Preconditioning strategies related to multiphysics problems

Trygve Bærland

Thesis submitted for the degree of Ph.D.

Department of Mathematics
University of Oslo

December, 2018

Preface

This thesis is submitted to the Department of Mathematics, University of Oslo in partial fulfillment of the requirements for the degree of Ph.D. The work that lead to this thesis was carried out under the supervision of professors Ragnar Winther and Kent-Andre Mardal.

The thesis is structured as an introduction followed by a collection of papers. A lot of research starts out from relatively simple, tangible motivations, but as one's work proceeds one strays further into abstractions, simplifications, and/or generalizations. That is not to say that the work doesn't have value, of course, but that it can be easy to lose sight of the proverbial forest for the trees. The papers that make up this thesis are no exceptions to this. Therefore, the introduction chapter of this thesis will attempt to draw a line from basic motivations to the problems that lead to the papers, as well as to provide a larger scientific context in which the work sits. I have opted to put an emphasis on motivating the ideas, rather than mathematical rigour. That being said, I hope I have managed to strike some sort of balance between the two. With hopefully some of the readers safely guided through the motivation, we end the chapter on brief summaries of each paper, and try to see how these trees look in the forest.

Acknowledgements

" Dreaming about being an actress is more exciting than being one. "

— Marilyn Monroe

There are a number of people that have contributed in making these past years, at work or otherwise, not only endurable but also enjoyable, for which I am tremendously grateful. Too many to all thank by name here, but there are some I would be remiss if I did not show my deep appreciation for. So I will allow myself these few paragraphs that will teeter towards the sentimental.

I would like to begin by thanking my supervisors, professors Ragnar Winther and Kent-Andre Mardal. Our discussions have been enlightening, fun, engaging, frustrating, and embarrasing, but never failing to be inspiring.

Deep gratitude goes to my family as well. Your love was never in doubt, but still feels miraculous. I also want to express my appreciation of my parents, Ingunn and Jan, for their absolute support, patience and safety; and to my brother Tarjei and sister Hedda for seeing me at my lowest and still seeing me.

To all my friends outside of work, I am thankful that you only rarely asked "How's your thesis going?", but also: To Haakon, Martin, and Halvor, for tremendous humour, initiative and shared experiences; Timmy, for his one-handed backhand and always drumming in his own time; Tobias, for talking coherently about incoherent philosophy, and in general not being afraid to get technical.

Thanks is also due to my colleagues and friends at the university, in particular: Espen and Adrian, for imbuing their sarcasm with just enough kindness to make it funny; Fabian, for showing me that it's never too late to take up new hobbies and enjoy it, however poorly done; and Miro, for many fun and stimulating conversations, as well as not looking down on me when I come with simple IT problems.

Lastly, I will always be grateful to Martha Johanna for her unwavering kindness and for distractions that made up a life. To the rest of her family, I thank them for their seemingly endless capacity for love, and letting me share in it.

Oslo, December 2018
Trygve Bærland

Introduction

Broadly speaking, this thesis is about preconditioning. That is, the design and analysis of algorithms for solving linear algebraic systems that result from discretizations of partial differential equations (henceforth PDEs). More specifically, the thesis is about the design of preconditioners suitable for problems concerning multiple interacting physical phenomena.

To describe the role preconditioners play in research and industry, it seems appropriate to start from the perspective of PDEs. The importance PDEs enjoy are oft-repeated and rarely understated. PDEs are powerful in their variety. For instance, they play a vital role in our understanding of such disparate phenomena as e.g. gravity, electromagnetics, fluid mechanics and thermodynamics. Moreover, PDEs are flexible in that one PDE, or a system of PDEs, can be used to described different things. As a particularly relevant example, Biot's consolidation model, first introduced in its general form in [17], attempts to model the fluid flow in a porous media. As such, it has found applications in for example reservoir engineering ([68, 96]), environmental engineering ([82]), and biomechanics ([75, 80]). In a sense, the study of PDEs is a way to view the world. As a consequence the mathematical analysis of PDEs is a wide topic of research, that brings to bear tools from branches of mathematics ranging from functional analysis and measure theory to topology and homological algebra.

Unfortunately, even for PDEs' power and flexibility as a lens through which to view the world, they can be difficult to solve. A purely mathematical analysis of a given problem can in general only get you so far. One might be able to determine the existence and uniqueness of a solution, and hopefully even that the solution is stable with regards to small pertubations of the data. However, finding a formula for the solution will be impossible in all but the simplest PDEs, on the simplest geometries. So with the hope of finding an exact solution abandoned, we turn to numerical mathematics and computers to find discrete approximations of the continuous problem. There are numerous ways a PDE can be discretized, popular choices being by finite difference methods, finite elements, spectral methods. Regardless of choice of discretization, it is natural that the numerical method, and its analysis, should be informed by the analysis and mathematical tools employed on the continuous problem. That is, the discrete problem should preserve as much of the continuous PDE as possible. Such structure preserving discretizations can be exemplified by finite element exterior calculus ([9, 11] are key references), which is a unification and generalization of much of the theory on finite element methods for saddle point problems — a class of system of PDEs we will become more acquainted with later on.

Another recurring theme with many discretization methods is that the continuous problem is replaced by a large number of simpler, algebraic equations. This invariably leads to the desire to solve large, and often sparse, matrix equations on a computer, and the number of unknowns will grow with the fidelity of the numerical approximation. At this point the issue of scalability of how one solves these matrix equations appears. Let N denote the order of the matrix equation $\mathsf{A}\mathsf{x} = \mathsf{b}$, arising from some discretization of a PDE, where in particular A is a sparse matrix with Z non-zero entries. Then the cost of a solving the set of equations using Gaussian elimination will be of the order $\mathcal{O}(ZN \log N)$, [54]. As such, the solving time will become infeasible for the large values of N encountered in applications. Iterative solution methods, on the other hand, can offer a computational complexity of $\mathcal{O}(N)$ per iteration. Of

particular relevance to this thesis is the family of Krylov subspace methods, which we will discuss in more depth later on in the introduction. Although these methods are order-optimal per iteration, a caveat is the number of iterations needed to convergence within a given tolerance for error. Specifically, the iteration count will typically increase as N increases. This is tied to the spectral properties of A, and as a consequence the structure of discretization method as well as the PDE being considered.

To get iterative methods that achieve a prescribed error tolerance within a number of iterations independent of N, we introduce preconditioners. Here, we consider the modified system $\mathsf{BAx} = \mathsf{Bb}$, where B is the preconditioner matrix. The goal is to design B so that BA has nice spectral properties, while also keeping the application of B at most $\mathcal{O}(N)$ in complexity. This will then lead to order-optimal solution methods. With these design goals in mind, it comes as little surprise that B should be related to A, and in turn be informed by the foregoing analysis done for both the continuous PDE and its discretization. In some sense, the preconditioner can be viewed as a culmination of the preceeding analysis. It attempts to make explicit in the solution method more of the structure of the PDE than what more naive iterative methods take into account. This rather vague argument will be further substantiated throughout the introduction.

Mathematical background

In the following sections we will cover in more detail the mathematical theory most relevant to the papers that make up this thesis. In particular, the discussions regarding abstract preconditioning draw from [63]. Thus, we refer the reader there for further details and examples. There, the design of efficient preconditioners relies on the specific properties of the underlying PDE and the chosen discretization method. See also [32, 40, 61, 91] for further expositions on this perspective.

Krylov Subspace Methods

A popular family of iterative methods for solving linear equations are the Krylov subspace methods. In this section we shall give a brief introduction to the hallmark Krylov methods, viz. the conjugate gradient method, [38], with emphasis on error estimation. We will also discuss how to generalize the conjugate gradient method to get other Krylov subspace methods. More in-depth analyses of Krylov subspace methods can be found in e.g. [33, 55, 76].

Let X be a Hilbert space, possibly infinite-dimensional, with inner product and norm denoted by $(\cdot,\cdot)$ and $||\cdot||$, respectively. We consider $\mathscr{A} : X \to X$ a linear, symmetric and positive definite operator. That is, $\mathscr{A}$ is linear and satisfies

$$(\mathscr{A}x, y) = (\mathscr{A}y, x), \quad \text{and} \quad (\mathscr{A}x, x) > 0 \tag{1}$$

for every $x, y \in X$. For a given $f \in X$ we aim to find an $x \in X$ so that

$$\mathscr{A}x = f, \tag{2}$$

using an iterative method. The iterates will be chosen from the Krylov subspaces

$$V_k = V_k(\mathscr{A}, f) = \text{span}(f, \mathscr{A}f, \ldots, \mathscr{A}^{k-1}f)$$

We associate with (2) the energy functional

$$\Phi(x) := \frac{1}{2}(\mathscr{A}x, x) - (f, x).$$

We will now see that the solution $x \in X$ to (2) minimizes Φ over X. Let $y, z \in X$, using the definition of Φ we see that

$$\Phi(y+z) - \Phi(y) = (\mathscr{A}y - f, z) + \frac{1}{2}(\mathscr{A}z, z) \geq (\mathscr{A}y - f, z).$$

Thus, we see that $x \in X$ solves (2) if and only if

$$\Phi(x) \leq \Phi(y), \quad \forall y \in X. \tag{3}$$

The conjugate gradient method is based on this minimizing property, where the iterates x_k are implicitly defined by

$$\Phi(x_k) \leq \Phi(y), \quad \forall y \in V_k. \tag{4}$$

Some of the power of the conjugate gradient lies in the fact that the iterates x_k defined by (4) can be computed relatively cheaply. In fact, to calculate x_k, only the previous iterate x_{k-1} and f are required, together with computing the action of $\mathscr{A}$ on any elements of X. More details on implementation can be found, for instance, in [33, Sec. 10.2.2].

To estimate the error $x - x_k$, we begin by noting that similar to how we derived the minimization property in (3), we can derive that the iterates x_k satisfy

$$(\mathscr{A}x_k, y) = (f, y), \quad \forall y \in V_k. \tag{5}$$

From (5) and (2), the error $x - x_k$ satisfies the orthogonality condition

$$(\mathscr{A}(x - x_k), y) = 0, \quad \forall y \in V_k.$$

From this, we can verify that x_k is the element in V_k, which minimizes the error to x in the $\mathscr{A}$-norm. That is,

$$\|x - x_k\|_{\mathscr{A}} = \inf_{y \in V_k} \|x - y\|_{\mathscr{A}}, \tag{6}$$

where $\|\cdot\|_{\mathscr{A}}^2 = (\mathscr{A}\cdot, \cdot)$.

Next, we observe that since $x_k \in V_k$, it follows that $x_k = p_k(\mathscr{A})x$ for some polynomial p_k of order k with $p_k(0) = 0$. Thus, $x - x_k = q_k(\mathscr{A})x$, where q_k is a polynomial of order k with $q_k(0) = 1$. It follows that

$$\|x - x_k\|_{\mathscr{A}} \leq \sup_{\lambda \in \sigma(\mathscr{A})} |q_k(\lambda)| \|x\|_{\mathscr{A}}, \tag{7}$$

where $\sigma(\mathscr{A}) \subset \mathbb{R}$ denotes the spectrum of $\mathscr{A}$. This observation, together with the energy minimization property in (6), gives us a bound on the error as

$$\|x - x_k\|_{\mathscr{A}} \leq \inf_{\substack{p \in \mathscr{P}_k \\ p(0)=1}} \sup_{\lambda \in \sigma(\mathscr{A})} |p(\lambda)| \|x\|_{\mathscr{A}}. \tag{8}$$

The polynomial minimizing the right hand side of (8) can be expressed in terms of scaled Chebyshev polynomials (see for instance [83, Thm. 38.5]), in which case one can derive the bound

$$\|x - x_k\|_{\mathscr{A}} \leq 2\left(\frac{\sqrt{K(\mathscr{A})} - 1}{\sqrt{K(\mathscr{A})} + 1}\right)^k \|x\|_{\mathscr{A}}. \tag{9}$$

Here, $K(\mathscr{A}) := \|\mathscr{A}\|_{\mathscr{L}(X,X)} \|\mathscr{A}^{-1}\|_{\mathscr{L}(X,X)}$ denotes the condition number of $\mathscr{A}$, which in the

case of $\mathscr{A}$ being symmetric can be expressed as

$$K(\mathscr{A}) = \frac{\sup_{\lambda\in\sigma(\mathscr{A})}|\lambda|}{\inf_{\lambda\in\sigma(\mathscr{A})}|\lambda|}.$$

We first observe that from the error estimate (9) that the conjugate gradient method is guaranteed to converge. Secondly, the convergence factor depends adversely on the condition number of $\mathscr{A}$, in the sense that higher condition numbers yield slower convergence of the iterations.

If we assume that $\mathscr{A}$ is only symmetric indefinite, i.e. we no longer assume the inequality in (1), the associated energy functional Φ will no longer have a unique minimizer, and we cannot use conjugate gradient iterations to solve (2). Instead, we seek iterates $x_k \in V_k$ characterized by

$$x_k := \arg\min_{y\in V_k} ||\mathscr{A}y - f||^2. \tag{10}$$

This leads to the so-called minimal residual method, [67]. As with the conjugate gradient method, the iterates x_k can be efficiently computed recursively based only on evaluations of $\mathscr{A}$, and the iterates x_k satisfy

$$||\mathscr{A}(x-x_k)|| \leq \inf_{y\in V_k} ||\mathscr{A}(x-y)||.$$

Moreover, we have analogously to the conjugate gradient method that

$$||\mathscr{A}(x-x_k)|| \leq \inf_{\substack{p\in\mathscr{P}_k \\ p(0)=1}} \sup_{\lambda\in\sigma(\mathscr{A})} |p(\lambda)|\,||\mathscr{A}x||.$$

Since the spectrum of $\mathscr{A}$ is no longer contained in the positive half of the real line, we cannot proceed as before by using Chebyshev polynomials. However, the convergence rate will sill depend on the condition number of $\mathscr{A}$. For instance, by restricting to only symmetric polynomials in the above, we get the bound

$$||\mathscr{A}(x-x_k)|| \leq 2\left(\frac{K(\mathscr{A})-1}{K(\mathscr{A})+1}\right)^k ||\mathscr{A}x||,$$

[56]. We observe that the convergence of the minimal residual iteration is slower than for the conjugate gradient iteration, but the convergence behaves similarly in that an increase in $K(\mathscr{A})$ results in slower convergence.

Remark 1. *In addition to the above mentioned minimal residual method, the conjugate gradient method has generated numerous generalizations and offspring. Most notably GMRES (Generalized minimal residual), [77], and Bi-CGstab (stabilized Bi-conjugate gradient), [87], which are applicable for nonsymmetric problems, but come at a higher memory- and computational cost. There have also been developed Krylov subspace methods for eigenvalue problems, in particular the Arnoldi-, [12], and Lanczos iterations, [51].*

Preconditioning

In the previous section we saw that the convergence of most Krylov methods depends adversely on the conditioning of the operator $\mathscr{A}$. In many applications the conjugate gradient method, or another Krylov subspace method, is used to solve matrix equations that come from discretizations of PDEs. It is common in such applications that the underlying contin-

uous operator is unbounded from a Hilbert space into itself, making the resulting iterative method slowly converging.

The variational formulation of a PDE will often be well-posed on an infinite-dimensional Hilbert space X, densely and continuously embedded in some other Hilbert space Y. Specifically, let $(\cdot,\cdot)$ denote the inner product on Y, with corresponding norm $||\cdot||$, and let X^* denote the dual space of X, with duality pairing as an extension of $(\cdot,\cdot)$. Then, if we identify Y with its own dual, Y will be densely and continuously embedded in X^*. The general problem we consider is for a given $f \in X^*$, to find $u \in X$ so that

$$\mathscr{A}u = f, \tag{11}$$

where we assume $\mathscr{A} : X \to X^*$ to be a homeomorphism. That is, $\mathscr{A}$ is bounded and invertible from X to X^* with bounded inverse. We see that a Krylov space method as described previously based on (11) will not be well-defined, since we cannot identify X with its dual space. In a discrete setting, this will often lead to a deterioration of the convergence of the iterative method as the problem size increase.

The remedy is to introduce an additional operator, called the preconditioner, $\mathscr{B} : X^* \to X$, which we assume to be a symmetric positive-definite isomorphism. Then, the operator $\mathscr{B}\mathscr{A} : X \to X$ is symmetric in the preconditioned inner product $\left(\mathscr{B}^{-1}\cdot,\cdot\right)$, and thus Krylov space methods are applicable to the preconditioned system

$$\mathscr{B}\mathscr{A}u = \mathscr{B}f.$$

The convergence will then depend on the condition number

$$K(\mathscr{B}\mathscr{A}) = ||\mathscr{B}\mathscr{A}||_{\mathscr{L}(X,X)} \left\|(\mathscr{B}\mathscr{A})^{-1}\right\|_{\mathscr{L}(X,X)}.$$

A natural choice is $\mathscr{B} : X^* \to X$ as the Riesz mapping defined by

$$(\mathscr{B}f, v)_X = (f, v), \quad \forall v \in X, f \in X^*.$$

We will henceforth refer to this choice of $\mathscr{B}$ as the *canonical preconditioner*. Then it is straightforward to show that

$$K(\mathscr{B}\mathscr{A}) \leq ||\mathscr{A}||_{\mathscr{L}(X,X^*)} \left\|\mathscr{A}^{-1}\right\|_{\mathscr{L}(X^*,X)}, \tag{12}$$

which means that the convergence of a Krylov space method will be based solely on the mapping properties of $\mathscr{A}$.

Finite element methods, and many other numerical methods for the numerical solution of PDEs, are based on choosing a family of finite-dimensional subspaces $X_h \subset X$, indexed by h. Here, it is natural to assume that the spaces X_h provide better and better approximations of the space X. Therefore, we will in the following assume that

$$\lim_{h\to 0} \inf_{v\in X_h} ||w - v|| = 0, \quad \forall w \in X. \tag{13}$$

For a fixed h, let us solve the discrete problem

$$\mathscr{A}_h u_h = f_h, \tag{14}$$

where $\mathscr{A}_h : X_h \to X_h^*$ and $f_h \in X_h^*$ is the restriction of $\mathscr{A}$ and f to X_h, respectively. Since X_h is finite dimensional, the norms inherited from X and Y are equivalent on X_h, and we can make the identification of X_h with its dual. This makes (14) amenable to the Krylov methods discussed in the previous section.

N	32	64	128	256	512	1024
k	31	63	126	251	506	996

Table 1: Number of conjugate gradient iterations before reaching an error tolerance of 10^{-6} for the Poisson equation on the unit interval divided into N subintervals.

However, the norm equivalence on X_h will deteriorate as $h \to 0$, which will lead to a blow-up of the operator norms

$$\|\mathscr{A}_h\|_{\mathscr{L}(X_h,X_h)}, \left\|\mathscr{A}_h^{-1}\right\|_{\mathscr{L}(X_h,X_h)}. \tag{15}$$

This in turn will lead to $K(\mathscr{A}_h)$ increasing as $h \to 0$, and the convergence of a Krylov method will be slow.

Example 1 (Poisson equation)**.** To illustrate the above issue, consider the Poisson equation with homogeneous Dirichlet boundary conditions. We set $\Omega = (0,1)$ to be the unit interval on $\mathbb{R}$, with boundary $\partial\Omega$. Further, let $X = H_0^1(\Omega)$, and $Y = L^2(\Omega)$, and define the operator $\mathscr{A} : X \to X^*$ by

$$(\mathscr{A}v, w) := (\nabla v, \nabla w), \quad \forall v, w \in X.$$

We discretize X by partitioning Ω into $N+1$ subintervals of equal length $h = \frac{1}{N+1}$. We then define the discrete space X_h to be the subspace of X consisting of functions that are linear polynomials on each subinterval. See then that $\dim X_h = N$. If we have a basis for X_h, which we denote by $\{\phi_h^i\}_{i=1}^N$, we can assemble the matrix $\mathsf{A} \in \mathbb{R}^{N\times N}$ and vector $\mathsf{f} \in \mathbb{R}^N$ defined by

$$\mathsf{A}_{i,j} = (\mathscr{A}\phi_h^j, \phi_h^i), \quad \text{and} \quad \mathsf{f}_i = \left(f, \phi_h^i\right).$$

Let $\mathsf{u}_h \in \mathbb{R}^N$ be the solution to the matrix equation

$$\mathsf{A}\mathsf{u}_h = \mathsf{f}. \tag{16}$$

We now try to solve (16) on a computer using the conjugate gradient method for different values of N, and random right hand side. We stop after k iterations if the relative residual $\frac{\|r_k\|^2}{\|r_0\|^2} < 10^{-6}$. The results can be viewed in Table 1, where we see that indeed the number of iterations increases along with increasing N.

To overcome this deterioration of convergence we do as in the continuous case, and introduce a preconditioner $\mathscr{B}_h : X_h^* \to X_h$ by

$$(\mathscr{B}_h f, v)_X = (f, v), \forall v \in X_h, f \in X_h^*,$$

and consider the preconditioned discrete problem

$$\mathscr{B}_h\mathscr{A}_h u_h = \mathscr{B}_h f_h.$$

As already noted, the convergence of a Krylov method will then depend only on

$$\|\mathscr{A}_h\|_{\mathscr{L}(X_h,X_h^*)}, \left\|\mathscr{A}_h^{-1}\right\|_{\mathscr{L}(X_h^*,X_h)},$$

which are assumed bounded independently of h.

The canonical preconditioner $\mathscr{B}_h$ as defined above, will generally be expensive to compute. For instance, in Example 1 above, $\mathscr{B}_h = \mathscr{A}_h^{-1}$, and if $\mathscr{A}_h^{-1}$ already was cheap to compute, then we would already have an efficient solution algorithm available. To avoid this problem, we can replace $\mathscr{B}_h$ by a second symmetric, positive definite operator $\mathscr{B}_{h,2} : X_h^* \to X_h$, which we assume is easier to compute than $\mathscr{B}_h$, and is spectrally equivalent to $\mathscr{B}_h$. That is, there are constants C_1, C_2, which are independent of h so that

$$C_1(\mathscr{B}_h f, f) \leq \left(\mathscr{B}_{h,2} f, f\right) \leq C_2(\mathscr{B}_h f, f), \quad \forall f \in X_h^*. \tag{17}$$

Then $\mathscr{B}_{h,2}\mathscr{B}_h^{-1}$ is a linear isomorphism on X_h with $K(\mathscr{B}_{h,2}\mathscr{B}_h^{-1}) \leq \frac{C_2}{C_1}$. It then follows that if we use $\mathscr{B}_{h,2}$ as preconditioner instead,

$$K(\mathscr{B}_{h,2}\mathscr{A}_h) \leq \frac{C_2}{C_1} \|\mathscr{A}_h\|_{\mathscr{L}(X_h, X_h^*)} \left\|\mathscr{A}_h^{-1}\right\|_{\mathscr{L}(X_h^*, X_h)},$$

which is bounded independently of h.

From the above discussion, we observe that preconditioning can be seen as a way to introduce the norms under which $\mathscr{A}_h$ is a homemorphism into the implementation. Preferably with norms independent of the discretization parameter h. As such, the preconditioner should be informed by the analysis of both the continuous operator and its discretization. Having the norms, and thus the canonical projection $\mathscr{B}_h$, is rarely enough, though, and a computationally tractible preconditioner $\mathscr{B}_{h,2}$ must be constructed. This is by no means a simple matter in general, and since the preconditioner should be informed by the underlying PDE, we cannot hope for a universal preconditioner that works for every problem. However, we will later discuss multigrid methods, which have turned out to inspire quite general frameworks for constructing preconditioners for a wide variety of problems.

Saddle Point Problems

In the previous section we saw that a preconditioner for a given problem should be chosen based on the mapping properties of the underlying operator defining the PDE. We will now expand on this point, and see that the design of preconditioner for a system of PDEs can be broken into a series of simpler subproblems. When preconditioners for each subproblem have been established, a preconditioner for the original can easily be constructed. This approach has been used in for instance [6, 78], but we emphasize its wider applicability here by considering abstract saddle point systems. To that end, let V and Q be Hilbert spaces. We consider the problem for given $f \in V^*$ and $g \in Q^*$: Find $(u, p) \in V \times Q$ so that

$$\begin{aligned} a(u,v) + b(v,p) &= (f,v), \quad \forall v \in V \\ b(u,q) &= (g,q), \quad \forall q \in Q. \end{aligned} \tag{18}$$

Here, $a : V \times V \to \mathbb{R}$ and $b : V \times Q \to \mathbb{R}$ are bounded, bilinear forms. We can state the problem in our more abstract framework by introducing operators $A : V \to V^*$ and $B : V \to Q^*$ defined by

$$(Av, w) = a(v, w), \quad \text{and} \quad (Bv, q) = b(v, q),$$

for each $v, w \in V$ and $q \in Q$. Introducing the product space $W = V \times Q$, for which we have $W^* = V^* \times Q^*$, we can define $L := f \times g \in W^*$ by

$$(L, (v, q)) = (f, v) + (g, q),$$

and $\mathscr{A} : W \to W^*$ by

$$\mathscr{A} = \begin{pmatrix} A & B^* \\ B & 0 \end{pmatrix},$$

where $B^* : Q \to V^*$ is the adjoint of B. With $x = (u, p) \in W$, the saddle point system (18) can then be reformulated as

$$\mathscr{A} x = L. \tag{19}$$

The Brezzi conditions are sufficient conditions for the well-posedness of (18). Equivalently, they are sufficient for $\mathscr{A}$ to be an isomorphism. The conditions are as follows: There are constants $\alpha, \beta > 0$ so that

$$a(v, v) \geq \alpha \|v\|_V^2, \quad \forall v \in \ker B, \tag{20}$$

and

$$\inf_{q \in Q} \sup_{v \in V} \frac{b(v, q)}{\|v\|_V \|q\|_Q} \geq \beta. \tag{21}$$

We note that as a consequence of Banach's closed range theorem, the condition (21) is equivalent to the operator $B : V \to Q^*$ being surjective.

The canonical choice for preconditioner for $\mathscr{A}$ is then the Riesz mapping $\mathscr{B} : W^* \to W$, which takes the form

$$\mathscr{B} = \begin{pmatrix} \mathscr{B}_V & 0 \\ 0 & \mathscr{B}_Q \end{pmatrix},$$

where $\mathscr{B}_V : V^* \to V$ and $\mathscr{B}_Q : Q^* \to Q$ are the canonical Riesz mappings for V and Q, respectively. Thus, when designing preconditioners for problems of the form (18), we should determine spaces V and Q so that the Brezzi conditions are satisfied. We can then consider V and Q separately when constructing a preconditioner.

We note that the block-diagonal preconditioner $\mathscr{B}$ is not the only choice of structure for an abstract preconditioner of saddle point systems. For instance, if the problem is non-symmetric, or preserving symmetry is not of great importance, block-triangular preconditioners can be considered. These types of preconditioners were considered in e.g. [59, 99]. The point still remains, however, that the preconditioner should reflect norms under which the saddle point problem is well-posed.

Parameter-robust preconditioning

Many PDEs used to model the physical world depend on a variety of physical parameters whose values can range from the very small to the very large. When designing algorithms for solving linear systems arising from discretizations of such PDEs, we want the convergence to be indendepent of whatever value these parameters have, in addition to the fidelity of the discretization. Recall that the canonical preconditioner is given by the Riesz mapping on a Hilbert space to its dual for a Hilbert space in which the PDE is well-posed. For parameter-dependent PDEs this translates to identifying *parameter-dependent* norms so that the underlying operator defining the PDE is bounded uniformly in these norms, [100].

To be more specific, let $\varepsilon \in \mathbb{R}^m$ be a set of m parameter values. Further, let $\mathscr{A}_\varepsilon : X \to X^*$ be a linear homeomorphism, characterized by the parameters ε. Generally, we cannot expect the operator norms

$$\|\mathscr{A}_\varepsilon\|_{\mathscr{L}(X, X^*)}, \quad \text{and} \quad \left\|\mathscr{A}_\varepsilon^{-1}\right\|_{\mathscr{L}(X^*, X)}$$

to be bounded independent of ε. In view of (12), we cannot expect the condition number of $\mathscr{B}\mathscr{A}_\varepsilon$, where $\mathscr{B} : X^* \to X$ is the canonical preconditioner, to be bounded independently of

ε. The goal is then to define an ε-dependent norm, $\|\cdot\|_\varepsilon$, on X so that if we define $X_\varepsilon := X$ as a set, but with norm given by $\|\cdot\|_\varepsilon$, then the norms

$$\|\mathscr{A}_\varepsilon\|_{\mathscr{L}(X_\varepsilon, X_\varepsilon^*)}, \quad \text{and} \quad \|\mathscr{A}_\varepsilon^{-1}\|_{\mathscr{L}(X_\varepsilon^*, X_\varepsilon)}$$

are bounded indepedently of ε. It follows then that the canonical preconditioner $\mathscr{B}_\varepsilon : X_\varepsilon^* \to X_\varepsilon$ yields condition numbers for $\mathscr{B}_\varepsilon \mathscr{A}_\varepsilon$ that are bounded independently of ε.

Example 2 (Biot's consolidation model)**.** As an example of a parameter-dependent PDE we consider Biot's consolidation model, which we recall is popular in modelling fluid flow through elastic, porous media.

Let Ω be a bounded, n-dimensional domain. In a homogeneous, isotropic, linear elastic porous media, the quasi-static Biot's model is

$$\begin{aligned} -\operatorname{div}(2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda(\operatorname{div}\mathbf{u})\mathbf{I}) + \alpha\nabla p_F &= \mathbf{f}, \\ s_0 \dot{p}_F + \alpha \operatorname{div}\dot{\mathbf{u}} - \operatorname{div}(\kappa\nabla p_F) &= g \end{aligned} \tag{22}$$

in Ω. Here the unknowns are the displacement of the elastic medium, $\mathbf{u}$, and the fluid pressure, p_F. In particular, $\mathbf{u}$ is a vector field, and p_F is a scalar-valued function, both defined on Ω. In addition, $\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla + \nabla^\top)$ denotes the symmetric gradient. The right hand sides $\mathbf{f}$ and g are the given momentum- and mass sources, respectively. The system (22) should of course be paired with a suitable set of boundary conditions. As for the physical parameters, μ and λ are the Lamé parameters, which characterize the solid. The Biot-Willis constant, α, couples the two PDEs, while s_0 is the constrained specific storage coefficient and is defined as fluid content per change in pressure constrained by a fixed strain (cf. [90]). Lastly, κ is the symmetric permeability tensor, which roughly speaking measures the ease at which the fluid flows through the porous medium. The values of all the parameters mentioned above are assumed to be positive, and $\alpha < 1$. In the case of κ we assume that it is unformly symmetric positive definite, i.e. there are constants $0 < K_0 \leq K_1$ so that for every $x \in \Omega$ and $\xi \in \mathbb{R}^n$

$$K_0|\xi|^2 \leq \xi^\top \kappa(x)\xi \leq K_1|\xi|^2.$$

We will now consider a variational formulation of (22) with homogeneous Dirichlet boundary conditions, i.e. we impose $\mathbf{u} = 0$ and $p_F = 0$ on the boundary of Ω. Further, we will disregard the time derivatives in the second equation of (22). This is justified from the fact that any implicit time discretization of (22) will lead to the need to solve systems like (22) without time derivatives. For sake of further simplicity, we further assume $\mu = 1$, which can be achieved through a parameter rescaling, and that $s_0 = \frac{\alpha^2}{\lambda}$. The variational problem we now consider is then to seek $\mathbf{u} \in \boldsymbol{H}_0^1 = \boldsymbol{H}_0^1(\Omega)$ and $p_F \in H_0^1 = H_0^1(\Omega)$ so that

$$\begin{aligned} &(\boldsymbol{\varepsilon}(\mathbf{u}), \boldsymbol{\varepsilon}(\mathbf{v})) + \lambda(\operatorname{div}\mathbf{u}, \operatorname{div}\mathbf{v}) - \alpha(p_F, \operatorname{div}\mathbf{v}) = (\mathbf{f}, \mathbf{v}), \quad \forall \mathbf{v} \in \boldsymbol{H}_0^1 \\ &-\alpha(\operatorname{div}\mathbf{u}, q_F) - \frac{\alpha^2}{\lambda}(p_F, q_F) - (\kappa\nabla p_F, \nabla q_F) = (g, q_F), \quad \forall q_F \in H_0^1. \end{aligned} \tag{23}$$

In coefficient matrix form, we can write (23) as

$$\mathscr{A}\begin{pmatrix}\mathbf{u}\\ p_F\end{pmatrix} := \begin{pmatrix}-\operatorname{div}\boldsymbol{\varepsilon} - \lambda\nabla\operatorname{div} & \alpha\nabla \\ -\alpha\operatorname{div} & -\frac{\alpha^2}{\lambda} + \operatorname{div}\kappa\nabla\end{pmatrix}\begin{pmatrix}\mathbf{u}\\ p_F\end{pmatrix} = \begin{pmatrix}\mathbf{f}\\ g\end{pmatrix}, \tag{24}$$

where we clearly see that $\mathscr{A}$ depends on λ, α, and κ. As such, the operator norm of $\mathscr{A} : \boldsymbol{H}_0^1 \times H_0^1 \to \boldsymbol{H}^{-1} \times H^{-1}$ and its inverse will not be independent of these parameters. So we

introduce the parameter-dependent norms

$$\|\mathbf{v}\|_V^2 := \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2 + \lambda\|\operatorname{div}\mathbf{v}\|^2, \quad \mathbf{v} \in \boldsymbol{H}_0^1$$
$$\|q_F\|_Q^2 := \frac{\alpha^2}{\lambda}\|q_F\|^2 + (\kappa\nabla q_F, \nabla q_F), \quad q_F \in H_0^1,$$

and set $V = \boldsymbol{H}_0^1$ and $Q = H_0^1$ as sets imbued with the above norms, respectively. Then it can be shown that $\mathscr{A} : V \times Q \to V^* \times Q^*$ is a homeomorphism with itself and its inverse bounded independently of the parameters. With these norms the, canonical preconditioner $\mathscr{B} : V^* \times Q^* \to V \times Q$ for $\mathscr{A}$ then reads

$$\mathscr{B} = \begin{pmatrix} (-\operatorname{div}\boldsymbol{\varepsilon} - \lambda\nabla\operatorname{div})^{-1} & 0 \\ 0 & \left(\frac{\alpha^2}{\lambda} - \operatorname{div}\kappa\nabla\right)^{-1} \end{pmatrix}. \tag{25}$$

In preconditioning discretizations of (23), the non-zero blocks in (25) must be replaced by spectrally equivalent and computationally cheap operators. The second block of (25) can be replaced by suitable multigrid operators, which we will discuss in the next section. The first block is more troublesome, where common preconditioning strategies fail in the incompressible limit $\lambda \to \infty$ ([72]). See also [42], where a nonnested multigrid preconditioner is proposed for a discontinuous Galerkin discretization of an operator closely related to the V-norm.

In addition to the complicated first block of $\mathscr{B}$, many discretizations of Biot's consolidation model based on (23) will be numerically unstable, [34]. As such, there have been developed several different discretizations of (22). Most pertinent to our discussion so far are different mixed methods studied in e.g. [64, 69, 70, 95]. We remark that also primal formulations, e.g. [98], and least squares methods, [45], have also been studied, but we make no claim that this list is exhaustive. With a plethora of discretization methods comes also the need to develop several different preconditioners, each suitable for a particular discretization. In particular, block preconditioners for different discretizations of poroelasticity have been studied in e.g. [71, 73], and with an emphasis on parameter-robustness in [3, 29, 35, 43, 53, 74].

Multigrid methods

Up to this point, we have treated preconditioning in more abstract terms. That is, having established well-posedness of a PDE under certain norms yields a canonical preconditioner a fortiori. Similarly, canonical preconditioners for discretizations of PDEs are also borne out from considering the mapping properties of the discrete operator. As already remarked however, these canonical preconditioners are rarely computationally tractible, and in computer code should then be replaced by operators that are cheap to compute and spectrally equivalent to the canonical preconditioner according to (17). Optimally, the implemented preconditioner should also be computed in $\mathcal{O}(N)$ time, where N is the number of unknowns to be solved for. We are then left with the task of constructing such operators.

Multigrid methods is a framework that has proven effective in constructing preconditioners adressing the above design goals for a wide class of problems. As such, there is a large body of theory concerning these methods. Cf. [24, 84] and references therein. Moreover, multigrid methods have spawned a number of generalizations, and they can be seen as a special case of an even more general framework for constructing preconditioners, [93]. Therefore, we will in this section discuss a simple multigrid method, primarily from a motivational perspective, while also devoting time to point to some important generalizations.

To motivate multigrid methods, we revisit Example 1. Recall that $X_h \subset H_0^1(\Omega)$ is the space

of continuous, piecewise linear functions relative to a uniform partition of the unit interval $\Omega = (0,1)$. Then we define the discrete Laplacian $\mathscr{A}_h : X_h \to X_h$ as the restriction of the Laplacian $\mathscr{A}$ to X_h.

For a given $f \in X_h$ we seek $u \in X_h$ so that

$$\mathscr{A}_h u = f. \tag{26}$$

We begin by considering one of the simplest iterative methods for solving (26), viz. the Richardson's iteration[1]. Let $u_0 \in X_h$ be an initial guess, and $\tau \in \mathbb{R}$ a constant parameter. Then a subsequent iterate $u_{k+1} \in X_h$ is obtained from the previous iterate u_k by

$$u_{k+1} = u_k + \tau(f - \mathscr{A}_h u_k). \tag{27}$$

It is easy to see from (27) that the error $e_k := u - u_k$ satisfies

$$e_k = (I - \tau\mathscr{A}_h)e_{k-1} = (I - \tau\mathscr{A}_h)^{k-1}e_0.$$

For the iteration (27) to converge (in $\mathscr{A}_h$-norm), we must require that $||I - \tau\mathscr{A}_h||_{\mathscr{A}_h} < 1$. If $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and greatest eigenvalues of $\mathscr{A}_h$, respectively, then this means that the Richardson's iteration converges only if

$$0 < \tau < \frac{2}{\lambda_{\max}},$$

and the optimal choice is given by $\tau = \frac{2}{\lambda_{\min}+\lambda_{\max}}$. However, even with a choice of τ so that the Richardson's iteration converges, the convergence might be slow.

To see this, let now $0 < \lambda_{\min} = \lambda_1 < \lambda_2 < \cdots < \lambda_N = \lambda_{\max} < \infty$ be the eigenvalues of $\mathscr{A}_h$, and $v_1, \ldots, v_N$ the corresponding set of orthonormal eigenfunctions. Suppose now that we make the particular choice of $\tau = \frac{1}{\lambda_N}$, which ensures that the Richardson's iteration converges. We decompose the initial error as

$$e_0 = \sum_{i=1}^{N} c_i v_i,$$

and its norm can be written as $||e_0||^2_{\mathscr{A}_h} = \sum_{i=1}^{N} \lambda_i c_i^2$. The decomposition of the k'th error is then

$$e_k = \sum_{i=1}^{N} c_i \left(1 - \frac{\lambda_i}{\lambda_N}\right)^k v_i. \tag{28}$$

At this point, it is worthwhile to make a few observations. First, we see from (28) that the Richardson's iteration quickly reduces the components of the error that correspond to the large eigenvalues. On the other hand, if $\lambda_i << \lambda_N$, the corresponding error component is left virtually untouched, and so we should expect that convergence is slow on these components. The second observations we make is that for elliptic operators — as is the case for $\mathscr{A}$ — and discretization of such operators, large eigenvalues have highly oscillatory eigenfunctions, whereas smaller eigenvalues have smoother eigenfunctions.

From these two observations we can expect that the error after k iterations of (27) is relatively smooth compared to the initial error. Figure 1 validates these observations. Here, we have set $f = 0$, and we see that while the error e_5 is slightly smaller in magnitude to the random

[1] A preconditioner can be derived from any linear iterative method in a straightforward manner. See [24, Prop. 1.1].

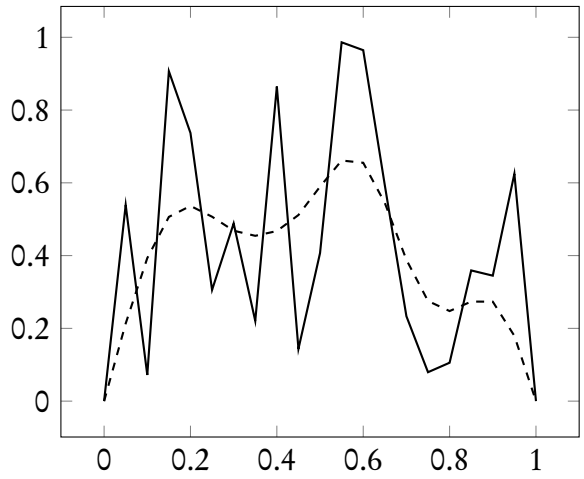


Figure 1: The solid line shows the initial error, e_0, while the dashed line shows the error, e_5, after 5 applications of (27) with $\tau = \frac{1}{\lambda_{\max}}$.

initial error, the most striking feature is that the error is much smoother.

At this point we can formulate one of the key ideas in multigrid. The smooth errors we get from a few Richardson's iterations, or any other simple smoothing procedure, can be well-represented on a coarser grids. On this coarser grid, a coarse solution is cheaper to compute, but roughly speaking what were low frequency components on the fine mesh are relatively high frequency on the coarser mesh, making them susceptible to a simple iterative method.

So suppose X_H is the space of continuous, piecewise linear functions relative to coarser partition of the unit interval. For instance, if $N+1 = \frac{1}{h}$ is even, we can take $H = 2h$. Then, $X_H \subset X_h$, and we denote by $I_H : X_H \to X_h$ the inclusion operator, while $Q_H = I_H^* : X_h \to X_H$ denotes the L^2 projection onto X_H. We define the action of our multigrid operator $\mathscr{B}_h : X_h \to X_h$ for a given $f \in X_h$ as

$$\begin{aligned} u^0 &= 0 \\ u^1 &= u^1 + \mathscr{R}_h(f - \mathscr{A}_h u^0) \\ u^2 &= u^1 + I_H \mathscr{B}_H Q_H (f - \mathscr{A}_h u^1) \\ u^3 &= u^2 + \mathscr{R}_h(f - \mathscr{A}_h u^2), \\ \mathscr{B}_h f &:= u^3. \end{aligned} \tag{29}$$

Here, $\mathscr{R}_h : X_h \to X_h$ represents the application of a number of Richardson's iterations, or another smoothing procedure. The relatively smooth residual, $f - \mathscr{A}_h u^1$, is then projected onto the coarser space X_H, where $\mathscr{B}_H$ can represent either an exact inversion of $\mathscr{A}_H$, or another level in a multigrid procedure. The last smoothing step can be seen to make $\mathscr{B}_h$ a symmetric operator.

An iterative method for solving (26) based on $\mathscr{B}_h$ will now take an initial guess $u_0 \in X_h$, and subsequent iterates u_{k+1} are then obtained from the previous iterative u_k by

$$u_{k+1} = u_k + \mathscr{B}_h(f - \mathscr{A}_h u_k). \tag{30}$$

Similar to Richardson's iteration, (30) converges if and only if $\|I - \mathscr{B}_h \mathscr{A}_h\|_{\mathscr{A}_h} < 1$. Perhaps unsurprisingly, the convergence of the iterative method (30) requires some interplay between the chosen smoothing procedure and the coarse space. More specifically, the coarse grid correction should handle the components of the error not covered by the smoother. This heuristic

N	32	64	128	256	512	1024
k	31	62	126	248	502	1005
k_{MG}	4	4	3	3	2	2

Table 2: Number of preconditioned conjugate gradient iterations before reaching an error tolerance of 10^{-6} for the Poisson equation on the unit interval divided into N subintervals. k is number of iterations until reaching error tolerance with no preconditioner. k_{MG} is the corresponding iteration count when using a multigrid preconditioner with 4 grid levels and a Jacobi smoother on each level.

can be substantiated further into a number of conditions on $\mathscr{R}_h$ and X_H that must be verified. We settle here with referring to [24, Chapter 2] for further details.

Suppose that (30) provides a convergent iteration in $\mathscr{A}_h$, so there is some $0 \leq \delta < 1$ such that

$$0 \leq ||I - \mathscr{B}_h \mathscr{A}_h||_{\mathscr{A}_h} \leq \delta.$$

It then follows by the definition of the $\mathscr{A}_h$-norm that for every $u \in X_h$

$$(1-\delta)(\mathscr{A}_h u, u) \leq (\mathscr{B}_h \mathscr{A}_h u, \mathscr{A}_h u) \leq (\mathscr{A}_h u, u). \tag{31}$$

Replacing u by $\mathscr{A}_h^{-1} u$, we recover the spectral equivalence (17). Thus, if $\mathscr{B}_h$ provides a convergent linear iterative method, it is also suitable as preconditioner. Moreover, if the convergence factor δ is independent of h, then the condition number of $\mathscr{B}_h \mathscr{A}_h$ will be uniformly bounded in h.

To numerically validate this discussion, we revisit the tests we made in Example 1. We now use $\mathscr{B}_h$, defined by (29) as a preconditioner. For the smoother $\mathscr{R}_h$, we have used a scaling of the inverse of the diagonal entries of the matrix realization of $\mathscr{A}_h$, i.e. a Jacobi smoother. The smoother is applied on 3 succesively coarser grids, while on the coarsest grid we do an exact inversion. The results are shown in Table 2, where the iteration counts for the unpreconditioned conjugate gradient method are also displayed for comparison. Here we see that the iteration count for the preconditioned conjugate gradient method stays bounded independently of the problem size, in accordance with our discussion.

Remark 2. *In this section we have only considered a simple multigrid algorithm on a simple problem. But multigrid methods have been effectively applied to quite general elliptic problems, cf. [19, 62] and references therein. For $H(\mathrm{div})$-, or $H(\mathbf{curl})$ problems multigrid methods have been constructed, and proved efficient, in i.a. [7, 8, 39, 41].*

One major drawback of the multigrid framework as discussed in this section, is the assumption that we have a nested sequence of grids on which the coarser discrete spaces can be defined. In our above example, we used uniform partitions of the unit interval, and getting coarser partitions was relatively straightforward. However, in many applications such a sequence of grids is not readily available. To adress this issue, so-called algebraic multigrids (AMG) methods were developed, where the stiffness matrix, and its sparsity pattern, is used to construct coarse spaces. See [81, 89, 94] for general theory on AMG and an overview of its various flavours.

Further, abstract multigrid theory has also seen several generalizations. Of particular relevance to this thesis are the parallell multigrid method, [22], and multigrid methods for non-nested bilinear forms, [23]. In the former, the input f is projected onto each grid level, instead of the residual $f - \mathscr{A}_h u^1$ as was done in (29)*. This has the benefit that computing the action of $\mathscr{A}_h$ is not required, but generally leads to a worse spectral equivalence than standard multigrid operators. See also [97]. For the latter, the implicit assumption made in* (29) *that $\mathscr{A}_H$ is the restriction of $\mathscr{A}_h$ is lifted, and can be more general. This framework has proven useful in e.g. constructing preconditioners for*

discontinuous Galerkin discretizations, [31, 44]. Later, we will see the framework of [23] put to use when considering the fractional Laplacian in [14].

Fractional Sobolev spaces

So far, we have seen that when constructing preconditioners for saddle point systems, the task can be broken up into constructing preconditioners for PDEs with only one unknown. Moreover, we have also seen that multigrid operators can provide efficient preconditioners for a multitude of problems, in particular problems posed in H^1, $H(\mathrm{div})$, or $H(\mathbf{curl})$. The fractional Sobolev spaces, denoted H^s where s is a real parameter, have also received some attention from a preconditioning perspective, and are the main topic in two of the papers that make up this thesis. This, paired with the fact that H^s enter naturally into a variational formulation of a class of saddle point systems that we will discuss in the next section, justifies a brief discussion of fractional Sobolev spaces here.

Again, we let $\Omega \subset \mathbb{R}^n$ be a bounded domain. For $s \in [0,1]$, the space $H^s = H^s(\Omega)$ is an intermediate Hilbert space, containing H^1 and contained in L^2. There are a couple of ways to interpolate between Hilbert spaces[2], notably the real- and complex method, [1, Ch. 7]. See also [50] for an overview of the various definitions of fractional Laplacian operators, which naturally relates to fractional Sobolev spaces. However, the method of interpolation most pertinent to this thesis is the method based on spectral decomposition, as presented in [57]. To that end, let the inner product on H^1 be realized by the operator $\mathscr{A} := I - \Delta$, as

$$(u,v)_1 = (\mathscr{A}u, v) = (u,v) + (\nabla u, \nabla v), \quad u,v \in H^1.$$

Then $\mathscr{A}$ is unbounded as an operator mapping L^2 to L^2. However, $\mathscr{A}$ is well-defined on the set

$$D(\mathscr{A}) = \left\{ u \in L^2(\Omega) : \mathscr{A}u \in L^2(\Omega) \right\},$$

which is a dense subspace of L^2. On $D(\mathscr{A})$, $\mathscr{A}$ is symmetric positive-definite. From spectral theory (cf. e.g. [47, Ch. 7 and 9]) the fractional powers of $\mathscr{A}$, $\mathscr{A}^\theta$ for $\theta \in \mathbb{R}$, are well-defined. Note that in the particular case $\theta = \frac{1}{2}$,

$$\left\| \mathscr{A}^{\frac{1}{2}} u \right\|^2 = (\mathscr{A}u, u) = \|u\|_1 .$$

For $s \in [0,1]$, we define the fractional Sobolev spaces as

$$H^s = \left\{ u \in L^2 : \mathscr{A}^{\frac{s}{2}} u \in L^2 \right\}, \tag{32}$$

which is a Hilbert space with inner product given by

$$(u,v)_s = (\mathscr{A}^s u, v), \quad u,v \in H^s,$$

and we denote the corresponding norm by $\|\cdot\|_s$.

We define H_0^s to be the closure of $C_0^\infty(\Omega)$, the space of infinitely smooth functions with compact support in Ω, in the norm of H^s. We note that if $s \leq \frac{1}{2}$, the spaces $H_0^s(\Omega)$ and $H^s(\Omega)$ coincide (cf. [57, Theorem 11.1]). For $s \in [-1,0]$, we define a family of fractional Sobolev spaces using the dual of H_0^s. That is,

$$H^s = \left(H_0^{-s}\right)^* .$$

[2]These definitions are often equivalent in that they yield the same spaces as sets with equivalent norms.

Alternatively, replacing H^1 with H_0^1 and setting $\mathscr{A} = -\Delta$ in the above construction, will again yield the space H_0^s, with equivalent norm, for all s except when $s = \frac{1}{2}$. In this case, interpolation between H_0^1 and L^2 results in a space that is strictly contained in $H_0^{\frac{1}{2}}$.

One way to discretize $\mathscr{A}^s$ follows similar lines to the definition of the fractional Sobolev spaces. To see this, let $X_h \subset H_0^1$ be a finite-dimensional subspace with $\dim X_h = N$. We define the operator $\mathscr{A}_h : X_h \to X_h$ as the restriction of $\mathscr{A}$ to X_h. That is,

$$(\mathscr{A}_h u, v) = (\mathscr{A} u, v) = (\nabla u, \nabla v), \quad u, v \in X_h. \tag{33}$$

Using the fractional powers of $\mathscr{A}_h$, we define for $s \in \mathbb{R}$ the discrete fractional inner product on X_h by

$$(u, v)_{s,h} = \left(\mathscr{A}_h^s u, v\right), \quad u, v \in X_h,$$

and denoted the corresponding norm by $\|\cdot\|_{s,h}$.

A point worth making is that $X_h \subset H_0^s$ and $X_h \subset H^s$ for $s \in [0, 1]$, and so $\|v\|_s$ is well-defined for every $u \in X_h$. In fact, $\|\cdot\|_s$ and $\|\cdot\|_{s,h}$ are equivalent, [5]. Therefore, discrete operators $\mathscr{B}_h^s$ that are spectrally equivalent to $\mathscr{A}_h^{-s}$ can be used as preconditioners for PDEs well-posed in H^s.

However, in computer code, the construction of $\mathscr{A}_h^s$ in the way described above requires solving a potentially large eigenvalue problem, which is computationally expensive. Moreover, in the next section we will consider an application of H^s that does not involve evaluating the fractional Laplacian. As a consequence, we want to avoid any computation of the action of $\mathscr{A}_h^s$ when designing a preconditioner for H^s.

Remark 3. *Due to the numerous definitions of the fractional Laplace operators, there are likewise numerous discretizations and solution strategies for problems involving $(-\Delta)^s$. We refer to [58], and the references therein, for a discussion of various discretizations. As for general solution strategies of problems involving $(-\Delta)^s$, see e.g. [18, 37, 86].*

There have also been developed several preconditioners for the fractional Laplacian, both with positive- and negative fractionality. Multilevel preconditioners for an integral representation of the fractional Laplacian when $s = -\frac{1}{2}$ were studied in [20, 30]. For more general values of s, a hierarchical basis preconditioner was analyzed in [66] and further developed [21]. More recently, an auxiliary space preconditioner for when $s < 0$ was constructed in [79], but there the authors needed to assume a discretization of the fractional Laplacian of positive fractionality was available.

Multiphysics problems

Fractional Sobolev spaces, and the need for efficient preconditioners for them, appear naturally in many multiphysics problems. For example, when two different physical phenomena are modelled on separate domains, but coupled through an interaction on a common interface of lower geometrical dimensionality. One could also consider systems of PDEs where interacting physics are modelled on domains of different dimension. See for example [49].

Depending on the problem, and interface conditions one imposes, H^s with s being either positive or negative may appear. Thus, it is well-worth studying preconditioners for H^s with s both positive and negative.

To make the discussion more clear, we begin by considering the following example.

Example 3 (A simple trace constraint problem)**.** Let $\Omega = (0, 1)^2$ be the unit square in $\mathbb{R}^2$ and $\Omega_1 = \Omega \setminus \bar{\Omega}_2$, where $\Omega_2 = \left(\frac{1}{4}, \frac{3}{4}\right)^2$. Then, Ω_1 and Ω_2 form a partition of Ω, and we denote by Γ the interface between them. Further we define n_i, for $i = 1, 2$ to be the unit normal vector on

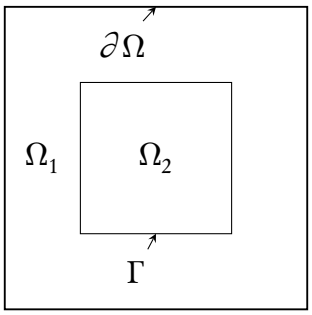


Figure 2: Geometry considered in Example 3.

Γ in the direction outward from Ω_i, which means that $n_2 = -n_1$. See Figure 2. We consider the partitioned Poisson problem of finding functions u_1 and u_2 so that

$$\begin{aligned} -\Delta u_1 &= f_1 \quad \text{in } \Omega_1, \\ -\Delta u_2 &= f_2 \quad \text{in } \Omega_2, \\ u_1 &= 0 \quad \text{on } \partial\Omega, \\ \nabla u_1 \cdot n_1 + \nabla u_2 \cdot n_2 &= 0 \quad \text{on } \Gamma, \\ u_1 &= u_2, \quad \text{on } \Gamma, \end{aligned} \tag{34}$$

where the right hand sides f_1 and f_2 are assumed given. The condition for u_1 on $\partial\Omega$ is simply homogeneous Dirichlet conditions, and is present to yield a well-posed problem. The interface conditions posed on Γ in (34) can be viewed as imposing continuity. A variational formulation of (34) states that the solution (u_1, u_2) minimizes the energy functional given by

$$F(u_1, u_2) = \frac{1}{2}||\nabla u_1||_{\Omega_1} + \frac{1}{2}||\nabla u_2||_{\Omega_2} - (f_1, u_1)_{\Omega_1} - (f_2, u_2)_{\Omega_2},$$

under the constraint that $T_1 u_1 - T_2 u_2 = 0$, where T_i, for $i = 1, 2$, are the trace operators onto Γ from Ω_i. Further, the subscript Ω_i denotes restriction of the L^2-norm and -inner product to Ω_i. The minimum is sought for $(u_1, u_2) \in H^1_{\partial\Omega}(\Omega_1) \times H^1(\Omega_2)$, where $H^1_{\partial\Omega}(\Omega_1) = \left\{v \in H^1(\Omega_1) : v\big|_{\partial\Omega} = 0\right\}$.

We now aim to enforce the trace constraint weakly through the use of a Lagrange multiplier. To that end, we note that T_i extends to a continuous, surjective operator $H^1(\Omega_i) \to H^{\frac{1}{2}}(\Gamma)$, cf. [57, Thm. 8.3]. This means that $(T_1 u_1 - T_2 u_2, \mu)_\Gamma = 0$ for every $\mu \in H^{-\frac{1}{2}}(\Gamma)$. Thus, by introducing a Lagrange multiplier $\lambda \in H^{-\frac{1}{2}}(\Gamma)$, we seek a critical point $(u_1, u_2, \lambda) \in W := H^1_{\partial\Omega}(\Omega_1) \times H^1(\Omega_2) \times H^{-\frac{1}{2}}(\Gamma)$ to the auxiliary energy functional

$$\tilde{F}(u_1, u_2, \lambda) = F(u_1, u_2) + (T_1 u_1 - T_2 u_2, \lambda)_\Gamma.$$

Differentiating $\tilde{F}$ with respect to each of its arguments, we find that (u_1, u_2, λ) satisfies

$$\mathscr{A}\begin{pmatrix} u_1 \\ u_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} -\Delta_{\Omega_1} & 0 & T_1^* \\ 0 & -\Delta_{\Omega_2} & -T_2^* \\ T_1 & -T_2 & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ 0 \end{pmatrix}. \tag{35}$$

We see that (35) exhibits a saddle point structure like (19). Hence, in proving that $\mathscr{A} : W \to W^*$ is a homeomorphism it is sufficient to verify the Brezzi conditions (20) and (21). In par-

ticular, the coercivity of $-\Delta_{\Omega_i}$ over the kernel of T_i follows from the Poincaré inequality

$$\|v\|^2_{\Omega_i} \leq C_i \|\nabla v\|^2_{\Omega_i} = C_i \left((-\Delta_{\Omega_i})v, v\right)_{\Omega_i}$$

whenever $v \in H^1(\Omega_i)$ satisfies $T_i v = 0$. The inf-sup condition (21) follows from the surjectivity of the trace operators T_i.

The canonical preconditioner $\mathscr{B} : W^* \to W$ is then given by

$$\mathscr{B} = \begin{pmatrix} (I - \Delta_{\Omega_1})^{-1} & 0 & 0 \\ 0 & (I - \Delta_{\Omega_2})^{-1} & 0 \\ 0 & 0 & (I - \Delta_\Gamma)^{\frac{1}{2}} \end{pmatrix}.$$

Discrete operators to replace the first two blocks of $\mathscr{B}$ are well-studied, while preconditioners for the third block were discussed in the previous section. However, note that neither $(I - \Delta_\Gamma)^{\frac{1}{2}}$, nor its inverse, appear in the definition of $\mathscr{A}$. We recall that a construction of fractional powers of operators required the solving of an computationally expensive eigenvalue problem. As such, one would in general like the preconditioner replacing the third block of $\mathscr{B}$ to not involve any evaluation of $(I - \Delta_\Gamma)^{\pm\frac{1}{2}}$.

We end the example on the following remark. If in (34), we introduce the auxiliary variables $\sigma_i = -\nabla u_i$, for $i = 1, 2$. Instead of the continuity of potential condition on the interface in (34), we impose that $\sigma_1 \cdot n_1 + \sigma_2 \cdot n_2 = 0$ using a Lagrange multiplier. Then, the Lagrange multiplier $\lambda \in H^{\frac{1}{2}}(\Gamma)$, and the canonical preconditioner for the resulting saddle point system will contain a block of the form $(I - \Delta_\Gamma)^{-\frac{1}{2}}$. Cf. for instance [85]. That is, even for relatively simple trace constraint problems there is a need for preconditioners for H^s both with positive- and negative values of s.

We remark that in the above, we considered solving (34) monolithically, i.e. we solve for both subproblems on Ω_1 and Ω_2 simultaneously. This is not the only available solution strategy. In particular, sequential solvers would solve first for one subproblem, followed by the solution of the other, using the first solution as boundary data. This process can then be repeated until the solution has converged within a certain error tolerance. Solvers of this type are considered for the Darcy-Stokes problem in i.a. [25, 28, 88]. This approach circumvents the need for fractional Sobolev spaces. However, following [92], successive solvers can be re-interpreted as block-preconditioned Richardson's iteration, which suggests that monolithic solvers are more efficient.

Example 3, though simple, hopefully motivates the appearence of fractional Sobolev spaces in more complicated systems of PDEs. For instance, the above example can be generalized to non-overlapping domain decomposition methods, [46, 60], or imposing Dirichlet boundary conditions weakly, [13]. Similar approaches can be found in multiphysics problem, e.g. the coupled Darcy-Stokes system in [52], or Stokes flow coupled with Biot's consolidation model in [4].

We can also generalize Example 3 to consider an additional PDE with domain on Γ, as opposed to only continuity conditions. Moreover, Γ can have codimension two. In the above problem, this translates to Γ being a point, while in 3D, Γ would be a curve embedded in Ω. Such problems were considered in [48, 49], and are particularly relevant in biomedical applications (cf. e.g. [16, 26, 27]).

Overview of papers

Paper I T. BÆRLAND, J. J. LEE, K.-A. MARDAL, AND R. WINTHER,
Weakly Imposed Symmetry and Robust Preconditioners for Biot's Consolidation Model,
De Gruyter Computational Methods in Applied Mathematics 2017; 17(3):377-396.

Paper II T. BÆRLAND, M. KUCHTA, AND K.-A. MARDAL,
Multigrid Methods for Discrete Fractional Sobolev Spaces,
SIAM Journal on Scientific Computing, (2018). Submitted.

Paper III T. BÆRLAND,
An Auxiliary Space Preconditioner for Fractional Laplacian of Negative Order,
In preparation.

Paper IV T. BÆRLAND, K.-A. MARDAL, AND T. THOMPSON,
Uniform Preconditioners for the Mixed Darcy Problem,
Numerische Mathematik, (2018). Submitted.

Paper I

In this paper we consider the construction of robust preconditioners for finite element approximations of Biot's consolidation model (see (22)). In particular we introduce the stress tensor $\sigma := 2\mu\varepsilon(\mathbf{u}) + \lambda \operatorname{tr}\varepsilon(\mathbf{u})\mathbb{I} - \alpha p_F \mathbb{I}$, which we note is a symmetric tensor field. We note that introducing the stress as a primary variable has potential benefits in problems where the Biot model is coupled with other physics and there are trace conditions for the stress on the interface. A problem of this type was, for instance, considered in [4]. The discretization of symmetric tensor fields using finite element spaces has historically been a difficult problem, and the first stable element in three dimensions was proposed in [2]. This element uses quartic shape functions and has 162 local degrees of freedom. To avoid this problem, we follow the method in [10], and impose the symmetry of σ weakly by introducing a Lagrange multiplier, γ.

In coefficient matrix form, the full system then reads

$$\mathscr{A}\begin{pmatrix}\sigma\\ p_F\\ \mathbf{u}\\ \gamma\end{pmatrix} = \begin{pmatrix} A & K^* & -\nabla & \operatorname{skw}^* \\ K & B - \operatorname{div}(\kappa\nabla) & 0 & 0 \\ \operatorname{div} & 0 & 0 & 0 \\ \operatorname{skw} & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix}\sigma\\ p_F\\ \mathbf{u}\\ \gamma\end{pmatrix} = \begin{pmatrix}0\\ g\\ -\mathbf{f}\\ 0\end{pmatrix}. \tag{36}$$

Here, A is the compliance tensor, K is defined by $K\tau = \frac{\alpha}{2\mu+n\lambda}\operatorname{tr}\tau$ for a tensor field τ, while $B = \frac{\alpha^2}{\lambda}\left(1 + \frac{n\lambda}{2\mu+n\lambda}\right)$, and skw is the skew-operator on tensor fields.

We establish suitable norms in which $\mathscr{A}$, defined by (36), is a homeomorphism. The norms depend on the boundary conditions considered, but lead to bounds on $\mathscr{A}$ and its inverse that are independent of all physical parameters, in particular in the incompressible limit when λ is large. We then propose a finite element discretization of (36) based on stable discretizations of the elasticity- and reaction-diffusion subproblems. With an argument similar to the continuous case, we prove that this discretization is stable in all physical parameters as well as the discretization parameter.

Motivated by the preceeding analyses, we then propose block-diagonal preconditioners to efficiently solve (36) numerically. An advantage of this work is that each block can be replaced by already well-studied preconditioners. Specifically, one can use properly scaled $H(\operatorname{div})$ preconditioners for σ, H^1 preconditioners for p_F, and L^2 preconditioners for $\mathbf{u}$ and

γ. One caveat is the fully clamped case when $\mathbf{u} = 0$ is imposed on the entire boundary of Ω. In this case, we show that an efficient preconditioner for σ can be constructed from a scaled $H(\mathrm{div})$ preconditioner together with correction term on the one-dimensional subspace spanned by the constant $\mathbb{I}$-tensor fields.

The paper is concluded by a series of numerical experiments validating the theoretical findings.

Paper II

In this paper we consider the construction of preconditioners for the problem

$$(-\Delta)^s u = f,$$

with $s \in [-1, 1]$. We observe that when $s \geq 0$, $(-\Delta)^s$ behaves spectrally similar to $(-\Delta)$ in that high frequency eigenmodes correspond to high eigenvalues, which motivates the use of multigrid techniques to construct our preconditioner.

Our main motivational application are trace constraint- and multiphysics problems as discussed earlier in the introduction. As such, we assume that no discretization of $(-\Delta)^s$ is readily available. We therefore develop an additive multigrid operator, [22], for the case $s \geq 0$. Another issue is the non-nestedness of the bilinear forms on each grid level, and so we resort to the framework laid out in [23], in which one key assumption that needs verification is

$$\left(A_h^s v, v\right) \leq (A_H^s v, v), \quad \forall v \in V_H.$$

Here, A_h^s and A_H^s are spectral discretizations of $(-\Delta)^s$ over finite-dimensional subspaces $V_H \subset V_h \subset H_0^1$. In our context, the discrete spaces are the standard continuous, piecewise linear finite element spaces, and the above inequality is proved using Jensen's operator inequality, [36].

Further, we propose additive Schwarz operators based on overlapping domain decomposition as smoothers. The additive multigrid preconditioner, B_h^s, is then shown to be spectrally equivalent to A_h^{-s} using standard techniques for $s = 0$ and $s = 1$, from which the intermediate cases follows, again by use of Jensen's inequality. The proven spectral equivalence depends on the number of grid levels used, but otherwise indepedent of the discretization parameter. A great advantage of the proposed preconditioners is that they can be implemented by only minor changes to standard multigrid implementation, and the smoothers are natural interpolations of Jacobi smoother for $s = 0$ and $s = 1$.

For the case $s < 0$, the spectral similarity between $(-\Delta)^s$ and $(-\Delta)$ breaks down. In this case the highly oscillatory eigenmodes of $(-\Delta)^s$ correspond to relatively low eigenvalues, and vice versa. As such, hoping that the same additive multigrid framework we used for $s > 0$ will work here seems naive at best. Instead, we propose a preconditioner based on the multiplicative decomposition

$$A_h^{-s} = A_h^{-\frac{1+s}{2}} A_h A_h^{-\frac{1+s}{2}}.$$

The left- and right factors of the above factorization are then replaced by $B_h^{\frac{1+s}{2}}$, where we see that $\frac{1+s}{2} \geq 0$. The appropriate spectral equivalence is then shown under an additional regularity assumption on $B_h^{\frac{1+s}{2}}$. The resulting condition number will then behave like the square of the condition number for $B_h^{\frac{1+s}{2}} A_h^{\frac{1+s}{2}}$.

The paper is concluded by a series of numerical experiments. In the first set of experiments we validate the theoretical results established for when $s \geq 0$ and provide some backing for the more heuristical arguments made for $s < 0$. Lastly, we provide numerical tests showing that

the proposed preconditioners are efficient as part of solution algorithms for a trace constrained problem.

Future work

In paper II we only consider geometrical multigrid. That is, we assume a hierarchy of nested meshes are given. However, the analysis suggests that whenever the subspace hierarchy used is effective for both $s = 0$ and $s = 1$, the same should hold true for $s \in (0,1)$. As such, it would be interesting to consider subspaces used in popular algebraic multigrid procedures to precondition the fractional Laplacian.

By a similar reasoning, we expect the additive multigrid framework to be effective when considering discontinuous finite element spaces. That is, multigrid preconditioners have been developed for discontinuous finite element discretizations of the Laplace operator, as well as the identity operator. Thus, it is reasonable to consider similar subspace decompositions when constructing preconditioners for nonconforming discretizations of $(-\Delta)^s$, $s \in (0,1)$.

Paper III

One of the main disadvantages of the work in Paper II was the unsatisfactory theoretical grounding of proposed preconditioners for when $s < 0$. Moreover, the condition numbers, although bounded in h, could become quite large. In Paper III we propose a novel family of auxiliary space preconditioners to precondition $(-\Delta)^s$ when $s \in [-1,0]$.

The construction is based on the observation that the canonical preconditioner, $(-\Delta)^{-s}$, is the positive fractional power of a differential operator. Suppose then that our preconditioner consists of applications of ordinary differential operators. Then a correction is needed to compensate for the "overshoot" in fractionality, and this correction should then behave like a preconditioner for a fractional operator of positive fractionality. In particular, in this work we consider the gradient operator, ∇, which leads to the preconditioner

$$\mathscr{B}^s = \nabla^* \Lambda^{-(1+s)} \nabla, \tag{37}$$

where Λ is the operator realizing the $H(\mathrm{div})$ inner product,

$$(\Lambda \mathbf{v}, \mathbf{w}) = (\mathbf{v}, \mathbf{w}) + (\mathrm{div}\,\mathbf{v}, \mathrm{div}\,\mathbf{w}), \quad \mathbf{v}, \mathbf{w} \in H(\mathrm{div}).$$

What is worth noting in (37) is that since $1 + s \geq 0$, and multigrid operators to approximate Λ^{-1} are popular in the literature, we may hope that the middle middle factor in (37) can be replaced by an additive multigrid operator, similar to the ones studied in Paper II.

In Paper III, we proceed by proving that $\mathscr{B}^s$ in the continuous setting is indeed spectrally equivalent to $(-\Delta)^{-s}$. This is proven by fairly well-known bounds on the gradient operator, together with standard interpolation theory. We then extend these results to the discrete setting, where H^s is discretized by discontinuous finite element spaces and $H(\mathrm{div})$ is discretized by either Raviart-Thomas- (RT) or Brezzi-Douglas-Marini (BDM) spaces. The arguments made in the discrete setting are fairly similar to the continuous case, but we make abundant use of Jensen's inequality to get sharper bounds.

The remainder of the paper is then devoted to analyzing additive multigrid operators to precondition fractional powers of Λ. The analysis follows closely the analysis in Paper II, but we need to make assumptions on some two-grid error bounds for Λ to complete the analysis. Proving these assumption falls beyond the scope of the current work, but we provide suggestions of techniques to proving them, as well as substantiating their veracity.

As with the preceeding papers, Paper III is concluded with a series of numerical experiments to validate our theoretical results. The numerical tests also suggest that the additive

multigrid preconditioner is effective for fractional $H(\mathrm{div})$ problems and as a replacement of the middle factor in (37).

Future work

An obvious continuation of this work is to get a firmer theoretical grounding for the additive multigrid operator fractional $H(\mathrm{div})$ operators. As outlined in the paper, techniques used in [18] for elliptic operators might prove fruitful in this regard.

Another shortcoming of the current work is that we only consider discontinuous finite element spaces to discretize $(-\Delta)^s$. This was mainly due to the fact that the discrete gradient operator is well-known to be injective into the considered discrete $H(\mathrm{div})$ spaces. The same does not quite hold true for continuous finite element spaces. To remedy this, there are at least two options available. As one option, you can increase the degree on the $H(\mathrm{div})$ discretization so that ∇ becomes injective. This means that size of fractional $H(\mathrm{div})$ preconditioner becomes large, which might not be desirable in implementation. The second option is to discretize $H(\mathrm{div})$ using $H(\mathbf{curl})$-conforming finite element spaces for which injectivity of the gradient operator is known. Then one has to come up with a suitable discrete operator to represent Λ, and design efficient preconditioners for fractional powers of it. This falls outside the framework we have put forth in Paper III, as we did not consider non-conforming methods.

As an additional direction for generalization, one could consider the full de Rham complex. The auxiliary space preconditioner constructed in Paper III relies on the inf-sup stability of the divergence operator from the RT- or BDM spaces onto the space of discontinuous, piecewise polynomials. It is well-known that this property is a special case of cohomology preserving discretizations of the de Rham complex, [9]. As a consequence, we can expect similar auxiliary space preconditioners to prove effective in preconditioning e.g. Λ^s when $s \in [-1, 0]$. Granted though that applications of this to physical models seems more scarce.

Paper IV

This paper is a note on robust preconditioners for the mixed formulation of the Darcy problem. That is, we consider the problem of finding $\mathbf{u}$ and p so that

$$\begin{aligned} \frac{1}{K}\mathbf{u} - \nabla p &= \mathbf{f}, \quad \text{in } \Omega \\ \mathrm{div}\,\mathbf{u} &= g, \quad \text{in } \Omega. \end{aligned} \tag{38}$$

paired with suitable boundary conditions. System (38) is one of the simplest models for pressure-driven fluid flow through a porous medium. Here, p plays the role of pressure, while $\mathbf{u}$ is the fluid flux. The permeability is denoted by K, and models the resistance in the medium to the flow.

A variational formulation of (38) yields an operator characterizing the left-hand side given by

$$\mathscr{A} = \begin{pmatrix} K^{-1} & -\nabla \\ \mathrm{div} & 0 \end{pmatrix}.$$

In this note, we introduce the spaces

$$\mathbf{V} = K^{-1/2}\mathbf{L}^2 \cap H(\mathrm{div}), \quad \text{and} \quad Q = L^2 + K^{1/2}H^1$$

with norms given by

$$\|\mathbf{v}\|_{\mathbf{V}}^2 = K^{-1}\|\mathbf{v}\|^2 + \|\operatorname{div}\mathbf{v}\|^2, \quad \mathbf{v} \in H(\operatorname{div})$$
$$\|q\|_Q^2 = \inf_{\phi \in H^1} \left(\|q-\phi\|^2 + K\|\nabla\phi\|^2\right), \quad q \in L^2.$$

See [15] for further details on sum- and interpolation spaces. We further show that $\mathscr{A} : \mathbf{V} \times Q \to \mathbf{V}^* \times Q^*$ is a homeomorphism with bounds independent of $K \in (0, \infty)$. This suggests that the canonical preconditioner $\mathscr{B}$ takes the forms

$$\mathscr{B} = \begin{pmatrix} \left(K^{-1}I - \nabla\operatorname{div}\right)^{-1} & 0 \\ 0 & I + (-K\Delta)^{-1} \end{pmatrix}. \tag{39}$$

We continue by considering the discretization of (38) using mixed finite elements. In particular, we consider discretizing $\mathbf{V}$ using either Raviart-Thomas- or Brezzi-Douglas-Marini elements, and piecewise polynomial, discontinuous elements to discretize Q. The subsequent analysis is analogous to the continuous case, but we introduce a discrete gradient operator, ∇_h, which facilitates the definition of a discrete Q-norm. We establish that the discrete operator $\mathscr{A}_h$ satisfies the appropriate K-independent bounds.

The note is concluded on some numerical tests validating the proposed preconditioner structure.

A point worth making at this point is that in this work our analysis is concerned with the variational formulation of (38) seeking $\mathbf{u} \in H(\operatorname{div})$ and $p \in L^2$. However, (38) is also well-posed when considering $\mathbf{u} \in \mathbf{L}^2$ and $p \in H^1$. See examples 3.2 and 3.3 in [63]. Of course, also in this formulation the norms must be modified to achieve the desired K-robustness. This approach leads to the block-diagonal preconditioner

$$\begin{pmatrix} K & 0 \\ 0 & (-K\Delta)^{-1} \end{pmatrix}.$$

The second block in the above can be viewed as the inverse of the Schur complement of $\mathscr{A}$, and as a consequence, discrete preconditioners based on this approach will be robust in K, [65]. Comparing this with (39) we see that in either approach, we cannot do away with the Schur complement.

Bibliography

[1] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev spaces*, vol. 140 of Pure and Applied Mathematics (Amsterdam), Elsevier/Academic Press, Amsterdam, second ed., 2003.

[2] S. ADAMS AND B. COCKBURN, *A mixed finite element method for elasticity in three dimensions*, J. Sci. Comput., 25 (2005), pp. 515–521.

[3] J. H. ADLER, F. J. GASPAR, X. HU, C. RODRIGO, AND L. T. ZIKATANOV, *Robust block preconditioners for Biot's model*, arXiv preprint arXiv:1705.08842, (2017).

[4] I. AMBARTSUMYAN, E. KHATTATOV, I. YOTOV, AND P. ZUNINO, *A Lagrange multiplier method for a Stokes-Biot fluid-poroelastic structure interaction model*, Numer. Math., 140 (2018), pp. 513–553.

[5] M. ARIOLI AND D. LOGHIN, *Discrete interpolation norms with applications*, SIAM J. Numer. Anal., 47 (2009), pp. 2924–2951.

[6] D. N. ARNOLD, R. S. FALK, AND R. WINTHER, *Preconditioning discrete approximations of the Reissner-Mindlin plate model*, RAIRO Modél. Math. Anal. Numér., 31 (1997), pp. 517–557.

[7] ——, *Preconditioning in $H(\mathrm{div})$ and applications*, Math. Comp., 66 (1997), pp. 957–984.

[8] ——, *Multigrid in $H(\mathrm{div})$ and $H(\mathrm{curl})$*, Numer. Math., 85 (2000), pp. 197–217.

[9] ——, *Finite element exterior calculus, homological techniques, and applications*, Acta Numer., 15 (2006), pp. 1–155.

[10] ——, *Mixed finite element methods for linear elasticity with weakly imposed symmetry*, Math. Comp., 76 (2007), pp. 1699–1723.

[11] ——, *Finite element exterior calculus: from Hodge theory to numerical stability*, Bull. Amer. Math. Soc. (N.S.), 47 (2010), pp. 281–354.

[12] W. E. ARNOLDI, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[13] I. BABUŠKA, *The finite element method with Lagrangian multipliers*, Numer. Math., 20 (1972/73), pp. 179–192.

[14] T. BÆRLAND, M. KUCHTA, AND K.-A. MARDAL, *Multigrid methods for discrete fractional sobolev spaces*, arXiv preprint arXiv:1806.00222, (2018).

[15] J. BERGH AND J. LÖFSTRÖM, *Interpolation spaces. An introduction*, Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[16] S. BERTOLUZZA, V. CHABANNES, C. PRUD'HOMME, AND M. SZOPOS, *Boundary conditions involving pressure for the stokes problem and applications in computational hemodynamics*, Computer Methods in Applied Mechanics and Engineering, 322 (2017), pp. 58–80.

[17] M. A. BIOT, *General theory of three-dimensional consolidation*, Journal of applied physics, 12 (1941), pp. 155–164.

[18] A. BONITO AND J. E. PASCIAK, *Numerical approximation of fractional powers of elliptic operators*, Math. Comp., 84 (2015), pp. 2083–2110.

[19] J. H. BRAMBLE, *On the development of multigrid methods and their analysis*, in Mathematics of Computation 1943–1993: a half-century of computational mathematics (Van-

couver, BC, 1993), vol. 48 of Proc. Sympos. Appl. Math., Amer. Math. Soc., Providence, RI, 1994, pp. 5–19.
[20] J. H. BRAMBLE, Z. LEYK, AND J. E. PASCIAK, *The analysis of multigrid algorithms for pseudodifferential operators of order minus one*, Math. Comp., 63 (1994), pp. 461–478.
[21] J. H. BRAMBLE, J. E. PASCIAK, AND P. S. VASSILEVSKI, *Computational scales of Sobolev norms with application to preconditioning*, Math. Comp., 69 (2000), pp. 463–480.
[22] J. H. BRAMBLE, J. E. PASCIAK, AND J. XU, *Parallel multilevel preconditioners*, Math. Comp., 55 (1990), pp. 1–22.
[23] ——, *The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms*, Math. Comp., 56 (1991), pp. 1–34.
[24] J. H. BRAMBLE AND X. ZHANG, *The analysis of multigrid methods*, in Handbook of numerical analysis, Vol. VII, Handb. Numer. Anal., VII, North-Holland, Amsterdam, 2000, pp. 173–415.
[25] Y. CAO, M. GUNZBURGER, X. HE, AND X. WANG, *Robin-Robin domain decomposition methods for the steady-state Stokes-Darcy system with the Beavers-Joseph interface condition*, Numer. Math., 117 (2011), pp. 601–629.
[26] L. CATTANEO AND P. ZUNINO, *Computational models for fluid exchange between microcirculation and tissue interstitium.*, Networks & Heterogeneous Media, 9 (2014).
[27] C. D'ANGELO AND A. QUARTERONI, *On the coupling of 1D and 3D diffusion-reaction equations. Application to tissue perfusion problems*, Math. Models Methods Appl. Sci., 18 (2008), pp. 1481–1504.
[28] M. DISCACCIATI, A. QUARTERONI, AND A. VALLI, *Robin-Robin domain decomposition methods for the Stokes-Darcy coupling*, SIAM J. Numer. Anal., 45 (2007), pp. 1246–1268.
[29] M. FERRONATO, L. BERGAMASCHI, AND G. GAMBOLATI, *Performance and robustness of block constraint preconditioners in finite element coupled consolidation problems*, Internat. J. Numer. Methods Engrg., 81 (2010), pp. 381–402.
[30] S. A. FUNKEN AND E. P. STEPHAN, *The bpx preconditioner for the single layer potential operator*, Applicable Analysis, 67 (1997), pp. 327–340.
[31] J. GOPALAKRISHNAN AND G. KANSCHAT, *A multilevel discontinuous Galerkin method*, Numer. Math., 95 (2003), pp. 527–550.
[32] A. GÜNNEL, R. HERZOG, AND E. SACHS, *A note on preconditioners and scalar products in Krylov subspace methods for self-adjoint problems in Hilbert space*, Electron. Trans. Numer. Anal., 41 (2014), pp. 13–20.
[33] W. HACKBUSCH, *Iterative solution of large sparse systems of equations*, vol. 95 of Applied Mathematical Sciences, Springer, [Cham], second ed., 2016.
[34] J. B. HAGA, H. OSNES, AND H. P. LANGTANGEN, *On the causes of pressure oscillations in low-permeable and low-compressible porous media*, International Journal for Numerical and Analytical Methods in Geomechanics, 36 (2012), pp. 1507–1522.
[35] ——, *A parallel block preconditioner for large-scale poroelasticity with highly heterogeneous material parameters*, Comput. Geosci., 16 (2012), pp. 723–734.
[36] F. HANSEN AND G. K. PEDERSEN, *Jensen's operator inequality*, Bull. London Math. Soc., 35 (2003), pp. 553–564.
[37] S. HARIZANOV, R. LAZAROV, S. MARGENOV, P. MARINOV, AND Y. VUTOV, *Optimal solvers for linear systems with fractional powers of sparse spd matrices*, Numer. Linear Algebra Appl., (2018), p. e2167.
[38] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953).

[39] R. HIPTMAIR, *Multigrid method for* **H**(div) *in three dimensions*, Electron. Trans. Numer. Anal., 6 (1997), pp. 133–152. Special issue on multilevel methods (Copper Mountain, CO, 1997).
[40] ——, *Operator preconditioning*, Comput. Math. Appl., 52 (2006), pp. 699–706.
[41] R. HIPTMAIR AND J. XU, *Nodal auxiliary space preconditioning in* **H**(**curl**) *and* **H**(div) *spaces*, SIAM J. Numer. Anal., 45 (2007), pp. 2483–2509.
[42] Q. HONG AND J. KRAUS, *Uniformly stable discontinuous Galerkin discretization and robust iterative solution methods for the Brinkman problem*, SIAM J. Numer. Anal., 54 (2016), pp. 2750–2774.
[43] ——, *Parameter-robust stability of classical three-field formulation of Biot's consolidation model*, Electron. Trans. Numer. Anal., 48 (2018), pp. 202–226.
[44] Q. HONG, J. KRAUS, J. XU, AND L. ZIKATANOV, *A robust multigrid method for discontinuous Galerkin discretizations of Stokes and linear elasticity equations*, Numer. Math., 132 (2016), pp. 23–49.
[45] J. KORSAWE AND G. STARKE, *A least-squares mixed finite element method for Biot's consolidation problem in porous media*, SIAM J. Numer. Anal., 43 (2005), pp. 318–339.
[46] M. KOČVARA, D. LOGHIN, AND J. TURNER, *Constraint interface preconditioning for topology optimization problems*, SIAM J. Sci. Comput., 38 (2016), pp. A128–A145.
[47] E. KREYSZIG, *Introductory functional analysis with applications*, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1989.
[48] M. KUCHTA, K.-A. MARDAL, AND M. MORTENSEN, *On preconditioning saddle point systems with trace constraints coupling 3d and 1d domains–applications to matching and nonmatching fem discretizations*, arXiv preprint arXiv:1612.03574, (2016).
[49] M. KUCHTA, M. NORDAAS, J. C. G. VERSCHAEVE, M. MORTENSEN, AND K.-A. MARDAL, *Preconditioners for saddle point systems with trace constraints coupling 2d and 1d domains*, SIAM J. Sci. Comput., 38 (2016), pp. B962–B987.
[50] M. KWAŚNICKI, *Ten equivalent definitions of the fractional Laplace operator*, Fract. Calc. Appl. Anal., 20 (2017), pp. 7–51.
[51] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.
[52] W. J. LAYTON, F. SCHIEWECK, AND I. YOTOV, *Coupling fluid flow with porous media flow*, SIAM J. Numer. Anal., 40 (2002), pp. 2195–2218 (2003).
[53] J. J. LEE, K.-A. MARDAL, AND R. WINTHER, *Parameter-robust discretization and preconditioning of Biot's consolidation model*, SIAM J. Sci. Comput., 39 (2017), pp. A1–A24.
[54] X. S. LI AND J. W. DEMMEL, *Superlu_dist: A scalable distributed-memory sparse direct solver for unsymmetric linear systems*, ACM Transactions on Mathematical Software (TOMS), 29 (2003), pp. 110–140.
[55] J. LIESEN AND Z. STRAKOŠ, *Krylov subspace methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013. Principles and analysis.
[56] J. LIESEN AND P. TICHÝ, *Convergence analysis of Krylov subspace methods*, GAMM Mitt. Ges. Angew. Math. Mech., 27 (2004), pp. 153–173 (2005).
[57] J.-L. LIONS AND E. MAGENES, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
[58] A. LISCHKE, G. PANG, M. GULIAN, F. SONG, C. GLUSA, X. ZHENG, Z. MAO, W. CAI, M. M. MEERSCHAERT, M. AINSWORTH, ET AL., *What is the fractional laplacian?*, arXiv preprint arXiv:1801.09767, (2018).
[59] D. LOGHIN AND A. J. WATHEN, *Analysis of preconditioners for saddle-point problems*, SIAM J. Sci. Comput., 25 (2004), pp. 2029–2049.

[60] F. Magoulès and F.-X. Roux, *Lagrangian formulation of domain decomposition methods: A unified theory*, Applied Mathematical Modelling, 30 (2006), pp. 593–615.
[61] J. Málek and Z. Strakoš, *Preconditioning and the conjugate gradient method in the context of solving PDEs*, vol. 1 of SIAM Spotlights, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.
[62] J. Mandel, S. McCormick, and R. Bank, *Variational multigrid theory*, in Multigrid methods, vol. 3 of Frontiers Appl. Math., SIAM, Philadelphia, PA, 1987, pp. 131–177.
[63] K.-A. Mardal and R. Winther, *Preconditioning discretizations of systems of partial differential equations*, Numer. Linear Algebra Appl., 18 (2011), pp. 1–40.
[64] M. A. Murad, V. Thomée, and A. F. D. Loula, *Asymptotic behavior of semidiscrete finite-element approximations of Biot's consolidation problem*, SIAM J. Numer. Anal., 33 (1996), pp. 1065–1083.
[65] 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.
[66] P. Oswald, *Multilevel norms for $H^{-1/2}$*, Computing, 61 (1998), pp. 235–255.
[67] C. C. Paige and M. A. Saunders, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
[68] W. K. Pao, R. W. Lewis, and I. Masters, *A fully coupled hydro-thermo-poro-mechanical model for black oil reservoir simulation*, Int. J. Numer. Anal. Methods Geomech., 25 (2001), pp. 1229–1256.
[69] P. J. Phillips and M. F. Wheeler, *A coupling of mixed and continuous Galerkin finite element methods for poroelasticity. I. The continuous in time case*, Comput. Geosci., 11 (2007), pp. 131–144.
[70] ——, *A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity*, Comput. Geosci., 12 (2008), pp. 417–435.
[71] K. Phoon, K. Toh, S. Chan, and F. Lee, *An efficient diagonal preconditioner for finite element solution of biot's consolidation equations*, International Journal for Numerical Methods in Engineering, 55 (2002), pp. 377–400.
[72] S. Rhebergen, G. N. Wells, R. F. Katz, and A. J. Wathen, *Analysis of block preconditioners for models of coupled magma/mantle dynamics*, SIAM J. Sci. Comput., 36 (2014), pp. A1960–A1977.
[73] S. Rhebergen, G. N. Wells, A. J. Wathen, and R. F. Katz, *Three-field block preconditioners for models of coupled magma/mantle dynamics*, SIAM J. Sci. Comput., 37 (2015), pp. A2270–A2294.
[74] ——, *Three-field block preconditioners for models of coupled magma/mantle dynamics*, SIAM J. Sci. Comput., 37 (2015), pp. A2270–A2294.
[75] T. Roose, P. A. Netti, L. L. Munn, Y. Boucher, and R. K. Jain, *Solid stress generated by spheroid growth estimated using a linear poroelasticity model*, Microvascular research, 66 (2003), pp. 204–212.
[76] Y. Saad, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.
[77] Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[78] 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.
[79] R. Stevenson and R. van Venetië, *Optimal preconditioning for problems of negative order*, arXiv preprint arXiv:1803.05226, (2018).

[80] K. H. Støverud, M. Alnæs, H. P. Langtangen, V. Haughton, and K.-A. Mardal, *Poro-elastic modeling of syringomyelia–a systematic study of the effects of pia mater, central canal, median fissure, white and gray matter on pressure wave propagation and fluid movement within the cervical spinal cord*, Computer methods in biomechanics and biomedical engineering, 19 (2016), pp. 686–698.
[81] K. Stüben, *A review of algebraic multigrid*, J. Comput. Appl. Math., 128 (2001), pp. 281–309. Numerical analysis 2000, Vol. VII, Partial differential equations.
[82] P. Teatini, M. Ferronato, G. Gambolati, and M. Gonella, *Groundwater pumping and land subsidence in the emilia-romagna coastland, italy: Modeling the past occurrence and the future trend*, Water Resources Research, 42 (2006).
[83] L. N. Trefethen and D. Bau, III, *Numerical linear algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[84] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, Inc., San Diego, CA, 2001. With contributions by A. Brandt, P. Oswald and K. Stüben.
[85] A. Tveito, K. H. Jæger, M. Kuchta, K.-A. Mardal, and M. E. Rognes, *A cell-based framework for numerical modeling of electrical conduction in cardiac tissue*, Frontiers in Physics, 5 (2017), p. 48.
[86] P. N. Vabishchevich, *Numerically solving an equation for fractional powers of elliptic operators*, J. Comput. Phys., 282 (2015), pp. 289–302.
[87] H. A. van der Vorst, *Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–644.
[88] D. Vassilev, C. Wang, and I. Yotov, *Domain decomposition for coupled Stokes and Darcy flows*, Comput. Methods Appl. Mech. Engrg., 268 (2014), pp. 264–283.
[89] P. S. Vassilevski, *Multilevel block factorization preconditioners*, Springer, New York, 2008. Matrix-based analysis and algorithms for solving finite element equations.
[90] H. F. Wang, *Theory of linear poroelasticity with applications to geomechanics and hydrogeology*, Princeton University Press, 2017.
[91] A. J. Wathen, *Preconditioning*, Acta Numer., 24 (2015), pp. 329–376.
[92] J. A. White, N. Castelletto, and H. A. Tchelepi, *Block-partitioned solvers for coupled poromechanics: a unified framework*, Comput. Methods Appl. Mech. Engrg., 303 (2016), pp. 55–74.
[93] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613.
[94] J. Xu and L. Zikatanov, *Algebraic multigrid methods*, Acta Numer., 26 (2017), pp. 591–721.
[95] S.-Y. Yi, *Convergence analysis of a new mixed finite element method for Biot's consolidation model*, Numer. Methods Partial Differential Equations, 30 (2014), pp. 1189–1210.
[96] S. Yin, M. B. Dusseault, and L. Rothenburg, *Thermal reservoir modeling in petroleum geomechanics*, Int. J. Numer. Anal. Methods Geomech., 33 (2009), pp. 449–485.
[97] X. Zhang, *Multilevel Schwarz methods*, Numer. Math., 63 (1992), pp. 521–539.
[98] O. Zienkiewicz and T. Shiomi, *Dynamic behaviour of saturated porous media; the generalized biot formulation and its numerical solution*, International Journal for Numerical and Analytical Methods in Geomechanics, 8 (1984), pp. 71–96.
[99] W. Zulehner, *Analysis of iterative methods for saddle point problems: a unified approach*, Math. Comp., 71 (2002), pp. 479–505.
[100] ——, *Nonstandard norms and robust estimates for saddle point problems*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 536–560.

Paper I

Weakly Imposed Symmetry and Robust Preconditioners for Biot's Consolidation Model

T. Bærland, J. J. Lee, K.-A. Mardal, and R. Winther

De Gruyter Computational Methods in Applied Mathematics 2017; 17(3):377-396

WEAKLY IMPOSED SYMMETRY AND ROBUST PRECONDITIONERS FOR BIOT'S CONSOLIDATION MODEL

TRYGVE BÆRLAND[†], JEONGHUN J. LEE[‡], KENT-ANDRE MARDAL[†], AND RAGNAR WINTHER[†]

ABSTRACT. We discuss the construction of robust preconditioners for finite element approximations of Biot's consolidation model in poroelasticity. More precisely, we study finite element methods based on generalizations of the Hellinger-Reissner principle of linear elasticity, where the stress tensor is one of the unknowns. The Biot model has a number of applications in science, medicine, and engineering. A challenge in many of these applications is that the model parameters range over several orders of magnitude. Therefore, discretization procedures which are well behaved with respect to such variations are needed. The focus of the present paper will be on the construction of preconditioners, such that the preconditioned discrete systems are well-conditioned with respect to variations of the model parameters as well as refinements of the discretization. As a biproduct, we also obtain preconditioners for linear elasticity that are robust in the incompressible limit.

1. INTRODUCTION

The purpose of this paper is to discuss a family of finite element methods for Biot's consolidation model, with a focus on the construction of preconditioners for the discrete systems. The Biot model describes the deformation of an elastic porous medium saturated by a viscous fluid, leading to a system which describes the coupling between the elastic behaviour of the medium and the fluid flow. Therefore, the finite element systems will contain discrete versions of linear elasticity and porous medium flow as proper subsystems. The methods studied here are based on mixed finite element methods with weakly imposed symmetry for the elasticity part. In this respect, the methods presented here are generalizations of the methods for linear elasticity discussed in [4].

With Ω being an open domain in $\mathbb{R}^n$, the Biot model is a coupled system of partial differential equations of the form

$$\begin{aligned} -\mathbf{div}\,\mathscr{C}\boldsymbol{\epsilon}(\boldsymbol{u}) + \alpha \operatorname{grad} p &= f \quad \text{in } \Omega, \\ s_0 \dot{p} + \alpha \operatorname{div} \dot{\boldsymbol{u}} - \operatorname{div}(\boldsymbol{\kappa} \operatorname{grad} p) &= g \quad \text{in } \Omega, \end{aligned} \tag{1.1}$$

where the dots denote time derivation. The unknowns are the displacement of the structure $\boldsymbol{u}$, and the pore pressure p. The differential operator $\boldsymbol{\epsilon}$ is the symmetric gradient and $\mathscr{C}$ is the stiffness tensor which describes the strain-stress relation. The parameters s_0 and α are the so-called constrained specific storage coefficient and the Biot-Willis constant, respectively. Finally, $\boldsymbol{\kappa}$ is the hydraulic conductivity, determined by the permeability of the medium and the fluid viscosity, while f and g are given momentum- and mass sources, respectively.

In this paper, we will consider linear, isotropic elasticity, in which case the stiffness tensor is modelled as

$$\mathscr{C}\boldsymbol{\epsilon}(\boldsymbol{u}) = 2\mu\boldsymbol{\epsilon}(\boldsymbol{u}) + \lambda \operatorname{tr}\boldsymbol{\epsilon}(\boldsymbol{u})\mathbb{I} \equiv 2\mu\boldsymbol{\epsilon}(\boldsymbol{u}) + \lambda(\operatorname{div}\boldsymbol{u})\mathbb{I}, \tag{1.2}$$

[†]DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OSLO, BLINDERN, OSLO, 0316 NORWAY

[‡]THE INSTITUTE FOR COMPUTATIONAL AND ENGINEERING SCIENCES, UNIVERSITY OF TEXAS AUSTIN, 201 EAST 24TH ST, STOP C0200 POB 4.102, AUSTIN, TEXAS 78712, USA

E-mail addresses: trygveba@math.uio.no, jeonghun@ices.utexas.edu, kent-and@math.uio.no, rwinther@math.uio.no.

The research leading to these results has received funding the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339643. The work of Kent-Andre Mardal has also been supported by the Research Council of Norway through grant no. 209951 and a Center of Excellence grant awarded to the Center for Biomedical Computing at Simula Research Laboratory.

where μ, λ are the Lamé coefficients. We will allow the parameters μ, λ, and s_0 to be spatially varying, scalar valued functions, $\boldsymbol{\kappa}$ is a symmetric positive definite matrix-valued function, while $\alpha \in (0, 1]$ is constant. The well-posedness of system (1.1), with appropriate boundary and initial conditions, is discussed in [29].

The Biot system arises as a key model in many practical applications, such as in geoscience and in the modelling of soft tissues of the central nervous system. For many of these applications, the variations of the parameters will be quite large. For example, in geophysical applications the permeability may vary in the range from 10^{-9} to 10^{-21} m^2, [12, 34], while the Lamé coefficient λ can vary between 500 and 10^6 Pa in neurological applications [30, 32]. For a further discussion of relevant properties of the model parameters of the system (1.1), we refer to [20] and references given there.

Due to the wide range of physical applications of the Biot model, there is a need for numerical methods which behave robustly with respect to these variations of the model parameters. A number of finite element methods for the Biot model have previously been proposed in the literature. These studies include various primal methods [27, 33, 37], mixed methods [6, 22, 23, 36], and a discontinuous Galerkin method [10]. Combinations of these methods have also been proposed, see for example [18, 25, 26, 24, 35], while parameter-robust preconditioners are discussed in [5, 16, 28]. In fact, this was also the main theme of the paper [20], where the discretization is based on a standard H^1 formulation of the flow, combined with discretizing the elasticity part using stable mixed finite elements for the Stokes equation. A standard approach to obtain a locking free displacement method for linear elasticity, i.e., a method which behaves well for large Lamé parameters λ, is to introduce "solid pressure" as an additional unknown. This approach leads to a three-field formulation for the Biot model, where the unknowns are the displacement of the medium and the two pressures. The discussion in [20] shows that, in contrast to the situation for linear elasticity, this approach may not lead to a robust discretization of the Biot system. However, by introducing a new unknown, the so-called "total pressure", a robust discretization is obtained. In fact, the robustness of the discretization both with respect to the model parameters λ, $\boldsymbol{\kappa}$, and the discretization parameter h are obtained. Furthermore, robust preconditioners are constructed, i.e., preconditioners that behave uniformly well with respect to variations of the model parameters and refinements of the discretization.

The present paper can be seen as a continuation of [20], where the discretization of the elasticity part of the system is based on the mixed methods proposed in [4]. The mixed finite element methods studied in [4] are based on the Hellinger-Reissner variational principle of linear elasticity. An advantage of this approach is that robustness of the methods with respect to the Lamé parameter λ is more or less obtained automatically and that the stress tensor, which is of more interest in some applications, is computed directly. On the other hand, a difficulty of these methods is to construct stable finite element function spaces of exactly symmetric stress tensors. Therefore, the methods proposed in [4], based on weakly symmetric stresses, are employed. In the present paper, we generalize these methods to the Biot model. This leads to a four-field formulation where the unknowns are the stress tensor, the displacement of the structure, the pore pressure, and additionally a Lagrange multiplier which results from the weakly imposed symmetry constraint. The main purpose of the present paper is to discuss the properties of these finite element systems. In particular, as in [20], we will focus on the construction of robust preconditioners for the stationary systems obtained from a time discretization of the evolution problem (1.1).

This paper is organized as follows. In Section 2 we establish the notation that will be used throughout the paper and we give a brief description of the main strategy on how to construct preconditioners that are robust with respect to model parameters and mesh refinement. A proper weak formulation of a semidiscrete version of the Biot model, with four primary unknowns, is also stated in Section 2. Section 3 is devoted to parameter-robust stability results for both the continuous and discrete version of this problem, while more detailed discussions

of the construction of the corresponding preconditioners are given in Section 4. Finally, in Section 5 we present a few numerical experiments aimed at validating the theoretical results, followed by some concluding remarks in Section 6.

2. Preliminaries

We will denote by Ω a bounded domain in $\mathbb{R}^n$, with $n = 2$ or 3, and boundary $\partial\Omega$. The space of column n-vectors is written $\mathbb{V} = \mathbb{R}^n$, and $\mathbb{M}$ will denote the space of $n \times n$ real matrices. Then, $\mathbb{S}$ and $\mathbb{K}$ are the subspaces of symmetric- and skew symmetric matrices, respectively.

In the following, $H^k = H^k(\Omega)$ will denote the Sobolev spaces of functions on Ω with all derivates of order up to, and including, k in $L^2(\Omega)$, and its norm is denoted by $\|\cdot\|_k$. In addition, H_0^k will denote the closure of $C_0^\infty(\Omega)$ in H^k. If $\mathbb{X}$ is an inner product space, $L^2(\Omega;\mathbb{X})$ denotes the space of $\mathbb{X}$-valued, square integrable functions, and its norm and inner product will be denoted by $\|\cdot\|_0$ and $(\cdot,\cdot)$, respectively.

Next, $H(\mathrm{div},\Omega) = H(\mathrm{div},\Omega;\mathbb{V})$ will denote the Sobolev space of vector fields on Ω in $L^2(\Omega;\mathbb{V})$ with divergence in $L^2(\Omega)$, and its norm is denoted by $\|\cdot\|_{\mathrm{div}} := \left(\|\cdot\|_0^2 + \|\mathrm{div}\,\cdot\|_0^2\right)^{1/2}$. Similarly, $H(\mathbf{div},\Omega;\mathbb{M})$ will be functions in $L^2(\Omega;\mathbb{M})$ with divergence in $L^2(\Omega;\mathbb{V})$, where the divergence is taken by rows.

For a Hilbert space X, we denote its inner product by $\langle\cdot,\cdot\rangle_X$, except in the special case of $X = L^2(\Omega)$ already described, in which case $(\cdot,\cdot)$ is the inner product. If we let X^* denote a representation of the dual of X, the duality pairing between X and X^* will be denoted by $\langle\cdot,\cdot\rangle$. We will in the context of Sobolev spaces choose the representation X^* so that the duality pairing is an extension of the L^2 inner product. If Y denotes an additional Hilbert space, $\mathscr{L}(X,Y)$ denotes the space of bounded, linear operators from X to Y. If $T \in \mathscr{L}(X,Y^*)$, we denote its adjoint by T^*, which is an element of $\mathscr{L}(Y,X^*)$.

2.1. Abstract preconditioning of parameter dependent systems.

To motivate the analysis below, we will briefly discuss an abstract framework for preconditioning systems of partial differential equations and their discrete counterparts. For a more thorough discussion of this framework, we refer to [20, 21].

Let X be a real, separable Hilbert space. Suppose that $\mathscr{A} \in \mathscr{L}(X,X^*)$ is a linear and bounded operator, which is invertible with bounded inverse. Assume further that $\mathscr{A}$ is symmetric, i.e.,

$$\langle \mathscr{A}x, y\rangle = \langle x, \mathscr{A}y\rangle, \quad \forall x,y \in X.$$

We then consider the problem of finding $x \in X$ so that

$$\mathscr{A}x = f \tag{2.1}$$

in X^* for a given $f \in X^*$. Applying a symmetric, positive definite operator $\mathscr{B} \in \mathscr{L}(X^*,X)$ to problem (2.1) gives the preconditioned problem of finding $x \in X$ so that

$$\mathscr{B}\mathscr{A}x = \mathscr{B}f$$

in X. The convergence rate of a Krylov subspace method applied to the preconditioned problem is controlled by the condition number

$$K(\mathscr{B}\mathscr{A}) = \|\mathscr{B}\mathscr{A}\|_{\mathscr{L}(X,X)} \left\|(\mathscr{B}\mathscr{A})^{-1}\right\|_{\mathscr{L}(X,X)}$$

in the way that a large value of $K(\mathscr{B}\mathscr{A})$ will generally lead to slow convergence.

We note that one possible choice of the operator $\mathscr{B}$ is the Riesz map from X^* to X, or in fact, any operator spectrally equivalent to it. For linear systems arising as discretizations of partial differential equations an effective preconditioner also has to be easy to evaluate, i.e., we require that the action of the operator can be evaluated cheaply. For systems of partial differential equations, this point of view naturally leads to block diagonal preconditioners, where the blocks correspond to preconditioners of simpler and more canonical operators. For example, in the case of operators corresponding to stable discretizations of the inner products of Sobolev spaces like $X = H^1$, $X = H(\mathbf{curl})$, and $X = H(\mathrm{div})$, efficient algorithms that

are spectrally equivalent to the Riesz map from X^* to X can be constructed with multilevel algorithms, cf. e.g., [3, 9, 17].

Preconditioning of parameter dependent problems follows in a similar manner. Let $\mathscr{A}_\epsilon$ denote an operator depending on some collection of parameters ϵ. To construct a preconditioner for $\mathscr{A}_\epsilon$, we determine an ϵ-dependent Hilbert space, X_ϵ, such that $\mathscr{A}_\epsilon$ is a linear, symmetric map from X_ϵ to X_ϵ^*. Furthermore, the corresponding operator norms $\|\mathscr{A}_\epsilon^{-1}\|_{\mathscr{L}(X_\epsilon^*,X_\epsilon)}$ and $\|\mathscr{A}_\epsilon\|_{\mathscr{L}(X_\epsilon,X_\epsilon^*)}$ should be bounded independently of ϵ. Having determined X_ϵ, a suitable preconditioner is then a symmetric, positive definite operator $\mathscr{B}_\epsilon$ from X_ϵ^* to X_ϵ, where the operator norms of $\|\mathscr{B}_\epsilon\|_{\mathscr{L}(X_\epsilon^*,X_\epsilon)}$ and $\|\mathscr{B}_\epsilon^{-1}\|_{\mathscr{L}(X_\epsilon,X_\epsilon^*)}$are bounded independently of ϵ. We are then guaranteed that the condition number $K(\mathscr{B}_\epsilon\mathscr{A}_\epsilon)$ is bounded independently of ϵ, and, as a consequence, the performance of a Krylov subspace method will be ϵ independent.

2.2. **Variational formulation.** An implicit time discretization of the system (1.1), with time step Δt, will typically lead to a stationary system of the form

$$\begin{aligned} -\mathbf{div}\,\mathscr{C}\boldsymbol{\epsilon}(\boldsymbol{u}) + \alpha\,\mathrm{grad}\,p &= f \quad \text{in } \Omega, \\ s_0 p + \alpha\,\mathrm{div}\,\boldsymbol{u} - \Delta t\,\mathrm{div}(\boldsymbol{\kappa}\,\mathrm{grad}\,p) &= g \quad \text{in } \Omega. \end{aligned} \tag{2.2}$$

Here g encapsulates information about both the mass source and previous time steps. Furthermore, $\Delta t\boldsymbol{\kappa}$ can be regarded as one single parameter, which carries information about both the time discretization and the conductivity. Therefore, Δt is set equal to one in the discussion below, while the matrix valued function $\boldsymbol{\kappa}$ is assumed to be symmetric positive definite, but can be arbitrarily small. For parameter ranges of practical problems, it is typical that $\alpha > 0$ is of order 1, $1 \lesssim \mu \lesssim \lambda$, and $\mu \ll \lambda$ holds if the elastic matrix is nearly incompressible, i.e. if λ is large.

Since rescaling of the spectrum of (2.2) does not influence preconditioning, we rescale the system with a constant of order μ, say μ_0. More specifically, λ, α, s_0, f, $\boldsymbol{\kappa}$, g are replaced by λ/μ_0, α/μ_0, ..., g/μ_0. However, we shall use same symbols μ, λ, etc. for the rescaled parameters for simplicity of notation. As a consequence, in the rest of the paper, we assume that the first Lamé coefficient μ is of order 1, and other parameters are in the ranges

$$1 \lesssim \lambda < +\infty, \quad 0 < \alpha \leq 1, \quad 0 < \boldsymbol{\kappa} \leq 1. \tag{2.3}$$

Because we consider a formulation different from the one used in [20], some unknowns are also rescaled physical quantities in this paper. For example, $\sigma := 2\mu\boldsymbol{\epsilon}(\boldsymbol{u}) + (n\lambda\,\mathrm{div}\,\boldsymbol{u} - \alpha p)\mathbb{I}$ is the stress tensor in the original constitutive law. However, in the rest of the paper, μ, λ, α are parameters rescaled by μ_0, therefore σ is the stress tensor rescaled by μ_0.

The condition on $\boldsymbol{\kappa}$ given in (2.3) means that the pointwise eigenvalues of $\boldsymbol{\kappa}$ are uniformly bounded below by 0 and above by 1. The constrained specific storage coefficient is assumed to satisfy the relation $s_0 = \frac{\alpha^2}{\lambda}$. This assumption is mostly for sake of brevity, and the following analysis will work even if s_0 is only bounded from below by a constant times $\frac{\alpha^2}{\lambda}$. We refer to [20] for a more detailed discussion of scaling of the Biot system.

For (2.2) to be well-posed, it needs to be augmented with a set of boundary conditions. To that end, we introduce two separate partitions of the boundary, $\partial\Omega = \Gamma_p \cup \Gamma_f = \Gamma_d \cup \Gamma_t$, where Γ_p and Γ_d should have positive meaure, i.e., $|\Gamma_p|, |\Gamma_d| > 0$. General boundary conditions can then be posed as

$$\begin{aligned} p(t) &= p_0(t) && \text{on } \Gamma_p, \\ (\boldsymbol{\kappa}\,\mathrm{grad}\,p(t))\cdot\hat{\nu} &= z_{\hat{\nu}}(t) && \text{on } \Gamma_f, \\ \boldsymbol{u}(t) &= \boldsymbol{u}_0(t) && \text{on } \Gamma_d, \\ \sigma(t)\hat{\nu} &= (\mathscr{C}\boldsymbol{\epsilon}(\boldsymbol{u}) - \alpha p\mathbb{I})\hat{\nu} = \sigma_{\hat{\nu}}(t) && \text{on } \Gamma_t. \end{aligned}$$

For simplicity, we will in this paper only consider homogeneous boundary conditions. That is, $p_0, z_{\hat{\nu}}, \boldsymbol{u}_0, \sigma_{\hat{\nu}} = 0$.

For the weak formulation, we introduce a new unknown, the stress tensor, defined as

$$\sigma := \mathscr{C}\boldsymbol{\epsilon}(\boldsymbol{u}) - \alpha p \mathbb{I}, \tag{2.4}$$

and we denote the inverse of the stiffness tensor by $A = A_{\mu,\lambda} := \mathscr{C}^{-1}$, which is an operator acting on $\mathbb{S}$. With the stiffness tensor given by (1.2), we obtain

$$A\sigma = \frac{1}{2\mu}\left(\sigma - \frac{\lambda}{2\mu + n\lambda}\operatorname{tr}(\sigma)\mathbb{I}\right). \tag{2.5}$$

Furthermore, we note that the trace of (2.5) is given by

$$\operatorname{tr} A\sigma = \frac{1}{2\mu + n\lambda}\operatorname{tr}\sigma. \tag{2.6}$$

By using (2.4) and (2.6), we can express the term $\alpha \operatorname{div} \boldsymbol{u}$ in the second equation of (2.2) as a function of σ and p as

$$\alpha \operatorname{div} \boldsymbol{u} = \alpha \operatorname{tr} A(\sigma + \alpha p \mathbb{I}) = K\sigma + \frac{n\alpha^2}{2\mu + n\lambda} p, \tag{2.7}$$

where $K = K_{\alpha,\mu,\lambda} : \mathbb{M} \to \mathbb{R}$ is the operator defined pointwise by

$$K\tau := \frac{\alpha}{2\mu + n\lambda}\operatorname{tr}\tau. \tag{2.8}$$

After introducing σ defined by (2.4), and using (2.7), (2.2) becomes

$$\begin{aligned} A\sigma + K^* p - \boldsymbol{\epsilon}(\boldsymbol{u}) &= 0 \quad \text{in } \Omega, \\ K\sigma + Bp - \operatorname{div}(\boldsymbol{\kappa} \operatorname{grad} p) &= g \quad \text{in } \Omega, \\ -\mathbf{div}\sigma &= f \quad \text{in } \Omega. \end{aligned}$$

Here, K^* denotes the operator $p \mapsto \frac{\alpha}{2\mu+n\lambda} p\mathbb{I}$, while $B = B_{\alpha,\mu,\lambda}$ is the operator defined by

$$Bp := \left(s_0 + \frac{n\alpha^2}{2\mu + n\lambda}\right) p \equiv \frac{\alpha^2}{\lambda}\left(1 + \frac{n\lambda}{2\mu + n\lambda}\right) p. \tag{2.9}$$

To complete the formulation, we enforce the symmetry of the stress tensor in a weak manner, i.e., σ is now $\mathbb{M}$-valued, instead of $\mathbb{S}$, and we require that

$$(\sigma, \eta) = 0 \quad \forall \eta \in L^2(\Omega; \mathbb{K}).$$

The trade off is that we need to introduce a Lagrange multiplier, γ, which will also play the role of the skew symmetric part of $\mathbf{grad}\boldsymbol{u}$. This relaxation of the symmetry on σ also requires us to extend the definition of A from $\mathbb{S}$ to all tensors $\mathbb{M}$. We denote this extension by A as well, since it will also be given by formula (2.5).

The system now reads

$$\begin{aligned} A\sigma + K^* p - \mathbf{grad}\boldsymbol{u} + \gamma &= 0 \quad \text{in } \Omega, \\ K\sigma + Bp - \operatorname{div}(\boldsymbol{\kappa} \operatorname{grad} p) &= g \quad \text{in } \Omega, \\ -\mathbf{div}\sigma &= f \quad \text{in } \Omega, \\ (\sigma, \eta) &= 0 \quad \forall \eta \in L^2(\Omega; \mathbb{K}). \end{aligned} \tag{2.10}$$

Defining the function spaces

$$\begin{aligned} \Sigma &= \left\{\tau \in H(\mathbf{div}, \Omega; \mathbb{M}) : \tau \cdot \hat{\nu}|_{\Gamma_t} = 0\right\}, \\ Q &= \left\{q \in H^1(\Omega) : q|_{\Gamma_p} = 0\right\}, \\ \boldsymbol{V} &= L^2(\Omega; \mathbb{V}), \\ \Gamma &= L^2(\Omega; \mathbb{K}), \end{aligned} \tag{2.11}$$

an appriopriate weak formulation of (2.10) is: Find $(\sigma, p, \boldsymbol{u}, \gamma) \in \Sigma \times Q \times V \times \Gamma$ so that

$$
\begin{aligned}
(A\sigma,\tau)+(p,K\tau)+(\boldsymbol{u},\mathbf{div}\tau)+(\gamma,\tau)&=0 && \forall \tau\in\Sigma,\\
(K\sigma,q)+(Bp,q)+(\boldsymbol{\kappa}\operatorname{grad} p,\operatorname{grad} q)&=(g,q) && \forall q\in Q,\\
(\mathbf{div}\sigma,v)&=-(f,v) && \forall v\in V,\\
(\sigma,\eta)&=0 && \forall \eta\in\Gamma.
\end{aligned}
\tag{2.12}
$$

In matrix-vector form, the system (2.12) reads

$$
\mathscr{A}\begin{pmatrix}\sigma\\ p\\ \boldsymbol{u}\\ \gamma\end{pmatrix}:=\begin{pmatrix} A & K^* & -\mathbf{grad} & \mathrm{skw}^*\\ K & B-\operatorname{div}(\boldsymbol{\kappa}\operatorname{grad}) & 0 & 0\\ \mathbf{div} & 0 & 0 & 0\\ \mathrm{skw} & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\sigma\\ p\\ \boldsymbol{u}\\ \gamma\end{pmatrix}=\begin{pmatrix}0\\ g\\ -f\\ 0\end{pmatrix},
\tag{2.13}
$$

where $\mathrm{skw}:\mathbb{M}\to\mathbb{K}$ is the operator returning the skew-symmetric part of a tensor, in which case $\mathrm{skw}^*:\mathbb{K}\to\mathbb{M}$ is simply the inclusion operator. From (2.13), we see that the system exhibits a saddle point structure, and so well-posedness is ensured if the provided function spaces satisfies the stability conditions in Brezzi's theory of mixed methods (cf. [8]). We introduce the inner products

$$
\begin{aligned}
\langle\sigma,\tau\rangle_\Sigma&=\left(\frac{1}{2\mu}\sigma,\tau\right)+(\mathbf{div}\sigma,\mathbf{div}\tau) && \forall\sigma,\tau\in\Sigma,\\
\langle p,q\rangle_Q&=(Bp,q)+(\boldsymbol{\kappa}\operatorname{grad} p,\operatorname{grad} q) && \forall p,q\in Q,\\
\langle \boldsymbol{u},v\rangle_V&=(\boldsymbol{u},v) && \forall \boldsymbol{u},v\in V,\\
\langle\gamma,\eta\rangle_\Gamma&=(\gamma,\eta) && \forall\gamma,\eta\in\Gamma,
\end{aligned}
\tag{2.14}
$$

and define $\mathcal{X}:=\Sigma\times Q\times V\times\Gamma$ with inner products inherited from (2.14). With this notation the left-hand side of (2.12) can alternatively be written as $\langle\mathscr{A}(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,v,\eta)\rangle$, where the operator $\mathscr{A}:\mathcal{X}\to\mathcal{X}^*$ will be bounded. In fact, in the case when $|\Gamma_t|>0$ the operator $\mathscr{A}$ will be bounded independently of α, λ, and $\boldsymbol{\kappa}$, and to establish this uniform bound will be a main topic of the next section. However, in the clamped case, i.e., the case when $|\Gamma_t|=0$, we need to alter the norm of the space Σ to obtain a corresponding uniform bound. This discussion will also be given in the following section.

We end this section with the following remark.

Remark 1. As already noted, the coefficient matrix form in (2.13) exposes the saddle point structure of the system. However, worth noting is that a simple rearrangement of the terms leads to the system

$$
\mathscr{A}\begin{pmatrix}\sigma\\ \boldsymbol{u}\\ \gamma\\ p\end{pmatrix}:=\begin{pmatrix} A & -\mathbf{grad} & \mathrm{skw}^* & K^*\\ \mathbf{div} & 0 & 0 & 0\\ \mathrm{skw} & 0 & 0 & 0\\ K & 0 & 0 & B-\operatorname{div}(\boldsymbol{\kappa}\operatorname{grad})\end{pmatrix}\begin{pmatrix}\sigma\\ \boldsymbol{u}\\ \gamma\\ p\end{pmatrix}.
\tag{2.15}
$$

From this, we can consider the system as a coupling between a mixed formulation of linear elasticity with weakly imposed symmetry in the unknown $(\sigma,\boldsymbol{u},\gamma)$, and a reaction-diffusion equation in the pore pressure p. We will see that this observation will bear out stable finite element discretizations of this system.

3. Parameter robust stability

The purpose of this section is to establish stability bounds for the system (2.12) or equivalently (2.13). Note that this system depends on the parameters α,μ,λ implicitly through the definition of the operators A, B, and K, and explicitly of the hydraulic conductivity κ. However, our goal is to establish stability bounds where the stability constant is independent of these parameters, as long as they vary as specified in the beginning of Section 2.2. On the

other hand, we will allow the norms to depend on these parameters. More precisely, for the case $|\Gamma_t| > 0$ we will use the norms given by the inner products specified in (2.14), while the inner product of the space Σ has to be altered slightly in the clamped case, i.e., when $|\Gamma_t| = 0$. As we will see in the next section this perturbation will also have an effect on the construction of robust preconditioners.

3.1. **The continuous case.** We will first consider the case when $|\Gamma_t| > 0$. We introduce the two projections in $L^2(\Omega;\mathbb{M})$

$$P_0\tau := \tau - \frac{1}{n}\left(\frac{1}{|\Omega|}\int_\Omega \operatorname{tr}\tau \mathrm{d}x\right)\mathbb{I}, \quad P_D\tau := \tau - \frac{1}{n}\operatorname{tr}\tau\mathbb{I}.$$

That is, P_0 projects τ to its mean trace-free part, whereas P_D projects τ to its pointwise trace-free part. It then follows by algebraic considerations that

$$P_0P_D = P_DP_0 = P_D. \tag{3.1}$$

It is worthwhile to note that since $\mathbf{div}P_0 = \mathbf{div}$ on Σ, P_0 is also an orthogonal projection on Σ, not only on $L^2(\Omega;\mathbb{M})$. An algebraic manipulation gives

$$\begin{aligned}(A\tau,\tau) &= \left(\frac{1}{2\mu}\tau,\tau\right) - \left(\frac{\lambda}{2\mu(2\mu+n\lambda)}\operatorname{tr}\tau,\operatorname{tr}\tau\right)\\ &= \left(\frac{1}{2\mu}P_D\tau,P_D\tau\right) + \left(\frac{1}{2\mu+n\lambda}(I-P_D)\tau,(I-P_D)\tau\right)\end{aligned}$$

where the second equality follows from $\tau = P_D\tau + (I-P_D)\tau$ and the pointwise orthogonality of $P_D\tau$ and $(I-P_D)\tau = \frac{1}{n}\operatorname{tr}\tau\mathbb{I}$. From this, a two-sided bound of $(A\tau,\tau)$

$$\left(\frac{1}{2\mu}P_D\tau,P_D\tau\right) \le (A\tau,\tau) \le \left(\frac{1}{2\mu}\tau,\tau\right) \tag{3.2}$$

follows. We will use the following bound

$$\left(\frac{1}{2\mu}\tau,\tau\right) \le C\left(\left(\frac{1}{2\mu}P_D\tau,P_D\tau\right) + \|\mathbf{div}\tau\|_0^2\right), \quad \tau\in\Sigma, \tag{3.3}$$

where the constant C is independent of τ and λ. This bound leads to stability of linear elasticity in the incompressible limit, i.e., when $\lambda = +\infty$, and was used already in [2] to obtain robust stability of mixed finite element methods for such problems. The bound (3.3), along with (3.5) given below, will also be crucial for the construction of robust preconditioners for the Biot model. We therefore provide proofs of these estimates in Appendix A for the sake of completeness.

As a consequence of the (3.3), we observe that the following equivalence follows.

Lemma 3.1. *Assume Σ is given by the first definition in (2.11) with $|\Gamma_t| > 0$. There is a constant $C > 0$ such that*

$$(A\tau,\tau) + \|\mathbf{div}\tau\|_0^2 \le \langle\tau,\tau\rangle_\Sigma \le C((A\tau,\tau) + \|\mathbf{div}\tau\|_0^2) \tag{3.4}$$

for every $\tau\in\Sigma$. In particular, the constant C is independent of λ.

Proof. The first inequality of (3.4) follows immediately from (3.2). The second inequality follows as

$$\langle\tau,\tau\rangle_\Sigma \le C\left(\left(\frac{1}{2\mu}P_D\tau,P_D\tau\right) + \|\mathbf{div}\tau\|_0^2\right) \le C((A\tau,\tau) + \|\mathbf{div}\tau\|_0^2),$$

using (3.3) and (3.2). □

In the case of $|\Gamma_t| = 0$, the constant matrix field $\tau = \mathbb{I}$ is an element of Σ, and $(A\tau,\tau) + \|\mathbf{div}\tau\|_0^2 \to 0$ as $\lambda \to +\infty$. Therefore, we cannot hope to extend the λ-robust equivalence of Lemma 3.1 to the case $\Gamma_d = \partial\Omega$. In fact, $\tau = c\mathbb{I}$, for any nonzero $c \in \mathbb{R}$ is the only case that the equivalence fails. Excluding the span of $\{\mathbb{I}\}$ from Σ, we can still have a bound similar to (3.4) as

$$\left(\frac{1}{2\mu}P_0\tau, P_0\tau\right) \le C\left(\left(\frac{1}{2\mu}P_D\tau, P_D\tau\right) + \|\mathbf{div}\tau\|_0^2\right), \quad \tau \in \Sigma, \tag{3.5}$$

which is also proved in [2]. We can use (3.5) to establish that the operator $A - \mathbf{graddiv}$ is spectrally equivalent to the μ-scaled $H(\mathbf{div})$ inner product over $P_0(\Sigma)$, i.e., the subspace of Σ consisting of matrix fields with zero mean trace. On the other hand, for $\tau \in (I - P_0)(\Sigma)$, τ is a constant multiple of identity matrix field, so

$$(A\tau,\tau) = \left(\frac{1}{2\mu + n\lambda}(I - P_0)\tau, (I - P_0)\tau\right).$$

This gives a motivation to define an auxiliary inner product $\langle\cdot,\cdot\rangle_{\tilde{\Sigma}}$ on Σ as

$$\langle\sigma,\tau\rangle_{\tilde{\Sigma}} := \left(\frac{1}{2\mu}P_0\sigma, P_0\tau\right) + \left(\frac{1}{2\mu + n\lambda}(I - P_0)\sigma, (I - P_0)\tau\right) + (\mathbf{div}\sigma, \mathbf{div}\tau), \tag{3.6}$$

for $\sigma, \tau \in \Sigma$. The following lemma states that this inner product is spectrally equivalent to the inner product derived from $A - \mathbf{graddiv}$.

Lemma 3.2. *Assume $|\Gamma_t| = 0$. There exists a positive constant C such that*

$$C^{-1}((A\tau,\tau) + \|\mathbf{div}\tau\|_0^2) \le \langle\tau,\tau\rangle_{\tilde{\Sigma}} \le C((A\tau,\tau) + \|\mathbf{div}\tau\|_0^2). \tag{3.7}$$

In particular, the constant C is independent of λ.

Proof. Since P_0 is an orthogonal projection on Σ, in both inner products, it is sufficient to consider $\tau \in P_0(\Sigma)$ and $\tau \in (I - P_0)(\Sigma)$ separately.

If $\tau \in (I - P_0)(\Sigma)$, then $P_0\tau = 0$ and τ is a constant multiple of the identity matrix field, so

$$\langle\tau,\tau\rangle_{\tilde{\Sigma}} = (A\tau,\tau) + \|\mathbf{div}\tau\|_0^2,$$

which verifies (3.7) in this case.

Next, if $\tau \in P_0(\Sigma)$, i.e., $\tau = P_0\tau$, we have from (3.2) and (3.5) that

$$(A\tau,\tau) = (AP_0\tau, P_0\tau) \le \left(\frac{1}{2\mu}P_0\tau, P_0\tau\right) \le C\left(\left(\frac{1}{2\mu}P_D\tau, P_D\tau\right) + \|\mathbf{div}\tau\|_0^2\right)$$

and from the pointwise orthogonality of $(I - P_D)\tau = (P_0 - P_D)\tau$ and $P_D\tau$ that

$$\langle\tau,\tau\rangle_{\tilde{\Sigma}} = \left(\frac{1}{2\mu}P_0\tau, P_0\tau\right) + \|\mathbf{div}\tau\|_0^2 \ge \left(\frac{1}{2\mu}P_D\tau, P_D\tau\right) + \|\mathbf{div}\tau\|_0^2.$$

The left inequality of (3.7) easily follows from the above two inequalities. Furthermore, using (3.5) and (3.2), we obtain

$$\begin{aligned}\langle\tau,\tau\rangle_{\tilde{\Sigma}} &= \left(\frac{1}{2\mu}P_0\tau, P_0\tau\right) + \|\mathbf{div}\tau\|_0^2 \\ &\le C\left(\left(\frac{1}{2\mu}P_D\tau, P_D\tau\right) + \|\mathbf{div}\tau\|_0^2\right) \\ &\le C((A\tau,\tau) + \|\mathbf{div}\tau\|_0^2)\end{aligned}$$

which is the right inequality of (3.7).

□

We recall that the space $\mathcal{X} = \Sigma \times Q \times V \times \Gamma$ was introduced in Section 2.2 for the case when $|\Gamma_t| > 0$. For the clamped case, i.e., when $|\Gamma_t| = 0$, we consider the modified space given by $\tilde{\mathcal{X}} := \tilde{\Sigma} \times Q \times V \times \Gamma$, where $\tilde{\Sigma} = H(\mathbf{div}, \Omega; \mathbb{M})$, and with inner product given by (3.6). As a consequence of the spectral equivalences (3.4) and (3.7), we obtain that the operator $\mathscr{A}$ is bounded as an operator in $\mathscr{L}(\mathcal{X}, \mathcal{X}^*)$ when $|\Gamma_t| > 0$, and as an operator in $\mathscr{L}(\tilde{\mathcal{X}}, \tilde{\mathcal{X}}^*)$ in the clamped case when $|\Gamma_t| = 0$. More precisely, we have the following result.

Theorem 3.1. *Assume that the parameters λ, α, and $\boldsymbol{\kappa}$ satisfies condition (2.3). Let $X = \mathcal{X}$ if $|\Gamma_t| > 0$, and $X = \tilde{\mathcal{X}}$ if $|\Gamma_t| = 0$. Then, for the system (2.12) there is a constant $\beta > 0$, independent of λ, α, and $\boldsymbol{\kappa}$, so that the following inf-sup condition holds:*

$$\inf_{(\sigma,p,\boldsymbol{u},\gamma)\in X} \sup_{(\tau,q,\boldsymbol{v},\eta)\in X} \frac{\langle \mathscr{A}(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle}{\|(\sigma,p,\boldsymbol{u},\gamma)\|_X \, \|(\tau,q,\boldsymbol{v},\eta)\|_X} \geq \beta. \tag{3.8}$$

Proof. Consider first the case with $|\Gamma_t| > 0$, so that $X = \mathcal{X}$. To prove (3.8), we will show that there exist positive constants C_1, C_2, so that for every $0 \neq (\sigma, \boldsymbol{u}, \gamma, p) \in \mathcal{X}$ there are $(\tau, \boldsymbol{v}, \eta, q) \in \mathcal{X}$ so that

$$\begin{aligned} \|(\tau,q,\boldsymbol{v},\eta)\|_\chi &\leq C_1 \|(\sigma,p,\boldsymbol{u},\gamma)\|_\chi, \\ \langle \mathscr{A}(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle &\geq C_2 \|(\sigma,p,\boldsymbol{u},\gamma)\|_\chi^2, \end{aligned} \tag{3.9}$$

the key being that C_1 and C_2 will be independent of λ, α, and $\boldsymbol{\kappa}$.

To verify (3.9), let $(\sigma, \boldsymbol{u}, \gamma, p) \in \mathcal{X}$ be nonzero but otherwise arbitrary. From the theory of mixed elasticity with weakly enforced symmetry (see e.g., [4, 7]), there exists a $\beta_0 > 0$, and $\tilde{\tau} \in \Sigma$ so that

$$\begin{aligned} &\operatorname{div} \tilde{\tau} = \boldsymbol{u}, \\ &(\tilde{\tau}, \eta) = (\gamma, \eta) \quad \forall \eta \in L^2(\Omega; \mathbb{K}), \\ &\|\tilde{\tau}\|_\Sigma^2 \leq \beta_0^2 \big(\|\boldsymbol{u}\|_0^2 + \|\gamma\|_0^2 \big), \end{aligned} \tag{3.10}$$

with β_0 depending only on Ω. From (3.4), we see that

$$(A\tilde{\tau}, \tilde{\tau}) \leq \beta_0^2 \big(\|\boldsymbol{u}\|_0^2 + \|\gamma\|_0^2 \big). \tag{3.11}$$

By setting $\tau = \sigma + \delta_0 \tilde{\tau}$, $\boldsymbol{v} = -\boldsymbol{u} + \delta_1 \mathbf{div}\sigma$, $\eta = -\gamma$, and $q = p$, we find that

$$\|(\tau,q,\boldsymbol{v},\eta)\|_\chi \leq \sqrt{2(1 + \max(\delta_0^2\beta_0^2, \delta_1^2))} \, \|(\sigma,p,\boldsymbol{u},\gamma)\|_\chi,$$

which verifies the first inequality in (3.9). To prove the second inequality in (3.9), we begin by observing that after cancelling terms we obtain the identity

$$\begin{aligned} \langle \mathscr{A}(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle = (A\sigma,\sigma) + \delta_0(A\sigma,\tilde{\tau}) + 2(p,K\sigma) + \|p\|_Q^2 \\ + \delta_0(p,K\tilde{\tau}) + \delta_0\big(\|\boldsymbol{u}\|_0^2 + \|\gamma\|_0^2\big) + \delta_1 \|\mathbf{div}\sigma\|_0^2, \end{aligned} \tag{3.12}$$

where we have used the properties of $\tilde{\tau}$. To bound the three cross terms, we use Cauchy-Schwarz and Young's inequalities in a standard way. For the term $(A\sigma, \tilde{\tau})$, this and (3.11) yield

$$(A\sigma,\tilde{\tau}) \leq \frac{\epsilon_1}{2}(A\sigma,\sigma) + \frac{1}{2\epsilon_1}(A\tilde{\tau},\tilde{\tau}) \leq \frac{\epsilon_1}{2}(A\sigma,\sigma) + \frac{\beta_0^2}{2\epsilon_1}(\|\boldsymbol{u}\|_0^2 + \|\gamma\|_0^2) \tag{3.13}$$

for any $\epsilon_1 > 0$. We can derive similar bounds for the two terms involving the operator K. From the definition of K and Young's inequality, we obtain

$$(p,K\sigma) \leq \frac{\epsilon_2}{2}\left(\frac{n\alpha^2}{2\mu + n\lambda} p, p\right) + \frac{1}{2\epsilon_2}\left(\frac{1}{2\mu + n\lambda}\operatorname{tr}\sigma, \frac{1}{n}\operatorname{tr}\sigma\right). \tag{3.14}$$

For the first term in (3.14), the definition of B in (2.9) yields

$$\left(\frac{n\alpha^2}{2\mu+n\lambda}p,p\right)\leq\frac{1}{2}(Bp,p)\leq\frac{1}{2}\|p\|_Q^2.$$

Inserting this into (3.14), and using the properties of A, we obtain

$$\begin{aligned}(p,K\sigma)&\leq\frac{\epsilon_2}{4}\|p\|_Q^2+\frac{1}{2\epsilon_2}\left(\operatorname{tr}A\sigma,\frac{1}{n}\operatorname{tr}\sigma\right)\\&=\frac{\epsilon_2}{4}\|p\|_Q^2+\frac{1}{2\epsilon_2}\left(A\sigma,\frac{1}{n}\operatorname{tr}\sigma\mathbb{I}\right)\leq\frac{\epsilon_2}{4}\|p\|_Q^2+\frac{1}{2\epsilon_2}(A\sigma,\sigma),\end{aligned}\tag{3.15}$$

where $\epsilon_2>0$ is arbitrary. Furthermore, we have a similar bound

$$(p,K\tilde{\tau})\leq\frac{\epsilon_3}{4}\|p\|_Q^2+\frac{1}{2\epsilon_3}(A\tilde{\tau},\tilde{\tau})\leq\frac{\epsilon_3}{4}\|p\|_Q^2+\frac{\beta_0^2}{2\epsilon_3}(\|\boldsymbol{u}\|_0^2+\|\gamma\|_0^2).$$

As a consequence, after using (3.13) and (3.15) in (3.12) and collecting terms, together with using the properties of $\tilde{\tau}$, we end up with

$$\begin{aligned}\langle\mathscr{A}(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle\geq&\left(1-\frac{\delta_0\epsilon_1}{2}-\frac{1}{\epsilon_2}\right)(A\sigma,\sigma)\\&+\delta_0\left(1-\frac{\beta_0^2}{2\epsilon_1}-\frac{\beta_0^2}{2\epsilon_3}\right)(\|\boldsymbol{u}\|_0^2+\|\gamma\|_0^2)\\&+\left(1-\frac{\epsilon_2}{2}-\frac{\delta_0\epsilon_3}{4}\right)\|p\|_Q^2+\delta_1\|\mathbf{div}\sigma\|_0^2.\end{aligned}$$

If we can choose δ_0, ϵ_1, ϵ_2, and ϵ_3 so that all the coefficients above are positive, this will prove the second inequality in (3.9), because of (2.12). For instance, choosing $\delta_0=\frac{1}{6\beta_0^2}$, $\epsilon_1=\epsilon_3=2\beta_0^2$, $\epsilon_2=\frac{3}{2}$, and $\delta_1=\frac{1}{6}$ yields

$$\langle\mathscr{A}(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle\geq\frac{C}{6}\|\sigma\|_\Sigma+\frac{1}{12\beta_0^2}\left(\|\boldsymbol{u}\|_0^2+\|\gamma\|_0^2\right)+\frac{1}{6}\|p\|_Q^2,$$

in which case the second inequality in (3.9) holds with $\beta=\frac{1}{6}\min\left(C,\frac{1}{2\beta_0^2}\right)$.

In the case that $X=\tilde{\mathcal{X}}$ the argument is almost completely analogous. In particular, (3.10) continues to hold with $\|\cdot\|_{\tilde{\Sigma}}$ instead of $\|\cdot\|_\Sigma$ since $\|\tau\|_{\tilde{\Sigma}}\leq\|\tau\|_\Sigma$ for every $\tau\in\Sigma$. When $X=\tilde{\mathcal{X}}$ we must also use (3.7) instead of (3.4). Other than that, the argument remains unchanged. □

3.2. **The discrete case.** If we discretize (2.13) with finite element spaces $\Sigma_h\subset\Sigma$, $Q_h\subset Q$, $\Gamma_h\subset\Gamma$, and $V_h\subset V$, and define $\mathcal{X}_h=\Sigma_h\times Q_h\times V_h\times\Gamma_h$, the discrete formulation becomes:

Find $(\sigma_h,p_h,\boldsymbol{u}_h,\gamma_h)\in\mathcal{X}_h$ so that

$$\begin{aligned}(A\sigma_h,\tau)+(p_h,K\tau)+(\boldsymbol{u}_h,\mathbf{div}\tau)+(\gamma_h,\tau)&=0&&\forall\tau\in\Sigma_h,\\(K\sigma_h,q)+(Bp_h,q)+(\boldsymbol{\kappa}\operatorname{grad}p_h,\operatorname{grad}q)&=(g,q)&&\forall q\in Q_h,\\(\mathbf{div}\sigma_h,\boldsymbol{v})&=-(f,\boldsymbol{v})&&\forall\boldsymbol{v}\in\boldsymbol{V}_h,\\(\sigma_h,\eta)&=0&&\forall\eta\in\Gamma_h.\end{aligned}\tag{3.16}$$

Note that Γ_h is only a subspace of Γ, so symmetry of σ_h is imposed only weakly by the last equation in (3.16).

The left-hand side of the system above can be written in the form $\langle\mathscr{A}_h(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle$, where $\mathscr{A}_h:\mathcal{X}_h\to\mathcal{X}_h^*$ is the corresponding discrete coefficient operator. Our goal is to establish a discrete version of Theorem 3.1, i.e., a stability bound where the stability constant is independent of the model parameters as well as the mesh parameter h. We observe that the key feature of the proof of Theorem 3.1 was the property (3.10), which corresponds to the

stability of the underlying elasticity problem. For the proof to carry over to the discrete case, the finite element spaces should satisfy a discrete variant of property (3.10). In other words, the triple $(\Sigma_h, V_h, \Gamma_h)$ has to be a stable elasticity element. Therefore, we make the following definition.

Definition 3.1. We say the function spaces Σ_h, V_h, and Γ_h are elasticity stable if $\mathbf{div}\Sigma_h = V_h$, and there exists a constant $C > 0$, independent of discretization parameter h, such that for any $(\boldsymbol{u}_h, \gamma_h) \in V_h \times \Gamma_h$, there exists $\tau \in \Sigma_h$ satisfying

$$\begin{aligned} \mathbf{div}\tau &= \boldsymbol{u}_h, \\ (\tau, \eta) &= (\gamma_h, \eta) \quad \forall \eta \in \Gamma_h, \\ \|\tau\|_{\mathrm{div}} &\leq C\left(\|\boldsymbol{u}_h\|_0 + \|\gamma_h\|_0\right). \end{aligned}$$

Examples of elasticity stable elements can be found in [4, 1, 7, 11, 13, 15, 19, 31].

Theorem 3.2. *Let $X = \mathcal{X}$ if Γ_t has positive measure, and if $|\Gamma_t| = 0$ let $X = \tilde{\mathcal{X}}$. Suppose that $(\Sigma_h, V_h, \Gamma_h)$ in the discrete formulation (3.16) is elasticity stable, and that the parameter ranges in (2.3) are satisfied. Setting $\mathcal{X}_h = \Sigma_h \times V_h \times \Gamma_h \times Q_h$, with the same norm as X, and defining $\mathscr{A}_h : \mathcal{X}_h \to \mathcal{X}_h^*$, then there exists $\beta > 0$ such that*

$$\inf_{(\sigma,p,\boldsymbol{u},\gamma)\in\mathcal{X}_h} \sup_{(\tau,q,\boldsymbol{v},\eta)\in\mathcal{X}_h} \frac{\langle \mathscr{A}_h(\sigma,p,\boldsymbol{u},\gamma),(\tau,q,\boldsymbol{v},\eta)\rangle}{\|(\sigma,p,\boldsymbol{u},\gamma)\|_X \, \|(\tau,q,\boldsymbol{v},\eta)\|_X} \geq \beta,$$

and β is independent of λ, α, κ, and the discretization parameter h.

Proof. Analogous to the proof of Theorem 3.1, it is sufficient to prove that there exist constants C_1 and C_2 so that for every $0 \neq (\sigma_h, p_h, \boldsymbol{u}_h, \gamma_h) \in \mathcal{X}_h$ there is $(\tau, q, \boldsymbol{v}, \eta) \in \mathcal{X}_h$ so that

$$\begin{aligned} \|(\tau,q,\boldsymbol{v},\eta)\|_{\mathcal{X}} &\leq C_1 \|(\sigma_h,p_h,\boldsymbol{u}_h,\gamma_h)\|_{\mathcal{X}_h}, \\ \langle \mathscr{A}_h(\sigma_h,p_h,\boldsymbol{u}_h,\gamma_h),(\tau,q,\boldsymbol{v},\eta)\rangle &\geq C_2 \|(\sigma_h,p_h,\boldsymbol{u}_h,\gamma_h)\|^2_{\mathcal{X}_h}. \end{aligned}$$

Fix $(\sigma_h, p_h, \boldsymbol{u}_h, \gamma_h) \in \mathcal{X}_h$. Since Σ_h, V_h, and Γ_h are elasticity stable, $\mathbf{div}\sigma_h \in V_h$ and we can choose $\tilde{\tau} \in \Sigma_h$ such that

$$\begin{aligned} \mathbf{div}\tilde{\tau} &= \boldsymbol{u}_h, \\ (\tilde{\tau}, \eta) &= (\gamma_h, \eta) \quad \forall \eta \in \Gamma_h, \\ \|\tilde{\tau}\|_{\mathrm{div}} &\leq C\left(\|\boldsymbol{u}_h\|_0 + \|\gamma_h\|_0\right), \end{aligned}$$

where the constant C is independent of h and model parameters. Setting $\tau = \sigma_h + \delta_0\tilde{\tau}$, $q = p_h$, $\boldsymbol{v} = -\boldsymbol{u}_h + \delta_1\mathbf{div}\sigma_h$, and $\eta = -\gamma_h$, we have that $(\tau, q, \boldsymbol{v}, \eta) \in \mathcal{X}_h$ and

$$\begin{aligned} \langle \mathscr{A}(\sigma_h,p_h,\boldsymbol{u}_h,\gamma_h),(\tau,q,\boldsymbol{v},\eta)\rangle = {} & (A\sigma_h,\sigma_h) + \delta_0(A\sigma_h,\tilde{\tau}) + 2(p_h,K\sigma_h) + \|p_h\|^2_Q \\ & + \delta_0(p_h,K\tilde{\tau}) + \delta_0\left(\|\boldsymbol{u}_h\|^2_0 + \|\gamma_h\|^2_0\right) + \delta_1\|\mathbf{div}\sigma_h\|^2_0. \end{aligned}$$

The rest of the proof is completely analogous to the proof of Theorem 3.1.

□

4. Preconditioning

In this section, we will derive order optimal parameter-robust preconditioners for the discretized system. In the case where $|\Gamma_t| > 0$ it was shown in the previous section that the continuous operator $\mathscr{A} : \mathcal{X} \to \mathcal{X}^*$ was an isomorphism, where $\mathcal{X} = \Sigma \times Q \times V \times \Gamma$. A parameter-robust preconditioner is then constructed as an isomorphism $\mathscr{B} : \mathcal{X}^* \to \mathcal{X}$. The

canonical choice, which is symmetric and positive definite, is:

$$\mathscr{B} = \begin{pmatrix} \left(\frac{1}{2\mu} - \mathbf{graddiv}\right) & 0 & 0 & 0 \\ 0 & B - \operatorname{div}\boldsymbol{\kappa}\operatorname{grad} & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}^{-1} \tag{4.1}$$

In the discrete case, order optimal and spectrally equivalent realizations of the preconditioner can be constructed by multigrid techniques. The first block requires $H(\mathrm{div})$-preconditioners such as, e.g., [3, 17]. The second block is a second order elliptic operator for which multilevel algorithms are well known. If V_h and Γ_h are discontinuous finite element spaces, the third and fourth blocks are block diagonal mass matrices and their exact inverses, which are cheaply computable, can be used as preconditioners. When Γ_h is a Lagrange finite element (e.g., [7, 13]), simple iterative methods such as Jacobi or symmetric Gauss-Seidel give preconditioners that are spectrally equivalent to the inverse of the mass matrix.

The case $|\Gamma_t| = 0$ is more challenging and we recall that $\mathscr{A}$ is no longer stable in $\mathcal{X} = \Sigma \times Q \times V \times \Gamma$. In fact, stability was obtained in the alternative space $\tilde{\mathcal{X}} = \tilde{\Sigma} \times Q \times V \times \Gamma$. Therefore, the canonical choice for a parameter-robust preconditioner is then the symmetric and positive definite operator $\tilde{\mathscr{B}} : \tilde{\mathcal{X}}^* \to \tilde{\mathcal{X}}$ defined by

$$\tilde{\mathscr{B}} = \begin{pmatrix} \left(\frac{1}{2\mu}P_0 + \frac{1}{2\mu+n\lambda}(I-P_0) - \mathbf{graddiv}\right) & 0 & 0 & 0 \\ 0 & B - \operatorname{div}\boldsymbol{\kappa}\operatorname{grad} & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}^{-1}. \tag{4.2}$$

Here, $\tilde{\Sigma}$ is not a function space with standard $H(\mathrm{div})$ norm, and it is not clear that the multilevel algorithms developed for standard $H(\mathrm{div})$ spaces result in efficient preconditioners in $\tilde{\Sigma}$. Therefore, we will use a technique similar to the one used in [20]. In the rest of this section, we assume that μ, λ are constant on Ω. We recall the $\tilde{\Sigma}$ inner product

$$\langle \sigma, \tau \rangle_{\tilde{\Sigma}} = \left(\frac{1}{2\mu}P_0\sigma, P_0\tau\right) + \left(\frac{1}{2\mu+n\lambda}(I-P_0)\sigma, (I-P_0)\tau\right) + (\mathbf{div}\sigma, \mathbf{div}\tau).$$

To construct a preconditioner for this inner product we rely on the fact that we have efficient preconditioners for the weighted $H(\mathbf{div}, \Omega; \mathbb{M})$ inner product,

$$\langle \sigma, \tau \rangle_{\Sigma} = \left(\frac{1}{2\mu}\sigma, \tau\right) + (\mathbf{div}\sigma, \mathbf{div}\tau).$$

Let $\{\phi_i\}_{i=1}^N$ be a basis for $\Sigma_h \subset \Sigma$. Then we introduce the following matrices:

$$\begin{aligned} \tilde{\mathbb{B}}_{i,j} &= \left\langle \phi_j, \phi_i \right\rangle_{\tilde{\Sigma}}, \\ \mathbb{B}_{i,j} &= \left\langle \phi_j, \phi_i \right\rangle_{\Sigma}, \\ (\mathbb{B}_0)_{i,j} &= \frac{1}{2\mu}\left(P_0\phi_j, P_0\phi_i\right) + \left(\mathbf{div}\phi_j, \mathbf{div}\phi_i\right), \\ (\mathbb{B}_t)_{i,j} &= \frac{1}{2\mu}\left((I-P_0)\phi_j, (I-P_0)\phi_i\right). \end{aligned} \tag{4.3}$$

From (4.3) and (3.6), we see that

$$\mathbb{B} = \mathbb{B}_0 + \mathbb{B}_t, \quad \tilde{\mathbb{B}} = \mathbb{B}_0 + \frac{2\mu}{2\mu+n\lambda}\mathbb{B}_t.$$

Hence,

$$\tilde{\mathbb{B}} = \mathbb{B} - \rho\mathbb{B}_t \qquad \text{where} \quad \rho = \frac{n\lambda}{2\mu+n\lambda}.$$

Considering the entries of $\mathbb{B}_t$ in more detail, we find that

$$\begin{aligned}(\mathbb{B}_t)_{i,j} &= \frac{1}{2\mu}\frac{1}{n|\Omega|}\left(\int_\Omega \operatorname{tr}\phi_j \mathrm{d}x\right)\left(\int_\Omega \operatorname{tr}\phi_i \mathrm{d}x\right)\\ &= \frac{1}{2\mu} m m^T,\end{aligned}$$

where $m \in \mathbb{R}^N$ is the column vector with entries

$$m_i = \frac{1}{\sqrt{n|\Omega|}}\int_\Omega \operatorname{tr}\phi_i \mathrm{d}x. \tag{4.4}$$

Thus, we have that

$$\tilde{\mathbb{B}} = \mathbb{B} - \frac{\rho}{2\mu} m m^T. \tag{4.5}$$

Next, we define $w \in \mathbb{R}^N$ to be so that

$$\sum_{i=1}^N w_i \phi_i = \mathbb{I}. \tag{4.6}$$

Lemma 4.1. *Let $\{\phi_i\}_{i=1}^N$ be a basis for the finite dimensional function space $\Sigma_h \subset \Sigma$ and assume that μ and λ are positive constants. With $m \in \mathbb{R}^N$ defined by (4.4), $w \in \mathbb{R}^N$ defined by (4.6), and $\mathbb{B}$ the $N \times N$ matrix defined by (4.3), the following identities hold:*

$$\mathbb{B}w = \frac{\sqrt{n|\Omega|}}{2\mu} m, \quad w^T m = \sqrt{n|\Omega|}. \tag{4.7}$$

Proof. For the first identity in (4.7), we use the definition of m, w and $\mathbb{B}$ to see that the i'th component of $\mathbb{B}w$ is

$$(\mathbb{B}w)_i = \sum_{j=1}^N \left\langle w_j \phi_j, \phi_i \right\rangle_\Sigma = \frac{1}{2\mu}(\mathbb{I}, \phi_i) = \frac{\sqrt{n|\Omega|}}{2\mu} m_i.$$

Similarly, the second identity of (4.7) follows by

$$w^T m = \frac{1}{\sqrt{n|\Omega|}}\int_\Omega \sum_{i=1}^N \operatorname{tr} w_i \phi_i \mathrm{d}x = \frac{1}{\sqrt{n|\Omega|}}\int_\Omega \operatorname{tr}\mathbb{I}\mathrm{d}x = \sqrt{n|\Omega|}.$$

□

Corollary 4.1. *Under the same assumptions as in Lemma 4.1, and $\tilde{\mathbb{B}}$ the $N \times N$ matrix defined by (4.3), it holds that*

$$\tilde{\mathbb{B}} = \mathbb{V}_\lambda^T \mathbb{B} \mathbb{V}_\lambda, \tag{4.8}$$

where

$$\mathbb{V}_\lambda = \mathbb{I} + a w m^T, \tag{4.9}$$

with $a = \frac{1}{\sqrt{n|\Omega|}}(-1 + \sqrt{1-\rho})$. Moreover, $\mathbb{V}_\lambda$ is invertible with inverse given by

$$\mathbb{V}_\lambda^{-1} = \mathbb{I} + b w m^T, \tag{4.10}$$

where $b = \frac{1}{\sqrt{n|\Omega|}} \cdot \frac{1-\sqrt{1-\rho}}{\sqrt{1-\rho}}$.

Proof. By matrix multiplication and the identities in (4.7), we get

$$\begin{aligned}\mathbb{V}_\lambda^T \mathbb{B} \mathbb{V}_\lambda &= (\mathbb{B} + \frac{a\sqrt{n|\Omega|}}{2\mu} m m^T)(\mathbb{I} + a w m^T) \\ &= \mathbb{B} + \frac{1}{2\mu}(2a\sqrt{n|\Omega|} + a^2 n|\Omega|) m m^T,\end{aligned}$$

and inserting the value of a yields

$$\mathbb{V}_\lambda^T \mathbb{B} \mathbb{V}_\lambda = \mathbb{B} - \frac{\rho}{2\mu} m m^T = \tilde{\mathbb{B}}.$$

This proves (4.8) and further, using the second identity in (4.7), we see that

$$(\mathbb{I} + b w m^T)\mathbb{V}_\lambda = \mathbb{I} + (a + b + ab\sqrt{n|\Omega|}) w m^T.$$

With the given values of a and b the second term vanishes, so (4.10) is proved. □

Lemma 4.2. *Suppose that $\mathbb{D}$ is a preconditioner for $\mathbb{B}$ with condition number $K(\mathbb{D}\mathbb{B})$, then*

$$\tilde{\mathbb{D}} = \mathbb{V}_\lambda^{-1} \mathbb{D} \mathbb{V}_\lambda^{-T}, \tag{4.11}$$

where $\mathbb{V}_\lambda^{-1}$ is given by (4.10), is a preconditioner for $\tilde{\mathbb{B}}$ and $K(\tilde{\mathbb{D}}\tilde{\mathbb{B}}) = K(\mathbb{D}\mathbb{B})$. In particular, the condition number is independent of λ.

Proof. It is seen, using (4.8), that

$$\tilde{\mathbb{D}}\tilde{\mathbb{B}} = \mathbb{V}_\lambda^{-1} \mathbb{D} \mathbb{V}_\lambda^{-T} \mathbb{V}_\lambda^T \mathbb{B} \mathbb{V}_\lambda = \mathbb{V}_\lambda^{-1} \mathbb{D} \mathbb{B} \mathbb{V}_\lambda.$$

We see from this that $\tilde{\mathbb{D}}\tilde{\mathbb{B}}$ and $\mathbb{D}\mathbb{B}$ are similar matrices, and so their eigenvalues coincide. □

Hence, the well-known preconditioners for the weighted $H(\mathbf{div}, \Omega; \mathbb{M})$ inner product $\mathbb{B}$ can be reused such that we obtain a preconditioner spectrally equivalent to $\tilde{\mathbb{B}}$. Furthermore, the preconditioner $\tilde{\mathbb{D}}$ can then be implemented efficiently by applying $\mathbb{V}_\lambda^{-T}$, $\mathbb{D}$, and $\mathbb{V}_\lambda^{-1}$ sequentially. Note that due to the presence of $w m^T$, $\mathbb{V}_\lambda^{-T}$ and $\mathbb{V}_\lambda^{-1}$ are both generally dense matrices. Therefore, the action of $w m^T$ on a vector $v \in \mathbb{R}^N$ should be implemented as $w(m^T v)$, i.e., the dot product with m and a scaling of w.

5. Numerical results

In this section, we present a series of experiments that demonstrate the performance of the proposed preconditioners. In all of following numerical experiments Ω is taken to be the unit square $(0,1)^2$ divided in $N \times N$ squares, where each square is divided in two triangles. The parameters α, $\boldsymbol{\kappa}$, μ, and λ are all constants throughout the domain, unless stated otherwise. We let $\mu = \frac{1}{2}$ be fixed but vary α, $\boldsymbol{\kappa}$, and λ in the experiments. Specifically, in Case 1 we will validate the spectral equivalences (3.4) and (3.7) for both fully clamped- and nonclamped boundary conditions. Case 2 is concerned with a linear elasticity system with weakly imposed symmetry under fully clamped conditions as this represents the hardest case. In Case 3 the full Biot formulation of (3.16) is preconditioned using a preconditioner based on (4.2). As a final numerical experiment, we consider in Case 4 system (3.16) with spatially varying $\boldsymbol{\kappa}$. The tests are conducted using random right-hand sides and initial guesses. Convergence is reached when the square root of the relative preconditioned residual, i.e., $\frac{(Br_k, r_k)}{(Br_0, r_0)}$, where r_k is the residual at the k-th iteration and B is the preconditioner, is below a given tolerance.

Case 1. In the first test case, we show the performance of the preconditioners for the weighted $H(\mathbf{div}, \Omega; \mathbb{M})$ inner product under nonclamped and clamped conditions. That is, for a given right-hand side f_h, we solve the problem: Find $\sigma_h \in \Sigma_h$ such that

$$(A\sigma_h, \tau) + (\mathbf{div}\sigma_h, \mathbf{div}\tau) = (f_h, \tau) \quad \forall \tau \in \Sigma_h. \tag{5.1}$$

λ \ N	4	8	16	32	64
10^{-4}	3	2	2	2	2
10^{-2}	3	2	2	2	2
10^{0}	6	6	5	5	4
10^{2}	13	12	11	9	8
10^{4}	13	13	11	10	8
10^{6}	13	12	11	9	8
10^{8}	12	12	11	10	7
10^{10}	13	13	12	10	8
10^{12}	12	13	11	10	8

(A) $|\Gamma_t| > 0$.

λ \ N	4	8	16	32	64
10^{-4}	3	2	2	2	2
10^{-2}	3	3	2	2	2
10^{0}	6	6	5	5	4
10^{2}	11	11	10	8	7
10^{4}	9	10	10	8	7
10^{6}	9	8	9	8	7
10^{8}	7	7	7	7	7
10^{10}	7	7	7	7	6
10^{12}	7	7	8	8	3

(B) $|\Gamma_t| = 0$.

TABLE 1. Number of iterations for solving (5.1) using preconditioned conjugate gradient method with error tolerance 10^{-9}.

We use piecewise linear, row-wise Brezzi-Douglas-Marini (BDM) elements, as described in [4]. The linear system (5.1) is solved using the preconditioned conjugate gradient method where the choice of preconditioner depends on the boundary conditions. In the case of $|\Gamma_t| > 0$, we use a geometric multigrid procedure with a domain decomposition smoother, c.f. [3]. Subsequently, this preconditioner will be referred to as the AFW preconditioner. When $|\Gamma_t| = 0$, we construct a preconditioner using (4.11) and the AFW preconditioner for $\mathbb{D}$. The results can be viewed in Table 1 where we see that the number of iterations remains bounded as N and λ vary under both clamped and non-clamped boundary conditions.

Case 2. Before testing the preconditioner on the full Biot system, we present some numerical tests on the reduced system of linear elasticity with weakly enforced symmetry. In our notation, this system takes the following form:

For a given f_h, find $(\sigma_h, \boldsymbol{u}_h, \gamma_h) \in \Sigma_h \times \boldsymbol{V}_h \times \Gamma_h$ so that

$$(A\sigma_h, \tau) + (\boldsymbol{u}_h, \mathbf{div}\tau) + (\gamma_h, \tau) = 0 \qquad \forall \tau \in \Sigma_h, \tag{5.2a}$$

$$(\mathbf{div}\sigma_h, \boldsymbol{v}) = -(f_h, \boldsymbol{v}) \qquad \forall \boldsymbol{v} \in \boldsymbol{V}_h, \tag{5.2b}$$

$$(\sigma_h, \eta) = 0 \qquad \forall \eta \in \Gamma_h. \tag{5.2c}$$

For discretization, we can use any of the stable elements for mixed elasticity with weakly enforced symmetry, see e.g., [4]. In particular, in these numerical experiments, we use the same piecewise linear BDM elements for Σ_h as in Case 1, and piecewise constants for $\boldsymbol{V}_h$ and Γ_h. Additionally, we only consider fully clamped conditions in this case. The system (5.2) is stable in the inner products in (3.6) for Σ_h, $\boldsymbol{V}_h$, and Γ_h, respectively. For preconditioning of the Σ_h-block, we again use (4.11) together with the AFW preconditioner for $\mathbb{D}$, and for the $\boldsymbol{V}_h$- and Γ_h blocks we use the inverse of the diagonal elements of the corresponding mass matrices. The numerical results can be seen in Table 2. N denotes the number of squares on one side as before. Again, the number of iterations remains bounded both as N and λ increase.

Case 3. Considering the full Biot system with weakly imposed symmetry (3.16) with fully clamped conditions, we discretize Σ_h, $\boldsymbol{V}_h$, and Γ_h using the same function spaces as in Case 2, and Q_h is the space of piecewise continuous linear functions over the triangulation of Ω. The boundary conditions for the pressure are homogeneous Neumann conditions, i.e., $|\Gamma_p| = 0$, and to remove the singularity, we fix the value of the pore pressure at a single point. The triple $(\Sigma_h, \boldsymbol{V}_h, \Gamma_h)$ is elasticity stable, which ensures the stability of Theorem 3.2, and consequently, we can use a preconditioner based on (4.2). The actual preconditioner is then constructed using geometrical multigrid with Jacobi smoother replacing the second block of (4.2) for the

λ \ N	4	8	16	32	64
10^{-4}	18	19	19	19	19
10^{-2}	18	19	19	19	19
10^{0}	28	28	28	28	28
10^{2}	38	41	40	41	42
10^{4}	35	36	40	41	43
10^{6}	28	31	36	40	43
10^{8}	22	24	31	38	38
10^{10}	20	21	24	35	28

TABLE 2. Numerical result for mixed elasticity with weakly enforced symmetry. Table shows the number of preconditioned minimal residual iterations until reaching error tolerance 10^{-9}.

			N				
κ	α	λ	4	8	16	32	64
10^{0}	10^{0}	10^{0}	24	28	30	40	23
		10^{4}	31	26	48	38	24
		10^{8}	31	35	46	36	26
	10^{-4}	10^{0}	25	30	28	29	26
		10^{4}	29	28	29	39	24
		10^{8}	29	36	43	31	24
10^{-4}	10^{0}	10^{0}	25	25	20	17	12
		10^{4}	29	26	22	18	13
		10^{8}	29	26	23	17	13
	10^{-4}	10^{0}	21	20	17	14	12
		10^{4}	28	28	23	18	13
		10^{8}	28	26	23	17	13
10^{-8}	10^{0}	10^{0}	25	24	20	17	13
		10^{4}	29	26	23	17	13
		10^{8}	29	26	22	17	13
	10^{-4}	10^{0}	22	20	17	14	12
		10^{4}	28	26	22	17	13
		10^{8}	31	28	23	17	13

TABLE 3. Numerical results for preconditioning (3.16). Table shows the number of preconditioned minimal residual iterations until reaching error tolerance 10^{-9}.

pore pressure, while the remaining blocks are treated as in Case 2. The results can be seen in Table 3, where we see that robustness in N and λ continue to hold as well as for $\boldsymbol{\kappa}$ and α.

Case 4. As the final experiment, we again consider (3.16), but now with hydraulic conductivity $\boldsymbol{\kappa} = \kappa\mathbb{I}$, where κ is variable in Ω and defined by

$$\kappa(x,y) = \begin{cases} \kappa, & \text{if } y \in (1/4, 3/4) \\ 1, & \text{otherwise.} \end{cases} \tag{5.3}$$

The results can be seen in Table 4, where we again see robustness in all parameters.

			N				
κ	α	λ	4	8	16	32	64
10^0	10^0	10^0	19	22	30	20	19
		10^4	28	31	36	28	20
		10^8	26	31	24	23	20
	10^{-4}	10^0	21	24	31	25	19
		10^4	26	31	34	23	19
		10^8	28	31	36	28	19
10^{-4}	10^0	10^0	22	25	25	28	17
		10^4	27	31	33	26	16
		10^8	25	31	22	28	16
	10^{-4}	10^0	21	23	29	25	19
		10^4	28	31	32	19	17
		10^8	26	25	24	25	17
10^{-8}	10^0	10^0	22	24	30	19	17
		10^4	25	31	34	28	16
		10^8	27	30	34	28	17
	10^{-4}	10^0	21	22	30	19	19
		10^4	27	31	33	19	19
		10^8	28	31	31	19	17

TABLE 4. Numerical results for system (3.16) with variable κ according to (5.3) using preconditioner based on (4.2). Table shows the number of preconditioned minimal residual iterations until reaching error tolerance 10^{-9}.

6. CONCLUSIONS:

We have proposed a new variational formulation of Biot's consolidation model based on stress, displacement, and pressure, where the symmetry of the stress is imposed weakly. The formulation is robustly bounded and stable in a set of parameter-dependent norms. This motivates two preconditioners of the system, depending on the type of boundary conditions considered. We also show that the parameter-robust stability continues to hold when the elasticity part is discretized with finite element spaces based on mixed linear elasticity with weakly imposed symmetry, leaving a lot of freedom in the choice of discretization of the pressure.

The theoretical results in this work are backed up by a number of numerical experiments, showing robustness in a wide range of values for the shear- and bulk elastic moduli, hydraulic conductivity, as well as time- and space discretization parameters.

APPENDIX A. PROOFS OF (3.3) AND (3.5)

Finally, we provide proofs of (3.3) and (3.5).

Proof of (3.3). Fix $\tau \in \Sigma$ and recall that $|\Gamma_t| > 0$ and $\tau \cdot \nu = 0$ on Γ_t. By the pointwise decomposition $\tau = P_D\tau + (I - P_D)\tau$, and the fact that $(I - P_D)\tau = \frac{1}{n}\operatorname{tr}\tau\mathbb{I}$, it suffices to show that

$$\|\operatorname{tr}\tau\|_0^2 \leq C\left(\left(\frac{1}{2\mu}P_D\tau, P_D\tau\right) + \|\mathbf{div}\tau\|_0^2\right)$$

for some constant C independent of τ. To prove this, we use a well-known result for the right inverse of the divergence operator: There exists $\phi \in H^1_{\Gamma_d}(\Omega;\mathbb{V}) := \left\{\varphi \in H^1(\Omega;\mathbb{V}) : \varphi|_{\Gamma_d} = 0\right\}$ such that

$$\operatorname{div}\phi = \operatorname{tr}\tau, \qquad \|\phi\|_1 \leq C\,\|\operatorname{tr}\tau\|_0 \tag{A.1}$$

with $C > 0$ independent of τ, cf. Appendix B. We then have that

$$||\mathrm{tr}\,\tau||_0^2 = (\mathrm{tr}\,\tau, \mathrm{div}\,\phi) = (\mathrm{tr}\,\tau\mathbb{I}, \mathbf{grad}\phi).$$

Since $\mathrm{tr}\,\tau\mathbb{I} = n(\tau - P_D\tau)$, we get

$$||\mathrm{tr}\,\tau||_0^2 = n(\tau, \mathbf{grad}\phi) - n(P_D\tau, \mathbf{grad}\phi) = -n(\mathbf{div}\tau, \phi) - n(P_D\tau, \mathbf{grad}\phi),$$

where the first term of the final form is a result of integration by parts. Next, we may use Cauchy-Schwarz, which results in

$$\begin{aligned}||\mathrm{tr}\,\tau||_0^2 &\leq n\left(||\mathbf{div}\tau||_0\,||\phi||_0 + ||P_D\tau||_0\,||\mathrm{grad}\,\phi||_0\right)\\ &\leq n\left(||\mathbf{div}\tau||_0^2 + ||P_D\tau||_0^2\right)^{\frac{1}{2}}||\phi||_1\\ &\leq C\left(||\mathbf{div}\tau||_0^2 + ||P_D\tau||_0^2\right)^{\frac{1}{2}}||\mathrm{tr}\,\tau||_0,\end{aligned}$$

and so the result follows after dividing by $||\mathrm{tr}\,\tau||_0$. □

When $|\Gamma_t| = 0$, i.e. $\Gamma_d = \partial\Omega$, (A.1) can only hold if $\mathrm{tr}\,\tau$ has mean value zero. However, with this constraint, we can prove (3.5) with almost the same argument as above.

Proof of (3.5). Fix any $\tau \in \Sigma$. From the decomposition $P_0\tau = P_D\tau + (P_0 - P_D)\tau$, it suffices to prove the estimate for $(P_0 - P_D)\tau$ component. Denoting the mean value of the trace by

$$\overline{\mathrm{tr}\,\tau} := \frac{1}{|\Omega|}\int_\Omega \mathrm{tr}\,\tau \mathrm{d}x,$$

we have that $(I - P_D)P_0\tau = (P_0 - P_D)\tau = \frac{1}{n}(\mathrm{tr}\,\tau - \overline{\mathrm{tr}\,\tau})\mathbb{I}$, and so it is sufficient to show that

$$||\mathrm{tr}\,\tau - \overline{\mathrm{tr}\,\tau}||_0^2 \leq C\left(\left(\frac{1}{2\mu}P_D\tau, P_D\tau\right) + ||\mathbf{div}\tau||_0^2\right).$$

Since $\mathrm{tr}\,\tau - \overline{\mathrm{tr}\,\tau}$ is mean-value zero, there exists $\phi \in \{\varphi \in H^1(\Omega;\mathbb{V}) : \varphi|_{\partial\Omega} = 0\}$ such that

$$\mathrm{div}\,\phi = \mathrm{tr}\,\tau - \overline{\mathrm{tr}\,\tau}, \qquad ||\phi||_1 \leq C\,||\mathrm{tr}\,\tau - \overline{\mathrm{tr}\,\tau}||_0$$

with $C > 0$ independent of τ, (cf. [14, Theorem 5.1]). The rest of the proof is completely analagous to the proof of (3.3) above. □

Appendix B. Right inverse of divergence operator

A result for the right inverse of the divergence operator, as expressed by (A.1), is closely related to the inf-sup condition for the Stokes problem, and therefore well-known. However, we are not aware of a proper reference for the case when $|\partial\Omega| > |\Gamma_t| > 0$, i.e. for the case when $|\Gamma_d| > 0$, but Γ_d is not all of $\partial\Omega$. Therefore, for completeness, we include a proof here.

Lemma B.1. *Assume* $|\Gamma_t| > 0$ *and set* $H^1_{\Gamma_d}(\Omega;\mathbb{V}) = \left\{\phi \in H^1(\Omega;\mathbb{V}) : \phi|_{\Gamma_d} = 0\right\}$. *Then there is a constant* $C > 0$ *so that for every* $f \in L^2(\Omega)$ *there is a* $\phi \in H^1_{\Gamma_d}(\Omega;\mathbb{V})$ *so that*

$$\mathrm{div}\,\phi = f, \quad ||\phi||_1 \leq C\,||f||_0.$$

Proof. Take any $f \in L^2(\Omega)$. We first decompose f into its mean value zero- and mean value part as $f = f_0 + f_c$ where $f_0 \in L^2_0(\Omega)$ and $f_c = a_f 1_\Omega$ for $a_f \in \mathbb{R}$. Further, we can decompose $H^1_{\Gamma_d}(\Omega;\mathbb{V}) = H^1_0(\Omega;\mathbb{V}) \oplus V_1$, where

$$V_1 := \left\{\phi \in H^1_{\Gamma_d}(\Omega;\mathbb{V}) : (\mathbf{grad}\phi, \mathbf{grad}\psi) = 0,\ \forall \psi \in H^1_0(\Omega;\mathbb{V})\right\}.$$

Consider then the problem of finding $\zeta \in V_1$ so that

$$(\mathbf{grad}\zeta, \mathbf{grad}\psi) = (\mathbb{I}, \mathbf{grad}\psi),\ \forall \psi \in V_1. \tag{B.1}$$

By the Lax-Milgram lemma (cf. e.g., [8, Theorem 4.1.6]), problem (B.1) has a unique solution ζ and $||\zeta||_1 \leq C_1$ for some constant $C_1 > 0$ depending on Ω. Taking $\psi = \zeta$ in (B.1), we obtain

$$\int_\Omega \operatorname{div} \zeta \, dx = ||\mathbf{grad}\zeta||_0^2.$$

Therefore, if we set $\omega = \frac{a_f}{||\mathbf{grad}\zeta||_0^2}\zeta$ we have $\int_\Omega \operatorname{div} \omega dx = a_f$ and $||\omega||_1 \leq C\,||f_c||_0$ for some constant C depending on ζ. It follows that $f - \operatorname{div}\omega \in L_0^2(\Omega)$, i.e., $f - \operatorname{div}\omega$ has mean value zero. From the theory of Stokes equation, we can thus find a $\omega_0 \in H_0^1(\Omega;\mathbb{V})$ so that

$$\operatorname{div}\omega_0 = f - \operatorname{div}\omega, \quad ||\omega_0||_1 \leq C_2\,||f - \operatorname{div}\omega||_0, \tag{B.2}$$

where the constant C_2 is independent of $f - \operatorname{div}\omega$ (cf. [14, Theorem 5.1]). We set $\phi = \omega_0 + \omega$, and it follows from (B.2), that $\operatorname{div}\phi = f$. Using the triangle inequality, (B.2) and the properties of ω we estimate $||\phi||_1$ as

$$||\phi||_1 \leq ||\omega_0||_1 + ||\omega||_1 \leq C(||f - \operatorname{div}\omega||_0 + ||f_c||_0) \leq C(||f||_0 + ||\omega||_1) \leq C\,||f||_0,$$

which completes the proof. □

References

[1] D. N. Arnold, F. Brezzi, and J. Douglas, Jr. PEERS: a new mixed finite element for plane elasticity. *Japan J. Appl. Math.*, 1(2):347–367, 1984.

[2] D. N. Arnold, J. Douglas, Jr., and C. P. Gupta. A family of higher order mixed finite element methods for plane elasticity. *Numer. Math.*, 45(1):1–22, 1984.

[3] D. N. Arnold, R. S. Falk, and R. Winther. Preconditioning in $H(\mathrm{div})$ and applications. *Math. Comp.*, 66(219):957–984, 1997.

[4] D. N. Arnold, R. S. Falk, and R. Winther. Mixed finite element methods for linear elasticity with weakly imposed symmetry. *Math. Comp.*, 76(260):1699–1723, 2007.

[5] O. Axelsson, R. Blaheta, and P. Byczanski. Stable discretization of poroelasticity problems and efficient preconditioners for arising saddle point type matrices. *Comput. Vis. Sci.*, 15(4):191–207, 2012.

[6] L. Berger, R. Bordas, D. Kay, and S. Tavener. Stabilized lowest-order finite element approximation for linear three-field poroelasticity. *SIAM J. Sci. Comput.*, 37(5):A2222–A2245, 2015.

[7] D. Boffi, F. Brezzi, and M. Fortin. Reduced symmetry elements in linear elasticity. *Commun. Pure Appl. Anal.*, 8(1):95–121, 2009.

[8] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013.

[9] J. H. Bramble. *Multigrid methods*, volume 294 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1993.

[10] Y. Chen, Y. Luo, and M. Feng. Analysis of a discontinuous Galerkin method for the Biot's consolidation problem. *Appl. Math. Comput.*, 219(17):9043–9056, 2013.

[11] B. Cockburn, J. Gopalakrishnan, and J. Guzmán. A new elasticity element made for enforcing weak stress symmetry. *Math. Comp.*, 79(271):1331–1349, 2010.

[12] O. Coussy. *Poromechanics*. John Wiley & Sons, 2004.

[13] M. Farhloul and M. Fortin. Dual hybrid methods for the elasticity and the Stokes problems: a unified approach. *Numer. Math.*, 76(4):419–440, 1997.

[14] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.

[15] J. Gopalakrishnan and J. Guzmán. A second elasticity element using the matrix bubble. *IMA J. Numer. Anal.*, 32(1):352–372, 2012.

[16] J. B. Haga, H. Osnes, and H. P. Langtangen. A parallel block preconditioner for large-scale poroelasticity with highly heterogeneous material parameters. *Comput. Geosci.*, 16(3):723–734, 2012.

[17] R. Hiptmair and J. Xu. Nodal auxiliary space preconditioning in $\mathbf{H}(\mathbf{curl})$ and $\mathbf{H}(\mathrm{div})$ spaces. *SIAM J. Numer. Anal.*, 45(6):2483–2509, 2007.

[18] J. J. Lee. Robust error analysis of coupled mixed methods for Biot's consolidation model. *J. Sci. Comput.*, 69(2):610–632, 2016.

[19] J. J. Lee. Towards a unified analysis of mixed methods for elasticity with weakly symmetric stress. *Adv. Comput. Math.*, 42(2):361–376, 2016.

[20] J. J. Lee, K.-A. Mardal, and R. Winther. Parameter-robust discretization and preconditioning of Biot's consolidation model. *SIAM J. Sci. Comput.*, 39(1):A1–A24, 2017.

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

[22] M. A. Murad and A. F. D. Loula. Improved accuracy in finite element analysis of Biot's consolidation problem. *Comput. Methods Appl. Mech. Engrg.*, 95(3):359–382, 1992.

[23] M. A. Murad and A. F. D. Loula. On stability and convergence of finite element approximations of Biot's consolidation problem. *Internat. J. Numer. Methods Engrg.*, 37(4):645–667, 1994.

[24] P. J. Phillips and M. F. Wheeler. A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity. *Comput. Geosci.*

[25] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity I: the continuous in time case. *Comput. Geosci.*, 11(2):131–144, 2007.

[26] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity II: the discrete-in-time case. *Comput. Geosci.*, 11(2):145–158, 2007.

[27] M. B. Reed. An investigation of numerical errors in the analysis of consolidation by finite elements. *Internat. J. Numer. Analyt. Methods Geomech.*, 8(3):243–257, 1984.

[28] S. Rhebergen, G. N. Wells, A. J. Wathen, and R. F. Katz. Three-field block preconditioners for models of coupled magma/mantle dynamics. *SIAM J. Sci. Comput.*, 37(5):A2270–A2294, 2015.

[29] R. E. Showalter. Diffusion in poro-elastic media. *J. Math. Anal. Appl.*, 251(1):310–340, 2000.

[30] J. H. Smith and J. A. Humphrey. Interstitial transport and transvascular fluid exchange during infusion into brain and tumor tissue. *Microvasc. Res.*, 73(1):58–73, 2007.

[31] R. Stenberg. A family of mixed finite elements for the elasticity problem. *Numer. Math.*, 53(5):513–538, 1988.

[32] K. H. Støverud, M. Alnæs, H. P. Langtangen, V. Haughton, and K.-A. Mardal. Poro-elastic modeling of syringomyelia – a systematic study of the effects of pia mater, central canal, median fissure, white and gray matter on pressure wave propagation and fluid movement within the cervical spinal cord. *Comput. Methods Biomech. Biomed. Engin.*, 19(6):686–698, 2016. PMID: 26176823.

[33] P. A. Vermeer and A. Verruijt. An accuracy condition for consolidation by finite elements. *Internat. J. Numer. Analyt. Methods Geomech.*, 5(1):1–14, 1981.

[34] H. F. Wang. Theory of linear poroelasticity. *Princeton Series in Geophysics, Princeton University Press, Princeton, NJ*, 2000.

[35] S.-Y. Yi. A coupling of nonconforming and mixed finite element methods for Biot's consolidation model. *Numer. Methods Partial Differ. Equ.*, 29(5):1749–1777, 2013.

[36] S.-Y. Yi. Convergence analysis of a new mixed finite element method for Biot's consolidation model. *Numer. Methods Partial Differ. Equ.*, 30(4):1189–1210, 2014.

[37] O. C. Zienkiewicz and T. Shiomi. Dynamic behaviour of saturated porous media; the generalized Biot formulation and its numerical solution. *Internat. J. Numer. Analyt. Methods Geomech.*, 8(1):71–96, 1984.

Paper II

Multigrid Methods for Discrete Fractional Sobolev Spaces

T. Bærland, M. Kuchta, and K.-A. Mardal

SIAM Journal on Scientific Computing, (2018). Submitted.

MULTIGRID METHODS FOR DISCRETE FRACTIONAL SOBOLEV SPACES

TRYGVE BÆRLAND[†], MIROSLAV KUCHTA[†], AND KENT-ANDRE MARDAL[†]

ABSTRACT. Coupled multiphysics problems often give rise to interface conditions naturally formulated in fractional Sobolev spaces. Here, both positive and negative fractionality are common. When designing efficient solvers for discretizations of such problems it would then be useful to have a preconditioner for the fractional Laplacian. In this work, we develop an additive multigrid preconditioner for the fractional Laplacian with positive fractionality, and show a uniform bound on the condition number. For the case of negative fractionality, we re-use the preconditioner developed for the positive fractionality and left-right multiply a regular Laplacian with a preconditioner with positive fractionality to obtain the desired negative fractionality. Implementational issues are outlined in details as the differences between the discrete operators and their corresponding matrices must be addressed when realizing these algorithms in code. We finish with some numerical experiments verifying the theoretical findings.

1. INTRODUCTION

Multiphysics or multiscale problems often involve coupling conditions at interfaces which are manifolds of lower dimensions. The coupling conditions are, because of the lower dimensionality, naturally posed in fractional Sobolev spaces, and this fact seemingly complicates discretization schemes and solution algorithms. Our focus here will be on the development of solution algorithms in terms of multilevel preconditioners that from an implementational point of view only require minor adjustments of standard multilevel algorithms.

As simplified examples of problems involving interface conditions, let us consider the following two prototype problems. First an elliptic problem with a trace constraint

$$\begin{aligned} -\Delta u + T^*\lambda &= f, \quad x \in \Omega, \\ Tu &= g, \quad x \in \Gamma, \end{aligned} \tag{1.1}$$

and second an elliptic problem in mixed form with a trace constraint

$$\begin{aligned} u - \nabla p + T^*\lambda &= f, \quad x \in \Omega, \\ \nabla \cdot u &= g, \quad x \in \Omega, \\ Tu &= h, \quad x \in \Gamma. \end{aligned} \tag{1.2}$$

Here, Γ is a sub-manifold either within Ω or at its boundary, T is a trace operator and T^* its adjoint. Both problems are assumed to be equipped with suitable boundary conditions. We remark that although these problems are single physics problems, they may easily be coupled to other problems through the Lagrange multiplier at the interface. As such, the problems represent well the challenge of handling the interface properly in a multiphysics setting.

We may write the above problems as

$$\begin{pmatrix} -\Delta & T^* \\ T & 0 \end{pmatrix}\begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} I & -\nabla & T^* \\ \nabla\cdot & 0 & 0 \\ T & 0 & 0 \end{pmatrix}\begin{pmatrix} u \\ p \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ g \\ h \end{pmatrix}.$$

[†]DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OSLO, BLINDERN, OSLO, 0316 NORWAY
E-mail addresses: trygveba@math.uio.no, mirok@math.uio.no, kent-and@math.uio.no.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339643. Work of MK has been funded by Research Council of Norway through grant 'RigSpray' (grant no. 421160).

A crucial challenge is to discretize and solve these problems in a scalable way such that the computations scale linearly with the number of unknowns. Our approach here is to consider iterative methods and develop preconditioners that are both spectrally equivalent with the involved operators and of order-optimal complexity. The main difficulty is the handling of the Lagrange multiplier which falls outside the scope of standard multilevel methods. To provide a general framework, we will consider preconditioners constructed in terms of the so-called operator preconditioning approach [32] to be used for iterative methods. As will be explained later, the block diagonal preconditioners constructed by this technique will be of the following form:

$$\begin{pmatrix} -\Delta^{-1} & 0 \\ 0 & (-\Delta)^{\frac{1}{2}} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} (I-\nabla\nabla\cdot)^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & (-\Delta)^{-\frac{1}{2}} \end{pmatrix},$$

respectively. Multilevel methods spectrally equivalent with both $(-\Delta)^{-1}$ and $(I-\nabla\nabla\cdot)^{-1}$ are well known. The challenging part in both cases is the construction of efficient preconditioning algorithms that approximate the inverse of the fractional Laplace problems on the form

$$(-\Delta)^s u = f, \quad x \in \Gamma \tag{1.3}$$

with $s = 1/2$ and $s = -1/2$, equipped with suitable boundary conditions. Furthermore, if Γ is of codimension 2, numerical simulations [26] indicate that $s \in (-0.2, -0.1)$ gives rise to efficient preconditioners. In this paper we therefore consider methods for $s \in [-1, 1]$.

There are many examples of applications of fractional Laplacians in the literature and we mention a few that motivate this work. Non-overlapping domain decomposition preconditioners are studied in [3], [25]. Here, they use (1.3) with $s = \frac{1}{2}$ to precondition the interface problem involving the related Steklov-Poincaré operator. In [27] the authors use (1.3) with $s = -\frac{1}{2}$ as part of a block diagonal preconditioner for multiphysics problem where the constraint coupling two domains of different topological dimension is enforced by the Lagrange multiplier. Therein the fractionality s is dictated by the mapping properties of the Schur complement operator. Some further examples of coupled systems with domains of different dimensionality include Babuška's problem for enforcing Dirichlet boundary conditions on an elliptic operator [5], flow stabilization by removal of tangential velocity at the boundary through Lagrange multipliers [8], the no-slip condition on the surface of a falling solid in the Navier-Stokes fluid [17], inextensibility constraint in the complex model of vesicle formation [1], and the potential jump on a membrane of a cardiac cell [35]. We note that in these applications the fractional Laplace problem has to be solved with both positive and negative exponent.

There are several alternative approaches that have been used in order to approximate fractional Laplacians. Polynomial approximations of A^s, where A is a discrete Laplacian, can be computed with standard Krylov subspace methods. However, without any preconditioner a Krylov subspace of large dimension is required for convergence, see e.g. Lanczos method in [24, Section 4]. Preconditioners based on fractional powers of A on preconditioned problems on small subspaces have been shown efficient for various applications in [3, 38]. The contour integral method of [19] and the extended Krylov method of [24] are here related to rational function approximations of A^s, while [22] consider the best uniform rational aproximations of the trasformed function $A \mapsto A^{\beta-s}$. In general, the approximation properties of these methods depend on the condition number of A and thus computations of extremal eigenvalues are often part of the algorithm. Further, the computational complexity of the methods based on rational approximations depends on efficient solvers for auxiliary linear systems, e.g. $(A - q_k I)x = b$ in [22] where $q_k \in \mathbb{R}$ is a shift parameter. Almost mesh independent preconditioners for systems arising in [19] and [24] are discussed in [16]. An alternative approach to the matrix transfer method is presented in [9] where the inverse of the fractional Laplacian is defined via the (integral) Balakrishnan formula [6].

Multilevel methods for fractional Laplacians have been considered in [12, 22, 33, 34], but there seems to be a significant untapped potential for advancement. Our work here is closely related to [12], where order-optimal preconditioners for A^s when $s \in \left(-\frac{3}{2}, \frac{3}{2}\right)$ were constructed using a hierarchical basis approach. The paper did, however, only consider smoothers based on level-dependent scaling and did not put much focus on the actual implementation. Here, we will develop and analyze a multilevel algorithm that is straightforward to implement in a standard multilevel software framework. In fact, the main change required is an adjustment of the smoothers. To illustrate the change, let us assume that we want to solve the system $\mathsf{A}\mathsf{x} = \mathsf{b}$, where A is a stiffness matrix corresponding to a discretized Laplacian. A standard Jacobi algorithm can then be written

$$\mathsf{x}_i^{n+1} = \mathsf{x}_i^n - \frac{1}{\mathsf{A}_{i,i}} \mathsf{r}_i^n,$$

where $\mathsf{A}_{i,i}$ are the diagonal entries of the stiffness matrix for a discretized Laplacian, and r^n is the residual of the n'th iterate, x^n . In our case, for $\mathsf{A}^s\mathsf{x} = \mathsf{b}$, the proposed Jacobi smoother may be implemented as

$$\mathsf{x}_i^{n+1} = \mathsf{x}_i^n - \left(\frac{1}{\mathsf{M}_{i,i}^{1-s} \mathsf{A}_{i,i}^s} \right) \mathsf{r}_i^n.$$

Here, $\mathsf{M}_{i,i}$ are the diagonal entries of the mass matrix. We notice here that for $s = 0$ the action is a Jacobi iteration on the mass matrix, and for $s = 1$ the action is a Jacobi iteration on the stiffness matrix, and for $0 < s < 1$ the action is an interpolation between these two extremes. From an implementational point of view, the restriction and interpolation operators used are the same as those used in standard multilevel algorithms. However, from a theoretical point of view, the fact that we use standard restriction and interpolation operators, means that the multilevel approach will be non-nested. In fact, the matrices on coarser levels do not correspond to $(-\Delta)^s$-Galerkin projections of the matrix on the finer levels. We therefore employ the framework of non-nested multilevel methods [15]. Furthermore, a multiplicative multilevel algorithm would require computing the residual and hence the evaluation of the exact $(-\Delta)^{-s}$ operator on every level. Since the evaluation of the exact $(-\Delta)^{-s}$ is a computationally expensive procedure, we instead rely on the additive multilevel algorithm proposed in [14], where the same residual is used on all levels. The additive variant is significantly less efficient than corresponding multiplicative variants in terms of the conditioning (in the sense that the conditioning depends on the number of levels). Still, this is a small price to pay (only logarithmic in the number of unknowns) to avoid exact evaluation of the residual. In this paper we will assume quasi-uniform mesh and continuous piecewise linear finite elements. This is mainly for simplicity, and the results can be generalized to higher order discretizations, as well as discontinuous Galerkin methods.

The paper is structured as follows. In section 2, we introduce notation, and some useful operator inequalities related to fractional powers of positive operators. We also give a brief discussion of fractional Sobolev spaces. Section 3 is devoted to the analysis of an abstract multilevel framework. In section 4 we use this framework to define operators that are spectrally equivalent to the inverse of the fractional Laplacian when the fractionality $s \geq 0$. We discuss some strategies for preconditioning when $s < 0$ in section 5, and in section 6 we discuss implementation of the preconditioners developed in the previous sections. Finally, we provide numerical results that verify our theoretical result in section 7.

2. Notation and preliminaries

Let Ω be a bounded, Lipschitz domain in $\mathbb{R}^n$, with boundary $\partial\Omega$. We denote by $L^2(\Omega)$ the space of square integrable functions over Ω, with inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$. For $k \in \mathbb{N}$, we denote by $H^k(\Omega)$ the usual Sobolev spaces of functions in $L^2(\Omega)$ with all derivatives up to order k in $L^2(\Omega)$. The norm and inner product in H^k is denoted by $\|\cdot\|_k$ and $(\cdot,\cdot)_k$,

respectively. The closure in H^k of smooth functions with compact support in Ω is denoted as $H_0^k(\Omega)$ and its dual space is H^{-k}. In general a Hilbert space X is equipped with a norm $||\cdot||_X$ and an inner product $(\cdot,\cdot)_X$ and the dual space is denoted X'. For two Hilbert spaces X and Y, we write $\mathscr{L}(X,Y)$ to mean the space of bounded linear operators $T : X \to Y$, which we equip with the usual operator norm

$$||T||_{\mathscr{L}(X,Y)} = \sup_{x\in X} \frac{||Tx||_Y}{||x||_X}.$$

Let A be a symmetric positive-definite operator on a finite-dimensional Hilbert space X with dimension N. Denote by $\{(\lambda_k,\phi_k)\}_{k=1}^N$ the set of eigenpairs of A, normalized so that

$$(\phi_k,\phi_l)_X = \delta_{k,l},$$

where $\delta_{k,l}$ is the Kronecker delta. Then ϕ_k, for $k = 1,\dots,N$ forms an orthonormal basis of X, and if $u \in X$ has the representation $u = \sum_{k=1}^N c_k\phi_k$, then

$$Au = \sum_{k=1}^N \lambda_k c_k \phi_k.$$

For $s \in \mathbb{R}$, we define the fractional power A^s of A by

$$A^s u = \sum_{k=1}^N \lambda_k^s c_k \phi_k.$$

If A is only positive semi-definite, then we must restrict to $s \geq 0$, and the eigenvectors corresponding to the nullspace of A are left out (also for $s = 0$). If B is another symmetric positive semi-definite operator on X, we write $A \leq B$ if for every $u \in X$

$$(Au,u)_X \leq (Bu,u)_X$$

holds. Note that $0 \leq A$ is equivalent to saying that A is positive semi-definite.

A result in operator theory is the Löwner-Heinz inequality, which states that if $A \leq B$, then

$$A^s \leq B^s, \quad s \in [0,1], \tag{2.1}$$

cf. for instance [23]. Inequality (2.1) means that the function x^s with $x \in [0,\infty)$ is operator monotone for $s \in [0,1]$. It follows that $-(x)^s$ is operator convex (cf. [20]), that is, for any two symmetric positive semi-definite operators A and B on a Hilbert space X, the inequality

$$\lambda A^s + (1-\lambda)B^s \leq (\lambda A + (1-\lambda)B)^s$$

holds for every $\lambda \in [0,1]$. A key result regarding operator convex functions is the Jensen's operator inequality (cf. [21, Theorem 2.1]). The version we will use in the current work states that for any $K \in \mathbb{N}$ and $s \in [0,1]$

$$\sum_{k=1}^K P_k^* A_k^s P_k \leq \left(\sum_{k=1}^K P_k^* A_k P_k\right)^s, \tag{2.2}$$

where for $k = 1,\dots,K$, A_k are symmetric positive semi-definite operators on X, and P_k are linear operators on X so that $\sum_{k=1}^K P_k^* P_k \leq I$ and I is the identity operator on X.

2.1. **Fractional Sobolev spaces.** We consider the interpolation spaces between $H^1(\Omega)$ and $L^2(\Omega)$ as defined in [30]. Let the inner product on $H^1(\Omega)$ be realized by the operator $A := I-\Delta$, as

$$(u,v)_1 = (Au,v) = (u,v) + (\nabla u,\nabla v), \quad u,v \in H^1(\Omega).$$

A is unbounded as an operator mapping $L^2(\Omega)$ to $L^2(\Omega)$. However, A is well-defined on the set

$$D(A) = \left\{u \in L^2(\Omega) : Au \in L^2(\Omega)\right\},$$

which is a dense subspace of $L^2(\Omega)$. On $D(A)$, A is symmetric and positive-definite, and so the fractional powers of A, A^θ for $\theta \in \mathbb{R}$, are well-defined. Note that in the particular case $\theta = \frac{1}{2}$,

$$\left\| A^{\frac{1}{2}} u \right\|^2 = (Au, u) = \|u\|_1 .$$

For $s \in [0,1]$, we define the fractional Sobolev spaces as

$$H^s(\Omega) = \left\{ u \in L^2(\Omega) : A^{\frac{s}{2}} u \in L^2(\Omega) \right\}, \tag{2.3}$$

which is a Hilbert space with inner product given by

$$(u,v)_s = (A^s u, v), \quad u, v \in H^s(\Omega),$$

and we denote the corresponding norm by $\|\cdot\|_s$.

We define $H_0^s(\Omega)$ to be the closure of $C_0^\infty(\Omega)$, the space of infinitely smooth functions with compact support in Ω, in the norm of $H^s(\Omega)$. We note that if $s \leq \frac{1}{2}$, the spaces $H_0^s(\Omega)$ and $H^s(\Omega)$ coincide (cf. [30, Theorem 11.1]). For $s \in [-1,0]$, we define a family of fractional Sobolev spaces using the dual of $H_0^s(\Omega)$. That is,

$$H^s(\Omega) = \left(H_0^{-s}(\Omega) \right)' .$$

Replacing $H^1(\Omega)$ with $H_0^1(\Omega)$ and setting $A = -\Delta$ in the above construction, will again yield the space $H_0^s(\Omega)$, with equivalent norm, for all s except when $s = \frac{1}{2}$. In this case, interpolation between $H_0^1(\Omega)$ and $L^2(\Omega)$ results in a space that is strictly contained in $H_0^{\frac{1}{2}}(\Omega)$. The subsequent analysis is valid for both $H_0^s(\Omega)$ and $H^s(\Omega)$.

We remark that the above defined fractional space $H^s(\Omega)$ is equivalent to the fractional space $\hat{H}^s(\Omega)$ defined in terms of the norm

$$\|u\|^2_{\hat{H}^s(\Omega)} = \|u\|^2 + \int_{\Omega\times\Omega} \frac{|u(x)-u(y)|^2}{|x-y|^{n+2s}} \, \mathrm{d}x \mathrm{d}y.$$

A detailed overview of the various definitions of fractional Sobolev norms and their discretizations can be found in [31].

2.2. **Discrete fractional Sobolev spaces.** We will now consider a discretization of the fractional Sobolev spaces $H_0^s(\Omega)$ and $H^{-s}(\Omega)$ for $s \in [0,1]$. Let X_h be a finite-dimensional subspace of $H_0^1(\Omega)$, with $\dim X_h = N_h$. We define the operator $A_h : X_h \to X_h$ by

$$(A_h u, v) = (\nabla u, \nabla v), \quad u, v \in X_h. \tag{2.4}$$

Using the fractional powers of A_h, we define for $s \in \mathbb{R}$ the discrete fractional inner product on X_h by

$$(u,v)_{s,h} = \left(A_h^s u, v \right), \quad u, v \in X_h,$$

and denoted the corresponding norm by $\|\cdot\|_{s,h}$. It is clear that for $s = 0$ and $s = 1$, the two norms $\|\cdot\|_{s,h}$ and $\|\cdot\|_s$ coincide on X_h. Therefore, due to [4, Lemma 2.3], the norms $\|\cdot\|_{s,h}$ and $\|\cdot\|_s$, when $s \in [0,1]$, are equivalent on X_h, with constants of equivalence independent of N_h.

Let X_H be a subspace of X_h, and $A_H : X_H \to X_H$ be defined analogously to A_h in (2.4). If $I_H : X_H \to X_h$ is the inclusion map, we see that

$$A_H = I_H^* A_h I_H, \tag{2.5}$$

where I_H^* is the adjoint of I_H with respect to the L^2 inner product.

We may also define $A_H^s : X_H \to X_H$, but generally, $A_H^s \neq I_H^* A_h^s I_H$. However, by Jensen's operator inequality we have the following.

Lemma 2.1. *For every $s \in [0,1]$ we have*

$$I_H^* A_h^s I_H \leq A_H^s.$$

That is, for every $u \in X_H$,

$$\left(A_h^s u, u\right) \leq \left(A_H^s u, u\right). \tag{2.6}$$

Proof. For $s = 0$ and $s = 1$, (2.6) holds with equality, so let $0 < s < 1$. We start by noticing that since $I_H^* I_H$ is the identity on X_H,

$$A_H^2 = (I_H^* I_H A_H I_H^* I_H)^2 = I_H^* (I_H A_H I_H^*)^2 I_H.$$

By induction, we find that

$$A_H^k = I_H^* (I_H A_H I_H^*)^k I_H$$

for every nonnegative integer k. It follows that for any polynomial p

$$p(A_H) = I_H^* \, p\,(I_H A_H I_H^*)\, I_H. \tag{2.7}$$

Take now $\epsilon > 0$. The spectra of both A_H and $I_H A_H I_H^*$ are contained in some bounded, nonnegative interval $[0, b]$. By Weierstrass' approximation Theorem we can thus choose a polynomial p so that

$$||(I_H A_H I_H^*)^s - p(I_H A_H I_H^*)|| < \epsilon, \quad \text{and} \quad ||A_H^s - p(A_H)|| < \epsilon.$$

Using the triangle inequality and (2.7) we have that

$$||A_H^s - I_H^* \, (I_H A_H I_H^*)^s \, I_H|| \leq ||A_H^s - p(A_H)|| + ||I_H^* \, ((I_H A_H I_H^*)^s - p(I_H A_H I_H^*))\, I_H|| < 2\epsilon,$$

and since ϵ was arbitrary, this shows that

$$A_H^s = I_H^* (I_H A_H I_H^*)^s I_H. \tag{2.8}$$

Using (2.5) in (2.8), we get that

$$A_H^s = I_H^* (I_H I_H^* A_h I_H I_H^*)^s I_H. \tag{2.9}$$

Finally, $I_H I_H^*$ defines a symmetric operator on X_h with L^2 operator norm equal to 1. Since the function $x \mapsto -x^s$ is operator convex on $[0, \infty)$ we can use Jensen's operator inequality (2.2) in (2.9) to get

$$\begin{aligned} A_H^s &\geq I_H^* I_H I_H^* A_h^s I_H I_H^* I_H \\ &= I_H^* A_h^s I_H, \end{aligned}$$

where we have used that $I_H^* I_H$ is the identity on X_H. □

3. Abstract multilevel theory

In order to analyze and implement a multigrid preconditioner for the fractional Laplacian there are three main issues that need to be dealt with. First, we need to derive and implement a smoother with the desired properties. As already mentioned in the introduction, this step only requires a minor modification to standard smoothing algorithms. We will discuss the details concerning implementation later. Second, the restriction/interpolation operators do not result in a nested hierarchy of operators in our fractional setting as $A_H^s \neq I_H^* A_h^s I_H$. For this reason we will employ the framework for non-nested multilevel algorithms developed in [15]. Third, our main motivation for developing fractional multilevel solvers is their application to multiphysics and multiscale problems where the preconditioner for the fractional Laplacian is utilized at the interfaces. As such, the fractional Laplacian operator is not part of the original problem and we may therefore not assume that this operator has been implemented. Furthermore, implementing this operator in an efficient manner is a challenge, but is currently a very active research field, c.f. e.g [31] for an overview. To avoid the application of the fractional Laplacian on the various levels we employ additive multilevel schemes which enable the residual of the problem to be used at all levels and remove the need for implementing a global fractional Laplacian operator. That said, the theory developed here extends to multiplicative

algorithms for problems involving the fractional Laplacian, such as the standard V-cycle. In this section we will address the second and the third issues and outline a theory for an additive multilevel scheme applied to an abstract non-nested problem. As such, the analysis of this section is a synthesis of the papers [14] and [15].

Assume that we are given a nested sequence of finite-dimensional function spaces

$$V_1 \subset V_2 \subset \cdots \subset V_J = V, \quad J \geq 2.$$

We further assume that V, and consequently all subspaces of V, is endowed with an inner product $(\cdot,\cdot)$, with corresponding induced norm $\|\cdot\|$. Moreover, for each $k = 1,\ldots,J$, we assume that we are given a symmetric positive definite operator $A_k : V_k \to V_k$, and we set $A = A_J$. Note that we do not assume that the A_k operators are nested.

For the development and analysis of our multilevel algorithm, it will be useful to define a number of operators on each level k. First, we define $P_{k,k-1} : V_k \to V_{k-1}$ by

$$\left(A_{k-1}P_{k,k-1}v, w\right) = (A_k v, w), \quad \forall v \in V_k, w \in V_{k-1}. \tag{3.1}$$

We remark that in a nested setting, $P_{k,k-1}$ is the A-projection, while since the A_k operators are not nested, the $P_{k,k-1}$ operators are not projections. Next, we define $Q_k : V \to V_k$ by

$$(Q_k v, w) = (v, w), \quad \forall v \in V, w \in V_k. \tag{3.2}$$

It follows by the above definitions that

$$A_{k-1}P_{k,k-1} = Q_{k-1}A_k, \tag{3.3}$$

and $Q_l Q_k = Q_k Q_l = Q_l$ whenever $l \leq k$. For the sake of brevity, it will also be useful to define $P_k : V \to V_k$ by $P_k = P_{k+1,k}P_{k+2,k+1}\cdots P_{J,J-1}$. Using the definition of $P_{j+1,j}$ for $j = k,\ldots,J-1$ we see that

$$(A_k P_k v, w) = (Av, w), \quad \forall v \in V, w \in V_k.$$

Furthermore, applying (3.3) repeatedly, we find that

$$A_k P_k = Q_k A. \tag{3.4}$$

Finally, suppose we are given for each k a smoother, which is a symmetric positive definite operator $R_k : V_k \to V_k$ and in some sense should approximate A_k^{-1} on $V_k \backslash V_{k-1}$. We can now define an additive multilevel operator $B : V \to V$ by

$$B = \sum_{k=1}^{J} R_k Q_k. \tag{3.5}$$

As remarked in [14], B can be viewed as an additive version of the standard multiplicative V-cycle multigrid algorithm, where R_k plays the role of smoother. Because of this, it is reasonable that the assumptions we need to make to establish spectral equivalence between A^{-1} and B are similar to those made for standard multigrid algorithms.

We assume that for $k = 2,\ldots,J$

$$(A_k v, v) \leq \left(A_{k-1}v, v\right), \quad \forall v \in V_{k-1}. \tag{A.1}$$

Under assumption (A.1) and the definition of $P_{k,k-1}$ we see that for any $v \in V_k$

$$\begin{aligned}\left(A_{k-1}P_{k,k-1}v, P_{k,k-1}v\right) &= \left(A_k v, P_{k,k-1}v\right) \\ &\leq \left(A_k P_{k,k-1}v, P_{k,k-1}v\right)^{\frac{1}{2}} (A_k v, v)^{\frac{1}{2}} \\ &\leq \left(A_{k-1}P_{k,k-1}v, P_{k,k-1}v\right)^{\frac{1}{2}} (A_k v, v)^{\frac{1}{2}}.\end{aligned}$$

Thus, (A.1) implies

$$\left(A_{k-1}P_{k,k-1}v, P_{k,k-1}v\right) \leq (A_k v, v), \quad \forall v \in V_k. \tag{3.6}$$

Conversely, assume (3.6). Then, for any $v \in V_{k-1}$, by the definition of $P_{k,k-1}$,

$$\begin{aligned}(A_k v, v) &= \left(A_{k-1} P_{k,k-1} v, v\right) \\ &\leq \left(A_{k-1} P_{k,k-1} v, P_{k,k-1} v\right)^{\frac{1}{2}} \left(A_{k-1} v, v\right)^{\frac{1}{2}} \\ &\leq (A_k v, v)^{\frac{1}{2}} \left(A_{k-1} v, v\right)^{\frac{1}{2}},\end{aligned}$$

which implies (A.1). Thus, (A.1) and (3.6) are equivalent. Notice that a similar inequality to (3.6) would also hold for P_k, namely

$$(A_k P_k v, P_k v) \leq (Av, v), \quad \forall v \in V. \tag{3.7}$$

For the operators R_k, we assume there are constants $C_1, C_2 > 0$, independent of k so that

$$C_1 \frac{||v||^2}{\lambda_k} \leq (R_k v, v) \leq C_2 \left(A_k^{-1} v, v\right), \quad \forall v \in V_k, \tag{A.2}$$

where λ_k is the largest eigenvalue of A_k. Lastly, as is common in multigrid theory, we will use an approximation assumption to establish spectral equivalence between B and A^{-1}. In this work, we assume the following approximation property: There is an $\alpha \in (0, 1]$ and constant $C_3 > 0$, independent of k, so that

$$\left(A_k (I - P_{k,k-1}) v, v\right) \leq C_3^{\alpha} \left(\frac{||A_k v||^2}{\lambda_k}\right)^{\alpha} (A_k v, v)^{1-\alpha}, \quad \forall v \in V_k. \tag{A.3}$$

We are now in a position to state and prove the main theorem of this section. The proof closely resemble the proofs of Corollary 3 and Theorem 2 in [14], but is extended to handle the case of the non-nestedness of the operators.

Theorem 3.1. *Assume that* (A.1), (A.2), *and* (A.3) *hold. Then, with B given in* (3.5),

$$C_1 C_3^{-1} J^{1-\frac{1}{\alpha}} (Av, v) \leq (BAv, Av) \leq C_2 J (Av, v). \tag{3.8}$$

holds for every $v \in V$.

Proof. Fix $v \in V$. Using the definition of B together with (3.4) we find that

$$(BAv, Av) = \sum_{k=1}^{J} (R_k Q_k Av, Q_k Av) = \sum_{k=1}^{J} (R_k A_k P_k v, A_k P_k v).$$

Thus, applying the second inequality of (A.2) and (3.7) gives

$$(BAv, Av) \leq C_2 \sum_{k=1}^{J} (A_k P_k v, P_k v) \leq C_2 J (Av, v),$$

which proves the second inequality of (3.8).

For the first inequality of (3.8) we write

$$v = \sum_{k=1}^{J} (P_k - P_{k-1}) v,$$

where we interpret $P_0 = 0$ and $P_J = I$. By the definition of P_k, we have that $P_{k-1} = P_{k,k-1} P_k$, and so

$$v = \sum_{k=1}^{J} (I - P_{k,k-1}) P_k v.$$

It follows that

$$(Av, v) = \sum_{k=1}^{J} \left(A_k (I - P_{k,k-1}) P_k v, P_k v\right).$$

Using (A.3) and (3.7), gives

$$\begin{aligned}(Av,v) &\le C_3^{\alpha}\sum_{k=1}^{J}\left(\lambda_k^{-1}\,||A_kP_kv||^2\right)^{\alpha}(A_kP_kv,P_kv)^{1-\alpha}\\ &\le C_3^{\alpha}\,(Av,v)^{1-\alpha}\sum_{k=1}^{J}\left(\lambda_k^{-1}\,||A_kP_kv||^2\right)^{\alpha}.\end{aligned}$$

The first inequality of (A.2) then implies that

$$\begin{aligned}(Av,v) &\le (C_1^{-1}C_3)^{\alpha}\,(Av,v)^{1-\alpha}\sum_{k=1}^{J}(R_kA_kP_kv,A_kP_kv)^{\alpha}\\ &\le (C_1^{-1}C_3)^{\alpha}J^{1-\alpha}\,(Av,v)^{1-\alpha}\left(\sum_{k=1}^{J}(R_kA_kP_kv,A_kP_kv)\right)^{\alpha}\\ &\le (C_1^{-1}C_3)^{\alpha}J^{1-\alpha}\,(Av,v)^{1-\alpha}\,(BAv,Av)^{\alpha},\end{aligned}$$

where the second step follows by Hölder's inequality. The last step follows by the definition of B in (3.5), and (3.4). Dividing by $(C_1^{-1}C_3)^{\alpha}\,(Av,v)^{1-\alpha}J^{1-\alpha}$ on both sides and raising to the power $\frac{1}{\alpha}$ gives the first inequality of (3.8). □

Remark 1. Analagously to what was done in [14], we can replace the regularity assumption (A.3) with an assumption on the projections Q_k (cf. also [13]). In particular, if instead of (A.3), we assume that there is a constant $C_4 > 0$, independent of k so that

$$\left\|(I-Q_{k-1})v\right\|^2 \le C_4\lambda_k^{-1}(Av,v), \quad \forall v \in V,$$

then we can use an argument like what was made in [14, Theorem 1 and Corollary 1] to show that

$$C_4^{-1}C_1J^{-1}\,(Av,v) \le (BAv,Av) \le C_2J\,(Av,v) \tag{3.9}$$

for every $v \in V$.

4. Preconditioner for discrete fractional Laplacian

In this section we use the abstract theory developed in Section 3 to derive an order optimal preconditioner for the discrete fractional Laplacian A_h^s, described in Section 2, when $s \in [0,1]$.

Let Ω be a bounded, polygonal domain in $\mathbb{R}^n$ and suppose we are given a quasi-uniform triangulation of Ω, denoted by $\mathcal{T}_h$, where h denotes the characteristic mesh size. We restrict our discussion to the case when V_h is the space of continuous, piecewise linear functions relative to the triangulation $\mathcal{T}_h$ which vanish on $\partial\Omega$. To define a nested sequence of subspaces, we suppose that $\mathcal{T}_h$ is constructed by successive refinements. That is, we are given a sequence,

$$\mathcal{T}_1 \subset \cdots \mathcal{T}_J = \mathcal{T}_h,$$

of quasi-uniform triangulations, and $\mathcal{T}_k$ has characteristic mesh size h_k for $k = 1,\ldots,J$. In the following, we will assume the bounded refinement hypothesis, that is, $h_{k-1} \le \gamma h_k$ for $k = 2,\ldots,J$, where $\gamma \ge 1$ is a constant. In practice, γ is around 2. For each k we define V_k as the space of continuous, piecewise linear functions relative to $\mathcal{T}_k$ that vanish on $\partial\Omega$. Further, we define $A_k : V_k \to V_k$ by

$$(A_kv,w) = (\nabla v,\nabla w), \quad v,w \in V_k.$$

We now fix $s \in [0,1]$. Since A_k is symmetric positive definite, we can define A_k^s and corresponding norms

$$||v||_{s,k}^2 := \left(A_k^sv,v\right), \quad v \in V_k.$$

Note that if $s = 0$ or $s = 1$, the norm $||\cdot||_{s,k}$ coincides with the L^2- and H_0^1-norm, respectively. That is, $||\cdot||_{0,k} = ||\cdot||$, and $||\cdot||_{1,k} = ||\cdot||_1$.

Analogous to the discussion in Section 3 we also define operators $P^s_{k,k-1} : V_k \to V_{k-1}$ by

$$\left(A^s_{k-1}P^s_{k,k-1}v, w\right) = \left(A^s_k v, w\right), \quad \forall v \in V_k, w \in V_{k-1}. \tag{4.1}$$

$Q_k : V_J \to V_k$ as the L^2-projection, and $P^s_k := P^s_{k+1,k}P^s_{k+2,k+1}\cdots P^s_{J,J-1}$.

To complete the description of a multilevel preconditioner, we still need to define smoothers, R^s_k, for each k and s. In this work, we will define additive smoothers based on domain decomposition. To that end, let $\mathcal{N}_k$ be the set of vertices in the triangulation $\mathcal{T}_k$, and for each $\nu \in \mathcal{N}_k$, let $\mathcal{T}_{k,\nu}$ be the set of triangles meeting at the vertex ν. Then $\mathcal{T}_{k,\nu}$ forms a triangulation of a small subdomain $\Omega_{k,\nu}$. We define $V_{k,\nu}$ to be the subspace of functions in V_k with support contained in $\bar{\Omega}_{k,\nu}$. Analogously to what we did for V_k, we may define for each $\nu \in \mathcal{N}_k$ operators $A^s_{k,\nu} : V_{k,\nu} \to V_{k,\nu}$, and $Q_{k,\nu} : V_k \to V_{k,\nu}$. For $k = 2,\ldots,J$, we define

$$R^s_k := \sum_{\nu\in\mathcal{N}_k} A^{-s}_{k,\nu}Q_{k,\nu}, \tag{4.2}$$

while on the coarsest level we set $R^s_1 = A^{-s}_1$. We note that the smoothers are symmetric positive-definite, and their inverse satisfy

$$\left((R^s_k)^{-1}v, v\right) = \inf_{\substack{v=\sum_\nu v_\nu \\ v_\nu\in V_{k,\nu}}} \sum_{\nu\in\mathcal{N}_k} \left(A^s_{k,\nu}v_\nu, v_\nu\right), \quad v \in V_k. \tag{4.3}$$

With our particular choice of subspaces $V_{k,\nu}$, any $v \in V_k$ can be uniquely decomposed into $v = \sum_{\nu\in\mathcal{N}_k} v_\nu$, with $v_\nu \in V_{k,\nu}$. Moreover, it is well-known that this decomposition is L^2-stable. That is, there are constants $K_0, K_1 > 0$, independent of k and v so that

$$K_0 ||v||^2 \le \sum_{\nu\in\mathcal{N}_k} ||v_\nu||^2 \le K_1 ||v||^2. \tag{4.4}$$

Our preconditioner now reads

$$B^s_h := \sum_{k=1}^{J} R^s_k Q_k. \tag{4.5}$$

We want to apply Theorem 3.1 to the preconditioner defined by (4.5) and (4.2), so we need to verify assumptions (A.1)-(A.3).

Using Lemma 2.1, we immediately find that for every k,

$$\left(A^s_k v, v\right) \le \left(A^s_{k-1}v, v\right), \quad \forall v \in V_{k-1},$$

which verifies (A.1) in the current context.

We present the verification of (A.2) in the following Lemma.

Lemma 4.1. *For $k = 1,\ldots,J$, let $R^s_k : V_k \to V_k$ be defined by* (4.2)*. Then there are constants $C_1, C_2 > 0$, so that for every k,*

$$C_1 \frac{||v||^2}{\lambda^s_k} \le \left(R^s_k v, v\right) \le C_2\left(A^{-s}_k v, v\right), \quad \forall v \in V_k, \tag{4.6}$$

where λ^s_k is the largest eigenvalue of A^s_k.

Proof. It is evident that (4.6) holds on the coarsest level, i.e., for $k = 1$ (4.6) is satisfied with $C_1 = C_2 = 1$. So let $k \ge 2$, and fix $v \in V_k$. For $\nu \in \mathcal{N}_k$, let $\lambda^s_{k,\nu}$ denote the largest eigenvalue of $A^s_{k,\nu}$. To prove the first inequality in (4.6), we begin by noting that

$$\lambda^1_k = \sup_{w\in V_k} \frac{\left(A^1_k w, w\right)}{(w,w)} \ge \sup_{w\in V_{k,\nu}} \frac{\left(A^1_{k,\nu} w, w\right)}{(w,w)} = \lambda^1_{k,\nu}.$$

Thus, since $\lambda_k^s = (\lambda_k^1)^s$, we have that

$$\lambda_k^s \geq \lambda_{k,\nu}^s. \tag{4.7}$$

Let now $v = \sum_{\nu \in \mathcal{N}_k} v_\nu$ be the unique decomposition of v into $V_{k,\nu}$ for $\nu \in \mathcal{N}_k$. Using (4.7) and the second inequality of (4.4), together with the definition of $Q_{k,\nu}$ and R_k^s yields

$$\begin{aligned}
\frac{(v,v)}{\lambda_k^s} &= \frac{1}{\lambda_k^s} \sum_{\nu \in \mathcal{N}_k} (v, v_\nu) \\
&= \frac{1}{\lambda_k^s} \sum_{\nu \in \mathcal{N}_k} \left(Q_{k,\nu} v, v_\nu\right) \\
&\leq \left(\frac{1}{\lambda_k^s} \sum_{\nu \in \mathcal{N}_k} \left(Q_{k,\nu} v, Q_{k,\nu} v\right) \right)^{\frac{1}{2}} \left(\frac{1}{\lambda_k^s} \sum_{\nu \in \mathcal{N}_k} \|v_\nu\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\sum_{\nu \in \mathcal{N}_k} \frac{1}{\lambda_{k,\nu}^s} \left(Q_{k,\nu} v, Q_{k,\nu} v\right) \right)^{\frac{1}{2}} \left(\frac{K_1}{\lambda_k^s} \|v\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\sum_{\nu \in \mathcal{N}_k} \left(A_{k,\nu}^{-s} Q_{k,\nu} v, Q_{k,\nu} v\right) \right)^{\frac{1}{2}} \left(\frac{K_1}{\lambda_k^s} \|v\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(R_k^s v, v\right)^{\frac{1}{2}} \left(\frac{K_1}{\lambda_k^s} \|v\|^2 \right)^{\frac{1}{2}},
\end{aligned}$$

which proves the first inequality of (4.6) with $C_1 = K_1^{-1}$.

For the second inequality, we begin by noting that for $s = 1$, it was proven in [37, Lemma 7.2] that there is a constant C, independent of k so that

$$\left(R_k^1 v, v\right) \leq C \left(A_k^{-1} v, v\right), \quad \forall v \in V_k.$$

Since $s \in [0,1]$, it follows by the Löwner-Heinz inequality (2.1) that

$$\left((R_k^1)^s v, v\right) \leq C^s \left(A_k^{-s} v, v\right), \quad \forall v \in V_k. \tag{4.8}$$

Thus, if we can show that

$$\left(R_k^s v, v\right) \leq C \left((R_k^1)^s v, v\right), \tag{4.9}$$

for some constant C, which is independent of k, then (4.8) together with (4.9) imply the second inequality of (4.6).

We aim to prove (4.9) using Jensen's operator inequality. To that end, we need to scale R_k^s, so that (2.2) is applicable. From the characterization of $\left(R_k^0\right)^{-1}$ in (4.3) and the first inequality of (4.4) we have that

$$K_0 \|v\|^2 \leq \left(\left(R_k^0\right)^{-1} v, v\right),$$

which in turn implies that

$$\sum_{\nu \in \mathcal{N}_k} \left(Q_{k,\nu} v, Q_{k,\nu} v\right) = \left(R_k^0 v, v\right) \leq K_0^{-1} \|v\|^2.$$

If we now define $\tilde{Q}_{k,\nu} = K_0^{\frac{1}{2}} Q_{k,\nu}$, and $\tilde{R}_k^s = K_0 R_k^s$, we have that

$$\sum_{\nu \in \mathcal{N}_k} \left(\tilde{Q}_{k,\nu} v, \tilde{Q}_{k,\nu} v\right) \leq \|v\|^2 \text{ and } \left(\tilde{R}_k^s v, v\right) = \sum_{\nu \in \mathcal{N}_k} \left(A_{k,\nu}^{-s} \tilde{Q}_{k,\nu} v, \tilde{Q}_{k,\nu} v\right).$$

We can now use Jensen's operator inequality (2.2), together with an argument analogous to that in the proof of Lemma 2.1 to get

$$\begin{aligned} R_k^s &= K_0^{-1}\tilde{R}_k^s \\ &\leq K_0^{-1}(\tilde{R}_k^1)^s \\ &= K_0^{-(1-s)}(R_k^1)^s \end{aligned}$$

This, together with (4.8), proves the second inequality of (4.6) with $C_2 = K_0^{-(1-s)}C^s$. □

We observe that the proof of Lemma 4.1 shows that if the decomposition $V_k = \sum_{\eta\in\mathcal{N}_k} V_{k,\nu}$ is stable in both L^2- and H_0^1-norms, then it is also stable in the fractional norm $\|\cdot\|_{s,k}$. That is, if there are constants $c_0, c_1 > 0$ so that

$$\left(A_k^s v, v\right) \leq c_s\left((R_k^s)^{-1}v, v\right), \quad \forall v \in V_k,$$

with $s = 0$ and $s = 1$, then the same holds for every $s \in [0,1]$, with $c_s = c_0^{1-s}c_1^s$. In this way, the smoother defined by (4.2) is the natural interpolation between the corresponding smoothers for $s = 0$ and $s = 1$. As such, the results in Lemma 4.1 can be extended to more general overlapping domain decompositions than the one we consider here.

As noted in [11, Remark 5.1] the α in the approximation and regularity assumption (A.3) is closely related to the elliptic regularity of the continuous problem. Therefore, we make the following assumption:

Assumption 4.1. *There is an $\alpha' \in (0,1]$ so that A is a bounded operator from $H_0^1(\Omega)\bigcap H^{1+\alpha'}(\Omega)$ to $H^{-1+\alpha'}(\Omega)$, and A^{-1} is a bounded operator from $H^{-1+\alpha'}(\Omega)$ to $H_0^1(\Omega)\bigcap H^{1+\alpha'}(\Omega)$.*

Assumption 4.1 is standard for proving condition (A.3) in the case of $s = 1$ (cf. for instance [11]). In [9, Thm 4.3, and Rem. 4.1] Bonito et al. used Assumption 4.1 to prove the error estimate

$$\left\|(A^{-s} - A_k^{-s}Q_k)f\right\| \leq C h_k^{2s}\,\|f\|, \quad \forall f \in L^2(\Omega),$$

when $\alpha' > s$. By the triangle inequality and the bounded refinement hypothesis it then follows that

$$\left\|(A_k^{-s} - A_{k-1}^{-s}Q_{k-1})f\right\| \leq C h_k^{2s}\,\|f\|, \quad \forall f \in V_k, \tag{4.10}$$

for each k. This estimate is sufficient to verify (A.3) in our current context. The result is stated in the following Lemma.

Lemma 4.2. *Assume that Assumption 4.1 is satisfied with $\alpha' > s$. Then there is a constant $C_3 > 0$, so that for every k*

$$\left(A_k^s(I - P_{k,k-1}^s)v, v\right) \leq C_3\frac{\left\|A_k^s v\right\|^2}{\lambda_k^s}. \tag{4.11}$$

Proof. From the definition of $P_{k,k-1}^s$ in (4.1),

$$I - P_{k,k-1}^s = I - A_{k-1}^{-s}Q_{k-1}A_k^s = (A_k^{-s} - A_{k-1}^{-s}Q_{k-1})A_k^s,$$

and so, for any $v \in V_k$

$$\left(A_k^s(I - P_{k,k-1}^s)v, v\right) \leq \left((A_k^{-s} - A_{k-1}^{-s}Q_{k-1})A_k^s v, A_k^s v\right).$$

Using Cauchy-Schwarz inequality together with the error estimate (4.10), we get

$$\left(A_k^s(I - P_{k,k-1}^s)v, v\right) \leq \left\|(A_k^{-s} - A_{k-1}^{-s}Q_{k-1})A_k^s v\right\|\,\left\|A_k^s v\right\| \leq C h_k^{2s}\left\|A_k^s v\right\|^2. \tag{4.12}$$

By the quasi-uniformity of the mesh $h_k^2 \leq C\lambda_k^{-1}$, and it follows that $h_k^{2s} \leq C\lambda_k^{-s}$. Using this in (4.12) completes the proof. □

We are finally in a position to prove the main theorem of this section.

Theorem 4.1. *Let Assumption 4.1 be satisfied with $\alpha' > s$. Then, for $s \in [0,1]$ with B_h^s defined by* (4.2) *and* (4.5) *satisfies*

$$C_1 C_3^{-1} \left(A_h^s v, v\right) \leq \left(B_h^s A_h^s v, A_h^s v\right) \leq C_2 J \left(A_h^s v, v\right), \quad \forall v \in V, \tag{4.13}$$

where C_1, C_2, and C_3 are the same as in Lemmas 4.1 and 4.2.

Proof. This result is a straightforward application of Theorem 3.1 together with Lemmas 4.1 and 4.2. □

Theorem 4.1 shows that the condition number $K(B_h^s A_h^s) \leq CJ$, and so increases linearly with the number of mesh levels, but is independent of h.

With less regularity of the domain, we can still prove a slightly weaker form of spectral equivalence. By the assumed quasi-uniformity of $\mathcal{T}_k$, we have for $k = 2, \ldots, J$ that

$$\left\|(I - Q_{k-1})v\right\|^2 \leq C h_k^2 \|v\|_1^2, \quad \forall v \in V_k.$$

This, together with the boundedness of $I - Q_{k-1}$ and interpolation theory, yields

$$\left\|(I - Q_{k-1})v\right\|^2 \leq C h_k^{2s} \|v\|_{s,k}^2 \leq C_4 \lambda_k^{-s} \|v\|_{s,k}^2,$$

for some constant C_4, independent of k. By the discussion in Remark 1, we get that

$$C_4^{-1} C_1 J^{-1} \left(A_h^s v, v\right) \leq \left(B_h^s A_h^s v, A_h^s v\right) \leq C_2 J \left(A_h^s v, v\right), \quad \forall v \in V_h, \tag{4.14}$$

and the condition number is bounded by $K(B_h^s A_h^s) \leq CJ^2$.

5. Preconditioner when $s \in [-1, 0]$

For $s \in [-1, 0]$, the large eigenvalues of A_h^s correspond to smooth functions. In a multilevel setting this means that neither relaxation nor coarse grid correction will reduce the oscillatory components of the error. As a consequence, we cannot expect a direct multigrid approach to work. Moreover, when $s < 0$ the Löwner-Heinz' and Jensen's operator inequalities in (2.1) and (2.2) fail to hold, and the argument of Section 4 is no longer valid. In this section, we will therefore investigate an alternative approach for constructing preconditioners.

We will base the preconditioner for A_h^s when s is negative on our previously defined preconditioners B_h^t for $t \in [0,1]$ together with the multiplicative decomposition of A_h

$$A_h^{-s} = A_h^{-\frac{1+s}{2}} A_h A_h^{-\frac{1+s}{2}}. \tag{5.1}$$

We have for every $u \in V_h$ and $t \in \mathbb{R}$ that

$$\|u\|_{-s,h} = \left\| A_h^{-\frac{t+s}{2}} u \right\|_{t,h}.$$

The specific form we will use below is

$$\|u\|_{-\frac{1+s}{2}+\beta,h} = \left\| A_h^{-\frac{1+s}{2}} u \right\|_{\frac{1+s}{2}+\beta,h}, \tag{5.2}$$

which is valid for any $\beta \in \mathbb{R}$.

Replacing the left- and rightmost factor of the right hand side in (5.1) with a spectrally equivalent preconditioner $B_h^{\frac{1+s}{2}}$, yields a symmetric positive definite operator

$$\tilde{B}_h^s := B_h^{\frac{1+s}{2}} A_h B_h^{\frac{1+s}{2}}. \tag{5.3}$$

We want $\tilde{B}_h^s$ to be spectrally equivalent to A_h^{-s}. That is, there exist constants C_1, C_2 so that for every $u \in V_h$,

$$C_1 \left(A_h^s u, u\right) \leq \left(\tilde{B}_h^s A_h^s u, A_h^s u\right) \leq C_2 \left(A_h^s u, u\right) \tag{5.4}$$

holds. By the definition of $\tilde{B}_h^s$,

$$\left(\tilde{B}_h^s A_h^s u, A_h^s u\right) = \left(A_h B_h^{\frac{1+s}{2}} A_h^s u, B_h^{\frac{1+s}{2}} A_h^s u\right) = \left\| B_h^{\frac{1+s}{2}} A_h^s u \right\|_1^2,$$

and since $\left(A_h^s u, u\right) = ||u||_{s,h}^2$, we see that the spectral equivalence in (5.4) is equivalent to

$$C_1^{\frac{1}{2}} ||u||_{s,h} \leq \left\| B_h^{\frac{1+s}{2}} A_h^s u \right\|_1 \leq C_2^{\frac{1}{2}} ||u||_{s,h}, \quad \forall u \in V_h. \tag{5.5}$$

Using the preconditioner described in section 4, we have by the spectral equivalence established in Theorem 4.1 that there are constants $C_1, C_2 > 0$ so that

$$C_1 ||u||_{-\frac{1+s}{2},h} \leq \left\| B_h^{\frac{1+s}{2}} u \right\|_{\frac{1+s}{2},h} \leq C_2 J ||u||_{-\frac{1+s}{2},h}, \quad u \in V_h. \tag{5.6}$$

We assume now some additional regularity on $B_h^{\frac{1+s}{2}}$, similar to (5.2). That is, for some β, we have the norm equivalence

$$C_1 ||u||_{-\frac{1+s}{2}+\beta,h} \leq \left\| B_h^{\frac{1+s}{2}} u \right\|_{\frac{1+s}{2}+\beta,h} \leq C_2 J ||u||_{-\frac{1+s}{2}+\beta,h}. \tag{5.7}$$

In particular, with $\beta = \frac{1-s}{2} \in \left[\frac{1}{2}, 1\right]$, and replacing u by $A_h^s u$ in (5.7) we recover (5.5) and the spectral equivalence (5.4). We note also that if we assume the additional regularity of (5.7), we can bound the condition number of $\tilde{B}_h^s A_h^s$ by

$$K(\tilde{B}_h^s A_h^s) \leq K(B_h^{\frac{1+s}{2}} A_h^{\frac{1+s}{2}})^2. \tag{5.8}$$

We remark that while (5.5) may be a non-trivial property to validate because B_h is a discrete multigrid operator, similar conditions on the continuous differential operator are well-established. That is, for $(-\Delta)^{-s} : H^{-s} \to H^s$ the regularity conditions that enable a decomposition $(-\Delta)^{-s} = (-\Delta)^{-(s+t)}(-\Delta)^t$ such that

$$||(-\Delta)^{-s}||_{\mathcal{L}(H^{-s},H^s)} \leq ||(-\Delta)^{-(s+t)}||_{\mathcal{L}(H^{-s+t},H^s)} ||(-\Delta)^{-s}||_{\mathcal{L}(H^{-s},H^{-s+t})}$$

are well described, c.f. [31]. As such the regularity assumption (5.5) is a reasonable condition in the continuous setting, but the discrete setting is unclear.

6. Implementational concerns

The discrete operators discussed so far are related to, but are not the same as the matrices used in the implementation. In this section we will discuss how to implement these operators. We begin by discussing the matrix representation of the discrete fractional operators. We refer also to [32] for more details. While the discrete fractional operators satisify the group property $A_h^s A_h^t = A_h^{s+t}$, their matrix representations do not. In particular, for $t = -s$, $A_h^s A_h^{-s} = I_h$ and the finite element matrix representation of the identity is the mass matrix. Hence, in order to provide a precise description of the interpolation of the involved matrices, we let $\left\{\phi_h^i\right\}_{i=1}^{N_h}$ be the standard nodal basis for V_h, and we introduce the operators $\pi_h, \mu_h : V_h \to \mathbb{R}^{N_h}$, defined by

$$\begin{aligned} v &= \sum_{i=1}^{N_h} (\pi_h v)_i \, \phi_h^i, \quad \text{and} \\ (\mu_h v)_i &= \left(v, \phi_h^i\right), \quad i = 1, \ldots, N_h. \end{aligned} \tag{6.1}$$

Subsequently, we will refer to $\pi_h v$ and $\mu_h v$ as the primal and dual vector representation of v, respectively. The primal representation is sometimes called the nodal representation. We then have that

$$\mu_h^* = \pi_h^{-1}, \quad \text{and} \quad \pi_h^* = \mu_h^{-1} \tag{6.2}$$

To see this, take $\mathsf{v} \in \mathbb{R}^{N_h}$, and $u \in V_h$. Then,

$$\begin{aligned}\left(\mu_h^* \mathsf{v}, u\right) &= (\mathsf{v}, \mu_h u)_{l^2} \\ &= \sum_{i=1}^{N_h} \mathsf{v}_i \left(u, \phi_h^i\right) \\ &= \left(u, \sum_{i=1}^{N_h} \mathsf{v}_i \phi_h^i\right) \\ &= \left(u, \pi_h^{-1} \mathsf{v}\right),\end{aligned}$$

where $(\cdot,\cdot)_{l^2}$ is the standard Euclidean inner product on $\mathbb{R}^{N_h}$. This proves the first identity in (6.2). The second identity is proven similarly.

Using these operators, the stiffness matrix can then be expressed as

$$\mathsf{A}_h = \mu_h A_h \pi_h^{-1}, \quad \text{and} \quad (\mathsf{A}_h)_{i,j} = (A_h \phi_h^j, \phi_h^i), \quad 1 \le i, j \le N_h,$$

and the mass matrix is

$$\mathsf{M}_h = \mu_h I_h \pi_h^{-1} = \mu_h \pi_h^{-1}, \quad \text{and} \quad (\mathsf{M}_h)_{i,j} = (\phi_h^j, \phi_h^i), \quad 1 \le i, j \le N_h.$$

We see that for both the stiffness- and mass matrix, a matrix-vector product takes primal vectors as input and returns dual vectors.

For the matrix realization of A_h^s, let $\{(\lambda_i, \mathsf{u}_i)\}_{i=1}^{N_h} \subset \mathbb{R} \times \mathbb{R}^{N_h}$ be the eigenpairs of the generalized eigenvalue problem

$$\mathsf{A}_h \mathsf{u}_i = \lambda_i \mathsf{M}_h \mathsf{u}_i,$$

normalized so that $\mathsf{u}_j^\top \mathsf{M}_h \mathsf{u}_i = \delta_{i,j}$. Setting $\tilde{}_h = \operatorname{diag}(\lambda_1, \ldots, \lambda_{N_h})$, and $\mathsf{U} = [\mathsf{u}_1, \ldots, \mathsf{u}_{N_h}]$, we have that

$$\mathsf{U}^\top \mathsf{M}_h \mathsf{U} = \mathsf{I}, \quad \text{and} \quad \mathsf{U}^\top \mathsf{A}_h \mathsf{U} = \tilde{}_h. \tag{6.3}$$

We then define

$$\mathsf{A}_h^s = (\mathsf{M}_h \mathsf{U}) \tilde{}_h^s (\mathsf{M}_h \mathsf{U})^\top. \tag{6.4}$$

One can verify that the entries of A_h^s satisfy

$$(\mathsf{A}_h^s)_{i,j} = \left(A_h^s \phi_h^j, \phi_h^i\right),$$

in which case $\mathsf{A}_h^s = \mu_h A_h^s \pi_h^{-1}$. Using (6.3) we are also able to see that

$$(\mathsf{A}_h^s)^{-1} = \mathsf{U} \tilde{}_h^{-s} \mathsf{U}^\top = \pi_h A_h^{-s} \mu_h^{-1}, \tag{6.5}$$

making it the matrix realization of A_h^s viewed as an operator from X_h to X_h'. However, the group properties mentioned above only make sense when we consider A_h^s as operators on X_h. Thus, we see that for the matrices

$$\pi_h A_h^s \pi_h^{-1} = (\pi_h \mu_h^{-1}) \mu_h A_h^s \pi_h^{-1} = \mathsf{M}_h^{-1} \mathsf{A}_h^s,$$

the group properties are satisfied. This can also be verified using the definition of A_h^s in (6.4).

Since matrix-vector products involving A_h^s take primal vectors as input and return dual vectors, the matrix realization of B_h^s should take dual vectors as input and return primal vectors. Then the product $\mathsf{B}_h^s \mathsf{A}_h^s$ acts on primal vectors, and is thus suitable for a Krylov subspace method. See also [32, Section 6] and [10, Section 15]. Therefore, we define

$$\mathsf{B}_h^s = \pi_h B_h^s \mu_h^{-1}. \tag{6.6}$$

To see how B_h^s is implemented, we begin by supposing that $\dim V_k = N_k$ for $k = 1, \ldots, J$. Let $\left\{\phi_k^i\right\}_{i=1}^{N_k}$ be bases for V_k, and we define operators $\pi_k, \mu_k : V_k \to \mathbb{R}^{N_k}$ analogously to (6.1).

We then define mass and stiffness matrices on level k as $\mathsf{M}_k = \mu_k \pi_k^{-1}$ and $\mathsf{A}_k = \mu_k A_k \pi_k^{-1}$, respectively.

By assumption, for every k, $V_k \subset V_h$, and so there are matrices $\mathsf{I}_k : \mathbb{R}^{N_k} \to \mathbb{R}^{N_h}$ so that

$$\phi_k^i = \sum_{j=1}^{N_h} (\mathsf{I}_k)_{i,j} \phi_h^j, \quad i = 1, \ldots, N_k.$$

In fact, I_k is the matrix realization of the inclusion operator $I_k : V_k \to V_h$, i.e. $\mathsf{I}_k = \pi_h I_k \pi_k^{-1}$. Using that $Q_k = I_k^*$ and (6.2) we have that the transpose of I_k satisfies

$$\begin{aligned} \mathsf{I}_k^\top &= \left(\pi_h I_k \pi_k^{-1}\right)^* \\ &= (\pi_k^{-1})^* Q_k \pi_h^* \\ &= \mu_k Q_k \mu_h^{-1} \\ &=: \mathsf{Q}_k, \end{aligned}$$

which is the matrix realization of Q_k in dual representation. Thus, for the matrix B_h^s we have that

$$\begin{aligned} \mathsf{B}_h^s &= \pi_h B_h^s \mu_h^{-1} \\ &= \sum_{k=1}^{J} \pi_h I_k R_k^s Q_k \mu_h^{-1} \\ &= \sum_{k=1}^{J} (\pi_h I_k \pi_k^{-1})(\pi_k R_k^s \mu_k^{-1})(\mu_k Q_k \mu_h^{-1}) \\ &= \sum_{k=1}^{J} \mathsf{Q}_k^\top \mathsf{R}_k^s \mathsf{Q}_k, \end{aligned} \tag{6.7}$$

where we define $\mathsf{R}_k^s = \pi_k R_k^s \mu_k^{-1}$ as the matrix realization of R_k^s. We see that due to (6.5) $\mathsf{R}_1^s = (\mathsf{A}_1^s)^{-1}$. For $k \geq 2$ we define for $\nu \in \mathcal{N}_k$ operators $\pi_{k,\nu}, \mu_{k,\nu} : V_{k,\nu} \to \mathbb{R}^{\dim V_{k,\nu}}$ and matrices $\mathsf{Q}_{k,\nu} : \mathbb{R}^{N_k} \to \mathbb{R}^{\dim V_{k,\nu}}$, similarly to the above. The matrix realization of R_k^s then becomes

$$\mathsf{R}_k^s = \sum_{\nu \in \mathcal{N}_k} \mathsf{Q}_{k,\nu}^\top (\mathsf{A}_{k,\nu}^s)^{-1} \mathsf{Q}_{k,\nu}. \tag{6.8}$$

Here, $\mathsf{A}_{k,\nu}^s = \mu_{k,\nu} A_{k,\nu}^s \pi_{k,\nu}^{-1}$. By (6.5), the implementation of R_k^s will require solving many small eigenvalue problems. In the particular case of continuous, piecewise linear finite element functions, and subdomains $\Omega_{k,\nu}$ as described in Section 4, the subspaces $V_{k,\nu}$ are one-dimensional. The matrix R_k^s is then diagonal, with entries

$$(\mathsf{R}_k^s)_{i,i} = \frac{1}{(\mathsf{M}_k)_{i,i}^{1-s} (\mathsf{A}_k)_{i,i}^s}, \quad i = 1, \ldots, N_k.$$

That is, this is the smoother mentioned in the introduction.

Inserting (6.8) into (6.7) we get

$$\mathsf{B}_h^s = \mathsf{Q}_1^\top (\mathsf{A}_1^s)^{-1} \mathsf{Q}_1 + \sum_{k=2}^{J} \mathsf{Q}_k^\top \left(\sum_{\nu \in \mathcal{N}_k} \mathsf{Q}_{k,\nu}^\top (\mathsf{A}_{k,\nu}^s)^{-1} \mathsf{Q}_{k,\nu} \right) \mathsf{Q}_k. \tag{6.9}$$

We end this section by showing how to implement $\tilde{B}_h^s$, when $s \in [-1, 0]$. In this case, the matrix realization of $\tilde{B}_h^s$ can be found from $\mathsf{B}_h^{\frac{1+s}{2}}$ and A_h by

$$\begin{aligned}\tilde{\mathsf{B}}_h^s &:= \pi_h \tilde{B}_h^s \mu_h^{-1} \\ &= (\pi_h B_h^{\frac{1+s}{2}} \mu_h^{-1})(\mu_h A_h \pi_h^{-1})(\pi_h B_h^{\frac{1+s}{2}} \mu_h^{-1}) \\ &= \mathsf{B}_h^{\frac{1+s}{2}} \mathsf{A}_h \mathsf{B}_h^{\frac{1+s}{2}}.\end{aligned} \tag{6.10}$$

That is, $\tilde{B}_h^s$ is implemented as an application of $B_h^{\frac{1+s}{2}}$, followed by a multiplication of the stiffness matrix and a second application of $B_h^{\frac{1+s}{2}}$.

7. Numerical experiments

In this section we present a series of numerical experiments that aim to validate the theoretical results we established in previous sections. We also present numerical results for the case when $s < 0$, using $\tilde{B}_h^s$, defined in (5.3), as preconditioner. Specifically, in section 7.1 we solve

$$A_h^s u = f,$$

using preconditioned conjugate gradient method with B_h^s defined in (4.5) as preconditioner. Here, the main motivation is to validate the h-independence of $K(B_h^s A_h^s)$ implied by Theorem 4.1. In section 7.2, we consider a coupled multidomain problem, where the weakly imposed continuity on the interface leads to a Lagrange multiplier in $H^{\pm\frac{1}{2}}$.

The numerical experiments are conducted using random initial guess. Convergence in the iterative methods used is reached when the relative preconditioned residual, i.e. $\frac{(Br_k, r_k)}{(Br_0, r_0)}$, where r_k is the residual at the k'th iteration and B is the preconditioner, is below a given tolerance.

7.1. Preconditioning the fractional Laplacian.

In the first set of numerical experiments, we show the performance of the preconditioners B_h^s and $\tilde{B}_h^s$, defined in (4.5) and (5.3), respectively, depending on the sign of s for the A_h^s inner product. That is, for a given $f_h \in V_h$, we solve: Find $u_h \in V_h$ such that

$$\left(A_h^s u_h, v\right) = (f_h, v), \quad \forall v \in V_h, \tag{7.1}$$

where $s \in [-1, 1]$. We take $\Omega = [0, 1] \subset \mathbb{R}$, and $\mathcal{T}_h$ is a uniform partition of Ω consisting of $N = \frac{1}{h}$ elements. V_h is then the space of continuous, piecewise linear functions relative to $\mathcal{T}_h$ that vanish on $\partial\Omega$. The matrix representation of A_h^s we use here is provided by (6.4). This matrix is in general dense, and so matrix-vector multiplication will take $\mathcal{O}(N^2)$ operations. As such, the preconditioned iterative method will not be computationally optimal, but we stress that these experiments are designed only to validate the bounds on $K(B_h^s A_h^s)$.

We solve the linear system arising from (7.1) using preconditioned conjugate gradient, with B_h^s as preconditioner if $s \geq 0$, and $\tilde{B}_h^s$ if $s < 0$. For $s \geq 0$, iteration counts and estimated condition numbers can be viewed in Table 1. From these results we see that both the iteration counts and condition numbers stay uniformly bounded for each s.

The analogous results for $s \leq 0$ can be seen in Table 2. Here the situation is slightly more complicated. For each s, the iteration counts and condition numbers seem to increase for small N (large h), but ultimately stay bounded when N is increased. Worth noting is that the bound (5.8) is relatively sharp. For instance, for $s = -1$, the preconditioner $\tilde{B}_h^s$ does two applications of B_h^0, and has estimated condition numbers around 193. By Table 1, $K(B_h^0 A_h^0) \approx 13.8$, and so $K(\tilde{B}^{-1} A_h^{-1}) \approx K(B_h^0 A_h^0)^2$. Similar relations holds for other values of $s \leq 0$.

s \ N	32	64	128	256	512
0.0	20(13.5)	25(13.6)	28(13.8)	29(13.8)	29(13.9)
0.1	18(8.7)	21(8.9)	23(8.9)	24(8.9)	24(8.9)
0.2	16(5.8)	18(6.4)	19(6.5)	21(6.5)	21(6.6)
0.3	14(4.2)	15(4.7)	17(4.9)	18(5.0)	18(5.0)
0.4	12(3.4)	14(3.7)	15(3.8)	15(3.9)	16(3.9)
0.5	11(2.9)	12(3.0)	13(3.1)	13(3.1)	14(3.2)
0.6	12(2.9)	13(3.0)	13(3.0)	14(3.1)	14(3.0)
0.7	12(3.0)	13(3.0)	14(3.1)	14(3.1)	14(3.1)
0.8	13(3.2)	14(3.3)	14(3.3)	14(3.3)	14(3.3)
0.9	14(3.5)	15(3.6)	15(3.6)	15(3.6)	15(3.6)
1.0	14(4.0)	16(4.1)	16(4.1)	16(4.1)	16(4.1)

TABLE 1. Numerical results for $(-\Delta)^s$ with nonnegative s. Table shows the number of preconditioned conjugate gradient iterations until reaching error tolerance 10^{-15}. Estimated condition numbers are shown inside parentheses. N is number of elements on the finest mesh. $J = 5$ in all tests.

s \ N	32	64	128	256	512
−1.0	32(184.4)	47(192.4)	56(192.7)	64(193.8)	62(191.2)
−0.9	28(119.0)	43(118.9)	50(120.5)	54(120.7)	55(119.9)
−0.8	26(78.3)	37(82.6)	46(84.5)	48(83.8)	49(83.9)
−0.7	25(53.0)	33(60.1)	40(61.9)	42(62.1)	45(61.5)
−0.6	24(36.9)	31(43.8)	35(45.8)	38(46.2)	41(46.2)
−0.5	22(26.8)	25(31.9)	30(34.3)	34(34.9)	38(35.1)
−0.4	20(20.4)	24(24.8)	28(26.5)	32(27.0)	37(27.1)
−0.3	17(16.1)	21(19.3)	27(20.7)	30(21.1)	34(21.1)
−0.2	17(13.1)	21(15.3)	25(16.4)	29(16.7)	32(16.7)
−0.1	16(11.0)	20(12.4)	23(13.2)	27(13.5)	29(13.5)
0.0	14(9.4)	17(10.4)	20(11.0)	24(11.2)	27(11.1)

TABLE 2. Numerical results for $(-\Delta)^s$ with negative s. Table shows the number of preconditioned conjugate gradient iterations until reaching error tolerance 10^{-15}. Estimated condition numbers are shown inside parentheses. N is number of elements on the finest mesh. $J = 5$ in all tests.

7.2. **Multidomain preconditioning.** In this section we apply the multilevel algorithm (4.5) to construct mesh independent preconditioners for a coupled multidomain problem originating from a geometrically accurate model of electric signal propagation in cardiac tissue, the EMI model, [35].We remark that the EMI model is simple in a sense that it is a single-physics problem where two elliptic equations are coupled. However, the interface problems encountered here are identical to those found in multiphysics applications, e.g. the coupled Darcy-Stokes system [29] or the Stokes-Biot system [2].

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain decomposed into two non-overlapping subdomains Ω_1, Ω_2 with a common interface $\Gamma = \partial\Omega_1 \bigcap \partial\Omega_2$ forming a closed curve. Motivated by the application the subdomain Ω_1 is designated as the *exterior* domain, i.e. $\partial\Omega_2 \bigcap \partial\Omega = \emptyset$. With $\epsilon > 0$

and n the outer normal of the exterior domain we now aim to solve

$$\begin{aligned} u_1 - \Delta u_1 &= f_1, \quad x \in \Omega_1, \\ u_2 - \Delta u_2 &= f_2, \quad x \in \Omega_2, \\ n \cdot \nabla u_1 - n \cdot \nabla u_2 &= 0, \quad x \in \Gamma, \\ \epsilon(u_1 - u_2) + n \cdot \nabla u_1 &= g, \quad x \in \Gamma. \end{aligned} \tag{7.2}$$

The choice of boundary conditions for (7.2) shall be discussed shortly. We remark that in the EMI model the parameter ϵ plays a role of inverse time step and thus algorithms robust with respect to the parameter are of interest. However, here the system will be considered only for a fixed choice of the parameter.

Considering (7.2) with homogeneous Neumann boundary conditions $n \cdot \nabla u_1 = 0$ on $\partial\Omega$ and letting $W_1 = H^1(\Omega_1) \times H^1(\Omega_2) \times (H^{-1/2}(\Gamma) \bigcap \epsilon^{-1/2} L^2(\Gamma))$ the variational formulation of (7.2) defines an operator $\mathscr{A}_1 : W_1 \to W_1'$

$$\mathscr{A}_1 = \begin{pmatrix} I - \Delta & 0 & T_1^* \\ 0 & I - \Delta & -T_2^* \\ T_1 & -T_2 & -\epsilon^{-1} I \end{pmatrix}, \tag{7.3}$$

where T_i, $T_i v = v|_\Gamma$ for $v \in C(\bar{\Omega}_i)$, $i = 1,2$ are the trace operators on $H^1(\Omega_1)$ and $H^1(\Omega_2)$, respectively. Tveito et al. [35] further discuss a mixed formulation of the system (7.2) where additional unknowns $\sigma_i = -\nabla u_i$, $i = 1,2$ are introduced. If homogeneous Dirichlet boundary conditions $u_1 = 0$ on $\partial\Omega$ are assumed the mixed formulation leads to operator $\mathscr{A}_2 : W_2 \to W_2'$

$$\mathscr{A}_2 = \begin{pmatrix} I & \nabla & T^* \\ -\nabla \cdot & -I & 0 \\ T & 0 & -\epsilon I \end{pmatrix}, \tag{7.4}$$

with $W_2 = H(\mathrm{div}, \Omega) \times L^2(\Omega) \times (H^{1/2}(\Gamma) \bigcap \epsilon^{1/2} L^2(\Gamma))$ and T the normal trace operator $T\sigma = \sigma|_\Gamma \cdot n$ for $v \in [C(\Omega)]^2$. We remark that operators $\mathscr{A}_1$ and $\mathscr{A}_2$ also arise naturally in the analysis of non-overlapping domain decomposition methods for second order elliptic problems in the primal [36] and mixed formulation [18] respectively.

Assuming that the operators $\mathscr{A}_1$ and $\mathscr{A}_2$ are isomorphisms on their respective spaces[1] the preconditioners can be established within the framework of operator preconditioning [32]. In particular, the Riesz map preconditioner for (7.3) is

$$\mathscr{B}_1 = \begin{pmatrix} I - \Delta & & \\ & I - \Delta & \\ & & \epsilon^{-1} I + (-\Delta + I)^{-1/2} \end{pmatrix}^{-1}, \tag{7.5}$$

while (7.4) will be preconditioned by

$$\mathscr{B}_2 = \begin{pmatrix} I - \nabla\nabla \cdot & & \\ & I & \\ & & \epsilon I + (-\Delta + I)^{1/2} \end{pmatrix}^{-1}. \tag{7.6}$$

Note that the operator sums in $\mathscr{B}_1$, $\mathscr{B}_2$ are due to the fact that the interface spaces are intersection spaces [7].

In order to simplify the setting and focus only on the fractional operators in the preconditioners we remove the parameter dependence from the problems by settting $\epsilon = \infty$ in (7.3), (7.5) and similarly $\epsilon = 0$ for (7.4), (7.6). In turn, the interface spaces reduce to $H^{-1/2}(\Gamma)$ and $H^{1/2}(\Gamma)$ respectively and the multilevel algorithm is directly applicable to the related interface

[1] The proof of this result as well as stable finite element discretization of the problem are subject of current work and will be reported elsewhere. We remark that operator $\mathscr{A}_1$ in the limit case $\epsilon = \infty$ has been studied in [28] in the context of mortar finite element method.

h	dim W_h	#cells$_\Gamma$	#MG			#Eig
			$J=2$	$J=3$	$J=4$	
2.21E-02	4481	128	67	93	103	36
1.10E-02	17153	256	68	92	111	35
5.52E-03	67073	512	66	90	112	35
2.76E-03	265217	1024	64	90	112	34
1.38E-03	1054721	2048	64	88	108	33

TABLE 3. Number of MinRes itarations for the operator $B_1\mathscr{A}_1$ and $\epsilon = 10^{15}$ using multilevel algorithm with J levels as a preconditioner for $(-\Delta + I)^{-1/2}$. Realizing the fractional operator with spectral decomposition leads to iteration counts in the last column.

problems which now involve the operator $I - \Delta$, cf. the Laplacian operator in the previous sections.

Robustness of $\mathscr{B}_1$, $\mathscr{B}_2$, and in particular the fractional Sobolev space preconditioner, are finally demonstrated by observing the iteration counts of the preconditioned MinRes method. In the experiments we let $\Omega = [0,1]^2$ and $\Omega_2 = [0.25, 0.75]^2$. The finite element discretization of W_1 uses continuous linear Lagrange elements(P_1). For W_2 two different discretizations are considered; BDM_1-P_0-P_1 or RT_1-P_0-P_0. That is, the the first subspace of W_2 is constructed from linear Brezzi-Douglas-Marini element(BDM_1) or the lowest order Raviart-Thomas element(RT_1), while the the remaining subspaces use piecewise constant and piecewise linear Lagrange elements, respectively. Let us note that with RT_1-P_0-P_0 element the discretization is non-conforming owing to the piecewise constant space for the multiplier. Moreover, the fractional multigrid algorithm is then applied outside of the setting used in the analysis of Section 4. We remark that in this case the multigrid preconditioner uses the discrete operator

$$(A_h u, w) = (u, w) + \sum_{v \in \mathscr{N}} \{\{h\}\}_v^{-1} [\![u]\!]_v [\![w]\!]_v, \quad u, w \in V_h.$$

Here $\mathscr{N}$ is a set of all the vertices of the mesh and $\{\{u\}\}_v = \frac{1}{2}(u|_{K^+} + u|_{K^-})$, $[\![u]\!]_v = u|_{K^+} - u|_{K^-}$ are the average and the jump values computed from the two[2] cells $K^\pm$ connected to the vertex v. Moreover, $\dim V_{k,v} = 2$ so that the eigenvalue problems needed to be solved by the smoother 4.2 use 2 by 2 matrices. Compare this to the case of one-dimensional problems due to continuous, piecewise linear elements in section 6.

The discrete preconditioners shall use off-the-shelf methods for the first two blocks. More specifically, a single V cycle of algebraic multigrid is used for $\mathscr{B}_1$ while for $\mathscr{B}_2$ the action is computed exactly by a direct solver. The final block of the preconditioners is realized by the proposed multilevel preconditioner with different number of levels $J = 2, 3, 4$.

The number of MinRes iterations is shown in Table 3 and Table 4. Here, the iterations were started from a random initial vector and terminated once the relative preconditioned residual norm was less then 10^{-8} in magnitude. For both $\mathscr{B}_1$ and $\mathscr{B}_2$ the iterations are bounded in the discretization parameter. The tables further list iteration counts for preconditioners where the fractional operators were realized in terms of spectral decomposition. As expected from the theory and experiments for the Laplace problem the difference in iteration counts between the multilevel realization and specral realization is larger for $\mathscr{B}_1$ then it is for $\mathscr{B}_2$.

[2]Recall that the interface Γ is here taken as a simple closed curve.

h	#cells$_\Gamma$	BDM_1-P_0-P_1					RT_1-P_0-P_0				
		dim W_h	#MG			#Eig	dim W_h	#MG			#Eig
			$J=2$	$J=3$	$J=4$			$J=2$	$J=3$	$J=4$	
2.21E-02	128	33152	25	27	28	22	20736	27	28	27	22
1.10E-02	256	131840	25	27	29	22	82432	27	32	32	22
5.52E-03	512	525824	23	27	27	22	328704	27	33	36	22
2.76E-03	1024	2100224	22	27	29	22	1312768	27	33	40	22
1.38E-03	2048	8394752	22	25	29	22	5246976	25	35	40	22

TABLE 4. Number of MinRes itarations for the operator $B_2.\mathscr{A}_2$ and $\epsilon = 10^{-15}$ using multilevel algorithm with J levels as a preconditioner for $(-\Delta + I)^{1/2}$. Two different finite element discretizations are considered. Realizing the fractional operator with spectral decomposition leads to iteration counts in the #Eig column.

References

[1] S. Aland, S. Egerer, J. Lowengrub, and A. Voigt. Diffuse interface models of locally inextensible vesicles in a viscous fluid. *J. Comput. Phys.*, 277:32–47, 2014.

[2] I. Ambartsumyan, E. Khattatov, I. Yotov, and P. Zunino. A Lagrange multiplier method for a Stokes–Biot fluid–poroelastic structure interaction model. *Numer. Math.*, Apr 2018.

[3] M. Arioli, D. Kourounis, and D. Loghin. Discrete fractional Sobolev norms for domain decomposition preconditioning. *IMA J. Numer. Anal.*, 33(1):318–342, 2013.

[4] M. Arioli and D. Loghin. Discrete interpolation norms with applications. *SIAM J. Numer. Anal.*, 47(4):2924–2951, 2009.

[5] I. Babuška. The finite element method with Lagrangian multipliers. *Numer. Math.*, 20:179–192, 1972/73.

[6] A. V. Balakrishnan. Fractional powers of closed operators and the semigroups generated by them. *Pacific J. Math.*, 10:419–437, 1960.

[7] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[8] S. Bertoluzza, V. Chabannes, C. Prud'homme, and M. Szopos. Boundary conditions involving pressure for the Stokes problem and applications in computational hemodynamics. *Comput. Methods Appl. Mech. Engrg.*, 322:58–80, 2017.

[9] A. Bonito and J. E. Pasciak. Numerical approximation of fractional powers of elliptic operators. *Math. Comp.*, 84(295):2083–2110, 2015.

[10] J. H. Bramble. *Multigrid methods*, volume 294 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1993.

[11] J. H. Bramble and J. E. Pasciak. New convergence estimates for multigrid algorithms. *Math. Comp.*, 49(180):311–329, 1987.

[12] J. H. Bramble, J. E. Pasciak, and P. S. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69(230):463–480, 2000.

[13] J. H. Bramble, J. E. Pasciak, J. P. Wang, and J. Xu. Convergence estimates for product iterative methods with applications to domain decomposition. *Math. Comp.*, 57(195):1–21, 1991.

[14] J. H. Bramble, J. E. Pasciak, and J. Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55(191):1–22, 1990.

[15] J. H. Bramble, J. E. Pasciak, and J. Xu. The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms. *Math. Comp.*, 56(193):1–34, 1991.

[16] K. Burrage, N. Hale, and D. Kay. An efficient implicit FEM scheme for fractional-in-space reaction-diffusion equations. *SIAM J. Sci. Comput.*, 34(4):A2145–A2172, 2012.

[17] S. Court and M. Fournié. A fictitious domain finite element method for simulations of fluid–structure interactions: The Navier–Stokes equations coupled with a moving solid. *J. Fluids Struct.*, 55:398–408, 2015.

[18] L. C. Cowsar, J. Mandel, and M. F. Wheeler. Balancing domain decomposition for mixed finite elements. *Math. Comp.*, 64(211):989–1015, 1995.

[19] N. Hale, N. J. Higham, and L. N. Trefethen. Computing $\mathbf{A}^{\alpha}$, $\log(\mathbf{A})$, and related matrix functions by contour integrals. *SIAM J. Numer. Anal.*, 46(5):2505–2523, 2008.

[20] F. Hansen and G. Kjærgå rd Pedersen. Jensen's inequality for operators and Löwner's theorem. *Math. Ann.*, 258(3):229–241, 1982.

[21] F. Hansen and G. K. Pedersen. Jensen's operator inequality. *Bull. London Math. Soc.*, 35(4):553–564, 2003.

[22] S. Harizanov, R. Lazarov, P. Marinov, S. Margenov, and Y. Vutov. Optimal solvers for linear systems with fractional powers of sparse SPD matrices. *arXiv preprint arXiv:1612.04846*, 2016.

[23] T. Kato. Notes on some inequalities for linear operators. *Math. Ann.*, 125:208–212, 1952.

[24] L. Knizhnerman and V. Simoncini. A new investigation of the extended Krylov subspace method for matrix function evaluations. *Numer. Linear Algebra Appl.*, 17(4):615–638, 2010.
[25] M. Kočvara, D. Loghin, and J. Turner. Constraint interface preconditioning for topology optimization problems. *SIAM J. Sci. Comput.*, 38(1):A128–A145, 2016.
[26] M. Kuchta, K.-A. Mardal, and M. Mortensen. Preconditioning trace coupled 3d-1d systems using fractional Laplacian. *arXiv preprint arXiv:1612.03574*, 2016.
[27] M. Kuchta, M. Nordaas, J. C. G. Verschaeve, M. Mortensen, and K.-A. Mardal. Preconditioners for saddle point systems with trace constraints coupling 2d and 1d domains. *SIAM J. Sci. Comput.*, 38(6):B962–B987, 2016.
[28] B. P. Lamichhane and B. I. Wohlmuth. Mortar finite elements for interface problems. *Computing*, 72(3-4):333–348, 2004.
[29] W. J. Layton, F. Schieweck, and I. Yotov. Coupling fluid flow with porous media flow. *SIAM J. Numer. Anal.*, 40(6):2195–2218 (2003), 2002.
[30] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I*. Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
[31] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M. M. Meerschaert, M. Ainsworth, et al. What is the fractional Laplacian? *arXiv preprint arXiv:1801.09767*, 2018.
[32] K.-A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, 18(1):1–40, 2011.
[33] P. Oswald. Multilevel norms for H-1/2. *Computing*, 61(3):235–255, 1998.
[34] R. Stevenson. Stable three-point wavelet bases on general meshes. *Numer. Math.*, 80(1):131–158, 1998.
[35] A. Tveito, K. H. Jæger, M. Kuchta, K.-A. Mardal, and M. E. Rognes. A cell-based framework for numerical modeling of electrical conduction in cardiac tissue. *Front. Phys.*, 5:48, 2017.
[36] B. I. Wohlmuth. A mortar finite element method using dual spaces for the Lagrange multiplier. *SIAM J. Numer. Anal.*, 38(3):989–1012, 2000.
[37] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34(4):581–613, 1992.
[38] Q. Yang, I. Turner, F. Liu, and M. Ilić. Novel numerical methods for solving the time-space fractional diffusion equation in two dimensions. *SIAM J. Sci. Comput.*, 33(3):1159–1180, 2011.

Paper III

An Auxiliary Space Preconditioner for Fractional Laplacian of Negative Order

T. Bærland

In preparation.

AN AUXILIARY SPACE PRECONDITIONER FOR FRACTIONAL LAPLACIAN OF NEGATIVE ORDER

TRYGVE BÆRLAND[†]

ABSTRACT. Coupled multiphysics problems often give rise to interface conditions naturally formulated in fractional Sobolev spaces. Here, both positive and negative fractionality are common. When designing efficient solvers for discretizations of such problems it would then be useful to have a preconditioner for the fractional Laplacian, $(-\Delta)^s$, with $s \in [-1, 1]$. Previously, additive multigrid preconditioners for the case when $s \geq 0$ have been proposed. In this work we complement this construction with auxiliary space preconditioners suitable when $s \leq 0$. These preconditioners are shown to be spectrally equivalent to $(-\Delta)^{-s}$, but requires preconditioners for fractional $H(\mathrm{div})$ operators with positive fractionality. We design such operators based on an additive multigrid approach. We finish with some numerical experiments, verifying the theoretical results.

1. INTRODUCTION

In this paper we are concerned with the design and analysis of preconditioners for the fractional Laplacian with negative exponent. More specifically, let $\Omega \subset \mathbb{R}^n$ be a bounded n-dimensional domain, and $s \in [0, 1]$ a parameter. We then consider the problem of finding u satisfying

$$(-\Delta)^{-s} u = f, \tag{1.1}$$

where f is given. Here, $H_0^1(\Omega)$ denotes the usual Sobolev space of square-integrable functions with square-integrable first order derivatives and zero trace on the boundary of Ω, and $H^{-1}(\Omega)$ denotes its dual space. Then $(-\Delta)^{-s}$ is defined from the spectral decomposition of $(-\Delta) : H_0^1(\Omega) \to H^{-1}(\Omega)$. Our aim in this work is to design efficient preconditioners for discretizations of $(-\Delta)^{-s}$.

Due to the negative exponent, common preconditioning strategies will fail in this context. In particular, for positive s, $(-\Delta)^s$ behaves similarly to $-\Delta$ in that the eigenfunctions corresponding to high eigenvalues are oscillatory, and vice versa. As such, the error from simple iteration schemes, like Richardson's iteration, are relatively smooth and can be well-represented on a coarser function space. This observation suggests that multigrid operators can provide efficient preconditioners for $(-\Delta)^s$, and motivated the construction of additive multigrid preconditioners in [5]. However, in our current context the roles are reversed. The oscillatory eigenfunctions of $(-\Delta)^{-s}$ correspond to the lower end of the spectrum. Then, neither simple smoothing procedures nor coarse grid correction will eliminate the oscillatory part of the error, and therefore we cannot hope for a straightforward multigrid method to work.

The preconditioners proposed in this work will be based on the auxiliary space preconditioner framework, [32]. Of particular note is that the transfer operator, whose role is to relate the original space and the auxiliary space, will be a differential operator. Consequently, the preconditioner on the auxiliary space will have to be spectrally equivalent to the inverse of a differential operator raised to a positive, fractional power. To motivate this, let $H_0^s(\Omega)$ denote the spectral interpolation (see [27, Ch. 2]) between $L^2(\Omega)$ and $H_0^1(\Omega)$, and $H^{-s}(\Omega)$ the dual space of $H_0^s(\Omega)$. Then $(-\Delta)^{-s}$ is an isomorphism from $H^{-s}(\Omega)$ to $H_0^s(\Omega)$. Following the

[†]DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OSLO, BLINDERN, OSLO, 0316 NORWAY
E-mail address: trygveba@math.uio.no.

The research leading to these results has received funding the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339643.

operator preconditioning framework in [28], an efficient preconditioner for (1.1) should be based on a linear, symmetric isomorphism $B^s : H_0^s(\Omega) \to H^{-s}(\Omega)$, the canonical choice being the Riesz mapping $(-\Delta)^s$. Consequently, the preconditioner should behave like a differential operator raised to a positive, fractional power. Then, roughly speaking, if B^s consists of applications of any standard differential operator, a correction is needed to compensate for this overshoot in fractionality. This correction will then behave like the inverse of a fractional differential operator of positive order. In particular, we will see that $B^s = -\operatorname{div} \Lambda^{-(1-s)}\nabla$ is spectrally equivalent to $(-\Delta)^s$. Here, $\Lambda = I - \nabla \operatorname{div}$ is the operator realizing the $H(\operatorname{div})$ inner product. Thus, the problem of preconditioning $(-\Delta)^{-s}$ will be transferred to the problem of preconditioning Λ^{1-s}, which is amenable to an analysis similar to the one made in [5]. This is an attractive idea because, as we will see, Λ^{1-s} behaves similarly to Λ, where preconditioning strategies based on multilevel decompositions have proved efficient, [2, 3, 20, 21, 23, 26].

Preconditioners, and in particular preconditioners based on multilevel decompositions, for (1.1) have previously been studied. For $s = \frac{1}{2}$, Bramble et al. designed a V-cycle multigrid operator in [11]. Their construction was based on posing (1.1) in the weaker H^{-1} inner product, where the operator they considered had spectral properties suitable for multigrid analysis. In [17], similar ideas were used to construct and analyze an additive multigrid operator. Hierarchical basis preconditioners, suitable for (1.1) when $s \in \left(-\frac{3}{2}, \frac{3}{2}\right)$ were constructed in [29]. These preconditioners were based on an L^2-orthogonal decomposition into each level of the grid hierarchy, and thus restricting its use to wavelet spaces where such decompositions are feasible. This was remedied for finite element spaces of low order in [12] by replacing L^2-projections onto each level by more cheaply computed operators. In all the preconditioners mentioned above, one drawback is that only simple scaling smoothers can be used, which might be seen as too restrictive. Lastly, in [30] the authors constructed optimal auxiliary space preconditioners for (1.1), but they needed to presuppose that a discrete version of $(-\Delta)^s$ was easily computable in the auxiliary space. We will in this work not assume such a discrete operator to be at our disposable. That is, the proposed preconditioners will not require the computation of $(-\Delta)^{\pm s}$, or the fractional power of any positive definite operator for that matter.

The reason for this design choice is that our main motivational application are coupled multihysics- and trace constraint problems, where fractional Sobolev spaces are part of a well-posed variational formulation, but the fractional Laplacian is absent from the operator characterizing the problem. As an illustrative example, let Ω be a bounded domain $\mathbb{R}^n$, with $n = 2$ or 3, and Γ denotes a structure in Ω or on its boundary with codimension 1. Consider the Poisson equation, $-\Delta u = f$ in Ω, with the constraint conditions $u = g$ on Γ for given data f and g. Imposing the trace constraint weakly, similarly to how it was done in [4], yields a saddle point system of the form

$$\begin{aligned} -\Delta u + T^* \lambda &= f, \quad x \in \Omega \\ Tu &= g, \quad x \in \Gamma \end{aligned} \tag{1.2}$$

where $T : H^1(\Omega) \to H^{\frac{1}{2}}(\Gamma)$ is the trace operator. The solution (u, λ) is sought in $H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$. Rewriting (1.2) in matrix form, we have

$$\mathscr{A} \begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix},$$

where $\mathscr{A} = \begin{pmatrix} -\Delta & T^* \\ T & 0 \end{pmatrix}$ is an isomorphism from $H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ to $\left(H^1(\Omega)\right)' \times H^{\frac{1}{2}}(\Gamma)$. By the framework in [28], a preconditioner for a discretization of (1.2) should be based on a symmetric isomorphism $\mathscr{B} : \left(H^1(\Omega)\right)' \times H^{\frac{1}{2}}(\Gamma) \to H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$, with the canonical choice

being

$$\mathscr{B} = \begin{pmatrix} (I-\Delta)^{-1} & 0 \\ 0 & (-\Delta_\Gamma)^{\frac{1}{2}} \end{pmatrix}. \tag{1.3}$$

Cheaply computable operators, spectrally equivalent to $(I-\Delta)^{-1}$ are well known. The second block, $(-\Delta_\Gamma)^{\frac{1}{2}}$ is as such the challenging part when designing preconditioners based on (1.3). See also that the fractional Laplacian only appears in $\mathscr{B}$, and not in $\mathscr{A}$.

We remark that even if the above example is relatively simple, similar techniques can be used in problems where different PDEs are posed on separate domains and linked through some continuity conditions on a common interface Γ. One or more of these continuity conditions can then be enforced weakly by use of Lagrange multipliers, which often will posed in a fractional Sobolev space. When preconditioning the resultant system, the problem of establishing a computationally feasible operator, spectrally equivalent to $(-\Delta_\Gamma)^{\pm\frac{1}{2}}$ persists. For instance, in [25] the authors study a multiphysics problem posed on domains of different topological dimension, and continuity is imposed weakly using a Lagrange multiplier. Other applications can be found in [6], where the no-slip condition on the surface of a falling body in a fluid is imposed weakly, or in [31], where the potential jump on a membrane of a cardiac cell is treated similarly. If the embedded structure Γ in (1.2) instead has codimension 2, then numerical experiments in [24] suggests that block diagonal preconditioners where one block is based on $(-\Delta_\Gamma)^{-s}$, with $s \in (-0.2, -0.1)$, provide efficient preconditioners.

The current paper can in a couple of ways be viewed as continuation of [5]. Firstly, we define efficient preconditioners for the fractional Laplacian when the exponent $s \in [-1, 0]$, complementing the preconditioners introduced in the previous work. Secondly, in this work we generalize the results from [5] to positive fractional powers of Λ. The analysis will aim to substantiate the intuition that if additive multilevel methods are efficient for $s = 0$ and $s = 1$, then "by interpolation" it should be efficient for every $s \in (0, 1)$. We remark, however, that the analysis on these multilevel methods for fractional $H(\mathrm{div})$ operators assumes certain two-level error estimates on Λ^{1-s} that will go unproven in this work. This is an unsatisfactory state of affairs, but we do give an approach for how these error estimates can be proven, as well as motivate their veracity. The techniques we propose will borrow from [9], and would require a substantial additional toolset. As such, it is here left as future work.

The remainder of the current paper is structured as follows. In section 2 we describe the notation used throughout the paper, as well as give brief introductions to the theory of interpolation spaces and some useful results in functional analysis. Section 3 is devoted to substantiating the above heuristic argument, and show that provided we are given efficient preconditioners for fractional $H(\mathrm{div})$ operators with positive exponent, we can construct efficient preconditioners for the fractional Laplacian with negative exponent. Then, in section 4 we propose such preconditioners as additive multigrid operators and give sufficient conditions under which they are efficient. Lastly, in section 5 we provide a series of numerical experiments verifying the theoretical results obtained in this work.

2. Preliminaries

Let Ω be a bounded, polygonal domain in $\mathbb{R}^n$, with boundary $\partial\Omega$. We denote by $L^2(\Omega)$ the space of square integrable functions on Ω, with inner product $(\cdot,\cdot)$, and norm $\|\cdot\|$. We denote by $H^1(\Omega)$ the usual Sobolev space of functions in $L^2(\Omega)$ with all first-order derivatives also in $L^2(\Omega)$. The closure of smooth functions with compact support in Ω we denote by $H_0^1(\Omega)$, and its dual space is $H^{-1}(\Omega)$. For $k \in \{-1, 1\}$, the inner product and norm of $H^k(\Omega)$ we denote by $(\cdot,\cdot)_k$ and $\|\cdot\|_k$, respectively. Further, we let $H(\mathrm{div};\Omega)$ denote the Hilbert space of square-integrable vector fields on Ω with square-integrable divergence, while we write $H(\mathbf{curl};\Omega)$ to mean the space of square-integrable vector fields on Ω with square-integrable $\mathbf{curl}$. We let

$\Lambda(\cdot,\cdot)$ denote the standard inner product on $H(\mathrm{div};\Omega)$ defined by

$$\Lambda(\sigma,\tau) = (\sigma,\tau) + (\mathrm{div}\,\sigma, \mathrm{div}\,\tau), \quad \sigma,\tau \in H(\mathrm{div};\Omega).$$

In general, a Hilbert space X is equipped with an inner product and norm, which we denote by $(\cdot,\cdot)_X$ and $||\cdot||_X$, respectively, and its dual is denoted by X'. For two Hilbert spaces X and Y, we write $\mathscr{L}(X,Y)$ to mean the space of bounded linear operators $T : X \to Y$, which we equip with the usual operator norm

$$||T||_{\mathscr{L}(X,Y)} = \sup_{x\in X} \frac{||Tx||_Y}{||x||_X}.$$

Let now A be a symmetric positive definite operator on a Hilbert space X. For sake of simplicity, we assume the spectrum of A to be wholly discrete, i.e. A has empty continuous- and residual spectrum. Denote by $\{(\lambda_k,\phi_k)\}_{k=1}^{\infty}$ the set of eigenpairs of A, normalized so that

$$(\phi_k,\phi_l)_X = \delta_{k,l},$$

where $\delta_{k,l}$ is the Kronecker delta. Then ϕ_k, for $k = 1,2,\ldots$ forms an orthonormal basis of X, and if $u \in X$ has the representation $u = \sum_{k=1}^{\infty} c_k\phi_k$, then

$$Au = \sum_{k=1}^{\infty} \lambda_k c_k \phi_k.$$

For $s \in \mathbb{R}$, we define the fractional power A^s of A by

$$A^s u = \sum_{k=1}^{\infty} \lambda_k^s c_k \phi_k.$$

If A is only positive semi-definite, then we must restrict to $s > 0$. If B is another symmetric positive semi-definite operator on X, we write $A \le B$ if for every $u \in X$

$$(Au,u)_X \le (Bu,u)_X$$

holds. Note that $A \ge 0$ is equivalent to saying that A is positive semi-definite. In addition, we shall write $A \le 1$ to mean that $(Au,u)_X \le (u,u)_X$ for every $u \in X$.

A result in operator theory is the Löwner-Heinz inequality, which in our case states that if $A \le B$, then

$$A^s \le B^s, \quad s \in [0,1], \tag{2.1}$$

cf. for instance [22]. Inequality (2.1) means that the function x^s with $x \in [0,\infty)$ is operator monotone for $s \in [0,1]$. It follows that $-(x)^s$ is operator convex (cf. [18, Thm. 2.1 and 2.5]), that is, for any two symmetric positive semi-definite operators A and B on a Hilbert space X, the inequality

$$\lambda A^s + (1-\lambda)B^s \le (\lambda A + (1-\lambda)B)^s$$

holds for every $\lambda \in [0,1]$. A key result regarding operator convex functions is the Jensen's operator inequality (cf. [19, Theorem 2.1]). The version we will use in the current work states that for any bounded, symmetric positive semi-definite operator A on X, and $P : X \to X$ so that $P^*P \le 1$

$$P^*A^sP \le (P^*AP)^s. \tag{2.2}$$

We will at numerous times in this paper be in a position where we want to use (2.2), but where P is a contraction between different Hilbert spaces. Thus, we make the following slight generalization of (2.2).

Lemma 2.1. *Let X_1 and X_2 be two Hilbert spaces, and $T : X_1 \to X_2$ an operator satisfying $T^*T \le 1$ on X_1. Further, assume that A is a bounded, symmetric positive semi-definite operator on X_2. Then*

$$T^*A^sT \le (T^*AT)^s \tag{2.3}$$

for every $s \in [0,1]$.

Proof. See that (2.3) holds for $s = 0$ and $s = 1$, so fix $s \in (0,1)$. We define the auxiliary Hilbert space $X = X_1 \oplus X_2$, with inner product inherited from the inner products on X_1 and X_2. Now, define linear operators P and $\tilde{A}$ on X as

$$P = \begin{pmatrix} 0 & 0 \\ T & 0 \end{pmatrix}, \quad \text{and} \quad \tilde{A} = \begin{pmatrix} 0 & 0 \\ 0 & A \end{pmatrix}.$$

A simple calculation then shows that

$$P^*P = \begin{pmatrix} T^*T & 0 \\ 0 & 0 \end{pmatrix} \leq 1,$$

by the assumption on T. Similarly,

$$P^*\tilde{A}^\theta P = \begin{pmatrix} T^*A^\theta T & 0 \\ 0 & 0 \end{pmatrix}$$

for every $\theta > 0$. Then, we have from the standard Jensen's inequality in (2.2) that

$$\begin{pmatrix} T^*A^sT & 0 \\ 0 & 0 \end{pmatrix} = P^*\tilde{A}^sP \leq \left(P^*\tilde{A}P\right)^s = \begin{pmatrix} (T^*AT)^s & 0 \\ 0 & 0 \end{pmatrix}.$$

In particular, $T^*A^sT \leq (T^*AT)^s$, which completes the proof. □

2.1. **Interpolation spaces.** In defining fractional Sobolev spaces and fractional $H(\mathrm{div})$ spaces, we will use some results from interpolation theory, as presented in [27], and so we shall make a quick review.

Let X and Y be separable Hilbert spaces with inner products $(\cdot,\cdot)_X$ and $(\cdot,\cdot)_Y$, and corresponding norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. Furthermore, we assume that $X \subset Y$, with X dense in Y and continuous injection. In this case we call X and Y compatible.

Denote by $D(A)$ the set of $u \in Y$ so that the linear form

$$L_u(v) = (u,v)_X \quad v \in X$$

is continuous in Y. Following the discussion in [27], we note that $D(A)$ is dense in Y. Using Riesz' representation theorem, there is a $w \in Y$ so that

$$(w,v)_Y = (u,v)_X .$$

The mapping $u \mapsto w$ defines an unbounded linear operator $A : D(A) \to Y$, which is defined by

$$(Au,v)_Y = (u,v)_X . \tag{2.4}$$

Clearly, A is self-adjoint and positive. Using the spectral decomposition of self-adjoint operators, we may define the powers, A^θ, $\theta \in \mathbb{R}$, of A. We define interpolation spaces in the following way:

Definition 2.1. Let X and Y satisfy the above assumptions. For $\theta \in [0,1]$ we define the interpolation space

$$[Y,X]_\theta = D(A^{\frac{\theta}{2}}) = \left\{ u \in Y : A^{\frac{\theta}{2}}u \in Y \right\} \tag{2.5}$$

with norm given by the graph norm

$$\|u\|_{[Y,X]_\theta} := \left(\|u\|_Y^2 + \left(A^\theta u, u\right)_Y\right)^{\frac{1}{2}}. \tag{2.6}$$

It follows by the definition that

$$[Y,X]_0 = Y, \text{ and } [Y,X]_1 = X.$$

The following is a key Theorem in interpolation theory.

Theorem 2.1. *Let $\{X,Y\}$ and $\{\mathscr{X},\mathscr{Y}\}$ be two pairs of compatible Hilbert spaces. Further, let T be a continuous operator $\mathscr{L}(X,\mathscr{X})\cap\mathscr{L}(Y,\mathscr{Y})$, so that*

$$\|Tu\|_{\mathscr{X}} \leq M_0\|u\|_X,$$
$$\|Tu\|_{\mathscr{Y}} \leq M_1\|u\|_Y.$$

Then $T\in\mathscr{L}([Y,X]_\theta,[\mathscr{Y},\mathscr{X}]_\theta)$, and

$$\|Tu\|_{[\mathscr{Y},\mathscr{X}]_\theta} \leq CM_0^{1-\theta}M_1^{\theta}\|u\|_{[Y,X]_\theta}, \tag{2.7}$$

where C is a constant independent of T, $\mathscr{X}$, and $\mathscr{Y}$.

If we now make the identification $Y = Y'$, then $Y\subset X'$ is dense, with continuous embedding. Thus, the interpolation space $\left[X',Y\right]_\theta$ is well-defined for $\theta\in[0,1]$ according to definition 2.1. Moreover, we have that (cf. [27, Thm. 6.2])

$$\left[X',Y\right]_\theta = [Y,X]'_{1-\theta}. \tag{2.8}$$

It is well-known that $H^1(\Omega)$ is densely and continuously embedded in $L^2(\Omega)$, which implies that we can define the fractional Sobolev spaces $H^s(\Omega)$ for $s\in[0,1]$ as

$$H^s(\Omega) := \left[L^2(\Omega),H^1(\Omega)\right]_s$$

We go on to define $H_0^s(\Omega)$ as the closure in $H^s(\Omega)$ of smooth and compactly supported functions on Ω, while for $s\in[-1,0]$, we define

$$H^s(\Omega) = H_0^{-s}(\Omega)'$$

We note that this definition for negative fractional Sobolev spaces is equivalent to interpolation between $H^{-1}(\Omega)$ and $L^2(\Omega)$.

Similarly, we define the fractional $H(\mathrm{div};\Omega)$ space as

$$H^s(\mathrm{div};\Omega) := \left[L^2(\Omega),H(\mathrm{div};\Omega)\right]_s. \tag{2.9}$$

2.2. **Discrete interpolation spaces.** The discrete variant of fractional operators can be constructed analogously to the continuous setting. Suppose $X_h\subset X$ is a finite-dimensional subspace. We can define the operator $A_h : X_h\to X_h$ by

$$(A_h v,w)_Y = (v,w)_X.$$

We note that because X_h is finite-dimensional, all norms are equivalent, and in particular, A_h is a bounded operator. Since A_h is SPD, we can define its fractional powers A_h^θ for $\theta\in\mathbb{R}$, and discrete fractional norms $\|\cdot\|_{\theta,h}^2 := \left(A_h^\theta\cdot,\cdot\right)$. When $\theta=0$ and $\theta=1$, the norm $\|\cdot\|_{\theta,h}$ coincides with the Y- and X norm, respectively. Furthermore, for $\theta\in(0,1)$ the discrete norm is equivalent to the $[Y,X]_\theta$ norm, with constants of equivalence independent of X_h (cf. [1, Proposition 3.2])

Suppose now that we have an additional finite-dimensional subspace $X_H\subset X_h$. Analogously to before we can define the SPD operator $A_H : X_H\to X_H$, and its fractional powers A_H^θ, with $\theta\in\mathbb{R}$. In the case of $\theta=0$ or $\theta=1$ we have that

$$\left(A_H^\theta v,w\right)_Y = \left(A_h^\theta v,w\right)_Y,\quad v,w\in X_H.$$

However, this inheritance of bilinear forms fails when $\theta\in(0,1)$. Getting ahead of ourselves, the inheritance of bilinear forms is a common assumption in the design and analysis of multigrid algorithms. Therefore, that the inheritance fails to hold when $\theta\in(0,1)$ can be detrimental. The following lemma shows that we are able to recover one of the key inequalities used in [14] in the analysis of multigrid algorithms on non-inherited bilinear forms.

Lemma 2.2. *Let $\theta \in [0,1]$. We have that restricted to X_H*

$$A_h^\theta \leq A_H^\theta.$$

That is, for every $v \in X_H$

$$\left(A_h^\theta v, v\right)_Y \leq \left(A_H^\theta v, v\right)_Y. \tag{2.10}$$

Proof. As already noted, for $\theta = 0$ and $\theta = 1$ (2.10) holds with equality, so for the remainder of the proof let $0 < \theta < 1$.

Let $I_H : X_H \to X_h$ be the inclusion operator, and I_H^* its adjoint with respect to the Y-inner product. Then, $I_H^* I_H$ is the identity on X_H, so $I_H^* I_H \leq 1$ holds trivially. By Lemma 2.1, we thus have that

$$I_H^* A_h^\theta I_H \leq \left(I_H^* A_h I_H\right)^\theta. \tag{2.11}$$

The result follows from (2.11) and the observation that $A_H = I_H^* A_h I_H$. □

3. Preconditioner for fractional Laplacian

In this section we will establish a way to construct preconditioners for $(-\Delta)^{-s}$ when $s \in [0,1]$. We will begin by first considering the continuous setting, which will motivate the construction of preconditioners for a discretization of $(-\Delta)^{-s}$. We define $-\Delta : H_0^1(\Omega) \to H^{-1}(\Omega)$ by

$$((-\Delta)u, v) = (\nabla u, \nabla v), \quad u, v \in H_0^1(\Omega).$$

In view of the interpolation theory discussed in the previous section, it is evident that $(-\Delta)^s$ is well-defined for any $s \in [0,1]$, and it is an isomorphism from $H_0^s(\Omega)$ to $H^{-s}(\Omega)$. We denote its inverse by $(-\Delta)^{-s}$, and consider the problem of finding $u \in H^{-s}(\Omega)$ so that

$$(-\Delta)^{-s} u = f, \tag{3.1}$$

for a given $f \in H_0^s(\Omega)$. To precondition (3.1), we seek a self-adjoint isomorphism $B^s : H_0^s(\Omega) \to H^{-s}(\Omega)$, so that

$$C_1 \|u\|_{H^{-s}(\Omega)} \leq (B^s u, u) \leq C_2 \|u\|_{H^{-s}(\Omega)} \tag{3.2}$$

for some constant $C_1, C_2 > 0$.

Now, consider the gradient operator, ∇. It is clear that $\nabla \in \mathscr{L}(H_0^1(\Omega), L^2(\Omega))$. On $L^2(\Omega)$, we define

$$(\nabla u, \tau) = -(u, \operatorname{div} \tau), \quad u \in L^2(\Omega),\ \tau \in H(\operatorname{div}; \Omega).$$

Using integration by parts, this reduces to the standard ∇ when $u \in H_0^1(\Omega)$. Moreover, we have that

$$\|\nabla u\|_{H(\operatorname{div},\Omega)'} = \sup_{\tau \in H(\operatorname{div};\Omega)} \frac{(u, \operatorname{div} \tau)}{\|\tau\|_{H(\operatorname{div};\Omega)}} \leq \|u\|.$$

Thus,

$$\nabla \in \mathscr{L}(H_0^1(\Omega), L^2(\Omega)) \cap \mathscr{L}(L^2(\Omega), H(\operatorname{div};\Omega)'),$$

and Theorem 2.1 then implies that $\nabla \in \mathscr{L}\left(H_0^s(\Omega), \left[H(\operatorname{div};\Omega)', L^2(\Omega)\right]_s\right)$. In view of (2.8) and (2.9) we can rewrite this as

$$\nabla \in \mathscr{L}(H_0^s(\Omega), H^{1-s}(\operatorname{div};\Omega)'). \tag{3.3}$$

Suppose now that we are given a self-adjoint isomorphism $B_{\text{div}}^{1-s} : H^{1-s}(\operatorname{div};\Omega)' \to H^{1-s}(\operatorname{div};\Omega)$ which for every $\tau \in H^{1-s}(\operatorname{div};\Omega)'$ satisfies

$$C_{d,1} \|\tau\|_{H^{1-s}(\operatorname{div};\Omega)'}^2 \leq \left(B_{\text{div}}^{1-s} \tau, \tau\right) \leq C_{d,2} \|\tau\|_{H^{1-s}(\operatorname{div};\Omega)'}^2 \tag{3.4}$$

for some constants $C_{d,1}, C_{d,2} > 0$ independent of τ. We then define

$$B^s = \nabla^* B_{\text{div}}^{1-s} \nabla. \tag{3.5}$$

Our aim is to show that B^s defined by (3.5) satisfies (3.2). We begin by observing that B^s is self-adjoint and maps elements from $H_0^s(\Omega)$ to $H^{-s}(\Omega)$. Moreover, the mapping property of ∇ in (3.3) and the boundedness of B_{div}^{1-s} imply that $B^s \in \mathscr{L}(H_0^s(\Omega), H^{-s}(\Omega))$.

Establishing the lower bound of (3.2) is more difficult in that we want to interpolate between lower bounds on the gradient operator. However, Theorem 2.1 is not applicable in this setting. To overcome this problem, we will interpolate between bounds on a left-inverse, T, of ∇. In this work, we employ the Bogovskiĭ operator established in [15]. If Ω is star-shaped with respect to an open ball B, T takes for a vector field τ the explicit form

$$T\tau(x) = \int_\Omega K(x,y)(x-y)\cdot\tau(y)\mathrm{d}y, \quad \text{where } K(x,y) = \int_1^\infty (t-1)^{n-1}\theta(y+t(x-y))\mathrm{d}t.$$

Here, $\theta \in \mathscr{C}_0^\infty(\mathbb{R}^n)$ with support contained in B and integrates to 1. It can be checked that T is a left-inverse of ∇, and satisfies

$$T \in \mathscr{L}(L^2(\Omega), H_0^1(\Omega)) \cap \mathscr{L}(H(\mathrm{div};\Omega)', L^2(\Omega)), \tag{3.6}$$

see [15, Cor. 3.4]. We note that the definition of T can be extended to general Lipschitz domains — as such domains are finite unions of star-shaped domains — with the same mapping properties. From (3.6) and Theorem 2.1 we have that

$$T \in \mathscr{L}(H^{1-s}(\mathrm{div};\Omega)', H_0^s(\Omega)). \tag{3.7}$$

Finally, we are in a position to prove that B^s satisfies (3.2), and hence is a suitable preconditioner for (3.1). The result is stated in the following theorem.

Theorem 3.1. *Let $s \in [0,1]$, and $B_{\mathrm{div}}^{1-s} : H^{1-s}(\mathrm{div};\Omega)' \to H^{1-s}(\mathrm{div};\Omega)$ satisfy* (3.4). *Then B^s defined by* (3.5) *satisfies* (3.2) *with*

$$C_1 = C_{d,1}\|T\|^{-2}_{\mathscr{L}(H^{1-s}(\mathrm{div};\Omega)', H_0^s(\Omega))}, \quad \text{and} \quad C_2 = C_{d,2}\|\nabla\|^2_{\mathscr{L}(H_0^s(\Omega), H^{1-s}(\mathrm{div};\Omega)')}. \tag{3.8}$$

Proof. Fix $s \in [0,1]$, and take any $u \in H_0^s(\Omega)$. From the definition of B^s, see that

$$(B^s u, u) = \left(B_{\mathrm{div}}^{1-s}\nabla u, \nabla u\right).$$

From the second inequality of (3.4) and the mapping property of ∇ in (3.3) we deduce that

$$(B^s u, u) \le C_{d,2}\|\nabla u\|^2_{H^{1-s}(\mathrm{div};\Omega)'} \le C_{d,2}\|\nabla\|^2_{\mathscr{L}(H_0^s(\Omega), H^{1-s}(\mathrm{div};\Omega)')}\|u\|^2_{H_0^s(\Omega)},$$

which proves the second inequality of (3.2) with C_2 as given in (3.8).

We can treat the lower bound of (3.2) similarly, but now use the lower bound of (3.4) and (3.7). That is, we have

$$(B^s u, u) \ge C_{d,1}\|\nabla u\|^2_{H^{1-s}(\mathrm{div};\Omega)'}, \tag{3.9}$$

and, since $T\nabla$ is the identity on $H_0^s(\Omega)$,

$$\|u\|_{H_0^s(\Omega)} = \|T\nabla u\|_{H_0^s(\Omega)} \le \|T\|_{\mathscr{L}(H^{1-s}(\mathrm{div};\Omega)', H_0^s(\Omega))}\|\nabla u\|_{H^{1-s}(\mathrm{div};\Omega)'}. \tag{3.10}$$

Combining (3.9) and (3.10) yields

$$(B^s u, u) \ge C_{d,1}\|T\|^{-2}_{\mathscr{L}(H^{1-s}(\mathrm{div};\Omega)', H_0^s(\Omega))}\|u\|^2_{H_0^s(\Omega)}.$$

□

Remark 1. With the definition of B^s given in (3.5), we have essentially translated the problem of preconditioning $(-\Delta)^{-s}$ to the problem of preconditioning Λ^{1-s}. The advantage of this is that the latter problem has positive exponent, and so, as we will see, will have similar spectral properties to Λ, for which efficient preconditioning strategies have been studied earlier.

3.1. **Discrete setting.** We will now use the construction of B^s from the previous section as motivation to construct an analogous discrete operator. To that end, let $\mathscr{T}_h$ be a shape-regular triangulation of Ω, with characteristic mesh size h. For $r \geq 0$, we let S_h denote the space of all discontinuous, piecewise polynomials of degree at most r, subordinate to $\mathscr{T}_h$. That is,

$$S_h = \left\{ u \in L^2(\Omega) : u\big|_T \in P_r(T), \forall T \in \mathscr{T}_h \right\}.$$

We further let $V_h = \mathscr{RT}_r(\mathscr{T}_h) \subset H(\mathrm{div};\Omega)$ be the Raviart-Thomas space of index r, and $C_h = \mathscr{NE}_r(\mathscr{T}_h) \subset H(\mathbf{curl};\Omega)$ the Nedelec space of first kind of index r, both relative to the triangulation $\mathscr{T}_h$. It is then well-known that $\mathbf{curl}\,(C_h) \subset V_h$, and $\mathrm{div}\,(V_h) \subset S_h$. We define the discrete gradient operator $\nabla_h : S_h \to V_h$ by

$$(\nabla_h u, \tau) = -(u, \mathrm{div}\,\tau), \quad u \in S_h,\ \tau \in V_h, \tag{3.11}$$

and discrete curl operator $\mathbf{curl}_h : V_h \to C_h$ by

$$(\mathbf{curl}_h \tau, q) = (\tau, \mathbf{curl}\, q), \quad \tau \in V_h,\ q \in C_h. \tag{3.12}$$

With these definitions, we have the discrete Helmholtz decomposition $V_h = \mathbf{curl}\, C_h \oplus \nabla_h S_h$. That is, every $\tau \in V_h$ can be written as

$$\tau = \nabla_h u + \mathbf{curl}\, q, \tag{3.13}$$

for unique $u \in S_h$ and $q \in \mathbf{curl}_h V_h$. Cf. e.g. [3]. Moreover, this decomposition is orthogonal in both $(\cdot,\cdot)$ and $\Lambda(\cdot,\cdot)$.

To get a discrete analogue of the preconditioner B^s in (3.5), we further need to define discrete counterparts to the operators $-\Delta$ and Λ. To that end, we define the discrete Laplacian as $A_h := \nabla_h^* \nabla_h$, i.e. A_h is the symmetric operator on S_h that satisfies

$$(A_h u, v) = (\nabla_h u, \nabla_h v), \quad u, v \in S_h. \tag{3.14}$$

Lastly, since V_h is a conforming discretization of $H(\mathrm{div};\Omega)$, we simply take $\Lambda_h : V_h \to V_h$ to be the restriction of Λ to V_h. In other words,

$$(\Lambda_h \sigma, \tau) = \Lambda(\sigma, \tau), \quad \sigma, \tau \in V_h.$$

It is well-known that (cf. for instance [8]), with these particular choices of S_h and V_h, there is a $\beta > 0$ indepedent of h so that for every $u \in S_h$

$$\sup_{\tau \in V_h} \frac{(u, \mathrm{div}\,\tau)}{(\Lambda_h \tau, \tau)^{\frac{1}{2}}} \geq \beta \|u\|. \tag{3.15}$$

This implies that $\mathrm{div} : V_h \to S_h$ is surjective or, equivalently, that $\nabla_h : S_h \to V_h$ is injective. As a consequence, A_h is not only symmetric, but also positive-definite, and so A_h^s is well-defined for every $s \in \mathbb{R}$. The discrete counterpart to (3.1) is then to find, for $s \in [0,1]$ and $f \in S_h$, a $u \in S_h$ such that

$$A_h^{-s} u = f. \tag{3.16}$$

To precondition (3.16), we seek a symmetric positive definite operator $B_h^s : S_h \to S_h$ which is easy to compute and spectrally equivalent to A_h^s, with constants of equivalence independent of h. Using the previous continuous preconditioner defined in (3.5) as motivation, we will see that

$$B_h^s = \nabla_h^* B_{\mathrm{div},h}^{1-s} \nabla_h, \tag{3.17}$$

where $B_{\mathrm{div},h}^{1-s} : V_h \to V_h$ is a symmetric positive definite operator spectrally equivalent to $\Lambda_h^{-(1-s)}$, leads to an efficient preconditioner for A_h^{-s}. The key result in this section is given in Theorem 3.2 below, whose proof will resemble the argument we made in the continuous setting. In particular, we must ensure that ∇_h has the appropriate upper and lower bounds when $s = 0$ and $s = 1$. As we will see, the intermediate cases will then follow from Jensen's operator inequality.

For the upper bounds of ∇_h, we have from the definitions of ∇_h and Λ_h that

$$\begin{aligned}\left(\Lambda_h^{-1}\nabla_h u, \nabla_h u\right) &= \left\|\Lambda_h^{-\frac{1}{2}}\nabla_h u\right\|^2 = \sup_{\tau\in V_h}\frac{\left(\Lambda_h^{-\frac{1}{2}}\nabla_h u, \tau\right)^2}{\|\tau\|^2} \\ &= \sup_{\tau\in V_h}\frac{(\nabla_h u,\tau)^2}{(\Lambda_h\tau,\tau)} \le \|u\|^2,\end{aligned} \tag{3.18}$$

which is the discrete analogue to $\nabla \in \mathscr{L}(L^2(\Omega), H(\mathrm{div};\Omega)')$. The discrete analogue to $\nabla \in \mathscr{L}(H_0^1(\Omega), L^2(\Omega))$ is simply that $\|\nabla_h u\|^2 = (A_h u, u)$.

For the necessary lower bounds on ∇_h, we define $L : V_h \to S_h$ by $L\tau = u$ according to the discrete Helmholtz decomposition (3.13). It is then evident that $L\nabla_h$ is the identity on S_h. That L satisfies the discrete analogues to (3.6) is given in the following lemma.

Lemma 3.1. *With $L : V_h \to S_h$ as defined above, it holds for every $\tau \in V_h$ that*

$$\|L\tau\|^2 \le \beta^{-2}\left(\Lambda_h^{-1}\tau,\tau\right), \quad \text{and} \quad (A_h L\tau, L\tau) \le \|\tau\|^2, \tag{3.19}$$

where β is given by (3.15).

Proof. Fix $\tau \in V_h$, and let $u = L\tau$. From (3.15) and the decomposition (3.13), we have that

$$\beta\|u\| \le \sup_{\sigma\in V_h}\frac{(\nabla_h u,\sigma)}{(\Lambda_h\sigma,\sigma)^{\frac{1}{2}}} \le \sup_{\sigma\in V_h}\frac{(\tau,\sigma)}{(\Lambda_h\sigma,\sigma)^{\frac{1}{2}}}.$$

Replacing σ by $\Lambda_h^{-\frac{1}{2}}\sigma$ in the above yields

$$\beta\|u\| \le \sup_{\sigma\in V_h}\frac{\left(\Lambda_h^{-\frac{1}{2}}\tau,\sigma\right)}{\|\sigma\|} \le \left(\Lambda_h^{-1}\tau,\tau\right)^{\frac{1}{2}},$$

which proves the first inequality of (3.19).

The definitions of A_h and L, and the L^2-orthogonality of the decomposition (3.13) imply the second inequality of (3.19), since

$$(A_h L\tau, L\tau) = \|\nabla_h u\|^2 \le \|\tau\|^2.$$

□

We are now in a position to state and prove the main spectral equivalence result of this section, from which the spectral equivalence between B_h^s given in (3.17) and A_h^s will readily follow.

Theorem 3.2. *Let ∇_h, A_h and Λ_h be defined as above, and let $s \in [0,1]$. Then, for every $u \in S_h$*

$$\beta^{2(1-s)}\left(A_h^s u, u\right) \le \left(\Lambda_h^{-(1-s)}\nabla_h u, \nabla_h u\right) \le \left(A_h^s u, u\right), \tag{3.20}$$

where β is given by (3.15).

Proof. Fix $u \in S_h$ and $s \in [0,1]$. We begin by proving the second inequality of (3.20). Define $T_1 = \Lambda_h^{-\frac{1}{2}}\nabla_h : S_h \to V_h$. From (3.18) it follows that $T_1^* T_1 \le 1$. Thus, Lemma 2.1 implies that

$$T_1^*\Lambda_h^s T_1 \le \left(T_1^*\Lambda_h T_1\right)^s. \tag{3.21}$$

Inserting the definition of T_1 into (3.21) yields

$$\nabla_h^*\Lambda_h^{-(1-s)}\nabla_h \le \left(\nabla_h^*\nabla_h\right)^s = A_h^s,$$

which is equivalent to the second inequality of (3.20).

In proving the first inequality of (3.20), we will again make use of Lemma 2.1. To that end, we now set $T_2 = \beta L \Lambda_h^{\frac{1}{2}}$, and from Lemma 3.1 it follows that $T_2^* T_2 \leq 1$. Thus, an application of Lemma 2.1 yields

$$T_2^* A_h^s T_2 \leq (T_2^* A_h T_2)^s,$$

which after inserting the definition of T_2 becomes

$$\beta^{2(1-s)} \Lambda_h^{\frac{1}{2}} L^* A_h^s L \Lambda_h^{\frac{1}{2}} \leq \left(\Lambda_h^{\frac{1}{2}} L^* A_h L \Lambda_h^{\frac{1}{2}} \right)^s. \tag{3.22}$$

From Lemma 3.1 $L^* A_h L \leq 1$. Pre- and post multplying this inequality by $\Lambda_h^{\frac{1}{2}}$ and using the Löwner-Heinz inequality (2.1), we deduce that

$$\left(\Lambda_h^{\frac{1}{2}} L^* A_h L \Lambda_h^{\frac{1}{2}} \right)^s \leq \Lambda_h^s. \tag{3.23}$$

We now use (3.22) together with (3.23) and pre- and post multiply by $\Lambda_h^{-\frac{1}{2}}$ to get

$$\beta^{2(1-s)} L^* A_h^s L \leq \Lambda_h^{-(1-s)}.$$

Finally, multiplying from the left by ∇_h^* and from the right by ∇_h, and using that both $\nabla_h^* L^*$ and $L \nabla_h$ are the identity on S_h, we arrive at

$$\beta^{2(1-s)} A_h^s \leq \nabla_h^* \Lambda_h^{-(1-s)} \nabla_h,$$

which is the first inequality of (3.20). □

Corollary 3.1. *Under the same assumptions as in Theorem 3.2, suppose we are given a symmetric positive definite operator $B_{\mathrm{div},h}^{1-s} : V_h \to V_h$ spectrally equivalent to $\Lambda_h^{-(1-s)}$. That is, there are constants $C_1, C_2 > 0$ so that*

$$C_1 \left(\Lambda_h^{-(1-s)} \tau, \tau \right) \leq \left(B_{\mathrm{div},h}^{1-s} \tau, \tau \right) \leq C_2 \left(\Lambda_h^{-(1-s)} \tau, \tau \right) \tag{3.24}$$

for every $\tau \in V_h$. Then B_h^s defined by (3.17) *satisfies*

$$C_1 \beta^{2(1-s)} \left(A_h^s u, u \right) \leq \left(B_h^s u, u \right) \leq C_2 \left(A_h^s u, u \right) \tag{3.25}$$

for every $u \in S_h$.

Proof. Take any $s \in [0,1]$ and $u \in S_h$. By the definition of B_h^s, the second inequalities of (3.24) and (3.20)

$$\left(B_h^s u, u \right) \leq C_2 \left(\Lambda_h^{-(1-s)} \nabla_h u, \nabla_h u \right) \leq C_2 \left(A_h^s u, u \right),$$

which proves the second inequality of (3.25). The first inequality is proved similarly, using the lower bounds in (3.24) and (3.20). □

Remark 2. At this point it is worth remarking on the implementation of B_h^s. In computer code, a function $u \in S_h$ can have two distinct representations as vectors in $\mathbb{R}^{N_S}$, where $N_S = \dim S_h$. Let $\{\phi_h^i\}_{i=1}^{N_S}$ be a basis for S_h. Then, if $u = \sum_{i=1}^{N_S} c_i \phi_h^i$, we call the vector $\boldsymbol{u} = (c_1, \ldots, c_{N_S})^T \in \mathbb{R}^{N_S}$ the coefficient vector representation of u, while the vector $\tilde{\boldsymbol{u}} \in \mathbb{R}^{N_S}$ with entries $\tilde{\boldsymbol{u}}_i = \left(u, \phi_h^i \right)$, the dual vector representation of u. Cf. e.g. [10, Sec. 15] for more details. Let $\{\psi_h^i\}_{i=1}^{N_V}$, with $N_V = \dim V_h$, be a basis for V_h. For $\tau \in V_h$, let τ and $\tilde{\tau}$ be the analogous coefficient- and dual vector representations of τ. The most straightforward matrix realization of ∇_h is then the matrix $\mathsf{D}_h \in \mathbb{R}^{N_V \times N_S}$ with entries

$$(\mathsf{D}_h)_{i,j} = -\left(\phi_h^i, \operatorname{div} \psi_h^i \right).$$

We see that D_h takes coefficient vectors in $\mathbb{R}^{N_S}$ and returns dual vectors in $\mathbb{R}^{N_V}$. Conversely, the transpose D_h^T takes coefficient vectors in $\mathbb{R}^{N_V}$ as input and returns dual vectors in $\mathbb{R}^{N_S}$. If

$\mathsf{B}_{\mathrm{div},h}^{1-s}$ is the matrix realization of $B_{\mathrm{div},h}^{1-s}$ taking dual vectors as input and returning coefficient vectors, B_h^s can be realized by the matrix

$$\mathsf{B}_h^s = \mathsf{D}_h^T \mathsf{B}_{\mathrm{div},h}^{1-s} \mathsf{D}_h.$$

Then, B_h^s takes coefficient vectors as input and returns dual vectors, which is opposite to usual implementations of preconditioners. Thus, if this preconditioner should be used as part of a preconditioner for problems of the form (1.2), some care is needed. In particular, the Lagrange multiplier λ should be represented as a dual vector, while the trace constraint g should be represented by a coefficient vector. We see then that the matrix realization of the trace operator T should take coefficient vectors to coefficient vectors. That is, the matrix is simply a mapping of degrees of freedom from one space to another, and no numerical integration is needed.

By Corollary 3.1, we know that we can construct an efficient preconditioner for A_h^{-s}, provided we have an efficient preconditioner for Λ_h^{1-s} at our disposable. This is by no means a given. However, we will in the next section propose a construction of $B_{\mathrm{div},h}^{1-s}$ on V_h satisfying (3.24) based on an additive multigrid approach.

4. Additive multigrid methods for Λ_h^s

Recall that in section 3 we constructed an efficient preconditioner for A_h^{-s}, where A_h is a discrete Laplacian on S_h and $s \in [0,1]$ provided we are given an efficient preconditioner for Λ_h^{1-s} on V_h, which we denote by $B_{\mathrm{div},h}^{1-s}$. In this section we give one construction of $B_{\mathrm{div},h}^{1-s}$ based on a multigrid approach similar to that presented in [5].

To motivate the construction we note that multigrid methods, and other space decomposition methods, are popular and well-studied preconditioning strategies for $H(\mathrm{div})$ problem. A key observation is that Λ_h reduces to the identity operator on the kernel of div in V_h, while on the L^2-orthogonal complement Λ_h roughly behaves like an elliptic operator with a zero-order term. In particular, Λ_h can be decomposed into operators where subspace decomposition methods have proven to be efficient. We will now see that this line of reasoning continues to hold for Λ_h^s. To that end, consider the discrete Helmholtz decomposition of $v \in V_h$ given in (3.13),

$$\tau = \nabla_h u + \mathbf{curl}\, q, \tag{4.1}$$

where $u \in S_h$ and $q \in \mathbf{curl}_h V_h$. From the definition of ∇_h, we have that $\Lambda_h = I + \nabla_h \nabla_h^*$, which when applied to (4.1) yields

$$\Lambda_h \tau = \nabla_h (I + A_h) u + \mathbf{curl}\, q, \tag{4.2}$$

where we recall that $A_h = \nabla_h^* \nabla_h$ is a discrete Laplacian. We see that Λ_h is invariant in both $\nabla_h S_h$ and its orthogonal complement, $\mathbf{curl}\, C_h$. From (4.2) it is also evident that the projections $\tau \mapsto \nabla_h u$ and $\tau \mapsto \mathbf{curl}\, q$ both commute with Λ_h. In accordance with the discussion made in [16], it follows that Λ_h^s also leave the decomposition in (4.1) invariant. Thus, Λ_h^s reduces to the identity operator on $\mathbf{curl}\, C_h$, and behaves like $(I + A_h)^s$ on $\nabla_h S_h$. Multigrid methods were shown to be computationally effective for such operators in [5], and this motivates using a similar approach for constructing preconditioners for Λ_h^s.

Before proceeding, some issues need to be adressed. As shown in Lemma 2.2, the operators on each level will not be inherited. Therefore, the analysis will follow the framework of [14]. Another problem is that the computation of Λ_h^s requires solving a potentially large eigenvalue problem, which can be prohibitively expensive. As a consequence, we cannot assume that we can compute errors on each level. Standard multigrid algorithms, such as V-cycle, should then be excluded. For this reason, we design the operators as additive multigrid operators, [13], where the residual of the problem is transferred to every grid level, and no application of Λ_h^s is required.

In the following, we will use the same multilevel decomposition as was used in [2], but we emphasize that the analysis extends to other decompositions, such as that given in [20].

To construct our multigrid operator for Λ_h^s suppose $\mathscr{T}_h$ is the result of successive refinements. That is, we are given a sequence

$$\mathscr{T}_1 \subset \cdots \subset \mathscr{T}_J = \mathscr{T}_h,$$

of shape-regular triangulations of Ω, and $\mathscr{T}_k$ has charachteristic mesh size h_k for $k = 1,\dots,J$. We will assume that the refinements are bounded, in the sense that there is a constant $\gamma \geq 1$ so that $h_{k-1} \leq \gamma h_k$ for $k = 2,\dots,J$. We note that in applications γ is around 2. For each k, we set $V_k = \mathscr{RT}_r(\mathscr{T}_k)$ as the Raviart-Thomas space of index r relative to the mesh $\mathscr{T}_k$. We further define $S_k \subset S_h$ and $C_k \subset C_h$ analogously, as well as operators $\nabla_k : S_k \to V_k$ and $\mathbf{curl}_k : V_k \to C_k$ as the L^2-adjoint of div and **curl**, respectively.

For each k, we define $\Lambda_k : V_k \to V_k$ by

$$(\Lambda_k \sigma, \tau) = \Lambda(\sigma, \tau), \quad \sigma, \tau \in V_k.$$

It is evident that Λ_k is symmetric positive-definite, and so Λ_k^θ is well-defined for every $\theta \in \mathbb{R}$, and as a consequence of Lemma 2.2

$$\left(\Lambda_k^s \tau, \tau\right) \leq \left(\Lambda_{k-1}^s \tau, \tau\right) \tag{4.3}$$

for $s \in [0,1]$ and $\tau \in V_{k-1}$. For every k we define $Q_k : V \to V_k$ as the L^2-orthogonal projection and $P_{k,k-1}^s : V_k \to V_{k-1}$ by

$$\left(\Lambda_{k-1}^s P_{k,k-1}^s \sigma_k, \tau_{k-1}\right) = \left(\Lambda_k^s \sigma_k, \tau_{k-1}\right), \quad \sigma_k \in V_k,\ \tau_{k-1} \in V_{k-1},$$

with the interpretation that $P_{1,0}^s = 0$. We go on to define $P_k^s := P_{k+1,k}^s \cdots P_{J,J-1}^s : V_h \to V_k$, which satisfies

$$\left(\Lambda_k^s P_k^s \sigma, \tau_k\right) = \left(\Lambda_h^s \sigma, \tau_k\right),$$

for every $\sigma \in V_h$ and $\tau_k \in V_k$.

It follows by the definitions of P_k^s and Q_k that

$$\Lambda_k^s P_k^s = Q_k \Lambda_h^s. \tag{4.4}$$

Note that in general P_k^s is not a projection, except when $s = 0$ (in which case it coincides with Q_k) and $s = 1$. However, when $s \in (0,1)$ we have for any $\tau \in V_k$ that

$$\begin{aligned}\left(\Lambda_k^s P_k^s \tau, P_k^s \tau\right) &= \left(\Lambda_h^s \tau, P_k^s \tau\right) \\ &\leq \left(\Lambda_h^s \tau, \tau\right)^{\frac{1}{2}} \left(\Lambda_h^s P_k^s \tau, P_k^s \tau\right)^{\frac{1}{2}}.\end{aligned}$$

Applying (4.3) in the above, we deduce that

$$\left(\Lambda_k^s P_k^s \tau, P_k^s \tau\right) \leq \left(\Lambda_h^s \tau, \tau\right). \tag{4.5}$$

Suppose now that on each level k we are given symmetric positive definite operators $R_k^s : V_k \to V_k$. As is usual, we call these operators smoothers, and they should, in a sense to made clearer below, approximate Λ_k^{-s}. We then define our additive multigrid preconditioner $B_{\mathrm{div},h}^s : V_h \to V_h$ as

$$B_{\mathrm{div},h}^s = \sum_{k=1}^{J} R_k^s Q_k. \tag{4.6}$$

The following theorem gives sufficient conditions on the smoothers to establish spectral equivalence between $B_{\mathrm{div},h}^s$ and Λ_h^{-s}. The proof will mostly follow by standard techniques, but some care is needed since the operators are not inherited between grid levels.

Theorem 4.1. *Let $s \in [0,1]$ and suppose that for each $k = 1,\dots,J$, the operator R_k^s as defined above satisfies for every $\tau \in V_k$*

$$\left(R_k^s \tau, \tau\right) \leq C_1 \left(\Lambda_k^{-s} \tau, \tau\right), \tag{4.7}$$

and

$$\left(\left(R_k^s\right)^{-1} (I - P_{k,k-1}^s)\tau, (I - P_{k,k-1}^s)\tau\right) \leq C_2 \left(\Lambda_k^s (I - P_{k,k-1}^s)\tau, (I - P_{k,k-1}^s)\tau\right) \tag{4.8}$$

for some constants C_1 and C_2 that are independent of k. Then,

$$C_2^{-1} \left(\Lambda_h^s \tau, \tau\right) \leq \left(B_{\mathrm{div},h}^s \Lambda_h^s \tau, \Lambda_h^s \tau\right) \leq C_1 J \left(\Lambda_h^s \tau, \tau\right) \tag{4.9}$$

Proof. For the upper bound of (4.9), straightforward application of the definitions of $B_{\mathrm{div},h}^s$ and P_k^s show that

$$\left(B_{\mathrm{div},h}^s \Lambda_h^s \tau, \Lambda_h^s \tau\right) = \sum_{k=1}^{J} \left(R_k^s \Lambda_k^s P_k^s \tau, \Lambda_k^s P_k^s \tau\right). \tag{4.10}$$

Assumption (4.7) and the non-inheritance inequality (4.5) then imply that

$$\left(B_{\mathrm{div},h}^s \Lambda_h^s \tau, \Lambda_h^s \tau\right) \leq C_1 \sum_{k=1}^{J} \left(\Lambda_k^s P_k^s \tau, P_k^s \tau\right) \leq C_1 J \left(\Lambda_h^s \tau, \tau\right).$$

In proving the lower bound of (4.9), we consider the decomposition $\tau = \sum_{k=1} \tau_k$, with $\tau_k = (P_k^s - P_{k-1}^s)\tau = (I - P_{k,k-1}^s) P_k^s \tau \in V_k$, for $k = 1,\dots,J$. Here, we interpret $P_0 = 0$ and $P_J = I$. Then,

$$\left(\Lambda_h^s \tau, \tau\right) = \sum_{k=1}^{J} \left(\Lambda_k^s P_k^s \tau, \tau_k\right) = \sum_{k=1}^{J} \left(R_k^s \Lambda_k^s P_k^s \tau, \left(R_k^s\right)^{-1} \tau_k\right).$$

Since for every k, R_k^s is symmetric positive definite, we can use Cauchy-Schwarz' and assumption (4.8), resulting in

$$\begin{aligned}
\left(\Lambda_h^s \tau, \tau\right) &\leq \sum_{k=1}^{J} \left(R_k^s \Lambda_k^s P_k^s \tau, \Lambda_k^s P_k^s \tau\right)^{\frac{1}{2}} \left(\left(R_k^s\right)^{-1} \tau_k, \tau_k\right)^{\frac{1}{2}} \\
&\leq \left(\sum_{k=1}^{J} \left(R_k^s \Lambda_k^s P_k^s \tau, \Lambda_k^s P_k^s \tau\right)\right)^{\frac{1}{2}} \left(C_2 \sum_{k=1}^{J} \left(\Lambda_k^s \tau_k, \tau_k\right)\right)^{\frac{1}{2}} \\
&= \left(B_{\mathrm{div},h}^s \Lambda_h^s \tau, \Lambda_h^s \tau\right)^{\frac{1}{2}} \left(C_2 \sum_{k=1}^{J} \left(\Lambda_k^s \tau_k, \tau_k\right)\right)^{\frac{1}{2}},
\end{aligned} \tag{4.11}$$

where in the last step we have used (4.10). In view of (4.11), it only remains to show that

$$\sum_{k=1}^{J} \left(\Lambda_k^s \tau_k, \tau_k\right) \leq \left(\Lambda_h^s \tau, \tau\right) \tag{4.12}$$

to prove the lower bound of (4.9). Inserting the definition of τ_k and expanding factors, we find that

$$\left(\Lambda_k^s \tau_k, \tau_k\right) = \left(\Lambda_k^s P_k^s \tau, P_k^s \tau\right) - 2\left(\Lambda_k^s P_k^s \tau, P_{k-1}^s \tau\right) + \left(\Lambda_k^s P_{k-1}^s \tau, P_{k-1}^s \tau\right).$$

For the second term on the right hand side in the above, we have $\left(\Lambda_k^s P_k^s \tau, P_{k-1}^s \tau\right) = \left(\Lambda_{k-1}^s P_{k-1}^s \tau, P_{k-1}^s \tau\right)$ since $P_{k-1}^s = P_{k,k-1}^s P_k^s$, while for the third term we apply (4.3). Thus,

$$\left(\Lambda_k^s \tau_k, \tau_k\right) \leq \left(\Lambda_k^s P_k^s \tau, P_k^s \tau\right) - \left(\Lambda_{k-1}^s P_{k-1}^s \tau, P_{k-1}^s \tau\right).$$

It follows that

$$\sum_{k=1}\left(\Lambda_k^s\tau_k,\tau_k\right)\leq\sum_{k=1}^{J}\left[\left(\Lambda_k^sP_k^s\tau,P_k^s\tau\right)-\left(\Lambda_{k-1}^sP_{k-1}^s\tau,P_{k-1}^s\tau\right)\right]\leq\left(\Lambda_J^s\tau,\tau\right).$$

□

Now it remains to choose smoothers satisfying the assumptions in Theorem 4.1, and in this work we consider additive Schwarz operators based on the same space decomposition as in [2].

For $k\geq 2$, let $\mathcal{N}_k$ denote the set of vertices in $\mathcal{T}_k$, and for each $\nu\in\mathcal{N}_k$, let $\mathcal{T}_{k,\nu}$ be the set of simplices meeting at the vertex ν. Then $\mathcal{T}_{k,\nu}$ forms a triangulation of a small subdomain $\Omega_{k,\nu}$, and we define $V_{k,\nu}$ to be the subspace of functions in V_k with support contained in $\bar{\Omega}_{k,\nu}$. The operators $\Lambda_{k,\nu}:V_{k,\nu}\to V_{k,\nu}$ and $P_{k,\nu}^s$, $Q_{k,\nu}:V_k\to V_{k,\nu}$ are then defined analogously to the corresponding operators above. We then define

$$R_k^s=\sum_{\nu\in\mathcal{N}_k}\Lambda_{k,\nu}^{-s}Q_{k,\nu}, \tag{4.13}$$

while on the coarsest level we set $R_1^s=\Lambda_1^{-s}$. It is well-known that additive Schwarz operator of the form (4.13) are symmetric positive definite, and its inverse satisfies for $\tau\in V_k$

$$\left(\left(R_k^s\right)^{-1}\tau,\tau\right)=\inf_{\substack{\tau=\sum_\nu\tau_\nu\\ \tau_\nu\in V_{k,\nu}}}\sum_{\nu\in\mathcal{N}_k}\left(\Lambda_{k,\nu}^s\tau_\nu,\tau_\nu\right). \tag{4.14}$$

Moreover, the decomposition $V_k=\sum_{\nu\in\mathcal{N}_k}V_{k,\nu}$ is L^2-stable in the sense that for every $\tau\in V_k$ there are $\tau_\nu\in V_{k,\nu}$ so that $\tau=\sum_\nu\tau_\nu$ and

$$\sum_{\nu\in\mathcal{N}_k}\|\tau_\nu\|^2\leq c\,\|\tau\|^2, \tag{4.15}$$

for some constant c, independent of k and v. The analogue to (4.15) continues to hold if we define the decomposition $C_k=\sum_\nu C_{k,\nu}$ similarly (cf. [3]).

The verification of Assumption (4.7) in Theorem 4.1 is given in the following lemma.

Lemma 4.1. *For $k=1,\dots,J$ and $s\in[0,1]$, let R_k^s defined as above. Then there are constants $K_0,K_1\geq 0$, independent of k so that*

$$\begin{aligned}\left(R_k^0\tau,\tau\right)&\leq K_0(\tau,\tau)\\ \left(R_k^1\tau,\tau\right)&\leq K_1\left(\Lambda_k^{-1}\tau,\tau\right)\end{aligned} \tag{4.16}$$

for every $\tau\in V_k$. Moreover, for $s\in(0,1)$ and every $\tau\in V_k$

$$\left(R_k^s\tau,\tau\right)\leq K_0^{1-s}K_1^s\left(\Lambda_k^{-s}\tau,\tau\right), \tag{4.17}$$

where K_0 and K_1 are the same as in (4.16).

Proof. The assertions are evident when $k=1$, with $K_0=K_1=1$, so let $k\geq 2$. A proof of the second inequality of (4.16) can be found in e.g. [2, Theorem 4.1], so we limit ourselves only to sketch a proof here. Setting $P=R_k^1\Lambda_k=\sum_{\nu\in\mathcal{N}_k}P_{k,\nu}^1$, the uniform finite overlaps of the domains $\Omega_{k,\nu}$ ensure that

$$\Lambda(P\tau,P\tau)\leq K_1\Lambda(Pv,v),$$

for some K_1, independent of k. It then follows that

$$\left(R_k^s\Lambda_k\tau,\Lambda_k\tau\right)=\Lambda(P\tau,\tau)\leq\Lambda(P\tau,P\tau)^{\frac{1}{2}}\Lambda(\tau,\tau)^{\frac{1}{2}}\leq\left[K_1\Lambda(P\tau,\tau)\right]^{\frac{1}{2}}\Lambda(\tau,\tau)^{\frac{1}{2}}.$$

Replacing τ with $\Lambda_k^{-1}\tau$ in the above yields the second inequality of (4.16), and the first inequality of (4.16) can be proved similarly.

For the intermediate result when $s \in (0,1)$ we introduce the auxiliary Hilbert space $\mathbf{V} := \bigoplus_{\nu \in \mathcal{N}_k} V_{k,\nu}$, and define operators $\mathbf{Q} : V_k \to \mathbf{V}$ and $\tilde{} : \mathbf{V} \to \mathbf{V}$ given by

$$(\mathbf{Q}\tau)_\nu = Q_{k,\nu}\tau,$$

and

$$(\tilde{}\,\tau)_\nu = \Lambda_{k,\nu}\tau_\nu,$$

for $\tau \in V_k$, $\nu \in \mathcal{N}_k$, and $\tau \in \mathbf{V}$. In particular, we note that $\tilde{}$ is symmetric positive definite and diagonal on $\mathbf{V}$. Therefore, $(\tilde{}^{\theta}\tau)_\nu = \Lambda_{k,\nu}^{\theta}\tau_\nu$ for every $\theta \in \mathbb{R}$, and so we have $R_k^s = \mathbf{Q}^*\tilde{}^{-s}\mathbf{Q}$ for $s \in [0,1]$.

From the definition of R_k^0 and the first inequality of (4.16), $\mathbf{Q}^*\mathbf{Q} \le K_0$, and by scaling $\tilde{\mathbf{Q}} := K_0^{\frac{1}{2}}\mathbf{Q}$ we then have that

$$\tilde{\mathbf{Q}}^*\tilde{\mathbf{Q}} \le 1.$$

Then, by Lemma 2.1,

$$R_k^s = K_0\tilde{\mathbf{Q}}^*\tilde{}^{-s}\tilde{\mathbf{Q}} \le K_0\left(\tilde{\mathbf{Q}}^*\tilde{}^{-1}\tilde{\mathbf{Q}}\right)^s = K_0\left(K_0^{-1}R_k^1\right)^s = K_0^{1-s}\left(R_k^1\right)^s. \tag{4.18}$$

Now, the second inequality of (4.16) states that $R_k^1 \le K_1\Lambda_k^{-1}$. Inserting this into (4.18) yields

$$R_k^s \le K_0^{1-s}K_1^s\Lambda_k^{-s},$$

which is equivalent to (4.17). □

Establishing that Assumption (4.8) in Theorem 4.1 holds turns out to be a more complicated matter. In view of (4.14), we see that to prove (4.8) it is sufficient to find for every $\tau \in (I - P_{k,k-1}^s)(V_k)$ a decomposition $\tau = \sum_\nu \tau_\nu$, where $\tau_\nu \in V_{k,\nu}$ so that

$$\sum_{\nu \in \mathcal{N}_k}\left(\Lambda_{k,\nu}^s\tau_\nu, \tau_\nu\right) \le C\left(\Lambda_k^s\tau, \tau\right),$$

for some constant C that is independent of k and τ. In the following lemma, we verify this stable decomposition, assuming some error bounds on the discrete Helmholtz decomposition.

Lemma 4.2. *For $k = 2,\dots,J$, let $\tau \in (I - P_{k,k-1}^s)V_k$ have the discrete Helmholtz decomposition*

$$\tau = \nabla_k u + \mathbf{curl}\, q, \tag{4.19}$$

for some $u \in S_k$ and $q \in \mathbf{curl}_k V_k$. Assume there exists a constant c, independent of k and τ so that

$$\|\nabla_k u\|^2 \le c\,h_{k-1}^{2s}\left(\Lambda_k^s\tau, \tau\right), \quad \text{and} \quad \|q\| \le c\,h_{k-1}\|\tau\|. \tag{4.20}$$

Then there exists a decomposition $\tau = \sum_{\nu \in \mathcal{N}_k}\tau_\nu$ with $\tau_\nu \in V_{k,\nu}$, and a constant C so that

$$\sum_{\nu \in \mathcal{N}_k}\left(\Lambda_{k,\nu}^s\tau_\nu, \tau_\nu\right) \le C\left(\Lambda_k^s\tau, \tau\right). \tag{4.21}$$

Proof. Fix $\tau \in (I - P_{k,k-1}^s)V_k$, and let $u \in S_h$ and $q \in \mathbf{curl}_k V_k$ be the discrete Helmholtz decomposition according to (4.19). Further, for $\nu \in \mathcal{N}_k$, let $\tilde{\tau}_\nu \in V_{k,\nu}$ and $q_\nu \in C_{k,\nu}$ be L^2-stable decompositions of $\nabla_k u$ and q, respectively. That is, $q = \sum_\nu q_\nu$ and $\nabla_k u = \sum_\nu \tilde{\tau}_\nu$ satisfies

$$\sum_\nu \|\tilde{\tau}_\nu\|^2 \le c\,\|\nabla_k u\|^2, \quad \text{and} \quad \sum_\nu \|q_\nu\|^2 \le c\,\|q\|^2, \tag{4.22}$$

according to (4.15). Then $\tau = \sum_\nu \tau_\nu$, where we set $\tau_\nu = \tilde{\tau}_\nu + \mathbf{curl}\, q_\nu \in V_{k,\nu}$.

By standard inverse inequality, $\left(\Lambda_{k,\nu}\tilde{\tau}_\nu, \tilde{\tau}_\nu\right) \le c(1+h_k^{-2})\|\tilde{\tau}_\nu\|^2$, and the inequality $\left(\Lambda_{k,\nu}^s\tilde{\tau}_\nu, \tilde{\tau}_\nu\right) \le \left(\tilde{\tau}_\nu, \tilde{\tau}_\nu\right)^{1-s}\left(\Lambda_{k,\nu}\tilde{\tau}_\nu, \tilde{\tau}_\nu\right)^s$ (cf. e.g. [27, Ch. 2.5])

$$\left(\Lambda_{k,\nu}^s\tilde{\tau}_\nu, \tilde{\tau}_\nu\right) \le c(1+h_k^{-2})^s\,\|\tilde{\tau}_\nu\|^2. \tag{4.23}$$

Using the fractional inverse inequality (4.23) and a standard inverse inequality for q_ν, together with (4.22),

$$\begin{aligned}\sum_\nu \big(\Lambda_{k,\nu}^s \tau_\nu, \tau_\nu\big) &\le 2\sum_\nu \Big[\big(\Lambda_{k,\nu}^s \tilde{\tau}_\nu, \tilde{\tau}_\nu\big) + \|\mathbf{curl}\, q_\nu\|^2\Big] \\ &\le c\sum_\nu \Big[(1+h_k^{-2})^s \|\tilde{\tau}_\nu\|^2 + h_k^{-2}\|q_\nu\|^2\Big] \\ &\le c\Big[(1+h_k^{-2})^s \|\nabla_k u\|^2 + h_k^{-2}\|q\|^2\Big].\end{aligned}$$

Then, (4.21) follows from the above and assumption (4.20). □

Remark 3. Verifying the assumption of Lemma 4.2 is by no means a trivial matter, and falls beyond the scope of this paper. As such, we leave the additive multigrid operators $B^s_{\mathrm{div},h}$ on what may be deemed an unsure theoretical footing. However, we will here propose an approach to prove the assumptions made in Lemma 4.2. First off, the case $s = 1$ was proved in [2, 3], where the thrust of the argument relied on two-level error estimates and duality arguments.

From the identity (4.4) we see that

$$P_k^s - P_{k-1}^s = (\Lambda_k^{-s} - \Lambda_{k-1}^{-s} Q_{k-1}) Q_k \Lambda_h^s,$$

and so the assumptions in Lemma 4.2 are concerned with two-level error estimates for discretizations of fractional $H(\mathrm{div})$ problems. For the first estimate of (4.19), we recall the observation that Λ_k^s behaves like an elliptic operator on $\nabla_k S_k$, and so the required error estimate can be obtained using similar techniques as in [9, Thm. 4.3]. There, the authors proved error estimates, under some regularity assumptions on the domain Ω. The proof uses the integral formulation of the fractional Laplacian,

$$(-\Delta)^s = \frac{2\sin(\pi s)}{\pi}\int_0^\infty t^{2s-1}(I - t^2\Delta)^{-1}\mathrm{d}t.$$

See also [7, Sec. 10.4]. The advantage of this approach is that error estimates for the fractional Laplacian are transferred to error estimates for problems of the form

$$(I - t^2\Delta)u = f,$$

where an abundance of results are available.

5. Numerical experiments

We now present a series of numerical experiments, aimed at validating the theoretical results established in previous sections. Specifically, in section 5.1 we test the preconditioner $B^s_{\mathrm{div},h}$ defined in (4.6), and the spectral equivalence established in Theorem 4.1.

We then consider

$$A_h^s u = f, \tag{5.1}$$

for a given in $s \in [-1, 0]$ and $f \in S_h$. In section 5.2, (5.1) is first solved using $\nabla_h^* \Lambda_h^{-(1+s)} \nabla_h$ as preconditioner, before we use B_h^s defined in (3.17) as preconditioner. These experiments are to validate Theorem 3.2 and Corollary 3.1, respectively.

Where applicable, the numerical tests are conducted using preconditioned conjugate method, with random initial guess. Convergence of the iterations are reached when the relative preconditioned residual, i.e. $\frac{(Br_k, r_k)}{(Br_0, r_0)}$, where r_k is the k-th residual and B is the preconditioner, is below a given tolerance.

Note that in the following, all fractional powers of matrices are constructed by full spectral decomposition, requiring the solution of large generalized eigenvalue problems (see [25] for details). As such, the preconditioned iterative methods will not be computationally optimal, but the tests are designed only to validate the theoretical bounds on the condition numbers.

s \ N	208	800	3136	12416
0.0	20(4.9)	21(4.9)	21(4.9)	21(4.9)
0.1	20(4.6)	21(4.9)	22(5.2)	23(5.5)
0.2	22(5.6)	24(6.2)	25(6.8)	27(7.4)
0.3	24(6.6)	26(7.5)	27(8.1)	28(8.6)
0.4	26(8.0)	28(8.7)	29(9.2)	29(9.6)
0.5	27(9.2)	30(9.8)	30(10.3)	30(10.5)
0.6	29(10.4)	31(10.9)	31(11.3)	31(11.5)
0.7	30(11.6)	32(12.1)	32(12.4)	32(12.5)
0.8	31(13.0)	33(13.4)	33(13.5)	33(13.7)
0.9	32(14.5)	35(14.9)	34(14.9)	34(15.0)
1.0	33(16.1)	36(16.5)	36(16.6)	35(16.5)

TABLE 1. Numerical results preconditioning Λ_h^s. Table show number of conjugate gradient interations until reaching relative error tolerance 10^{-9}. Estimated condition numbers are shown inside parentheses. $N = \dim V_h$ and $J = 4$ in all tests.

This problem will not be encountered if B_h^s is used as part of a preconditioner for trace problems as presented in the introduction.

5.1. **Preconditioning for Λ_h^s.** In the first set of numerical experiments we consider the following problem: For a given $f \in V_h$ and $s \in [0,1]$, find $\sigma \in V_h$ so that

$$\Lambda_h^s \sigma = f. \tag{5.2}$$

We take $\Omega = [0,1]^2 \subset \mathbb{R}^2$, and $\mathcal{T}_h$ is a uniform partition of Ω. We take V_h to be the lowest order Raviart-Thomas space relative to the mesh $\mathcal{T}_h$. We solve the linear system arising from (5.2) using preconditioned conjugate gradient method, with $B_{\mathrm{div},h}^s$ given by (4.6) as preconditioner. The results can be seen in Table 1, from which we see that both iteration counts and condition numbers stay bounded independently of the dimension of V_h, in accordance with Theorem 4.1.

5.2. **Auxiliary space preconditioner.** We now consider (5.1) on the same computational domain as in the previous set of experiments. That is, $\Omega = [0,1]^2$, and $\mathcal{T}_h$ is a uniform triangulation of Ω. For the discrete space S_h we use piecewise constants relative to $\mathcal{T}_h$. In Table 2, we can view the calculated condition number of $\nabla_h^* \Lambda_h^{-(1+s)} \nabla_h A_h^s$, as well as the condition number expected from Theorem 3.2. The results show both uniform h-independence and is in good agreement with the theory.

Finally, we solve (5.1) using preconditioned conjugate gradient method, with $B_h^s = \nabla_h^* B_{\mathrm{div},h}^{1+s} \nabla_h$ defined in (3.17) as preconditioner. $B_{\mathrm{div},h}^{1+s}$ is chosen as the additive multigrid operator proposed in section 4, with V_h as the lowest order Raviart-Thomas space relative to $\mathcal{T}_h$. The results can be viewed in Table 3. Again, we see that both iteration counts and estimated condition numbers stay reasonably bounded, in agreement with Corollary 3.1, although a slight increase becomes pronounced as s approaches 0.

s \ N	512	2048	8192	$\beta^{-2(1+s)}$
−1.0	1.000	1.000	1.000	1.000
−0.9	1.005	1.005	1.005	1.005
−0.8	1.010	1.010	1.010	1.010
−0.7	1.015	1.015	1.015	1.015
−0.6	1.020	1.020	1.020	1.020
−0.5	1.025	1.025	1.025	1.025
−0.4	1.030	1.030	1.030	1.030
−0.3	1.035	1.035	1.035	1.035
−0.2	1.040	1.040	1.040	1.041
−0.1	1.045	1.045	1.045	1.046
0.0	1.050	1.051	1.051	1.051

TABLE 2. Numerical results for exact auxiliary space preconditioner. Table show condition number of $\nabla_h^* \Lambda_h^{-(1+s)} \nabla_h A_h^s$. N is dimension of S_h. The rightmost column shows expected condition number from Theorem 3.2 with $\beta^{-2} = 1.051$.

s \ N	128	512	2048	8192
−1.0	18(4.3)	19(4.4)	20(4.6)	21(4.6)
−0.9	17(3.7)	19(3.7)	19(3.7)	19(3.7)
−0.8	17(3.2)	18(3.2)	18(3.2)	18(3.2)
−0.7	17(2.9)	18(2.9)	18(2.9)	18(3.0)
−0.6	17(2.8)	18(3.0)	18(3.1)	19(3.1)
−0.5	18(3.2)	19(3.3)	20(3.4)	20(3.6)
−0.4	19(3.6)	21(3.8)	21(3.8)	22(4.4)
−0.3	19(4.0)	22(4.2)	22(4.2)	24(5.3)
−0.2	20(4.5)	23(4.8)	24(5.1)	26(6.2)
−0.1	21(5.1)	25(5.4)	26(6.1)	28(7.2)
0.0	22(5.8)	27(6.2)	28(7.4)	30(8.3)

TABLE 3. Numerical results for preconditioning A_h^s with B_h^s given by (3.17). Table show number of conjugate gradient interations until reaching error tolerance 10^{-10}. Estimated condition numbers are shown inside parentheses. N is dimension of S_h. $J = 4$ in all tests.

References

[1] M. Arioli and D. Loghin. Discrete interpolation norms with applications. *SIAM J. Numer. Anal.*, 47(4):2924–2951, 2009.

[2] D. N. Arnold, R. S. Falk, and R. Winther. Preconditioning in $H(\mathrm{div})$ and applications. *Math. Comp.*, 66(219):957–984, 1997.

[3] D. N. Arnold, R. S. Falk, and R. Winther. Multigrid in $H(\mathrm{div})$ and $H(\mathrm{curl})$. *Numer. Math.*, 85(2):197–217, 2000.

[4] I. Babuška. The finite element method with Lagrangian multipliers. *Numer. Math.*, 20:179–192, 1972/73.

[5] T. Bærland, M. Kuchta, and K.-A. Mardal. Multigrid methods for discrete fractional sobolev spaces. *arXiv preprint arXiv:1806.00222*, 2018.

[6] S. Bertoluzza, V. Chabannes, C. Prud'homme, and M. Szopos. Boundary conditions involving pressure for the Stokes problem and applications in computational hemodynamics. *Comput. Methods Appl. Mech. Engrg.*, 322:58–80, 2017.

[7] M. S. Birman and M. Z. Solomjak. *Spectral theory of selfadjoint operators in Hilbert space*. Mathematics and its Applications (Soviet Series). D. Reidel Publishing Co., Dordrecht, 1987. Translated from the 1980 Russian original by S. Khrushchëv and V. Peller.

[8] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013.

[9] A. Bonito and J. E. Pasciak. Numerical approximation of fractional powers of elliptic operators. *Math. Comp.*, 84(295):2083–2110, 2015.

[10] J. H. Bramble. *Multigrid methods*, volume 294 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1993.

[11] J. H. Bramble, Z. Leyk, and J. E. Pasciak. The analysis of multigrid algorithms for pseudodifferential operators of order minus one. *Math. Comp.*, 63(208):461–478, 1994.

[12] J. H. Bramble, J. E. Pasciak, and P. S. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69(230):463–480, 2000.

[13] J. H. Bramble, J. E. Pasciak, and J. Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55(191):1–22, 1990.

[14] J. H. Bramble, J. E. Pasciak, and J. Xu. The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms. *Math. Comp.*, 56(193):1–34, 1991.

[15] M. Costabel and A. McIntosh. On Bogovskiĭ and regularized Poincaré integral operators for de Rham complexes on Lipschitz domains. *Math. Z.*, 265(2):297–320, 2010.

[16] C. Davis. A Schwarz inequality for convex operator functions. *Proc. Amer. Math. Soc.*, 8:42–44, 1957.

[17] S. A. Funken and E. P. Stephan. The bpx preconditioner for the single layer potential operator. *Applicable Analysis*, 67(3-4):327–340, 1997.

[18] F. Hansen and G. Kjærgå rd Pedersen. Jensen's inequality for operators and Löwner's theorem. *Math. Ann.*, 258(3):229–241, 1982.

[19] F. Hansen and G. K. Pedersen. Jensen's operator inequality. *Bull. London Math. Soc.*, 35(4):553–564, 2003.

[20] R. Hiptmair. Multigrid method for $\mathbf{H}(\mathrm{div})$ in three dimensions. *Electron. Trans. Numer. Anal.*, 6(Dec.):133–152, 1997. Special issue on multilevel methods (Copper Mountain, CO, 1997).

[21] R. Hiptmair and J. Xu. Nodal auxiliary space preconditioning in $\mathbf{H}(\mathbf{curl})$ and $\mathbf{H}(\mathrm{div})$ spaces. *SIAM J. Numer. Anal.*, 45(6):2483–2509, 2007.

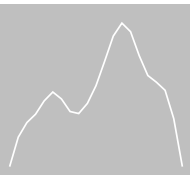
[22] T. Kato. Notes on some inequalities for linear operators. *Math. Ann.*, 125:208–212, 1952.

[23] T. V. Kolev and P. S. Vassilevski. Parallel auxiliary space AMG solver for $H(\mathrm{div})$ problems. *SIAM J. Sci. Comput.*, 34(6):A3079–A3098, 2012.

[24] M. Kuchta, K.-A. Mardal, and M. Mortensen. Preconditioning trace coupled 3d-1d systems using fractional Laplacian. *arXiv preprint arXiv:1612.03574*, 2016.

[25] M. Kuchta, M. Nordaas, J. C. G. Verschaeve, M. Mortensen, and K.-A. Mardal. Preconditioners for saddle point systems with trace constraints coupling 2d and 1d domains. *SIAM J. Sci. Comput.*, 38(6):B962–B987, 2016.
[26] I. V. Lashuk and P. S. Vassilevski. The construction of the coarse de Rham complexes with improved approximation properties. *Comput. Methods Appl. Math.*, 14(2):257–303, 2014.
[27] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I.* Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
[28] K.-A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, 18(1):1–40, 2011.
[29] P. Oswald. Multilevel norms for H-1/2. *Computing*, 61(3):235–255, 1998.
[30] R. Stevenson and R. van Venetië. Optimal preconditioning for problems of negative order. *arXiv preprint arXiv:1803.05226*, 2018.
[31] A. Tveito, K. H. Jæger, M. Kuchta, K.-A. Mardal, and M. E. Rognes. A cell-based framework for numerical modeling of electrical conduction in cardiac tissue. *Front. Phys.*, 5:48, 2017.
[32] J. Xu. The auxiliary space method and optimal multigrid preconditioning techniques for unstructured grids. *Computing*, 56(3):215–235, 1996. International GAMM-Workshop on Multi-level Methods (Meisdorf, 1994).

Paper IV



Uniform Preconditioners for the Mixed Darcy Problem

T. Bærland, K.-A. Mardal, and T. Thompson

Numerische Mathematik, (2018). Submitted.

UNIFORM PRECONDITIONERS FOR THE MIXED DARCY PROBLEM

TRYGVE BÆRLAND[†], KENT-ANDRE MARDAL[†], AND TRAVIS THOMPSON[‡]

ABSTRACT. Preconditioners for porous media flow problems in mixed form are frequently based on **H**(div) preconditioners rather than the pressure Schur complement. We show that when the permeability, K, is small the pressure Schur complement must also be addressed for **H**(div)-based preconditioners. The proposed approach is based on the operator preconditioning framework; thus, we construct the preconditioner by considering the mapping properties of the continuous operator, where the main challenge is a K-uniform inf-sup condition.

1. INTRODUCTION

In this paper we will consider the mixed formulation of the Darcy problem of the form

$$\frac{1}{K}\mathbf{u} - \nabla p = \mathbf{f}, \quad \text{in } \Omega, \tag{1}$$

$$\nabla \cdot \mathbf{u} = g, \quad \text{in } \Omega. \tag{2}$$

equipped with suitable boundary conditions. The variables $\mathbf{u}$ and p represent the fluid flux and pressure, respectively, and K denotes the permeability. Our purpose is to address efficient preconditioners exhibiting robustness for $K \in (0,1)$. We employ the operator preconditioning framework [15]; the general framework approach is predicated on establishing a well-posedness result for the continuous problem in K—weighted Sobolev spaces. For linear systems of the form

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix},$$

there are, generally speaking, two common approaches for constructing block-diagonal preconditioners; one may utilize a Schur complement for the first unknown or, alternatively, the second unknown. That is, the structural options for the preconditioner are:

$$\begin{bmatrix} A^{-1} & 0 \\ 0 & (BA^{-1}B^T)^{-1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} (A+B^TB)^{-1} & 0 \\ 0 & X? \end{bmatrix}.$$

The first approach results in three distinct unit sized eigenvalues [16] for any $0 < K < \infty$. On the other hand, to the authors knowledge, only partial explanations have been offered for the second approach for the mixed Darcy problem. In [1], **H**(div) preconditioners were constructed and applied to (1)-(2) in the case $K = 1$ for which X coincides with the inverse of a potentially diagonalized mass matrix. The same authors have also developed multilevel methods for weighted **H**(div) spaces in [2]; however, to our knowledge, the pressure Schur complement for the case $0 < K \ll 1$ has not be rigorously treated. The case with $X = -\gamma I$, and the first block equal to $(A + \frac{1}{\gamma}B^TB)^{-1}$, was investigated in [3, 11] and used successfully for the Oseen and Maxwell type problems, respectively. We here offer an alternative lower-right block where $X = I^{-1} + (B^TA^{-1}B)^{-1}$ and show that, for a mixed Darcy problem, this approach is robust with respect to K.

As a single-physics problem (1)-(2) the case $K \to 0$ is not particularly interesting; a simple scaling resolves the problem of a vanishing K. However, in a multi-physics setting, solution

[†]DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OSLO, BLINDERN, OSLO, 0316 NORWAY
[‡]DEPT. OF NUM. ANAL. AND SCI. COMPUT. SIMULA RESEARCH LABORATORY. FORNEBU, NORWAY
E-mail addresses: trygveba@math.uio.no, kent-and@math.uio.no, tthompson@simula.no.

algorithms are often constructed via decomposition into single-physics subproblems. Requiring a certain scaling of one of the problems may have a negative effect on the scaling of the other subproblems.

In the current work we construct block preconditioners based on the operator preconditioning framework introduced, briefly, in [1] and reviewed in [15]. Our preconditioners will, in the continuous case, take the following form:

$$\mathscr{B} = \begin{bmatrix} (K^{-1}I - \nabla\nabla\cdot)^{-1} & 0 \\ 0 & (I)^{-1} + (-K\Delta)^{-1} \end{bmatrix}.$$

It will be shown that preconditioners based on B are robust with respect to both the mesh size and the permeability K. We mention that this preconditioner and the analysis is closely related to [14] where the time-dependent Stokes problem was considered.

2. Preliminaries

Let Ω be a bounded Lipschitz domain in $\mathbb{R}^n$, $n = 2, 3$. Then $L^2 = L^2(\Omega)$ denotes the Sobolev space of square integrable functions, whereas $H^k = H^k(\Omega)$ denotes the space of functions with all derivatives up to order k in L^2. The space L^2_0 contains functions $g \in L^2$ with zero mean value. Vector valued functions, and Sobolev spaces of vector valued functions, are denoted by boldface. The space $\mathbf{H}(\mathrm{div})$ contains functions in $\mathbf{L}^2$ with divergence in L^2 and the subspace of functions $\mathbf{u} \in \mathbf{H}_0(\mathrm{div})$ are those with zero normal trace. The notation $(\cdot,\cdot)$ is used for the L^2 inner product of both scalar and vector fields as well as the duality paring between a space X and its dual space X'. The norm corresponding to the L^2 inner product is expressed with the canonical double-bar $||\cdot||_{L^2}$. For $\alpha > 0$, a fixed real value, the notation $||f||^2_{\alpha L^2} = \alpha^2(f,f)$ and $\alpha L^2 = \alpha L^2(\Omega)$ signifies the usual L^2 space equipped with the corresponding weighted inner product denoted by $(f,g)_\alpha = \alpha^2(f,g)$.

For two Hilbert spaces, X and Y, we denote by $\mathscr{L}(X,Y)$ the space of bounded linear maps from X to Y, with the standard operator norm

$$||T||_{\mathscr{L}(X,Y)} = \sup_{x \in X} \frac{||Tx||_Y}{||x||_X}, \quad T \in \mathscr{L}(X,Y).$$

In the subsequent analysis we employ both the intersection and sum of two Hilbert spaces X and Y. These composite spaces are formally defined as follows: Let X and Y be two Hilbert spaces both contained in some larger Hilbert space; the intersection space $X \cap Y$ and sum space $X + Y$ are also Hilbert spaces and their respective norms are

$$||z||^2_{X\cap Y} = ||z||^2_X + ||z||^2_Y$$

and

$$||z||^2_{X+Y} = \inf_{\substack{z=x+y \\ x\in X, y\in Y}} ||x||^2_X + ||y||^2_Y.$$

In addition, if $X \cap Y$ is dense in both X and Y, then

$$(X+Y)' = X' \cap Y'. \tag{3}$$

Lastly, suppose that $\{X_1, X_2\}$ and $\{Y_1, Y_2\}$ are pairs of Hilbert spaces such that both elements of each pairing are contained in a larger Hilbert space. If T is a bounded linear operator from X_i to Y_i for $i = 1, 2$, then

$$T \in \mathscr{L}(X_1 \cap X_2, Y_1 \cap Y_2) \cap \mathscr{L}(X_1 + X_2, Y_1 + Y_2), \tag{4}$$

and in particular

$$||T||_{\mathscr{L}(X_1+X_2,Y_1+Y_2)} \leq \max\left(||T||_{\mathscr{L}(X_1,Y_1)}, ||T||_{\mathscr{L}(X_2,Y_2)}\right).$$

Cf. [4, Ch. 2] for a further discussion of summation and intersection spaces.

3. Continuous stability and preconditioning

The weak formulation of (1)-(2) reads: Find $\mathbf{u} \in V$, $p \in Q$ such that

$$a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) = (\mathbf{f},\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}, \tag{5}$$

$$b(\mathbf{u},q) = (g,q), \quad \forall q \in Q. \tag{6}$$

where,

$$a(\mathbf{u},\mathbf{v}) = \left(K^{-1}\mathbf{u},\mathbf{v}\right) \text{ and } b(\mathbf{u},q) = (\nabla\cdot\mathbf{u},q).$$

The corresponding coefficient matrix reads,

$$\mathscr{A}\begin{bmatrix}\mathbf{u}\\ p\end{bmatrix} := \begin{bmatrix}K^{-1} & -\nabla\\ \operatorname{div} & 0\end{bmatrix}\begin{bmatrix}\mathbf{u}\\ p\end{bmatrix} = \begin{bmatrix}\mathbf{f}\\ g\end{bmatrix} \tag{7}$$

Below we show that the Brezzi conditions are satisfied uniformly in K in the following spaces for $\mathbf{u}$ and p, respectively:

$$\mathbf{V} = K^{-1/2}\mathbf{L}^2 \cap \mathbf{H}_0(\operatorname{div}), \quad Q = L^2_0 + K^{1/2}H^1 \cap L^2_0.$$

We remark that the below analysis is similar to [14], although the Sobolev spaces and bilinear forms involved are different. Let

$$\mathbf{Z} = \{\mathbf{u} \in \mathbf{V} \mid b(\mathbf{u},q) = 0, \quad \forall q \in Q\}.$$

Then clearly,

$$a(\mathbf{z},\mathbf{z}) = ||\mathbf{z}||^2_{K^{-1/2}\mathbf{L}^2} = ||\mathbf{z}||^2_{\mathbf{V}}, \quad \forall \mathbf{z} \in \mathbf{Z}$$

and hence coercivity of $a(\cdot,\cdot)$ is established. Furthermore, the boundedness of $a(\cdot,\cdot)$ follows from

$$a(\mathbf{u},\mathbf{v}) = (\mathbf{u},\mathbf{v})_{K^{-1/2}} \le ||\mathbf{u}||_{\mathbf{V}}||\mathbf{v}||_{\mathbf{V}} \quad \forall \mathbf{u},\mathbf{v} \in \mathbf{V}.$$

The boundedness of b follows from a decomposition argument. Let $q = q_0 + q_1$, $q_0 \in L^2$ and $q_1 \in H^1$ then

$$\begin{aligned} b(\mathbf{v},q) &= (\nabla\cdot\mathbf{v},q_0) - (\mathbf{v},\nabla q_1)\\ &= (\nabla\cdot\mathbf{v},q_0) - \left(K^{-1/2}\mathbf{v},K^{1/2}\nabla q_1\right)\\ &\le ||\nabla\cdot\mathbf{v}||_{L^2}||q_0||_{L^2} + ||\mathbf{v}||_{K^{-1/2}\mathbf{L}^2}||\nabla q_1||_{K^{1/2}\mathbf{L}^2}\\ &\le \left(\left(K^{-1}\mathbf{v},\mathbf{v}\right) + ||\nabla\cdot\mathbf{v}||^2_{L^2}\right)^{\frac{1}{2}}\left(||q_0||^2_{L^2} + (K\nabla q_1,\nabla q_1)\right)^{\frac{1}{2}}. \end{aligned}$$

Taking the infimum over all decompositions of q yields the desired bound.

The uniform inf-sup condition is established by a similar construction as in [14]; however, since the spaces involved are different a full argument is included for completeness. The argument proceeds by utilizing the Bogovskiĭ operator (cf. [7, 8, 10]) which, on a star shaped domain with respect to an open ball B, has the following explicit form:

$$\mathbf{S}g(\mathbf{x}) = \int_\Omega g(\mathbf{y})K(\mathbf{x}-\mathbf{y},\mathbf{y})\mathrm{d}\mathbf{y}, \text{ where } \mathbf{K}(\mathbf{z},\mathbf{y}) = \frac{\mathbf{z}}{|\mathbf{z}|^n}\int_{|\mathbf{z}|}^\infty \theta(\mathbf{y}+\mathbf{r}\frac{\mathbf{z}}{|\mathbf{z}|})\mathbf{r}^{n-1}\mathrm{d}\mathbf{r}.$$

Here $\theta \in C^\infty_0(\mathbb{R}^n)$, with

$$\operatorname{supp}\theta \subset B, \quad \text{and} \int_{\mathbb{R}^n}\theta(x) = 1.$$

The Bogovskiĭ operator $\mathbf{S}$ is a right inverse of the divergence operator and has the mapping properties

$$\mathbf{S} \in \mathscr{L}(L^2_0,\mathbf{H}^1_0) \cap \mathscr{L}(H^{-1},\mathbf{L}^2), \quad \text{and } \operatorname{div}\mathbf{S}g = g.$$

It follows directly for $g \in L^2_0$ that

$$||\mathbf{S}g||_{\mathbf{H}_0(\operatorname{div})} \le ||\mathbf{S}g||_{\mathbf{H}^1_0} \le C_{\mathbf{S}}||q||_{L^2_0}$$

for some constant $C_{\mathbf{S}}$. Hence,

$$\mathbf{S} \in \mathscr{L}(L^2_0,\mathbf{H}_0(\operatorname{div})) \cap \mathscr{L}(K^{-1/2}H^{-1},K^{-1/2}\mathbf{L}^2),$$

for any $K > 0$. From (4) and (3), we then have that

$$\mathbf{S} \in \mathscr{L}(K^{-1/2}H^{-1} \cap L_0^2, K^{-1/2}\mathbf{L}^2 \cap \mathbf{H}_0(\mathrm{div})) = \mathscr{L}(Q', \mathbf{V}),$$

and

$$\|\mathbf{S}g\|_{\mathbf{V}} \leq C_{\mathbf{S}} \|q\|_{Q'}.$$

The inf-sup condition follows directly. Take $p \in Q$ then

$$\begin{aligned} \sup_{\mathbf{v}\in\mathbf{V}} \frac{(\nabla\cdot\mathbf{v}, p)}{\|\mathbf{v}\|_{\mathbf{V}}} &\geq \sup_{g\in Q'} \frac{(\mathrm{div}\,\mathbf{S}g, p)}{\|\mathbf{S}g\|_{\mathbf{V}}} \\ &\geq C_S^{-1} \sup_{g\in Q'} \frac{(g,p)}{\|g\|_{Q'}} \\ &= C_S^{-1} \|p\|_Q. \end{aligned}$$

With K-uniform stability established, the framework put forth in [15] suggests that an efficient preconditioner for $\mathscr{A}$ can be constructed from preconditioners for the operators realizing the $\mathbf{V}$ and Q-norm. Preconditioners for the $\mathbf{V}$ inner product are well-known, c.f. e.g. [2, 12, 13]. For the Q-norm, we begin by recalling that

$$\|q\|_Q^2 = \inf_{\phi\in H^1\cap L_0^2} \left(\|q-\phi\| + K\|\nabla\phi\|^2\right), \tag{8}$$

where the infimum on the right hand side is attained with ϕ satisfying

$$(\phi,\psi) + K(\nabla\phi, \nabla\psi) = (q,\psi), \quad \forall \psi \in H^1 \cap L_0^2.$$

That is, $\phi = (I - K\Delta)^{-1}q$ and $q - \phi = -K\Delta\phi$. The Q-norm can then be characterized by

$$\begin{aligned} \|q\|_Q^2 &= ((-K\Delta)\phi, (I-K\Delta)\phi) \\ &= \left((-K\Delta)(I-K\Delta)^{-1}q, q\right). \end{aligned}$$

Thus, the canonical preconditioner $\mathscr{B}_Q : Q' \to Q$ is given by

$$\mathscr{B}_Q = \left[(-K\Delta)(I-K\Delta)^{-1}\right]^{-1} = I + (-K\Delta)^{-1}.$$

Hence, the operator $\mathscr{B} : \mathbf{V}' \times Q' \to \mathbf{V} \times Q$ defined as

$$\mathscr{B} = \begin{bmatrix} (K^{-1}I - \nabla\,\mathrm{div})^{-1} & 0 \\ 0 & I + (-K\Delta)^{-1} \end{bmatrix} \tag{9}$$

provides a robust preconditioner for (5)-(6) in the sense that the condition number of $\mathscr{B}\mathscr{A}$ is bounded uniformly in K.

4. Discrete Stability and preconditioning

In this section we describe the construction of a preconditioner for discretizations based on Brezzi-Douglas-Marini (BDM) and Raviart-Thomas (RT) elements [5, 17]. The discrete approach reflects many aspects of the continuous setting of section 3. However, due to the the discontinuous polynomial nature of the pressure elements, we first define a discrete H^1-norm to establish the Q-norm in the discrete case.

Let $\mathscr{T}_h$ be a shape regular simplicial mesh defined on the bounded, Lipschitz domain Ω and let $r \geq 0$. Let $\mathbf{V}_h$ be the $\mathbf{H}(\mathrm{div})$-conforming discrete space given by either the RT elements of order r or the BDM elements of order $r+1$. Define Q_h to be the usual corresponding space of discontinuous, piecewise polynomials of order r. Consider the discrete mixed Darcy problem given by: find $\mathbf{u}_h \in \mathbf{V}_h$ and $p_h \in Q_h$ such that

$$a(\mathbf{u}_h, \mathbf{v}) + b(\mathbf{v}, p_h) = (\mathbf{f}, \mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}_h, \tag{10}$$

$$b(\mathbf{u}_h, q) = (g, q), \quad \forall q \in Q_h. \tag{11}$$

The discrete Q_h-norm will be defined in terms of a discrete gradient which is the negative L^2-adjoint of the div operator on $\mathbf{V}_h$. First, $\nabla_h : Q_h \to \mathbf{V}_h$ is defined by

$$(\nabla_h q, \mathbf{v}) = -(q, \operatorname{div} \mathbf{v}). \tag{12}$$

It is well-known, [6, 9], that with these particular choices of Q_h and $\mathbf{V}_h$, there is an h-independent constant $\beta > 0$ such that

$$\sup_{\mathbf{v} \in \mathbf{V}_h} \frac{(q, \operatorname{div} \mathbf{v})}{\|\mathbf{v}\|_{\mathbf{H}(\operatorname{div})}} \geq \beta \|q\|_{\mathbf{L}^2} \tag{13}$$

for every $q \in Q_h$. It follows that ∇_h is injective and we can define the discrete H^1-norm on Q_h via

$$\|q\|_{1,h} = \|\nabla_h q\|_{\mathbf{L}^2}.$$

We denote the space H^1_h as the set Q_h equipped with the norm $\|\cdot\|_{1,h}$ and the space L^2_h as the set Q_h equipped with the usual L^2-norm. With these notations in hand, the discrete analogue of the Q-norm, given by (8), is defined as

$$\|q\|^2_{Q_h} = \inf_{\phi \in Q_h} \left(\|q - \phi\|^2_{L^2} + K \|\phi\|^2_{1,h} \right). \tag{14}$$

Defining $\mathbf{H}_{0,h}(\operatorname{div})$ and $\mathbf{L}^2_h$ analogously to L^2_h, we may then say that

$$\mathbf{V}_h = K^{-1/2} \mathbf{L}^2_h \cap \mathbf{H}_{0,h}(\operatorname{div}) \quad \text{and} \quad Q_h = L^2_h + K^{1/2} H^1_h,$$

and it is under these norms that we will show K-robust stability of (10)-(11).

Boundedness and coercivity of $a(\cdot,\cdot)$ and boundedness of $b(\cdot,\cdot)$ follows from the same arguments put forth in section 3. Verifying a K-independent inf-sup condition will therefore conclude the argument. To accomplish this a left-inverse of ∇_h will be constructed, satisfying appropriate bounds, allowing for a similar argument to that of section 3 for the operator $\mathbf{S}$.

Let $\mathbf{Z}_h$ denote the discrete kernel of the div operator; i.e. the set of $\mathbf{v}_h \in \mathbf{V}_h$ with

$$(\operatorname{div} \mathbf{v}_h, q_h) = 0, \quad \forall\, q_h \in Q_h \tag{15}$$

From (13) it follows [9] that $\nabla_h : Q_h \to \mathbf{Z}_h^{\perp}$ is a linear bijection. Furthermore every $\mathbf{v} \in \mathbf{V}_h$ can be uniquely decomposed as

$$\mathbf{v} = \nabla_h r + \tilde{\mathbf{v}}, \tag{16}$$

where $r \in Q_h$, with $\nabla_h r \in \mathbf{Z}_h^{\perp}$, and $\tilde{\mathbf{v}} \in \mathbf{Z}_h$. Since the spaces considered for $\mathbf{V}_h$ satisfy the relation $\operatorname{div} \mathbf{V}_h \subset Q_h$ it follows that $\operatorname{div} \tilde{\mathbf{v}} = 0$, for every $\tilde{\mathbf{v}} \in \mathbf{Z}_h$, and the decomposition (16) is orthogonal with respect to both the $\mathbf{H}(\operatorname{div})$ and $\mathbf{L}^2$ inner products.

We now define the lifting operator $\Theta_h : \mathbf{V}_h \to Q_h$ by $\Theta_h \mathbf{v} = r$, according to (16). It is evident that $\Theta_h \nabla_h$ is the identity operator on Q_h and that $\Theta_h \tilde{\mathbf{v}} = 0$ for all $\tilde{\mathbf{v}} \in \mathbf{Z}_h$. Moreover, the inf-sup condition (13) and the $\mathbf{H}(\operatorname{div})$-orthogonality of (16) implies that

$$\begin{aligned}
\|\Theta_h \mathbf{v}\| &\leq \beta^{-1} \sup_{\mathbf{w} \in \mathbf{V}_h} \frac{(\Theta_h \mathbf{v}, \operatorname{div} \mathbf{w})}{\|\mathbf{w}\|_{\mathbf{H}(\operatorname{div})}} \\
&= \beta^{-1} \sup_{\mathbf{w} \in \mathbf{V}_h} \frac{(\nabla_h \Theta_h \mathbf{v}, \mathbf{w})}{\|\mathbf{w}\|_{\mathbf{H}(\operatorname{div})}} \\
&\leq \beta^{-1} \sup_{\mathbf{w} \in \mathbf{V}_h} \frac{(\mathbf{v}, \mathbf{w})}{\|\mathbf{w}\|_{\mathbf{H}(\operatorname{div})}} \\
&= \beta^{-1} \|\mathbf{v}\|_{\mathbf{H}_{0,h}(\operatorname{div})'},
\end{aligned}$$

which means that $\Theta_h \in \mathscr{L}(\mathbf{H}_{0,h}(\mathrm{div})', L_h^2)$. From the L^2-orthogonality of (16) we also have that $\|\Theta_h \mathbf{v}\|_{1,h} \leq \|\mathbf{v}\|$, which implies that $\Theta_h \in \mathscr{L}(\mathbf{L}_h^2, H_h^1)$. From these bounds on Θ_h, together with (4), we deduce that

$$\Theta_h \in \mathscr{L}\left(K^{1/2}\mathbf{L}_h^2 + \mathbf{H}_{0,h}(\mathrm{div})', K^{1/2}H_h^1 + L_h^2\right) = \mathscr{L}(\mathbf{V}_h', Q_h). \tag{17}$$

Since $\Theta_h \nabla_h$ is the identity on Q_h, we get from (17) that for every $q \in Q_h$,

$$\begin{aligned} \|q\|_{Q_h} &\leq C \, \|\nabla_h q\|_{\mathbf{V}_h'} \\ &= C \sup_{\mathbf{v} \in \mathbf{V}_h} \frac{(q, \mathrm{div}\, \mathbf{v})}{\|\mathbf{v}\|_{\mathbf{V}_h}}, \end{aligned}$$

where $C = \max(1, \beta^{-1})$ and is thus independent of h.

We now consider the construction of suitable preconditioners for (10)-(11) that are robust in both K and h. If we define $\mathscr{A}_h$ as the coefficient matrix characterizing the left-hand side of (10)-(11), the above analysis implies that $\mathscr{A}_h$ is a homeomorphism from $\mathbf{V}_h \times Q_h$ to its dual. Moreover, the norms on $\mathscr{A}_h$ and its inverse are bounded independently of K and h. Using an argument analogous to the one we made to define $\mathscr{B}$ in section 3, we define the canonical preconditioner $B_h : \mathbf{V}_h' \times Q_h' \to \mathbf{V}_h \times Q_h$ as

$$B_h = \begin{bmatrix} \left(K^{-1} I_h - \nabla_h \, \mathrm{div}\right)^{-1} & 0 \\ 0 & I_h + (-K \, \mathrm{div}\, \nabla_h)^{-1} \end{bmatrix}. \tag{18}$$

For simplicitly, a small liberty has been taken for the notation of I_h in (18). Specifically, I_h in the top left block of (18) signifies the identity $I_h : \mathbf{V}_h \to \mathbf{V}_h$ while the use of the same symbol in the bottom right block signifies the identity on Q_h; recall that the discrete gradient, ∇_h, is defined by (12).

5. Numerical Experiments

Let Ω be a triangulation of the unit square such that the unit square is first divided in $N \times N$ squares of length $h = 1/N$. Each square is then divided into two triangles. Below we will consider the case of homogeneous Dirichlet conditions for the flux and compute the eigenvalues of the preconditioned system. As order optimal multilevel methods for both $\mathbf{H}(\mathrm{div})$ and H^1 problems are well-known we consider preconditioners based simply on exact inversion. It was shown in [18] that the following local $-\Delta_h$ operator is spectrally equivalent with $-\mathrm{div}\, \nabla_h$ for RT and BDM elements of arbitrary order:

$$(-\Delta_h p, q) = \sum_{T \in \mathscr{T}_h} \int_T \nabla p \cdot \nabla q \, dx + \sum_{E_i \in \mathscr{E}_{\mathscr{I}}} \int_{E_i} [p][q] \, dS + \sum_{E_i \in \mathscr{E}_{\mathscr{D}}} \int_{E_i} pq \, dS.$$

Here, $\mathscr{T}_h$ is a triangulation of the domain Ω; internal faces are signified by the set $\mathscr{E}_{\mathscr{I}}$ whereas $\mathscr{E}_{\mathscr{D}}$ denotes faces at the boundary associated with a Dirichlet condition. In Table 1 we compare the two different preconditioners:

$$\mathscr{B}_1 = \begin{bmatrix} (\frac{1}{K} I - \nabla\nabla\cdot)^{-1} & 0 \\ 0 & (I)^{-1} + (-K\Delta)^{-1} \end{bmatrix} \text{ and } \mathscr{B}_2 = \begin{bmatrix} (\frac{1}{K} I - \nabla\nabla\cdot)^{-1} & 0 \\ 0 & (I)^{-1} \end{bmatrix}.$$

We would argue that $\mathscr{B}_2$ represents a natural choice with the exception of employing operator conditioning combined with intersection and sum spaces. Table 1 shows that $\mathscr{B}_1$ yields robust results for any $K \in (0, 1)$.

Numerical experiments confirm that the robust behavior applies also to Neumann conditions and BDM elements.

	$K\backslash h$	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\mathscr{B}_1$	1	1.1	1.1	1.1	1.1
	10^{-2}	2.2	2.1	2.1	2.1
	10^{-4}	3.1	3.3	3.0	2.9
	10^{-6}	3.2	3.6	3.8	3.9
	10^{-8}	3.6	3.6	3.8	3.9
$\mathscr{B}_2$	1	1.0	1.1	1.1	1.1
	10^{-2}	6.0	6.0	6.1	6.1
	10^{-4}	500	505	507	507
	10^{-6}	$4.9 \cdot 10^4$	$4.9 \cdot 10^4$	$4.9 \cdot 10^4$	$4.9 \cdot 10^4$
	10^{-8}	$4.7 \cdot 10^6$	$5.1 \cdot 10^6$	$4.5 \cdot 10^6$	$4.7 \cdot 10^6$

TABLE 1. Condition numbers for the operators $\mathscr{B}_1 A$ and $\mathscr{B}_2 A$.

6. ACKNOWLEDGEMENT

The authors gratefully acknowledge helpful discussions with Ragnar Winther[1] in addition to productive exchanges with Marie E. Rognes[2].

REFERENCES

[1] D. Arnold, R. Falk, and R. Winther. Preconditioning in $H(\mathrm{div})$ and applications. *Mathematics of Computation of the American Mathematical Society*, 66(219):957–984, 1997.

[2] D. N. Arnold, R. S. Falk, and R. Winther. Multigrid in $H(\mathrm{div})$ and $H(\mathrm{curl})$. *Numerische Mathematik*, 85(2):197–217, 2000.

[3] M. Benzi and M. A. Olshanskii. An augmented Lagrangian-based approach to the Oseen problem. *SIAM Journal on Scientific Computing*, 28(6):2095–2113, 2006.

[4] J. Bergh and J. Löfström. *Interpolation Spaces*. A series of comprehensive studies in mathematics. Springer, New York, 1976.

[5] F. Brezzi, J. Douglas, and L. D. Marini. Two families of mixed finite elements for second order elliptic problems. *Numerische Mathematik*, 47(2):217–235, 1985.

[6] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, 1 edition, 1991.

[7] M. Costabel and A. McIntosh. On Bogovskiĭ and regularized Poincaré integral operators for de Rham complexes on Lipschitz domains. *Mathematische Zeitschrift*, 265(2):297–320, 2010.

[8] G. Galdi. *An introduction to the mathematical theory of the Navier-Stokes equations: Steady-state problems*. Springer Science & Business Media, 2011.

[9] G. Gatica. *A Simple Introduction to the Mixed Finite Element Method*. Springer, 2014.

[10] M. Geißert, H. Heck, and M. Hieber. On the equation div u= g and Bogovskii's operator in Sobolev spaces of negative order. In *Partial differential equations and functional analysis*, pages 113–121. Springer, 2006.

[11] C. Greif and D. Schötzau. Preconditioners for saddle point linear systems with highly singular (1, 1) blocks. *ETNA, Special Volume on Saddle Point Problems*, 22:114–121, 2006.

[12] R. Hiptmair and J. Xu. Nodal auxiliary space preconditioning in $\mathbf{H}(\mathbf{curl})$ and $\mathbf{H}(\mathrm{div})$ spaces. *SIAM J. Numer. Anal.*, 45(6):2483–2509, 2007.

[13] T. V. Kolev and P. S. Vassilevski. Parallel auxiliary space AMG solver for $H(\mathrm{div})$ problems. *SIAM J. Sci. Comput.*, 34(6):A3079–A3098, 2012.

[14] K.-A. Mardal, J. Schöberl, and R. Winther. A uniformly stable Fortin operator for the Taylor–Hood element. *Numerische Mathematik*, 123(3):537–551, 2013.

[1]Dept. of Mathematics, University of Oslo. Oslo, Norway.

[2]Dept. of Numerical Anal. and Sci. Comput. , Simula Research Laboratory. Fornebu, Norway

[15] K.-A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numerical Linear Algebra with Applications*, 18(1):1–40, 2011.
[16] M. F. Murphy, G. H. Golub, and A. J. Wathen. A note on preconditioning for indefinite linear systems. *SIAM Journal on Scientific Computing*, 21(6):1969–1972, 2000.
[17] P.-A. Raviart and J.-M. Thomas. A mixed finite element method for 2-nd order elliptic problems. In *Mathematical aspects of finite element methods*, pages 292–315. Springer, 1977.
[18] T. Rusten, P. Vassilevski, and R. Winther. Interior penalty preconditioners for mixed finite element approximations of elliptic problems. *Mathematics of Computation of the American Mathematical Society*, 65(214):447–466, 1996.