_V^2 \quad \textit{for all } w \in V$$

and

$$\underline{c}_q^2 \|r\|_Q^2 \le \|Cr\|_{Q^*}^2 + \|B^* r\|_{V^*}^2 \le \overline{c}_q^2 \|r\|_Q^2 \quad \textit{for all } r \in Q,$$

then there are constants $\underline{c}_x, \overline{c}_x > 0$ *such that*

$$\underline{c}_x \|z\|_X \le \|\mathcal{A} z\|_{X^*} \le \overline{c}_x \|z\|_X \quad \textit{for all } z \in X,$$

where $\underline{c}_x$ *and* $\overline{c}_x$ *depend only on* $\underline{c}_v$, $\underline{c}_q$, $\overline{c}_v$, *and* $\overline{c}_q$.

Proof. For $z = (w, r)$ we have:

$$\begin{aligned} \|\mathcal{A} z\|_{X^*} &= \left(\|Aw + B^* r\|_{V^*}^2 + \|Bw - Cr\|_{Q^*}^2 \right)^{1/2} \\ &\le \left(2\|Aw\|_{V^*}^2 + 2\|B^* r\|_{V^*}^2 + 2\|Bw\|_{Q^*}^2 + 2\|Cr\|_{Q^*}^2 \right)^{1/2} \\ &\le \left(2\overline{c}_v^2 \|w\|_V^2 + 2\overline{c}_q^2 \|r\|_Q^2 \right)^{1/2} \le \sqrt{2}\, \max(\overline{c}_v, \overline{c}_q)\, \|z\|_X, \end{aligned}$$

which proves the upper bound in (2.6) with $\overline{c}_x = \sqrt{2}\,\max(\overline{c}_v, \overline{c}_q)$.

For showing a lower bound, we start with the following estimate based on the triangle inequality in X^*:

$$\begin{aligned}\|\mathcal{A}z\|_{X^*} &= \left(\|Aw + B^*r\|_{V^*}^2 + \|Bw - Cr\|_{Q^*}^2\right)^{1/2} \\ &\geq \left(\|B^*r\|_{V^*}^2 + \|Bw\|_{Q^*}^2\right)^{1/2} - \left(\|Aw\|_{V^*}^2 + \|Cr\|_{Q^*}^2\right)^{1/2} \\ &= (\eta - \xi)\,\|z\|_X\end{aligned}$$

for $z = (w, r)$ with

$$\xi = \frac{\left(\|Aw\|_{V^*}^2 + \|Cr\|_{Q^*}^2\right)^{1/2}}{\|(w,r)\|_X}, \quad \eta = \frac{\left(\|B^*r\|_{V^*}^2 + \|Bw\|_{Q^*}^2\right)^{1/2}}{\|(w,r)\|_X}.$$

A second lower bound follows from:

$$\|\mathcal{A}z\|_{X^*} = \sup_{0\neq(v,q)\in X} \frac{\mathcal{B}((w,r),(v,q))}{\|(v,q)\|_X} \geq \frac{\mathcal{B}((w,r),(w,-r))}{\|(w,-r)\|_X} = \frac{a(w,w) + c(r,r)}{\|z\|_X}.$$

Since

$$a(w,v)^2 \leq a(w,w)\,a(v,v) \leq a(w,w)\,\|Av\|_{V^*}\|v\|_V \leq \overline{c}_v\,a(w,w)\,\|v\|_V^2,$$

we have

$$\|Aw\|_{V^*}^2 = \sup_{0\neq v\in V} \frac{a(w,v)^2}{\|v\|_V^2} \leq \overline{c}_v\,a(w,w).$$

Analogously, we obtain

$$\|Cr\|_{Q^*}^2 \leq \overline{c}_q\,c(r,r).$$

Hence

$$\begin{aligned}a(w,w) + c(r,r) &\geq \frac{1}{\overline{c}_v}\,\|Aw\|_{V^*}^2 + \frac{1}{\overline{c}_q}\,\|Cr\|_{Q^*}^2 \geq \frac{1}{\max(\overline{c}_v,\overline{c}_q)}\,\left(\|Aw\|_{V^*}^2 + \|Cr\|_{Q^*}^2\right) \\ &= \frac{\sqrt{2}}{\overline{c}_x}\,\xi^2\,\|(w,r)\|_X^2.\end{aligned}$$

With this estimate we obtain for the second lower bound:

$$\|\mathcal{A}z\|_{X^*} \geq \frac{\sqrt{2}}{\overline{c}_x}\,\xi^2\,\|z\|_X.$$

Taking the maximum of the two lower bounds we immediately obtain

$$\|\mathcal{A}z\|_{X^*} \geq \varphi(\xi,\eta)\,\|z\|_X \quad \text{with} \quad \varphi(\xi,\eta) = \max\left[\eta - \xi, \frac{\sqrt{2}}{\overline{c}_x}\,\xi^2\right].$$

Observe that

$$\xi^2 + \eta^2 \geq \underline{c}_v^2 + \underline{c}_q^2 > 0.$$

Elementary calculations show that

$$\min\left\{\varphi(\xi,\eta)\colon \xi^2+\eta^2 \geq \underline{c}_v^2+\underline{c}_q^2\right\} \;\geq \frac{3-\sqrt{5}}{4}\,\frac{\underline{c}_v^2+\underline{c}_q^2}{\max(\overline{c}_v,\overline{c}_q)} = \underline{c}_x,$$

which concludes the proof. □

In the following two lemmas it will be shown that the conditions (2.7) and (2.8) of the last theorem can be replaced by two other conditions which will turn out to be more easy to work with.

LEMMA 2.4. *If there are constants $\underline{\gamma}_v, \overline{\gamma}_v > 0$ such that*

$$\underline{\gamma}_v \,\|w\|_V^2 \leq a(w,w) + \|Bw\|_{Q^*}^2 \leq \overline{\gamma}_v \,\|w\|_V^2 \quad \text{for all } w \in V, \tag{2.9}$$

then (2.7) *is satisfied with constants $\underline{c}_v, \overline{c}_v > 0$ that depend only on $\underline{\gamma}_v, \overline{\gamma}_v$.*

And, vice versa, if there are constants $\underline{c}_v, \overline{c}_v > 0$ such that (2.7) *is satisfied, then* (2.9) *is satisfied with constants $\underline{\gamma}_v, \overline{\gamma}_v > 0$ that depend only on $\underline{c}_v, \overline{c}_v$.*

Proof. Assume that (2.9) is satisfied. Then we have

$$a(w,v)^2 \leq a(w,w)a(v,v) \leq \overline{\gamma}_v \, a(w,w)\|v\|_V^2,$$

which implies

$$\|Aw\|_{V^*}^2 \leq \overline{\gamma}_v \, a(w,w).$$

Therefore,

$$\begin{aligned}\|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2 &\leq \overline{\gamma}_v \, a(w,w) + \|Bw\|_{Q^*}^2 \\ &\leq \max(\overline{\gamma}_v, 1)\, \left(a(w,w) + \|Bw\|_{Q^*}^2\right) \leq \max(\overline{\gamma}_v, 1)\, \overline{\gamma}_v \,\|w\|_V^2.\end{aligned}$$

This shows the upper bound in (2.7) for $\overline{c}_v^2 = \max(\overline{\gamma}_v, 1)\, \overline{\gamma}_v$.

For the lower bound observe that, for all $\varepsilon > 0$:

$$a(w,w) \leq \|Aw\|_{V^*} \,\|w\|_V \leq \frac{1}{2\varepsilon}\,\|Aw\|_{V^*}^2 + \frac{\varepsilon}{2}\,\|w\|_V^2, \tag{2.10}$$

which implies

$$\underline{\gamma}_v \,\|w\|_V^2 \leq a(w,w) + \|Bw\|_{Q^*}^2 \leq \frac{1}{2\varepsilon}\,\|Aw\|_{V^*}^2 + \frac{\varepsilon}{2}\,\|w\|_V^2 + \|Bw\|_{Q^*}^2,$$

and, therefore,

$$\left(\underline{\gamma}_v - \frac{\varepsilon}{2}\right)\,\|w\|_V^2 \leq \frac{1}{2\varepsilon}\,\|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2 \leq \max\left(\frac{1}{2\varepsilon}, 1\right)\left(\|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2\right).$$

For $\varepsilon = \underline{\gamma}_v$ we obtain the lower bound in (2.7) with $\underline{c}_v = \min(\underline{\gamma}_v, 1/2)\underline{\gamma}_v$.

Now assume that (2.7) is satisfied. Then we have, see the proof of the last theorem:

$$\|Aw\|_{V^*}^2 \leq \overline{c}_v \, a(w,w)$$

and, therefore,

$$\begin{aligned}a(w,w) + \|Bw\|_{Q^*}^2 &\geq \overline{c}_v^{-1}\,\|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2 \\ &\geq \min(1, \overline{c}_v^{-1})\,\left(\|Aw\|_{V^*}^2 + \|Bw\|_{Q^*}^2\right) \geq \min(1, \overline{c}_v^{-1})\,\underline{c}_v^2\,\|w\|_V^2\end{aligned}$$

showing the lower bound in (2.9) for $\underline{\gamma}_v = \min(1, \overline{c}_v^{-1})\, \underline{c}_v^2$.

For the upper bound we use (2.10) for $\varepsilon = 1/2$ and obtain:

$$a(w,w) + \|Bw\|_{Q^*}^2 \le \|Aw\|_{V^*}^2 + \frac{1}{4}\|w\|_V^2 + \|Bw\|_{Q^*}^2 \le \left(\overline{c}_v^2 + \frac{1}{4}\right)\|w\|_V^2.$$

So, the upper bound in (2.9) is satisfied for $\overline{\gamma}_v = \overline{c}_v^2 + 1/4$. □

Completely analogously, we have

LEMMA 2.5. *If there are constants $\underline{\gamma}_q, \overline{\gamma}_q > 0$ such that*

$$\underline{\gamma}_q \|r\|_Q^2 \le c(r,r) + \|B^* r\|_{V^*}^2 \le \overline{\gamma}_q \|r\|_Q^2 \quad \textit{for all } r \in Q, \tag{2.11}$$

then (2.8) *is satisfied with constants $\underline{c}_q, \overline{c}_q > 0$ that depend only on $\underline{\gamma}_q, \overline{\gamma}_q$.*

And, vice versa, if there are constants $\underline{c}_q, \overline{c}_q > 0$ such that (2.8) *is satisfied, then* (2.11) *is satisfied with constants $\underline{\gamma}_q, \overline{\gamma}_q > 0$ that depend only on $\underline{c}_q, \overline{c}_q$.*

By summarizing the results of the last two theorems and lemmas we finally obtain

THEOREM 2.6. *If there are constants $\underline{\gamma}_v, \overline{\gamma}_v, \underline{\gamma}_q, \overline{\gamma}_q > 0$ such that*

$$\underline{\gamma}_v \|w\|_V^2 \le a(w,w) + \|Bw\|_{Q^*}^2 \le \overline{\gamma}_v \|w\|_V^2 \quad \textit{for all } w \in V \tag{2.12}$$

and

$$\underline{\gamma}_q \|r\|_Q^2 \le c(r,r) + \|B^* r\|_{V^*}^2 \le \overline{\gamma}_q \|r\|_Q^2 \quad \textit{for all } r \in Q, \tag{2.13}$$

then

$$\underline{c}_x \|z\|_X \le \|\mathcal{A} z\|_{X^*} \le \overline{c}_x \|z\|_X \quad \textit{for all } z \in X \tag{2.14}$$

is satisfied with constants $\underline{c}_x, \overline{c}_x > 0$ that depend only on $\underline{\gamma}_v, \overline{\gamma}_v, \underline{\gamma}_q, \overline{\gamma}_q$. And, vice versa, if the estimates (2.14) *are satisfied with constants $\underline{c}_x, \overline{c}_x > 0$, then the estimates* (2.12) *and* (2.13) *are satisfied with constants $\underline{\gamma}_v, \overline{\gamma}_v, \underline{\gamma}_q, \overline{\gamma}_q > 0$ that depend only on $\underline{c}_x, \overline{c}_x$.*

Remark 3. *In the case $C = 0$ (i.e. $c(v,v) \equiv 0$) the lower estimate in condition* (2.11) *has the special form*

$$\underline{\gamma}_q \|r\|_Q^2 \le \|B^* r\|_{V^*}^2 \quad \textit{for all } r \in Q. \tag{2.15}$$

From the lower estimate in (2.9) *it immediately follows that*

$$\underline{\gamma}_v \|w\|_V^2 \le a(w,w) \quad \textit{for all } w \in \ker B = \{v \in V : Bv = 0\}. \tag{2.16}$$

On the other hand, from (2.15) *and* (2.16) *the lower estimate in* (2.9) *easily follows, using the fact that* (2.15) *implies*

$$\underline{\gamma}_q \|w\|_V^2 \le \|Bw\|_{Q^*}^2 \quad \textit{for all } w \in (\ker B)^\perp,$$

where $(\ker B)^\perp$ *denotes the orthogonal complement of* $\ker B$. *So we have recovered a classical result by Brezzi [8], [9]: Let a and b bounded bilinear forms and $c \equiv 0$. Then the problem* (2.1) *is well-posed if and only if a is coercive on* $\ker B$, *see* (2.16), *and the inf-sup condition for b is satisfied, see* (2.15).

Next we want to rewrite (2.12) and (2.13) in a more convenient form. Using the definition of $\mathcal{I}_Q$ it follows that

$$\begin{aligned}\|Bw\|_{Q^*}^2 &= \sup_{0\neq q\in Q}\frac{\langle Bw,q\rangle^2}{\|q\|_Q^2} = \sup_{0\neq q\in Q}\frac{(\mathcal{I}_Q^{-1}Bw,q)_Q^2}{\|q\|_Q^2}\\ &= \|\mathcal{I}_Q^{-1}Bw\|_Q^2 = (\mathcal{I}_Q^{-1}Bw,\mathcal{I}_Q^{-1}Bw)_Q = \langle Bw,\mathcal{I}_Q^{-1}Bw\rangle = \langle B^*\mathcal{I}_Q^{-1}Bw,w\rangle\end{aligned}$$

and, similarly

$$\|B^*r\|_{V^*}^2 = \langle B\mathcal{I}_V^{-1}B^*r,r\rangle.$$

Then the conditions (2.12) and (2.13) read:

$$\underline{\gamma}_v\,\langle\mathcal{I}_Vw,w\rangle \le \langle(A+B^*\mathcal{I}_Q^{-1}B)w,w\rangle \le \overline{\gamma}_v\,\langle\mathcal{I}_Vw,w\rangle \quad \text{for all } w\in V$$

and

$$\underline{\gamma}_q\,\langle\mathcal{I}_Qr,r\rangle \le \langle(C+B\mathcal{I}_V^{-1}B^*)r,r\rangle \le \overline{\gamma}_q\,\langle\mathcal{I}_Qr,r\rangle \quad \text{for all } r\in Q$$

or, in short,

$$\mathcal{I}_V \sim A+B^*\mathcal{I}_Q^{-1}B \quad \text{and} \quad \mathcal{I}_Q\sim C+B\mathcal{I}_V^{-1}B^*, \tag{2.17}$$

by using the following notation:

Notation 2. *Let $M,N\in L(H,H^*)$ be two linear self-adjoint operators. Then*

1. *$M\le N$, if and only if*

$$\langle Mx,x\rangle \le \langle Nx,x\rangle \quad \textit{for all } x\in H.$$

2. *$M\lesssim N$, if and only if there is a constant $c\ge 0$ such that $M\le cN$.*
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 α and h, see the introduction), then we additionally assume that the involved constants are independent of those parameters.

It is clear that (2.17) is equivalent to

$$\mathcal{I}_V \sim A+B^*(C+B\mathcal{I}_V^{-1}B^*)^{-1}B \quad \text{and} \quad \mathcal{I}_Q\sim C+B\mathcal{I}_V^{-1}B^* \tag{2.18}$$

and also to

$$\mathcal{I}_Q \sim C+B(A+B^*\mathcal{I}_Q^{-1}B)^{-1}B^* \quad \text{and} \quad \mathcal{I}_V\sim A+B^*\mathcal{I}_Q^{-1}B. \tag{2.19}$$

The equivalent pairs of conditions (2.17), (2.18), and (2.19) are conditions for $\mathcal{I}_V$ and $\mathcal{I}_Q$, or in other words, for the inner products in V and Q. In (2.18) as well as in (2.19) the first condition involves only one unknown operator ($\mathcal{I}_V$ in (2.18) and $\mathcal{I}_Q$ in (2.19)). The second condition (in (2.18) as well as in (2.19)) serves as a sort of definition of the second operator in terms of the first one.

3. Special cases. In this section we will discuss the conditions (2.17) for several important cases. We will focus on the finite-dimensional case, on the one hand since we are mainly interested in the construction of preconditioners for discretized problems, on the other hand to avoid some technical difficulties in the infinite-dimensional case. Nevertheless, all results for the finite-dimensional case can be carried over to the infinite-dimensional case under appropriate conditions.

If V and Q are finite-dimensional, then all operators can be represented as matrices acting on vectors of real numbers representing the elements in V and Q with respect to some chosen bases. In this matrix-vector notation problem (2.5) becomes a linear system of the following form

$$\mathcal{A}\begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \quad \text{with} \quad \mathcal{A} = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}, \tag{3.1}$$

where B^T denotes the transposed matrix. For the matrix $\mathcal{I}_X$ representing the inner product in X we have the special form

$$\mathcal{I}_X = \begin{bmatrix} \mathcal{I}_V & 0 \\ 0 & \mathcal{I}_Q \end{bmatrix}$$

because of (1.4). Condition (2.6) is satisfied for all constants $\underline{c}_x, \overline{c}_x > 0$ with

$$\underline{c}_x \leq |\lambda_{\min}| \quad \text{and} \quad |\lambda_{\max}| \leq \overline{c}_x,$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are eigenvalues of the generalized eigenvalue problem

$$\mathcal{A}x = \lambda \mathcal{I}_X x$$

of maximal and minimal modulus, respectively. For the condition number we then obtain

$$\kappa(\mathcal{A}) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|} \leq \frac{\overline{c}_x}{\underline{c}_x}.$$

In the survey article [4] a wide range of preconditioners for linear systems of the form (3.1) are discussed, among other topics. Following the notation of [4] our focus is the case of block diagonal preconditioners $\mathcal{P} = \mathcal{I}_X$ and our particular interest is the issue of robustness.

We first consider two simple cases. Note that the preconditioners mentioned in the following two subsections as well as their analysis are well-known in general and in the context of various applications, see [4] and the many references there on block diagonal preconditioners. They are included here as preliminaries for Subsection 3.3.

3.1. A and $S = C + BA^{-1}B^T$ are non-singular. The matrix S is called the (negative) Schur complement. In this case

$$\mathcal{I}_V = A \quad \text{and} \quad \mathcal{I}_Q = S = C + BA^{-1}B^T \tag{3.2}$$

satisfy (2.17): Since

$$0 \leq B^T(C + B\mathcal{I}_V^{-1}B^T)^{-1}B \leq \mathcal{I}_V \quad \text{for all } \mathcal{I}_V,$$

we have for $\mathcal{I}_V = A$:

$$\mathcal{I}_V = A \leq A + B^T\mathcal{I}_Q^{-1}B = A + B^T(C + B\mathcal{I}_V^{-1}B^T)^{-1}B \leq A + \mathcal{I}_V = 2A = 2\mathcal{I}_V.$$

So, (2.12) holds for $\underline{\gamma}_v = 1$ and $\overline{\gamma}_v = 2$, while (2.13) trivially holds for $\underline{\gamma}_q = \overline{\gamma}_q = 1$. Then robust estimates directly follow from Theorem 2.6.

Preconditioners of this type were proposed and analyzed, e.g., in [26], [32], [29] in the context of mixed formulations of second-order elliptic equations and the Stokes problem.

Remark 4. *For $C = 0$ we have direct access to the constants in* (2.6). *The generalized eigenvalue problem*

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \lambda \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix}$$

has exactly 3 eigenvalues, namely 1, $(1+\sqrt{5})/2$ *and* $(1-\sqrt{5})/2$, *see, e.g., [14], [19]. Therefore,* (2.6) *is satisfied for*

$$\underline{c}_x = \frac{\sqrt{5}-1}{2} \quad \textit{and} \quad \overline{c}_x = \frac{\sqrt{5}+1}{2}.$$

3.2. C **and** $R = A + B^T C^{-1} B$ **are non-singular.** Analogous to the first case one can show that

$$\mathcal{I}_V = R = A + B^T C^{-1} B \quad \text{and} \quad \mathcal{I}_Q = C \tag{3.3}$$

satisfy (2.17).

3.3. A **and** C **are non-singular.** Then, of course, both solutions, (3.2) and (3.3), are available. It is easy to conclude from (2.18) and (2.19) that for all possible choices of $\mathcal{I}_V$ and $\mathcal{I}_Q$ satisfying (2.17) we have

$$A \lesssim \mathcal{I}_V \lesssim A + B^T C^{-1} B \quad \text{and} \quad C \lesssim \mathcal{I}_Q \lesssim C + B A^{-1} B^T.$$

So, the solutions (3.2) and (3.3) determine the lower and upper bounds for the range of all possible solutions.

We will now construct additional solutions to (2.17). With

$$\hat{\mathcal{I}}_V = A^{-1/2} \mathcal{I}_V A^{-1/2}, \quad \hat{\mathcal{I}}_Q = C^{-1/2} \mathcal{I}_Q C^{-1/2}, \quad \text{and} \quad \hat{B} = C^{-1/2} B A^{-1/2}$$

condition (2.17) takes the following form:

$$\hat{\mathcal{I}}_V \sim I + \hat{B}^T \hat{\mathcal{I}}_Q^{-1} \hat{B} \quad \text{and} \quad \hat{\mathcal{I}}_Q \sim I + \hat{B} \hat{\mathcal{I}}_V^{-1} \hat{B}^T. \tag{3.4}$$

We first discuss the special case where the equivalence relation $\sim$ is replaced by equality:

$$\hat{\mathcal{I}}_V = I + \hat{B}^T \hat{\mathcal{I}}_Q^{-1} \hat{B} \quad \text{and} \quad \hat{\mathcal{I}}_Q = I + \hat{B} \hat{\mathcal{I}}_V^{-1} \hat{B}^T.$$

Eliminating $\hat{\mathcal{I}}_Q$ from the first equation with the help of the second equation yields:

$$\hat{\mathcal{I}}_V = I + \hat{B}^T (I + \hat{B} \hat{\mathcal{I}}_V^{-1} \hat{B}^T)^{-1} \hat{B}. \tag{3.5}$$

By using the Sherman-Morrison-Woodbury formula, see, e.g., [11], we have

$$(I + \hat{B} \hat{\mathcal{I}}_V^{-1} \hat{B}^T)^{-1} = I - \hat{B} (\hat{\mathcal{I}}_V + \hat{B}^T \hat{B})^{-1} \hat{B}^T.$$

Therefore,

$$\begin{aligned}\hat{B}^T(I+\hat{B}\hat{\mathcal{I}}_V^{-1}\hat{B}^T)^{-1}\hat{B} &= \hat{B}^T\hat{B} - \hat{B}^T\hat{B}(\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})^{-1}\hat{B}^T\hat{B}\\ &= \hat{B}^T\hat{B}(\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})^{-1}\left[(\mathcal{I}_V+\hat{B}^T\hat{B})-\hat{B}^T\hat{B}\right] = \hat{B}^T\hat{B}(\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})^{-1}\mathcal{I}_V\\ &= \left[(\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})-\hat{\mathcal{I}}_V\right](\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})^{-1}\mathcal{I}_V = \mathcal{I}_V - \hat{\mathcal{I}}_V(\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})^{-1}\mathcal{I}_V.\end{aligned}$$

This allows to rewrite the condition (3.5):

$$\hat{\mathcal{I}}_V = I + \hat{\mathcal{I}}_V - \hat{\mathcal{I}}_V(\hat{\mathcal{I}}_V+\hat{B}^T\hat{B})^{-1}\hat{\mathcal{I}}_V,$$

which simplifies to the quadratic matrix equation

$$\left(\hat{\mathcal{I}}_V\right)^2 - \hat{\mathcal{I}}_V - \hat{B}^T\hat{B} = 0.$$

It is easy to see that

$$\hat{\mathcal{I}}_V = f(\hat{B}^T\hat{B}) \quad \text{with} \quad f(x) = \frac{1}{2} + \sqrt{\frac{1}{4}+x}$$

solves this equation, see [13] for matrix functions. For $\mathcal{I}_Q$ we obtain similarly:

$$\hat{\mathcal{I}}_Q = f(\hat{B}\hat{B}^T).$$

We now return from the equality conditions to the original equivalence conditions (3.4). Obviously, all matrices $\hat{\mathcal{I}}_V$ and $\hat{\mathcal{I}}_Q$ which are spectrally equivalent to $f(\hat{B}^T\hat{B})$ and $f(\hat{B}\hat{B}^T)$, respectively, satisfy the equivalence conditions (3.4). Since

$$\sqrt{1+x} \le f(x) \le \frac{2}{\sqrt{3}}\sqrt{1+x} \quad \text{for all } x \ge 0,$$

we have

$$f(\hat{B}^T\hat{B}) \sim (I+\hat{B}^T\hat{B})^{1/2} \quad \text{and} \quad f(\hat{B}\hat{B}^T) \sim (I+\hat{B}\hat{B}^T)^{1/2}.$$

Therefore,

$$\hat{\mathcal{I}}_V = (I+\hat{B}^T\hat{B})^{1/2} \quad \text{and} \quad \hat{\mathcal{I}}_Q = (I+\hat{B}\hat{B}^T)^{1/2}$$

satisfy (3.4). If expressed in terms on the original matrices, we obtain

$$\mathcal{I}_V = A^{1/2}\left(A^{-1/2}RA^{-1/2}\right)^{1/2}A^{1/2} \quad \text{and} \quad \mathcal{I}_Q = C^{1/2}\left(C^{-1/2}SC^{-1/2}\right)^{1/2}C^{1/2},$$

or, in short:

$$\mathcal{I}_V = [A,R]_{1/2} \quad \text{and} \quad \mathcal{I}_Q = [C,S]_{1/2},$$

using the following notation:

Notation 3. *Let M and N be symmetric and positive definite n-by-n matrices. Then, for all $\theta \in [0,1]$, the symmetric and positive definite matrix $[M,N]_\theta$ is given by*

$$[M,N]_\theta = M^{1/2}\left(M^{-1/2}NM^{-1/2}\right)^\theta M^{1/2}.$$

Remark 5. *Each of the symmetric and positive definite matrices M and N represents an inner product and, therefore, a Hilbert space structure on $\mathbb{R}^n$. For each $\theta \in (0,1)$ an intermediate Hilbert space structure on $\mathbb{R}^n$ can be defined using the so-called real method, see [5], whose inner product is represented by the symmetric and positive definite matrix $[M,N]_\theta$ up to a scaling factor.*

More generally, we obtain solutions to (3.4) of the form

$$\hat{\mathcal{I}}_V = (I + \hat{B}^T\hat{B})^\theta \quad \text{and} \quad \hat{\mathcal{I}}_Q = (I + \hat{B}\hat{B}^T)^{1-\theta}$$

for each $\theta \in [0,1]$. This is a direct consequence of the inequalities:

$$(1+x)^\theta \le 1 + \frac{x}{(1+x)^{1-\theta}} \le 2\,(1+x)^\theta \quad \text{for all } x \ge 0.$$

Translating this result to the untransformed quantities leads to the following solutions to (2.17):

$$\mathcal{I}_V = [A,R]_\theta \quad \text{and} \quad \mathcal{I}_Q = [C,S]_{1-\theta} = [R,C]_\theta. \tag{3.6}$$

Remark 6. *This family of solutions can also be derived from the solutions* (3.2) *and* (3.3) *by using the interpolation theorem, see [5].*

Since

$$\frac{1}{2^{1-\theta}}(1+x^\theta) \le (1+x)^\theta \le 1 + x^\theta \quad \text{for all } x \ge 0,$$

it easily follows that

$$(I + \hat{B}^T\hat{B})^\theta \sim I + (\hat{B}^T\hat{B})^\theta \quad \text{and} \quad (I + \hat{B}\hat{B}^T)^{1-\theta} \sim I + (\hat{B}\hat{B}^T)^{1-\theta},$$

which implies that

$$\mathcal{I}_V = A + [A, B^TC^{-1}B]_\theta \quad \text{and} \quad \mathcal{I}_Q = C + [C, BA^{-1}B^T]_{1-\theta}. \tag{3.7}$$

also satisfy (2.17). This form of a solution will turn out to be particularly useful for the optimal control problems in Section 4.

Summarizing the results in this section, we have identified several pairs $(\mathcal{I}_V, \mathcal{I}_Q)$ of matrices, see (3.2), (3.3), (3.6), and (3.7), representing inner products, which satisfy (2.17) under appropriate assumptions. Which of these pairs are more appropriate for preconditioning than others depend on the particular application. Two applications from optimal control will be discussed next.

4. Applications. Let Ω be an open and bounded domain in $\mathbb{R}^d$ for $d \in \{1,2,3\}$ with Lipschitz-continuous boundary Γ and let $L^2(\Omega)$, $H^1(\Omega)$, and $H^1_0(\Omega)$ be the usual Lebesgue space and Sobolev spaces of functions on Ω. The inner product and the norm in $L^2(\Omega)$ are denoted by $(\cdot,\cdot)_{L^2}$ and $\|\cdot\|_{L^2}$, respectively.

4.1. Distributed optimal control of elliptic equations. We consider the following optimal control problem: Find the state $y \in H^1_0(\Omega)$ and the control $u \in L^2(\Omega)$ that minimizes the cost functional

$$J(y,u) = \frac{1}{2}\,\|y - y_d\|^2_{L^2} + \frac{\alpha}{2}\,\|u\|^2_{L^2}$$

subject to the state equation

$$\begin{aligned} -\Delta y &= u \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Gamma, \end{aligned}$$

or, more precisely, subject to the state equation in its weak form, given by

$$(\operatorname{grad} y, \operatorname{grad} z)_{L^2} = (u, z)_{L^2} \quad \text{for all } z \in H_0^1(\Omega).$$

Here $y_d \in L^2(\Omega)$ is the given (desired) state and $\alpha > 0$ is a regularization parameter.

Remark 7. *For ease of notation we will use the symbols* $(\cdot,\cdot)_{L^2}$ *and* $\|\cdot\|_{L^2}$ *not only for the case of scalar functions but also for the case of vector-valued functions and later on also for the case of matrix-valued functions: For* $\sigma, \tau \in L^2(\Omega, \mathbb{R}^{d\times d})$ *with components* σ_{ij}, τ_{ij} *the* L^2 *inner product is given by:*

$$(\sigma, \tau)_{L^2} = \sum_{i,j=1}^{d} (\sigma_{ij}, \tau_{ij})_{L^2}$$

with associated norm $\|\sigma\|_{L^2} = (\sigma, \sigma)_{L^2}^{1/2}$.

The Lagrangian functional associated to this optimization problem is given by:

$$\mathscr{L}(y, u, p) = J(y, u) + (\operatorname{grad} y, \operatorname{grad} p)_{L^2} - (u, p)_{L^2},$$

leading to following optimality system

$$\begin{aligned} (y, z)_{L^2} \qquad\qquad + (\operatorname{grad} z, \operatorname{grad} p)_{L^2} &= (y_d, z)_{L^2} \quad \text{for all } z \in H_0^1(\Omega), \\ \alpha\,(u, v)_{L^2} \; - (v, p)_{L^2} \qquad\quad &= 0 \qquad\quad \text{for all } v \in L^2(\Omega), \\ (\operatorname{grad} y, \operatorname{grad} q)_{L^2} - (u, q)_{L^2} \qquad &= 0 \qquad\quad \text{for all } q \in H_0^1(\Omega), \end{aligned}$$

which characterizes the solution $(y, u) \in H_0^1(\Omega) \times L^2(\Omega)$ of the optimal control problem with Lagrangian multiplier (or co-state) $p \in H_0^1(\Omega)$, see, e.g., [15], [30].

From the second equation we learn that $u = \alpha^{-1}\, p$, which allows to eliminate the control resulting in the reduced optimality system

$$\begin{aligned} (y, z)_{L^2} \qquad + (\operatorname{grad} z, \operatorname{grad} p)_{L^2} &= (y_d, z)_{L^2} \quad \text{for all } z \in H_0^1(\Omega), \\ (\operatorname{grad} y, \operatorname{grad} q)_{L^2} - \alpha^{-1}\,(p, q)_{L^2} \quad &= 0 \qquad\quad \text{for all } q \in H_0^1(\Omega). \end{aligned}$$

As an example of a discretization method we discuss the finite element method on a simplicial subdivision of Ω with continuous and piecewise linear functions for both the state and the co-state. This leads to the linear system

$$\begin{bmatrix} M & K \\ K & -\alpha^{-1}\, M \end{bmatrix} \begin{bmatrix} \underline{y} \\ \underline{p} \end{bmatrix} = \begin{bmatrix} \underline{f} \\ \underline{0} \end{bmatrix}$$

for the unknown vectors $\underline{y}$, $\underline{p}$ of coefficients of the approximate solutions relative to the nodal basis. Here M denotes the mass matrix representing the L^2 inner product and K denotes the stiffness matrix representing the elliptic operator of the state equation on the finite element space.

This linear system fits into the general framework of Sections 2 and 3 with

$$A = M, \quad B = K, \quad C = \alpha^{-1}\, M.$$

One particular pair of matrices $\mathcal{I}_V$ and $\mathcal{I}_Q$ satisfying (2.17) is given by (3.7) with $\theta = 1/2$:

$$\mathcal{I}_V = A + \left[A, B^T C^{-1} B\right]_{1/2} = M + \alpha^{1/2} \left[M, K M^{-1} K\right]_{1/2}$$

and

$$\mathcal{I}_Q = C + \left[C, B A^{-1} B^T\right]_{1/2} = \alpha^{-1} M + \alpha^{-1/2} \left[M, K M^{-1} K\right]_{1/2}.$$

Now

$$\left[M, K M^{-1} K\right]_{1/2} = M^{1/2} \big(\underbrace{M^{-1/2} K M^{-1} K M^{-1/2}}_{= \left(M^{-1/2} K M^{-1/2}\right)^2} \big)^{1/2} M^{1/2} = K.$$

Hence we obtain

$$\mathcal{I}_V = M + \alpha^{1/2} K \quad \text{and} \quad \mathcal{I}_Q = \alpha^{-1} M + \alpha^{-1/2} K. \tag{4.1}$$

From the analysis in Sections 2 and 3 it follows that

$$\mathcal{P} = \begin{bmatrix} M + \alpha^{1/2} K & 0 \\ 0 & \alpha^{-1} M + \alpha^{-1/2} K \end{bmatrix}$$

is a robust block diagonal preconditioner for

$$\mathcal{A} = \begin{bmatrix} M & K \\ K & -\alpha^{-1} M \end{bmatrix}. \tag{4.2}$$

The application of the preconditioner $\mathcal{P}$ requires an efficient evaluation of $\mathcal{P}^{-1} r$ for some given vector r. Up to a scaling factor both diagonal blocks of $\mathcal{P}$ are of the form $\gamma M + K$. This matrix is the stiffness matrix of the second-order elliptic differential operator of the state equation perturbed by a zero-order term. Multigrid or multilevel preconditioners which work robustly in γ are well-known, see, e.g., [21], [7]. So, in practice, the block matrices of the theoretical preconditioner $\mathcal{P}$ are replaced by such efficient preconditioners.

The same analysis can be carried out not only on the discrete level but also on the continuous level leading to the following inner products in $V = H_0^1(\Omega)$ for the state

$$(y, z)_V = (y, z)_{L^2} + \alpha^{1/2} (\operatorname{grad} y, \operatorname{grad} z)_{L^2}$$

and in $Q = H_0^1(\Omega)$ for the co-state

$$(p, q)_Q = \alpha^{-1} (p, q)_{L^2} + \alpha^{-1/2} (\operatorname{grad} p, \operatorname{grad} q)_{L^2}.$$

Well-posedness with robust estimates then follows for the associated norms

$$\|z\|_V = \left(\|z\|_{L^2}^2 + \alpha^{1/2} \|\operatorname{grad} z\|_{L^2}^2 \right)^{1/2}$$

and

$$\|q\|_Q = \left(\alpha^{-1} \|q\|_{L^2}^2 + \alpha^{-1/2} \|\operatorname{grad} q\|_{L^2}^2 \right)^{1/2},$$

which differ from the standard H^1-norms.

The particular form of the elliptic operator for y in the state equation does not play an essential role as long as the associated bilinear form of the weak formulation is symmetric, bounded and coercive. See [28] for a related elliptic optimal control problem, where robust block-preconditioners have been constructed using the non-standard inner products and norms from above, and [27], where robust all-at-once multigrid methods were analyzed based on these non-standard norms. The form of the control and the cost functional, however, was essential for constructing the inner products. The results cannot be easily extended to other cases.

Remark 8. *If instead of* (3.7) *one uses the solution* (3.2)*, then one obtains*

$$\mathcal{I}_V = M \quad \textit{and} \quad \mathcal{I}_Q = \alpha^{-1} M + K M^{-1} K,$$

also leading to a robust block diagonal preconditioner. While $\mathcal{I}_V$ is here much easier than before, the matrix $\mathcal{I}_Q$ requires more work. It can be interpreted as matrix representation of a fourth-order differential operator. For preconditioners developed along this lines see, e.g., [23]. These preconditioners were shown to be robust with respect to h (but not with respect to the regularization parameter).

Remark 9. *More generally, for any 2-by-2 block matrix $\mathcal{A}$ of the form* (4.2)*, the matrix $\mathcal{P}$ of the form* (4.1) *is a robust preconditioner as long as M and K are symmetric and positive semi-definite with* $\ker M \cap \ker K = \{0\}$*. For generalized eigenvalue problem*

$$\mathcal{A} \begin{bmatrix} u \\ p \end{bmatrix} = \lambda \, \mathcal{P} \begin{bmatrix} u \\ p \end{bmatrix} \tag{4.3}$$

it can be shown that

$$|\lambda_{min}| \geq \frac{1}{\sqrt{2}} \quad \textit{and} \quad |\lambda_{min}| \leq 1,$$

resulting in a condition number estimate $\kappa(\mathcal{A}) \leq \sqrt{2}$. This can be seen by considering the generalized eigenvalue problem

$$Mz = \mu \, (M + \alpha^{1/2} \, K) z.$$

Since M is symmetric and positive semi-definite and $M + \alpha^{1/2} \, K$ is symmetric and positive definite, there is a basis $\{e_1, e_2, \dots, \}$ of eigenvectors e_i with corresponding eigenvalues $\mu_i \in [0, 1]$, which is orthonormal with respect to the $M + \alpha^{1/2} \, K$ inner product. By expanding u and p in terms of this basis with coefficients $\hat{u}_i$ and $\hat{p}_i$ the generalized eigenvalue problem (4.3) *reduces to*

$$\begin{bmatrix} \mu_i & \alpha^{-1/2}(1-\mu_i) \\ \alpha^{-1/2}(1-\mu_i) & -\alpha^{-1} \, \mu_i \end{bmatrix} \begin{bmatrix} \hat{u}_i \\ \hat{p}_i \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 \\ 0 & \alpha^{-1} \end{bmatrix} \begin{bmatrix} \hat{u}_i \\ \hat{p}_i \end{bmatrix}.$$

Since $(\hat{u}_i, \hat{p}_i)^T \neq 0$ for at least one index i, it follows for that index that

$$\det \left(\begin{bmatrix} \mu_i & \alpha^{-1/2}(1-\mu_i) \\ \alpha^{-1/2}(1-\mu_i) & -\alpha^{-1} \, \mu_i \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & \alpha^{-1} \end{bmatrix} \right) = 0,$$

i.e.:

$$\lambda^2 - \mu_i^2 - (1-\mu_i)^2 = 0.$$

So, $|\lambda| = \sqrt{\mu_i^2 + (1-\mu_i)^2}$ for some $\mu_i \in [0, 1]$, which immediately implies the estimates for the moduli of λ_{min} and λ_{max}.

4.2. Distributed optimal control of the Stokes equations. The second example of an optimal control problem is the so-called velocity tracking problem for Stokes flow with distributed control: Find the velocity $u \in H_0^1(\Omega)^d$, the pressure $p \in L_0^2(\Omega) = \{q \in L^2(\Omega) \colon \int_\Omega q \, dx = 0\}$ and the control $f \in L^2(\Omega)^d$ that minimizes the cost functional

$$J(u,f) = \frac{1}{2}\|u - u_d\|_{L^2}^2 + \frac{\alpha}{2}\,\|f\|_{L^2}^2$$

subject to the state equations

$$\begin{aligned} -\Delta u + \operatorname{grad} p &= f \quad \text{in } \Omega, \\ \operatorname{div} u &= 0 \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

or, more precisely again, subject to the state equations in its weak form:

$$\begin{aligned} (\operatorname{grad} u, \operatorname{grad} v)_{L^2} - (\operatorname{div} v, p)_{L^2} &= (f,v)_{L^2} \quad \text{for all } v \in H_0^1(\Omega)^d, \\ -(\operatorname{div} u, q)_{L^2} &= 0 \qquad\quad \text{for all } q \in L_0^2(\Omega). \end{aligned}$$

Here $u_d \in L^2(\Omega)^d$ is the desired velocity and $\alpha > 0$ is a regularization parameter.

The Lagrangian functional associated to this optimization problem is given by:

$$\begin{aligned} &\mathscr{L}(u,p,f,\hat{u},\hat{p}) \\ &\quad = J(u,f) + (\operatorname{grad} u, \operatorname{grad} \hat{u})_{L^2} - (\operatorname{div} \hat{u}, p)_{L^2} - (\operatorname{div} u, \hat{p})_{L^2} - (f,\hat{u})_{L^2} \end{aligned}$$

leading to following optimality system

$$\begin{aligned} (u,v)_{L^2} \qquad\qquad\qquad\quad &+ (\nabla v, \nabla \hat{u})_{L^2} - (\operatorname{div} v, \hat{p})_{L^2} &&= (u_d, v)_{L^2}, \\ &- (\operatorname{div} \hat{u}, q)_{L^2} &&= 0, \\ \alpha\,(f,\phi)_{L^2} \; &- (\phi, \hat{u})_{L^2} &&= 0, \\ (\nabla u, \nabla \hat{v})_{L^2} \quad - (\operatorname{div} \hat{v}, p)_{L^2} &- (f,\hat{v})_{L^2} &&= 0, \\ -(\operatorname{div} u, \hat{q})_{L^2} &&&= 0 \end{aligned}$$

for all test functions $v, \hat{v} \in H_0^1(\Omega)^d$, $q, \hat{q} \in L_0^2(\Omega)$, and $\phi \in L^2(\Omega)^d$. This system characterizes the solution $(y,p,f) \in H_0^1(\Omega)^d \times L_0^2(\Omega) \times L^2(\Omega)^d$ of the optimal control problem with Lagrangian multipliers $(\hat{u},\hat{p}) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$.

As in the elliptic case the control f can be eliminated (using the third equation) resulting in the reduced optimality system:

$$\begin{aligned} (u,v)_{L^2} \qquad\qquad\qquad\quad &+ (\nabla v, \nabla \hat{u})_{L^2} \; - (\operatorname{div} v, \hat{p})_{L^2} &&= (u_d, v)_{L^2}, \\ &- (\operatorname{div} \hat{u}, q)_{L^2} &&= 0, \\ (\nabla u, \nabla \hat{v})_{L^2} \quad - (\operatorname{div} \hat{v}, p)_{L^2} &- \alpha^{-1}\,(\hat{u}, \hat{v})_{L^2} &&= 0, \\ -(\operatorname{div} u, \hat{q})_{L^2} &&&= 0 \end{aligned}$$

for all test functions $v, \hat{v} \in H_0^1(\Omega)^d$ and $q, \hat{q} \in L_0^2(\Omega)$.

As an example of a discretization method we discuss the Taylor-Hood element on a simplicial subdivision of Ω consisting of continuous and piecewise quadratic functions

for u and $\hat{u}$ and continuous and piecewise linear functions for p and $\hat{p}$. This leads to the linear system

$$\begin{bmatrix} M & & K & -D^T \\ & 0 & -D & \\ K & -D^T & -\alpha^{-1} M & \\ -D & & & 0 \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{p} \\ \underline{\hat{u}} \\ \underline{\hat{p}} \end{bmatrix} = \begin{bmatrix} M \underline{u}_d \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Here M denotes the mass matrix representing the standard inner product in $L^2(\Omega)^d$ and K denotes the stiffness matrix representing the vector Laplace operator on the finite element space. Additionally, D denotes the matrix representation of the divergence operator on the involved finite element spaces. Observe that D is of full rank, since the Taylor-Hood element satisfies a discrete inf-sup condition (under mild conditions on the underlying mesh).

This linear system fits into the general framework of Section 2 with

$$A = \begin{bmatrix} M & \\ & 0 \end{bmatrix}, \quad B = \begin{bmatrix} K & -D^T \\ -D & \end{bmatrix}, \quad C = \begin{bmatrix} \alpha^{-1} M & \\ & 0 \end{bmatrix} = \alpha^{-1} A,$$

but it does not fit into any of the special cases discussed in Section 3. So we have to go back to the original conditions (2.17). Because of

$$C = \alpha^{-1} A \quad \text{and} \quad B^T = B,$$

it is natural to make the ansatz

$$\mathcal{I}_Q = \alpha^{-1} \mathcal{I}_V. \tag{4.4}$$

Then both conditions in (2.17) reduce to one condition for $\mathcal{I}_V$, namely:

$$A + \alpha B \mathcal{I}_V^{-1} B \sim \mathcal{I}_V. \tag{4.5}$$

We will now try to find such a matrix $\mathcal{I}_V$ which is of a block diagonal form:

$$\mathcal{I}_V = \begin{bmatrix} \mathcal{I}_u & \\ & \mathcal{I}_p \end{bmatrix}.$$

Then condition (4.5) reads

$$\begin{bmatrix} M + \alpha K \mathcal{I}_u^{-1} K + \alpha D^T \mathcal{I}_p^{-1} D & -\alpha K \mathcal{I}_u^{-1} D^T \\ -\alpha D \mathcal{I}_u^{-1} K & \alpha D \mathcal{I}_u^{-1} D^T \end{bmatrix} \sim \begin{bmatrix} \mathcal{I}_u & \\ & \mathcal{I}_p \end{bmatrix} \tag{4.6}$$

after expanding the matrix expression on the left-hand side. For discussing this equivalence relation the following lemma is very helpful:

LEMMA 4.1. *Let $\mathcal{M}$ be a symmetric and positive definite 2-by-2 block matrix and $\mathcal{D}$ a 2-by-2 block diagonal matrix with symmetric and positive definite diagonal blocks. Then*

$$\mathcal{M} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \sim \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} = \mathcal{D}$$

if and only if

$$M_{11} \sim D_{11}, \quad \text{and} \quad M_{22} \sim D_{22}, \quad \text{and} \quad M_{11} \lesssim M_{11} - M_{12} M_{22}^{-1} M_{21} \tag{4.7}$$

A proof of this result can be found in [1], where instead of the third condition in (4.7) a strengthened Cauchy-Schwarz inequality was used: There is a constant $\gamma < 1$ such that

$$y^T M_{21} x \le \gamma \sqrt{(x^T M_{11} x)} \sqrt{(y^T M_{22} y)} \quad \text{for all } x, y.$$

This condition is equivalent to

$$M_{11} \le \eta \, (M_{11} - M_{12} M_{22}^{-1} M_{21}) \quad \text{with } \eta = \frac{1}{1-\gamma^2},$$

see, e.g., [31].

If Lemma 4.1 is applied to (4.6), the first two conditions in (4.7) read

$$M + \alpha \, K \mathcal{I}_u^{-1} K + \alpha \, D^T \mathcal{I}_p^{-1} D^T \sim \mathcal{I}_u \quad \text{and} \quad \alpha \, D \mathcal{I}_u^{-1} D^T \sim \mathcal{I}_p. \tag{4.8}$$

The second condition in (4.8) is trivially satisfied for the choice

$$\mathcal{I}_p = \alpha \, D \mathcal{I}_u^{-1} D^T. \tag{4.9}$$

Then the first condition in (4.8) simplifies to

$$M + \alpha \, K \mathcal{I}_u^{-1} K \sim \mathcal{I}_u, \tag{4.10}$$

since

$$0 \le \alpha \, D^T \mathcal{I}_p^{-1} D = D^T \left(D \mathcal{I}_u^{-1} D^T \right)^{-1} D \le \mathcal{I}_u.$$

It is easy to check that

$$\mathcal{I}_u = M + \alpha^{1/2} \, K \tag{4.11}$$

satisfies Condition (4.10). It remains to verify the third condition in (4.7), which reads here:

$$\mathcal{I}_u \lesssim R_u \tag{4.12}$$

with

$$R_u = M + \alpha \, K \mathcal{I}_u^{-1} K + \alpha \, D^T \mathcal{I}_p^{-1} D - \alpha \, K \mathcal{I}_u^{-1} D^T \left(D \mathcal{I}_u^{-1} D^T \right)^{-1} D \mathcal{I}_u^{-1} K.$$

THEOREM 4.2. *Condition* (4.12) *is satisfied for the matrix* $\mathcal{I}_u$*, given by* (4.11).

Proof. We have

$$R_u = M + \alpha \, D^T \mathcal{I}_p^{-1} D + \alpha \left[K \mathcal{I}_u^{-1} K - K \mathcal{I}_u^{-1} D^T \left(D \mathcal{I}_u^{-1} D^T \right)^{-1} D \mathcal{I}_u^{-1} K \right].$$

Since

$$D^T \left(D \mathcal{I}_u^{-1} D^T \right)^{-1} D \le \mathcal{I}_u, \tag{4.13}$$

it follows that

$$K \mathcal{I}_u^{-1} K - K \mathcal{I}_u^{-1} D^T \left(D \mathcal{I}_u^{-1} D^T \right)^{-1} D \mathcal{I}_u^{-1} K \ge K \mathcal{I}_u^{-1} K - K \mathcal{I}_u^{-1} \mathcal{I}_u \mathcal{I}_u^{-1} K = 0.$$

Therefore, we obtain as a first estimate

$$R_u \geq M + \alpha\, D^T \mathcal{I}_p^{-1} D. \tag{4.14}$$

For deriving a second estimate we start with the following observation:

$$\begin{aligned} &K\mathcal{I}_u^{-1} D^T \left(D\mathcal{I}_u^{-1} D^T\right)^{-1} D\mathcal{I}_u^{-1} K = \alpha\, K\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} K \\ &\quad = (\alpha^{1/2}\, K)\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} (\alpha^{1/2}\, K) = (\mathcal{I}_u - M)\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} (\mathcal{I}_u - M) \\ &\quad = D^T \mathcal{I}_p^{-1} D - M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D - D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M + M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M. \end{aligned}$$

Therefore,

$$\begin{aligned} R_u = M + \alpha\, K\mathcal{I}_u^{-1} K - \alpha\, M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M \\ + \alpha\, \left(M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D + D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M\right) \end{aligned}$$

From (4.13) it follows that

$$\alpha\, \mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} = \mathcal{I}_u^{-1} D^T \left(D\mathcal{I}_u^{-1} D^T\right)^{-1} D\mathcal{I}_u \leq \mathcal{I}_u^{-1} \mathcal{I}_u \mathcal{I}_u^{-1} = \mathcal{I}_u^{-1}.$$

Hence

$$\begin{aligned} M + \alpha\, K\mathcal{I}_u^{-1} K - \alpha\, M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M \\ \geq M + \alpha\, K\mathcal{I}_u^{-1} K - M\mathcal{I}_u^{-1} M = \sqrt{\alpha}\, K. \end{aligned} \tag{4.15}$$

Furthermore, we have for all vectors v:

$$\begin{aligned} &\alpha\, v^T \left(M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D + D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M\right) v \\ &\quad = 2v^T M\mathcal{I}_u^{-1} \left(\alpha\, D^T \mathcal{I}_p^{-1} D\right) v \\ &\quad = 2v^T M\mathcal{I}_u^{-1/2} \mathcal{I}_u^{-1/2} \left(\alpha\, D^T \mathcal{I}_p^{-1} D\right)^{1/2} \left(\alpha\, D^T \mathcal{I}_Q^{-1} D\right)^{1/2} v \\ &\quad \geq -2 \left\|\mathcal{I}_u^{-1/2} M v\right\| \left\|\mathcal{I}_u^{-1/2} \left(\alpha\, D^T \mathcal{I}_p^{-1} D\right)^{1/2}\right\| \left\|\left(\alpha\, D^T \mathcal{I}_p^{-1} D\right)^{1/2} v\right\| \\ &\quad \geq -2 \left\|\mathcal{I}_u^{-1/2} M v\right\| \left\|\left(\alpha\, D^T \mathcal{I}_p^{-1} D\right)^{1/2} v\right\| \geq -\left\|\mathcal{I}_u^{-1/2} M v\right\|^2 - \left\|\left(\alpha\, D^T \mathcal{I}_p^{-1} D\right)^{1/2} v\right\|^2 \\ &\quad = -v^T M\mathcal{I}_u^{-1} M v - v^T \left(\alpha\, D^T \mathcal{I}_p^{-1} D\right) v \geq -v^T M v - v^T \left(\alpha\, D^T \mathcal{I}_p^{-1} D\right) v. \end{aligned}$$

This shows that

$$\alpha\, \left(M\mathcal{I}_u^{-1} D^T \mathcal{I}_p^{-1} D + D^T \mathcal{I}_p^{-1} D\mathcal{I}_u^{-1} M\right) \geq -\left(M + \alpha\, D^T \mathcal{I}_p^{-1} D\right). \tag{4.16}$$

From (4.15) and (4.16) we obtain the second estimate for R_u:

$$R_u \geq \sqrt{\alpha}\, K - \left(M + \alpha\, D^T \mathcal{I}_p^{-1} D\right).$$

Finally, by combining this estimate with (4.14) it follows that

$$\begin{aligned} 3R_u = 2R_u + R_u &\geq 2\left(M + \alpha\, D^T \mathcal{I}_p^{-1} D\right) + \sqrt{\alpha}\, K - \left(M + \alpha\, D^T \mathcal{I}_p^{-1} D\right) \\ &= M + \alpha\, D^T \mathcal{I}_p^{-1} D + \sqrt{\alpha}\, K \geq M + \sqrt{\alpha}\, K = \mathcal{I}_u, \end{aligned}$$

which completes the proof. □

Therefore, summarizing (4.4), (4.11), and (4.9), we have shown that

$$\mathcal{P} = \begin{bmatrix} \mathcal{I}_u & & & \\ & \alpha\, D\mathcal{I}_u^{-1}D^T & & \\ & & \alpha^{-1}\,\mathcal{I}_u & \\ & & & D\mathcal{I}_u^{-1}D^T \end{bmatrix} \quad \text{with} \quad \mathcal{I}_u = M + \alpha^{1/2}\,K \tag{4.17}$$

is a robust block-diagonal preconditioner of

$$\mathcal{A} = \begin{bmatrix} M & & K & -D^T \\ & 0 & D & \\ K & -D^T & -\alpha^{-1}\,M & \\ -D & & & 0 \end{bmatrix}.$$

Up to a scaling factor the diagonal blocks of $\mathcal{P}$ are of the form $\gamma\,M + K$ or $D(\gamma\,M + K)^{-1}D^T$. As already mentioned for the elliptic optimal control problem multigrid or multilevel preconditioners for $\gamma\,M + K$ are available which work robustly in γ. This is also the case for the matrix $D(\gamma\,M + K)^{-1}D^T$, which is the Schur complement of a discretized generalized Stokes problem, see [10],[6], [20], [16], [17]. So, in practice, the block matrices of the theoretical preconditioner $\mathcal{P}$ are replaced by such efficient preconditioners.

Remark 10. *As for the elliptic problem the same analysis can also be done on the continuous level leading the corresponding non-standard norms in* $H_0^1(\Omega) \times L_0^2(\Omega)$ *for* u *and* p *as well as for the Lagrangian multipliers* $\hat{u}$ *and* $\hat{p}$.

Remark 11. *If the objective functional of the optimal control problem for the Stokes equations contains an additional* L^2*-term for the pressure, then the (1,1) block of the system matrix becomes non-singular, and, as an alternative, the block preconditioner based on* (3.2) *could be used as a theoretical preconditioner. As in the elliptic case the Schur complement* S *can be interpreted as discretization matrix of a system of partial differential equations including fourth-order differential operators. For preconditioners developed along this lines see, e.g., [24]. These preconditioners were shown to be robust with respect to* h *(but not with respect to the regularization parameter).*

4.3. A numerical example. Numerical experiments for elliptic optimal control problems can be found in [28] and [27]. Here we present some numerical experiments for the velocity tracking problem for Stokes flow with distributed control on the unit square domain $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$. Following Example 1 in [12] we chose the desired (target) velocity $u_d(x,y) = (U(x,y), V(x,y))^T$, given by

$$U(x,y) = 10\frac{\partial}{\partial y}(\varphi(x)\varphi(y)) \quad \text{and} \quad V(x,y) = -10\frac{\partial}{\partial x}(\varphi(x)\varphi(y))$$

with

$$\varphi(z) = \big(1 - \cos(0.8\pi z)\big)(1-z)^2.$$

The velocity u_d is divergence free. Note that, contrary to the problem considered here, in [12] the velocity tracking problem was discussed for a time-dependent Navier-Stokes flow with distributed control.

The problem was discretized by the Taylor-Hood pair of finite element spaces consisting of continuous piecewise quadratic polynomials for the velocity and continuous piecewise linear polynomials for the pressure on a triangulation of Ω. The initial mesh

contains four triangles obtained by connecting the two diagonals. The final mesh was constructed by applying k uniform refinement steps to the initial mesh, leading to a mesh size $h = 2^{-k}$. Figure 4.1 shows the solution for the velocity u (left part) and the control f (right part) computed at the finest mesh ($k = 7$) for $\alpha = 10^{-6}$. The computed velocity u is optically indistinguishable from the target velocity u_d. The length of the largest arrow in Figure 4.1, left part, representing the velocity u corresponds to a value of 1. The values for the pressure p range from -5.7 to 5.7. The length of the largest arrow in Figure 4.2 representing the control f corresponds to a value of 57.7.

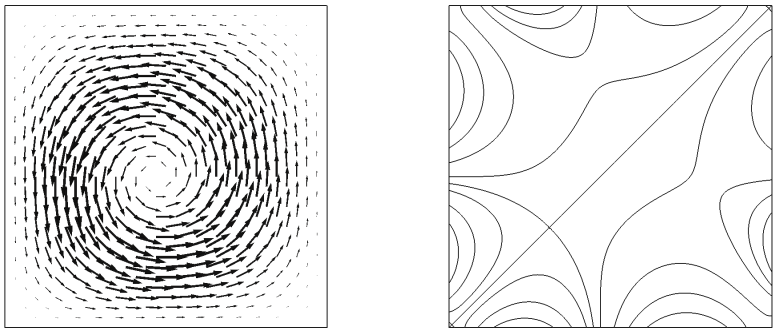

FIG. 4.1. *The velocity u (left part) and the pressure p (right part) at grid level $k = 7$ and $\alpha = 10^{-6}$.*

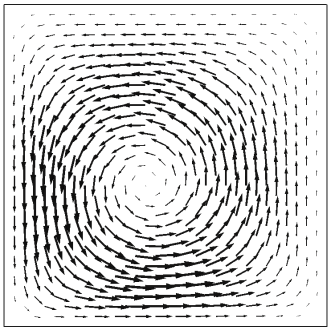

FIG. 4.2. *The control f at grid level $k = 7$ and $\alpha = 10^{-6}$.*

For constructing a practically realizable preconditioner we proceeded as follows: First we replaced the matrix $D(M + \alpha^{1/2}K)^{-1}D^T$ in (4.17) by $\left(\alpha^{1/2}\, M_p^{-1} + K_p^{-1}\right)^{-1}$ as proposed in [10], where M_p and K_p denote the mass matrix and the stiffness matrix for the pressure element, respectively. Then the application of the preconditioner would require the multiplication of vectors from the left by the inverses of the matrices $M + \alpha^{1/2}K$, M_p and K_p. These actions were replaced by one step of a V-cycle iteration for $M+\alpha^{1/2}K$ and K_p and by one step of a symmetric Gauss-Seidel iteration for M_p. The V-cycle was done with one step of a symmetric Gauss-Seidel iteration for the pre-smoothing process and for the post-smoothing process. The resulting realizable preconditioner is spectrally equivalent to the theoretical preconditioner (4.17) according to the analysis in [21], [6], [20], [16], [17].

Table 4.1 shows the condition number of the preconditioned system matrix for

various values of h and α, where k denotes the number of refinements (corresponding to the mesh size $h = 2^{-k}$), N is the total number of degrees of freedom of the discretized reduced optimality system.

TABLE 4.1
Condition numbers

k	N	α						
		10^{-12}	10^{-8}	10^{-4}	1	10^{4}	10^{8}	10^{12}
4	9 030	3.43	4.01	7.39	9.52	9.75	9.76	9.76
5	36 486	3.42	4.83	8.20	9.98	10.18	10.18	10.18
6	146 694	3.45	6.06	8.88	10.33	10.50	10.50	10.50
7	588 294	3.80	7.12	9.45	10.65	10.74	10.75	10.75

In Table 4.2 the number of MINRES iterations is shown, required for reducing the Euclidean norm of the initial (preconditioned) residual by a factor $\varepsilon = 10^{-8}$. The initial guess was 0.

TABLE 4.2
Number of MINRES iterations

k	N	α						
		10^{-12}	10^{-8}	10^{-4}	1	10^{4}	10^{8}	10^{12}
4	9 030	32	45	89	106	108	108	108
5	36 486	34	47	95	112	114	114	114
6	146 694	34	51	101	118	120	120	120
7	588 294	34	55	107	124	126	126	126

5. Concluding Remarks. The equivalence relations in (2.17) are necessary and sufficient for obtaining robust estimates, or, in the context of iterative methods, for obtaining robust block diagonal preconditioners. So, they can be used to check whether a particular preconditioner is robust or not. How to resolve these conditions, i.e., how to find $\mathcal{I}_V$ and $\mathcal{I}_Q$ satisfying (2.17), is a much harder problem. In Section 3 we have demonstrated how to resolve (2.17) in special cases. These cases cover the distributed optimal control problem for elliptic equations, but not the distributed optimal control problem for the Stokes equations. Nevertheless, also for the Stokes control problem robust preconditioners could be constructed by a specialized analysis of (2.17). The numerical experiments for the velocity tracking problem for a Stokes flow show condition numbers of moderate size.

Acknowledgment. A first preliminary version of the contents of this paper was presented in a talk given by the author at the BIRS Workshop Advances and Perspectives on Numerical Methods for Saddle Point Problems, April 2009, Banff, Canada. The work of the author on this topic has been strongly influenced by a series of three lectures given by R. Winther at the London Society Durham Symposium on Computational Linear Algebra for Partial Differential Equations, July 2008, see [18] for the corresponding survey article.

The author is very grateful to Markus Kollmann for providing him with the numerical results.

REFERENCES

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

[2] I. Babuška, *Error-bounds for finite element method*, Numer. Math., 16 (1971), pp. 322–333.

[3] I. Babuška and A. K. Aziz, *Survey lectures on the mathematical foundations of the finite element method*, in The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, A. K. Aziz, ed., New York-London: Academic Press, 1972, pp. 3–343.

[4] M. Benzi, G. H. Golub, and J. Liesen, *Numerical Solution of Saddle Point Problems*, Acta Numerica, 14 (2005), pp. 1–137.

[5] J. Bergh and J. Löfström, *Interpolation spaces. An introduction*, Grundlehren der mathematischen Wissenschaften. 223. Berlin-Heidelberg-New York: Springer-Verlag, 1976.

[6] J.H. Bramble and J.E. Pasciak, *Iterative techniques for time dependent Stokes problems*, Comput. Math. Appl., 33 (1997), pp. 13–30.

[7] J. H. Bramble, J. E. Pasciak, and P. S. Vassilevski, *Computational scales of Sobolev norms with application to preconditioning*, Math. Comput., 69 (2000), pp. 463–480.

[8] F. Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, R.A.I.R.O. Anal.Numer., 2 (1974), pp. 129–151.

[9] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, 1991.

[10] J. Cahouet and J.-P. Chabard, *Some fast 3D finite element solvers for the generalized Stokes problem*, Int. J. Numer. Methods Fluids, 8 (1988), pp. 865–895.

[11] G. H. Golub and C. F. Van Loan, *Matrix computations. 3rd ed.*, Texts and Readings in Mathematics 43. New Delhi: Hindustan Book Agency, 2007.

[12] M.D. Gunzburger and S. Manservisi, *Analysis and approximation of the velocity tracking problem for Navier-Stokes flows with distributed control*, SIAM J. Numer. Anal., 37 (2000), pp. 1481–1512.

[13] N. J. Higham, *Functions of matrices. Theory and computation*, Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2008.

[14] Yu.A. Kuznetsov, *Efficient iterative solvers for elliptic finite element problems on nonmatching grids*, Russ. J. Numer. Anal. Math. Model., 10 (1995), p. 187211.

[15] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Berlin-Heidelberg-New York: Springer-Verlag, 1971.

[16] K.-A. Mardal and R. Winther, *Uniform preconditioners for the time dependent Stokes problem*, Numer. Math., 98 (2004), pp. 305–327.

[17] ———, *Uniform preconditioners for the time dependent Stokes problem*, Numer. Math., 103 (2006), pp. 171–172.

[18] ———, *Preconditioning discretizations of systems of partial differential equations*, Numer. Linear Algebra Appl., 18 (2011), pp. 1–40.

[19] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972.

[20] M. A. Olshanskii, J. Peters, and A. Reusken, *Uniform preconditioners for a parameter dependent saddle point problem with application to generalized Stokes interface equations*, Numer. Math., 105 (2006), pp. 159–191.

[21] M. A. Olshanskii and A. Reusken, *On the convergence of a multigrid method for linear reaction-diffusion problems*, Computing, 65 (2000), pp. 193–202.

[22] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[23] T. Rees, H. S. Dollar, and A. J. Wathen, *Optimal solvers for PDE-constrained optimization*, SIAM J. Sci. Comput., 32 (2010), pp. 271 – 298.

[24] T. Rees and A. J. Wathen, *Preconditioning iterative methods for the optimal control of the Stokes equation*, Technical Report NA-10/04, University of Oxford, June 2010.

[25] R. T. Rockafellar, *Convex analysis*, Princeton Landmarks in Mathematics. Princeton, NJ: Princeton University Press, 1997.

[26] T. Rusten and R. Winther, *A preconditioned iterative method for saddle-point problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887 – 904.

[27] J. Schöberl, R. Simon, and W. Zulehner, *A robust multigrid method for elliptic optimal control problems*, NuMa-Report 2010-01, Institute of Computational Mathematics, Johannes Kepler University Linz, Austria, 2010.

[28] 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.

[29] D. Silvester and A. Wathen, *Fast iterative solution of stabilised Stokes systems. II: Using general block preconditioners*, SIAM J. Numer. Anal., 31 (1994), pp. 1352–1367.
[30] F. Tröltzsch, *Optimal control of partial differential equations. Theory, methods and applications*, Graduate Studies in Mathematics 112. Providence, RI: American Mathematical Society (AMS), 2010.
[31] P. S. Vassilevski, *Hybrid V-cycle algebraic multilevel preconditioners*, Math. Comput., 58 (1992), pp. 489–512.
[32] A. Wathen and D. Silvester, *Fast iterative solutions of stabilized Stokes systems. Part I: Using simple diagonal preconditioners*, SIAM J. Numer. Anal., 30 (1993), pp. 630 – 649.

Latest Reports in this series

2009

[..]

2009-07	Ulrich Langer, Huidong Yang and Walter Zulehner *A Grid-enabled Solver for the Fluid-Structure Interaction (FSI) Problem*	August 2009
2009-08	Stefan Takacs and Walter Zulehner *Multigrid Methods for Elliptic Optimal Control Problems with Neumann Boundary Control*	October 2009
2009-09	Dylan M. Copeland and Ulrich Langer *Domain Decomposition Solvers for Nonlinear Multiharmonic Finite Element Equations*	November 2009
2009-10	Huidong Yang and Walter Zulehner *A Newton Based Fluid-structure Interaction (FSI) Solver with Algebraic Multigrid Methods (AMG) on Hybrid Meshes*	November 2009
2009-11	Peter Gruber, Dorothee Knees, Sergiy Nesenenko and Marita Thomas *Analytical and Numerical Aspects of Time-dependent Models with Internal Variables*	November 2009
2009-12	Clemens Pechstein and Robert Scheichl *Weighted Poincaré Inequalities and Applications in Domain Decomposition*	November 2009
2009-13	Dylan Copeland, Michael Kolmbauer and Ulrich Langer *Domain Decomposition Solvers for Frequency-Domain Finite Element Equations*	December 2009
2009-14	Clemens Pechstein *Shape-Explicit Constants for Some Boundary Integral Operators*	December 2009
2009-15	Peter G. Gruber, Johanna Kienesberger, Ulrich Langer, Joachim Schöberl and Jan Valdman *Fast Solvers and A Posteriori Error Estimates in Elastoplasticity*	December 2009

2010

2010-01	Joachim Schöberl, René Simon and Walter Zulehner *A Robust Multigrid Method for Elliptic Optimal Control Problems*	Januray 2010
2010-02	Peter G. Gruber *Adaptive Strategies for High Order FEM in Elastoplasticity*	March 2010
2010-03	Sven Beuchler, Clemens Pechstein and Daniel Wachsmuth *Boundary Concentrated Finite Elements for Optimal Boundary Control Problems of Elliptic PDEs*	June 2010
2010-04	Clemens Hofreither, Ulrich Langer and Clemens Pechstein *Analysis of a Non-standard Finite Element Method Based on Boundary Integral Operators*	June 2010
2010-05	Helmut Gfrerer *First-Order Characterizations of Metric Subregularity and Calmness of Constraint Set Mappings*	July 2010
2010-06	Helmut Gfrerer *Second Order Conditions for Metric Subregularity of Smooth Constraint Systems*	September 2010
2010-07	Walter Zulehner *Non-standard Norms and Robust Estimates for Saddle Point Problems*	November 2010

From 1998 to 2008 reports were published by SFB013. Please see
`http://www.sfb013.uni-linz.ac.at/index.php?id=reports`
From 2004 on reports were also published by RICAM. Please see
`http://www.ricam.oeaw.ac.at/publications/list/`
For a complete list of NuMa reports see
`http://www.numa.uni-linz.ac.at/Publications/List/`