Preconditioners for singular problems and coupled problems with domains of different dimensionality

Miroslav Kuchta

Thesis submitted for the degree of Ph.D.
Department of Mathematics
University of Oslo

December 2016

Preface

This thesis is submitted to the Department of Mathematics, University of Oslo in partial fulfillment of the requirements for degree Ph.D. The work that led up to the thesis was carried under the supervision of Associate Professor Mikael Mortensen and Professor Kent-Andre Mardal in the period between September 2012 and November 2016.

The thesis consists of four research papers preceded by a short introduction which provides motivation for the problems studied in the articles and reviews some of the concepts that are central to the thesis.

During my master studies at Charles University I have had the pleasure to witness several talks by Professor Kumbakonam Rajagopal. As the great orator and presenter he is, he would always start his presentation by quotations "so that the audience had something to think about in case his talk was exceedingly boring". I shall follow his example here. The following are some of the quotes that have inspired me throughout the last four years.

> *"The road to wisdom?– Well, it's plain and simple to express: Err and err and err again, but always less and less and less."*
>
> — Piet Hein

> *"I learned very early the difference between knowing the name of something and knowing something."*
>
> — Richard Feynman

> *"Singular matrices, in a certain sense, do not exist at all. "*
>
> — Cornelius Lanczos

In case of Professor Rajagopal the quotes were never the most amusing part of the presentation. It is my hope that the reader will reach the same conclusion about the presented thesis.

Acknowledgments

I would first of all like to thank my two supervisors, Mikael Mortensen and Kent-Andre Mardal, for their guidance, kindness, patience and enthusiasm. I am especially indebted to Kent for introducing me to the wonderful world of preconditioning and for the numerous stories and weekend discussions about the aspects of life which are typically not viewed through the prism of partial differential equations.

I am grateful to Arne Bøckmann, Magnus Vartdal, Susanne Støle-Hentschel, Diako Darian, Tormod Landet and most recently Jakob Schreiner for many laughs, advice and stimulating discussions not only during the time that we have shared an office together. I would also like to thank my colleagues from the mechanics division as well as Biljana Dragišić and Terje Kvernes from the administration and IT services for creating an enjoyable working environment. I also need to thank Magne Nordaas and Joris Verschaeve with whom I have collaborated on the first paper.

For her energy, awesomeness, positive attitude and friendship which countless times have made a sad day better I am thankful to my friend and colleague Erika Kristina Lindstrøm. Thanks is also due to Jakub Kalus who has been my roommate for the past three years.

Finally, I would like to express my deepest gratitude to my family and in particular my parents. This thesis would not have been possible without their endless support, encouragement and trust.

Oslo, December 2016
Miroslav Kuchta

Introduction

This chapter begins by placing the problems studied in the thesis into a broader context. Afterwards, some relevant tools of analysis which were used throughout the work are presented. Finally, the chapter concludes with a summary of the papers.

Motivation

In his essay *The unreasonable effectiveness of mathematics in the natural sciences*, see [72], the Nobel prize winning physicist Eugen Wigner calls mathematics the correct language for formulating the laws of (inanimate[1]) nature. In the language of mathematics these laws are often expressed with an elegance and beauty that inspires awe much like the natural phenomena governed by the laws themselves. However, the aesthetics of notation is secondary to the fact that it is the mathematical language and reasoning that make consequences of these laws computable. Thus it can be seen how well the existing laws explain a given phenomena or whether, perhaps, a new law is needed.

In the mathematical language many laws of nature are formulated as, and numerous physical processes are described by, partial differential equations (PDEs). To give a few examples, the Newton's law of gravity can be formulated as a Poisson equation while the laws of conservation of mass and momentum for the incompressible materials are expressed as the Navier-Stokes equations. The Poisson equation models stationary heat distribution, distribution of charges in electrostatics or, under certain assumptions, deformation of an elastic body, see e.g. [23]. The applications of Navier-Stokes equations are then ranging from hemodynamics, e.g [52], through dynamics of Earth's mantle, e.g. [58], to formation of galaxies, e.g. [38]. It is fair to say that one learns about nature by solving PDEs.

In general, the exact solution to the equations can only be computed under special circumstances and therefore approximations of the solution, which can be obtained from simpler problems, are of interest. Such approximations can be constructed for example by the method of finite differences, the finite volume method or the finite element method[2], see [56] and the references therein. With the listed methods the approximation is obtained by solving a sparse linear system acquired by transformation/discretization of the underlying PDE. Even with simple equations, such as the Poisson problem, the requirements on the accuracy of the approximation can result in systems with millions of unknowns.

Nowadays a sparse linear system of order one million, i.e. 10^6 unknowns, is solved in a few seconds on an ordinary laptop. This is in a striking contrast to say 1960's when inverting a *dense* matrix with 200 unknowns was a task pushing the limits of the most powerful ma-

[1] For the sciences that involve human beings [29] points to the *unreasonable effectiveness of data*.

[2] The idea of discretizing the domain to obtain a linear system from a differential equation is certainly not recent. Already Daniel Bernoulli (1700-1782) computed the deflection of a chain by considering it as a collection of small segments [37, ch 4.].

chines (supercomputers) available at that time [70, ch 32.]. The speed-up has been enabled by advances in (super)computer hardware as well as advances in the field of computational methods. In fact, the contribution of the two is about the same [50] implying that the algorithms for solving linear systems, in some sense, keep up with Moore's law[3].

Let us denote by N the order of a linear system $\mathsf{Ax} = \mathsf{b}$ due to either of the discretization methods above. As noted before the system matrix is sparse with $\mathcal{O}(N)$ number of nonzero entries Z. In a general case, the performant direct methods known to date, e.g. [39, 40], can solve the system in at most $\mathcal{O}(ZN \log N)$ operations. In contrast, iterative methods can often compute the (approximate) solution only in $\mathcal{O}(N)$ operations [4]. A class of iterative methods which can offer linear complexity are the Krylov subspace methods such as the conjugate gradient method (CG) [31, 63] or the minimal residual method (MINRES) [51]. An overview of other methods can be found e.g. in [24, ch 11.4]. A cost of a single step of these algorithms is proportional to N as it typically involves only a fixed number of matrix-vectors products. For optimal complexity it is therefore necessary to guarantee that the solution with the prescribed error tolerance is obtained in a number of steps independent of N. To this end the methods are applied to a modified system $\mathsf{BAx} = \mathsf{Bb}$ instead of the original problem. Here, B is the preconditioner matrix introduced to improve convergence of the method while *not* increasing its complexity.

For many PDEs efficient preconditioners for the related linear systems are known; the Poisson equation is solved efficiently by CG employing multigrid as a preconditioner [68, 4], MINRES with preconditioner of [71, 65] is an optimal solver for the Stokes equations, while [64] present a preconditioner for solving linearized Navier-Stokes equations efficiently by generalized minimal residual method of [60]. However, there is a wide range of problems for which efficient numerical algorithms are missing. In this thesis such methods are established for two types of problems; a multiscale problem describing coupling of two diffusive processes on domains with different dimensionality (see (17) and (18)) and the Neumann problem of linear elasticity (see (19) and (20)).

Multiscale problem

The considered multiscale problem is particularly relevant in biomedical applications where a wide range of spatial scales is present. Here, resolving the smallest structures as three-dimensional objects embedded in the surrounding domain of interest might be prohibitively expensive, e.g. alveoli in the human lungs have a typical diameter of 200μm [49], for the capillaries in the cardiac muscle the diameter is even smaller 5μm [55]. One option to include these scales into the model is order reduction. Then, the three-dimensional structures will be modeled by PDEs posed on lower-dimensional manifolds. In case of capillaries, the radii are negligible in comparison to their lengths and the manifold is a curve. Finally, the complete model represents a coupling between a process in the three-dimensional bulk domain and a possibly different process on the reduced domain. If further modeling assumptions are made about the processes in the bulk a $2d$-$1d$ coupled problem can be obtained.

The $3d$-$1d$ models have been used e.g. by [25, 41, 22, 57] to study blood and oxygen transport in the brain, [15] to describe fluid exchange between microcirculation and tissue interstitium, [14] to study efficiency of cancer therapies delivered through microcirculation or

[3] This achievement should not be taken for granted. For example, improvements in the compiler technology lag significantly [62].

[4] Assuming $N = 10^8$, and a computer delivering 1TFLOPs (for simplicity let one flop correspond to one operation step of the method) the complexity of the algorithms translates respectively into a computational time of one day and less than one second.

[47] to investigate hyperthermia as a cancer treatment. Since the focus here is on handling the domain coupling rather than on the physical applications, the coupled problems considered in the thesis are simpler than those previously cited.

The proposed strategy of handling the coupling is to include the constraint by introducing a new unknown, the Lagrange multiplier. This way a symmetric, indefinite system suitable for MINRES method is obtained. A major advantage of the approach is that it allows for using the standard and efficient preconditioners for the PDEs that describe the processes on the involved domains. At the same time, the size of the linear system is only marginally increased. However, establishing an efficient preconditioner for the Lagrange multiplier is not trivial and the contribution of this work is showing how the challenge can be overcome.

Singular problem

The Neumann problem of linear elasticity describes deformation of an (elastic) body which is not anchored in space. As such, it is a natural model to use with objects that, in the broad sense of the word, float. Examples of such objects are ships, celestial objects [69] or human brain [21]. Unfortunately, the absence of anchoring renders the equations of the model singular and this fact presents an issue for the numerical methods.

In special cases, the singularity can be avoided by choosing the discretization method such that the singular modes (rigid motions) are not present, e.g. spherical body and discretization using spherical harmonics. A more universal approach is to artificially fix the body in space by modified boundary conditions or by constraining the point displacement. However, these approaches leave artifacts on the solution, e.g. [61], or result in poor convergence properties[5], see Paper IV.

This thesis discusses how the Neumann problem can be solved efficiently without resorting to any of the above tricks. More specifically, an orthonormal basis of the space of rigid motions is constructed and later employed to formulate well-posed problems suitable for MINRES or CG methods. For both methods efficient preconditioners are designed.

Methods

In the following, both the multiscale problem and the Neumann problem shall be discretized by the finite element method. Consequently, weak forms of the governing equations are considered and the problems are regarded as operator equations defined in terms of bilinear and linear forms over suitable Hilbert spaces. The considered problems fit into an abstract framework of operator preconditioning [44], see also [43], in which the structure of the *discrete* preconditioners is identified by considering the properties of the *continuous* operators. As the framework is an essential tool used throughout the thesis, we shall next briefly review the main ideas.

Operator preconditioning

Let V be a Hilbert space with norm $||\cdot||_V$. We shall denote as V' the space of linear functionals defined on V, while the action of $f \in V'$ on $v \in V$ is written as $\langle f, v\rangle$.

[5] In some sense, the statement of Archimedes *“Give me the place to stand, and I shall move the Earth.”* is wrong here. Fixed points are not sufficient for the Neumann problem.

The problems to be analyzed in the subsequent chapters are saddle point systems of the form: Given $f \in V'$, $h \in Q'$ find $u \in V$, $p \in Q$ such that for all test functions $v \in V$, $q \in Q$

$$a(u,v) + b(v,p) + b(u,q) = \langle f, v\rangle + \langle h, q\rangle. \tag{1}$$

Here, $a : V \times V \rightarrow \mathbb{R}$, $b : V \times Q \rightarrow \mathbb{R}$ and $L \in V' \times Q'$ are respectively the given bilinear and linear forms defined over a pair of Hilbert spaces V, Q. The problem (1) can be equivalently stated as an operator equation $\mathscr{A}x = L$ for $\mathscr{A} : W \rightarrow W'$, $W = V \times Q$ and $x \in W$, $L \in W'$ given as

$$\begin{bmatrix} A & B' \\ B & \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ h \end{bmatrix} \tag{2}$$

with the operators A, B defined in terms of the bilinear forms a, b from (1) as

$$A : V \rightarrow V', \langle Au, v\rangle = a(u,v) \text{ and } B : V \rightarrow Q', \langle Bu, q\rangle = b(u,q).$$

Note that $B' : Q \rightarrow V'$ is the adjoint of B, $\langle Bu, q\rangle = \langle B'q, u\rangle$.

Let now $\mathscr{A} : W \rightarrow W'$ be a symmetric isomorphism and suppose that we wish to find the solution of a well-posed problem (2) by Krylov iterations. We refer to e.g. [26] for formulation of CG and MINRES methods in the Hilbert space (infinite dimensional) setting and only note here that in order for the methods to be well-defined, the Krylov subspaces $K_k = \text{span}\{L, \mathscr{A}L, \mathscr{A}^2 L, \dots \mathscr{A}^{k-1}L\}$, in which the approximations are sought, must make sense. For $\mathscr{A} : W \rightarrow W'$, $L \in W'$ such a construction is not meaningful[6].

For the Krylov method to be well-defined the iterations must use an isomorphism $W \rightarrow W$. In the framework of operator preconditioning the isomorphism is constructed as $\mathscr{B}\mathscr{A}$ where the preconditioner $\mathscr{B} : W' \rightarrow W$ is a symmetric positive-definite. As such, its inverse $\mathscr{B}^{-1}$ defines an inner product on W, namely $(\cdot,\cdot)_{\mathscr{B}} = \langle \mathscr{B}^{-1}\cdot, \cdot\rangle$. Moreover the operator $\mathscr{B}\mathscr{A}$ is symmetric with respect to the induced inner product

$$(\mathscr{B}\mathscr{A}x, y)_{\mathscr{B}} = \langle \mathscr{B}^{-1}\mathscr{B}\mathscr{A}x, y\rangle = \langle \mathscr{A}x, y\rangle = \langle \mathscr{A}y, x\rangle = \langle \mathscr{B}^{-1}\mathscr{B}\mathscr{A}y, x\rangle = (\mathscr{B}\mathscr{A}y, x)_{\mathscr{B}}.$$

We note that the Krylov subspace constructed by applying $\mathscr{B}\mathscr{A}$ to $\mathscr{B}L \in W$ is well-defined.

The needed mapping $\mathscr{B}$ can be obtained by the Riesz representation theorem. Thus different norm-equivalent inner products on W' give rise to different spectrally equivalent preconditioners. The spectral equivalence property is of practical importance since the preconditioners are typically not equally cost efficient.

Given V, Q with their respected norms the operator $\mathscr{A}$ from (2) can be shown to be an isomorphism by verifying the Brezzi conditions [12]:

(i) There exists $\alpha^* > 0$ such that,

$$a(u,v) \leq \alpha^* ||u||_V ||v||_V \text{ for any } u, v \in V. \tag{3a}$$

(ii) There exists $\alpha_* > 0$ such that for any $u \in Z$, $Z = \{v \in V; b(v,q) = 0 \text{ for all } q \in Q\}$

$$a(u,u) \geq \alpha_* ||u||_V^2. \tag{3b}$$

[6] When the operator equation is discretized this fact is manifested by iterations being unbounded with respect to the size of the linear system.

(iii) There exists $\beta^* > 0$ such that,

$$b(u,q) \leq \beta^* ||u||_V ||q||_Q \text{ for any } u \in V, q \in Q. \tag{3c}$$

(iv) (inf-sup condition) There exists $\beta_* > 0$ such that

$$\sup_{v \in V} \frac{b(q,v)}{||v||_V} \geq \beta_* ||q||_Q \text{ for any } q \in Q. \tag{3d}$$

We remark that the Brezzi theorem is a special case of the result due to Babuška [5] and Nečas [48] (see also [10, ch 3.3], [18]). Therein, the bilinear form related to $\mathscr{A}$ is not assumed to have the structure (1). Note also, that if the Brezzi conditions hold, the preconditioner for the operator equations $\mathscr{A}x = L$ is such that $\langle \mathscr{B}^{-1}(u,p),(u,p)\rangle = ||u||_V^2 + ||p||_Q^2$, i.e. the preconditioner is diagonal. We finally note that the condition number of the preconditioned operator $\mathscr{B}\mathscr{A}$ is given in terms of the constants from the inequalities (3a)–(3d), see e.g. [35].

Assuming that the Brezzi conditions hold, we shall finally solve the problem (2) numerically. To this end let $V_h \subset V$, $Q_h \subset Q$ be the finite element subspaces and consider the problem: Find $u_h \in V_h$, $p_h \in Q_h$ such that

$$a(u_h,v_h) + b(v_h,p_h) + b(u_h,q_h) = \langle f, v_h\rangle + \langle h, q_g\rangle \quad v_h \in V_h, q_h \in Q_h$$

or equivalently $\mathscr{A}_h x_h = L_h$. Similarly, let $\mathscr{B}_h$ denote the Galerkin approximation of the preconditioner $\mathscr{B}$. It is well known that for the discretization to be stable, the finite elements must be chosen such that (3a)–(3d) are true on the discrete subspaces V_h, Q_h. In particular, discrete versions of the coercivity condition (3b) and the inf-sup condition (3d) do not follow automatically from the continuous case. However, if the conditions hold, the number of Krylov iterations on $\mathscr{B}_h\mathscr{A}_h$ will be bounded in the discretization parameter as the convergence of the method is estimated in terms of the (discrete) Brezzi constants. Such an estimate for the MINRES method can be found e.g. in [53, ch 4.]. We note that the action of $\mathscr{B}_h$ is typically expensive to compute and in the Krylov method the preconditioner is therefore replaced by a more practical, spectrally equivalent operator. The new preconditioner then leads to different Brezzi constants.

To demonstrate the power of the abstract framework let us apply it to a simple example of a $1d$-$0d$ coupled problem. Using the example we shall also illustrate some of the crucial concepts that appear in the $2d$-$1d$ and $3d$-$1d$ coupled problems studied in Paper I and Paper II, e.g. the trace and extension operators.

The $1d$-$0d$ coupled problem

Suppose $\Omega = (-1,1)$ and $\gamma \in \Omega$. For consistency of notation with the rest of the work we write $\Delta u = {}^{\mathrm{d}^2 u}/_{\mathrm{d}x^2}$ (and similarly for ∇) and consider the following problem

$$\begin{aligned} \Delta u + p\delta_\gamma &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \\ T_\gamma u &= h && \text{at } \gamma, \end{aligned} \tag{4}$$

with f, h the given data. Here u is the unknown function constrained at point γ, while the scalar p is the Lagrange multiplier associated with the point constraint. We shall come back to its physical meaning later. Finally the operators T_γ (trace), δ_γ (mean value) are respectively such that $T_\gamma v = v(\gamma)$ and $\int_\Omega (\delta_\gamma v)(x)\,\mathrm{d}x = v(\gamma)$ for v a continuous function.

To discuss the weak form of (4) let $V = H^1_0(\Omega)$ where the space shall be considered with the norm $||u||_V = \sqrt{(\nabla u, \nabla u)}$. Here $(\cdot,\cdot)$ denotes the $L^2(\Omega)$ inner product while the corresponding norm is denoted as $||\cdot||$. Note that the fact that $||\cdot||_V$ is indeed a norm on V follows from Poincaré inequality, e.g. [11, ch 5.3].

As the functions in V are continuous both T_γ and δ_γ are continuous operators on V. More precisely, by the fundamental theorem of calculus the estimate

$$\langle \delta_\gamma, v\rangle = \int_{-1}^{\gamma} \nabla v(x)\,\mathrm{d}x \leq C||\nabla v|| = C||v||_V \tag{5}$$

holds. Moreover, by Riesz representation theorem there exists a unique element $g_\gamma \in V$ such that $\left(\nabla g_\gamma, \nabla v\right) = \langle \delta_\gamma, v\rangle$ for any $v \in V$. Note that in turn $C = ||g_\gamma||_V$ in the proceeding estimate. The function g_γ is the Green's function satisfying

$$\begin{aligned} \Delta u &= \delta_\gamma \qquad &&\text{in } \Omega,\\ u &= 0 &&\text{on } \partial\Omega. \end{aligned} \tag{6}$$

The solution of (6) then reads $g_\gamma(x) = \frac{1}{2}(1 - |x-\gamma| - x\gamma)$ and it follows that $||g_\gamma||^2_V = \frac{1}{2}(1-\gamma^2)$. Using the Green's function we shall define extension operator $E_\gamma : \mathbb{R} \mapsto V$, $E_\gamma : q \to q g_\gamma$. Clearly, the operator is bounded and T_γ is its inverse.

Interestingly, the Green's function here can be linked to the famous Basel problem of summing $\sum_{k\geq 1}^{\infty} {}^1\!/\!_{k^2}$. The problem was first solved by Euler who showed the result to be ${}^{\pi^2}\!/\!_6$, see e.g. [17]

Remark 1 (Basel problem) *Let functions u_k, $k \geq 1$, $k \in \mathbb{N}$ be defined as*

$$u_k(x) = \begin{cases} \sin \frac{k\pi}{2} x & k \text{ even },\\ \cos \frac{k\pi}{2} x & k \text{ odd }. \end{cases}$$

Then $u_k \in V$ and pairs $(u_k, {}^{k^2\pi^2}\!/\!_4)$ solve the eigenvalue problem $-\Delta u_k = \lambda_k u_k$. Moreover, the set $\{u_k\}_{k=1}^{\infty}$ forms an L^2-orthonormal and V-orthogonal basis of V. Note that by expanding $u \in V$ in the basis, it is easy to see that the smallest eigenvalue $\lambda_1 = \pi^2/4$ is the constant of Poincaré lemma on the space V. Indeed $||u||^2 \leq \lambda_1^{-1}||u||^2_V$ can be seen to hold. A function which is particularly simple to represent in the basis of eigenfunctions is the Green's function. Here the expansion coefficients are computed simply by evaluating the eigenvectors

$$g_\gamma = \sum_{k=1}^{\infty} G_k u_k, \qquad G_k = \frac{\langle \delta_\gamma, u_k\rangle}{\lambda_k}.$$

Moreover, the norms are easily computable by Parseval's equality

$$||g_\gamma||^2_0 = \sum_{k=1}^{\infty} \frac{\langle \delta_\gamma, u_k\rangle^2}{\lambda_k^2}, \qquad ||g_\gamma||^2_V = \sum_{k=1}^{\infty} \frac{\langle \delta_\gamma, u_k\rangle^2}{\lambda_k}. \tag{7}$$

Finally, consider g_0, the Green's function at the origin. Then $g_0 = \sum_{k\ odd}^{\infty} \frac{4}{\pi^2 k^2} \cos \frac{k\pi}{2} x$ and upon

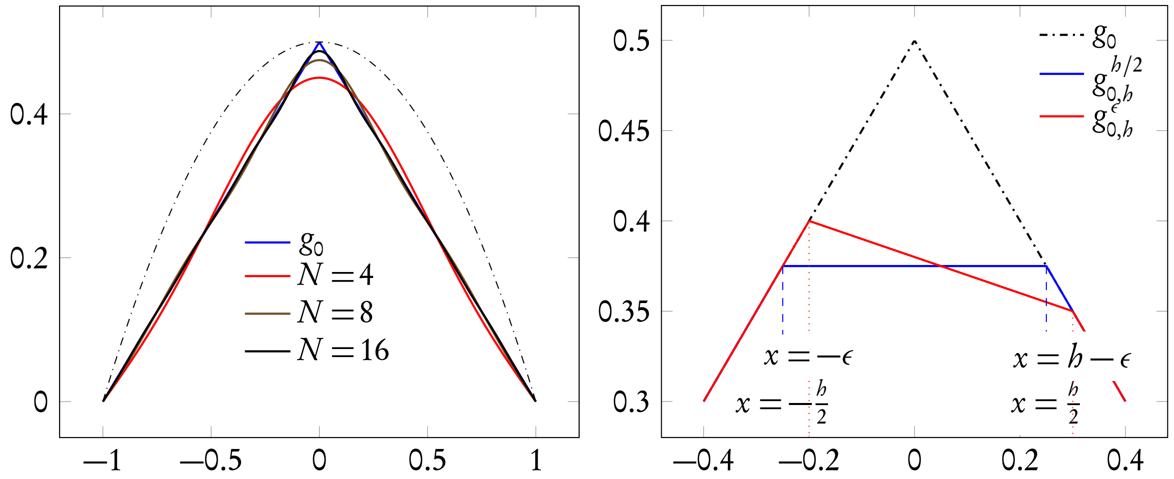

Figure 1: (Left) Partial sums of series of the Green's function $g_0 = \frac{1}{2}(1-|x|)$ for different values of N. (Right) The exact Green's function g_0 representing functional δ_0 and its continuous piecewise linear approximations $g^{\epsilon}_{0,h}$. On a nonmathching mesh the approximations cannot resolve the kink of the true solution within an element. The element boundaries are indicated by dashed and dotted vertical lines.

evaluating the right hand side in (7) we have $\sum_{k\ odd}^{\infty} 1/k^2 = \pi^2/8$. To illustrate convergence of the series for g_0, Figure 1 plots several partial sums $\sum_{k\ odd}^{N} \frac{4}{\pi^2 k^2}\cos\frac{k\pi}{2}x$.

To solve (4) the problem is recast as a saddle point system for $u \in V$ and a Lagrange multiplier $p \in Q$, $Q = \mathbb{R}$ satisfying

$$(\nabla u, \nabla v) + p\langle \delta_\gamma, v\rangle + q\langle \delta_\gamma, u\rangle = \langle f, v\rangle + hq, \quad v \in V, q \in Q. \tag{8}$$

Equivalently, (8) defines an operator equation (2) with

$$\langle Au, v\rangle = (\nabla u, \nabla v) \quad \text{and} \quad \langle Bq, u\rangle = q\langle T_\gamma, u\rangle. \tag{9}$$

To show that the system (8) is well posed, let us see if the Brezzi conditions hold. Clearly, $\alpha^* = \alpha_* = 1$ in (3a), (3b). Moreover $p\langle \delta_\gamma, v\rangle \leq |p|\,||g||_V ||v||_V$ follows from (5). Thus if the space Q is considered with norm $||q||_Q = |q|$ then $\beta^* = ||g_\gamma||_V$ in (3c). On the other hand, setting $||q||_Q = |q|\,||g_\gamma||_V$ yields $\beta^* = 1$. We are left with verifying the inf-sup condition. However, employing the extension $\langle E_\gamma, q\rangle = q\,g_\gamma$ it holds that

$$\sup_{v\in V} \frac{q\langle \delta_\gamma, v\rangle}{||v||_V} = \sup_{v\in V} \frac{q\left(\nabla g_\gamma, \nabla v\right)}{||v||_V} \geq \frac{q\left(\nabla g_\gamma, \nabla g_\gamma\right)}{||g_\gamma||_V} = q||g_\gamma||_V.$$

Thus $\beta_* = 1$ if $||q||_Q = |q|\,||g_\gamma||_V$ and $\beta^* = ||g_\gamma||_V$ if $||q||_Q = |q|$.

Using the two norms of the space Q with respect to which (8) was proved well-posed we define two preconditioners for the problem

$$\mathcal{B}_1 = \begin{bmatrix} -\Delta & \\ & 1 \end{bmatrix}^{-1} \quad \text{and} \quad \mathcal{B}_\gamma = \begin{bmatrix} -\Delta & \\ & ||g_\gamma||_V^2 \end{bmatrix}^{-1}. \tag{10}$$

Here, $\mathscr{B}_1$ corresponds to considering the space Q with $|\cdot|$ norm, while with $\mathscr{B}_\gamma$ the norm is $\|p\|_Q = |p|\|g_\gamma\|_V$. Obviously, the preconditioner $\mathscr{B}_\gamma$ is less practical of the two as it involves solving a global problem (6) in order to find the norm of the Green's function. However, here the quantity is easily obtained, $\|g_\gamma\|_V^2 = \frac{1}{2}(1-\gamma^2)$. Moreover, as is shown next, the preconditioner is independent of the location of the point constraint.

Consider now the eigenvalue problem for operator $\mathscr{A}$ with the preconditioners (10): Find $(u, p, \lambda) \in V \times Q \times \mathbb{R}$ such that

$$(\nabla u, \nabla v) + p\langle \delta_\gamma, v\rangle + q\langle \delta_\gamma, u\rangle = \lambda(\nabla u, \nabla v) + C\lambda pq, \quad v \in V, q \in Q. \tag{11}$$

Here C takes the values 1 or $\|g_\gamma\|_V^2$ depending on whether $\mathscr{B}_1$ or $\mathscr{B}_\gamma$ is considered. We show that there are at most three distinct eigenvalues. First suppose that $p = 0$. By testing (11) with $(v, q) = (0, 1)$ it can be seen that $u \in V$ must be such that $\langle \delta_\gamma, u\rangle = 0$, i.e. u is orthogonal to g_γ in the inner product of V. For such u the eigenvector $(u, 0)$ has an eigenvalue $\lambda = 1$. Next suppose $u = g_\gamma$. Then (11) reads

$$\langle \delta_\gamma, v\rangle + p\langle \delta_\gamma, v\rangle + q\langle \delta_\gamma, g_\gamma\rangle = \lambda\langle \delta_\gamma, v\rangle + C\lambda pq, \quad v \in V, q \in Q.$$

Testing by $(0, 1)$ yields $p = C^{-1}\lambda^{-1}\|g_\gamma\|_V^2$. Further, testing by $(g_\gamma, 0)$ and using p yields a quadratic equation with roots

$$\lambda = \frac{1 \pm \sqrt{1 + \frac{4\|g_\gamma\|_V^2}{C}}}{2}, \tag{12}$$

which for the considered choices of C become

$$\lambda = \frac{1 \pm \sqrt{1 + 2(1-\gamma^2)}}{2} \quad \text{and} \quad \lambda = \frac{1 \pm \sqrt{5}}{2}.$$

Note that for $C = 1$, i.e. $\mathscr{B}_1$ preconditioner, the spectrum depends on the location of γ. In fact, for γ approaching the boundary the smallest eigenvalue goes to zero. This is consistent with the fact that having $\gamma = \pm 1$ would enforce boundary conditions which are already built into the function space V and thus (8) becomes singular. However, the preconditioner $\mathscr{B}_1$ is never singular. For preconditioner $\mathscr{B}_\gamma$ the spectrum is constant and therefore independent of γ. We remark that the discrete version of $\mathscr{B}_\gamma$ is the Schur complement preconditioner of [46].

Before considering the discrete setting, we shall come back, as promised, to the meaning of the Lagrange multiplier p.

Remark 2 (Interpretation of Lagrange multiplier) *Observe that by integrating by parts*

$$(\nabla u, \nabla v) = \int_{-1}^{\gamma} \nabla u(x)\nabla v(x)\,\mathrm{d}x + \int_{\gamma}^{1} \nabla u(x)\nabla v(x)\,\mathrm{d}x = \{\nabla u(\gamma^-) - \nabla u(\gamma^+)\} + \langle -\Delta u, v\rangle.$$

Upon inserting the above into (8) *and testing the equation with* $(v, 0)$ *we then obtain*

$$\{\nabla u(\gamma^-) - \nabla u(\gamma^+)\}\, v(\gamma) - pv(\gamma) = 0, \quad v \in V, \gamma \in \Omega.$$

The Lagrange multiplier therefore represents a jump of the derivative of u at point γ. Thus, $p = 0$ is necessary for $u \in H^2(\Omega)$. We remark that the relation of p to flux of u is to be expected. In [6] *the use of Lagrange multipliers for enforcing boundary conditions is studied and it is shown that* $p=-\nabla u\cdot n$ *as functionals in* $H^{-\frac{1}{2}}(\partial\Omega)$*. Here* n *is the outer normal of the boundary.*

Finally, the remaining ingredient in the framework of operator preconditioning is a stable finite element discretization of the problem (1). Here we shall see that the spaces $V_h\subset V$ made of continuous linear Lagrange elements are suitable. To simplify the discussion we impose certain restrictions on the discretization of Ω. Namely, the mesh (subsequently referred to as *matching*) shall be such that γ is its vertex.

From the four discrete Brezzi conditions only the inf-sup condition is not immediately evident. To show that it indeed holds, let $u\in V$ be given, and suppose $u_h\in V_h$ is defined as a solution of

$$(\nabla u_h,\nabla v)=(\nabla u,\nabla v)\quad v\in V_h. \tag{13}$$

By Lax-Milgram theorem the problem (13) is well-posed and we have $||u_h||_V\leq ||u||_V$. Further, as the Green's function $g_\gamma\in V$ is piecewise linear with a kink at γ, it is exactly represented in V_h. More precisely, $g_\gamma=I_hg_\gamma$ where $I_h:V\rightarrow V_h$ is the nodal interpolant. Further, setting $v=g_\gamma$ in (13) it follows that

$$0=\Big(\nabla(u-u_h),\nabla g_\gamma\Big)=\langle\delta_\gamma,u-u_h\rangle. \tag{14}$$

Here, the last equality is due to property of the Green's function and $u-u_h\in V$. With mapping $\Pi_h:u\mapsto u_h$ the discrete inf-sup condition follows from the continuous one by the Fortin's criterion, e.g. [10, ch 4.4]

$$\sup_{u_h\in V_h}\frac{q_h\langle\delta_\gamma,u_h\rangle}{||u_h||_V}\geq\sup_{u\in V}\frac{q_h\langle\delta_\gamma,\Pi_hu\rangle}{||\Pi_hu||_V}=\sup_{u\in V}\frac{q_h\langle\delta_\gamma,u\rangle}{||\Pi_hu||_V}\geq\sup_{u\in V}\frac{q_h\langle\delta_\gamma,u\rangle}{||u||_V}\geq\beta_*||q_h||_Q.$$

Here the last inequality is due to the continuous inf-sup condition.

In case δ_γ is not supported by a mesh vertex (*nonmatching* mesh), u_h defined in (13) is not a Fortin projector since $\langle\delta_\gamma,u-u_h\rangle\neq 0$ in general. We mark that the discrete inf-sup condition still holds, but a modified projector is required to show the result. Moreover, convergence of the approximation may be suboptimal. One option to recover optimal convergence is then the extended finite element method, see e.g. [7].

The fact that (14) does not hold on a nonmatching mesh is illustrated in Figure 1 for special case of $u=g_0$. In the figure $\gamma=0$ is contained in a cell with volume h such that the distance from the point to the left edge is ϵ. We observe that the numerical Green's function $g^\epsilon_{0,h}$ defined by (13) matches g_0 everywhere but inside the intersected cell where it cannot resolve the kink of g_0. In turn the condition (14) cannot be met. However, the error $|u(\gamma)-u_h(\gamma)|=|\langle\delta_\gamma,u-u_h\rangle|$ can be controlled.

Remark 3 (Approximation error at γ) *The error of the numerical Green's function on a nonmatching mesh can be obtained by a direct calculation*

$$||g_\gamma-g^\epsilon_{\gamma,h}||^2_V=\frac{\epsilon(h-\epsilon)}{h}. \tag{15}$$

Note that the estimate $||g_\gamma-g^\epsilon_{\gamma,h}||_V\leq\frac{1}{2}\sqrt{h}$ *follows from the equality. Further, recall the property of the Green's function,* $\langle\delta_\gamma,v\rangle=v(\gamma)$, $v\in V$*, and the Galerkin orthogonality of the approximation*

i	0	1	2	3	4	5	6
$\varkappa_1$	3.732	4.442	6.416	10.438	18.462	34.479	66.489
$\varkappa_\gamma$	2.618	2.618	2.618	2.618	2.618	2.618	2.618

Table 1: Spectral condition number $\varkappa = {}^{\lambda_{\max}}/\lambda_{\min}$ of the eigenvalue problem (16) for $\gamma = -1 + 2^{-i}$. As the point approaches the boundary, the condition number of the $\mathscr{B}_1$-preconditioned system grows. With $\mathscr{B}_\gamma$ the condition number is constant at $\varkappa_\gamma = {}^{1+\sqrt{5}}/_{\sqrt{5}-1}$, cf. the eigenvalues (12).

$g^\epsilon_{\gamma,h}$, *that is* $\left(\nabla(g_\gamma - g^\epsilon_{\gamma,h}), \nabla v\right) = 0$, $v \in V_h$. *Then*

$$\begin{aligned}|u(\gamma) - u_h(\gamma)| &= \left(\nabla g_\gamma, \nabla(u - u_h)\right) = \left(\nabla(g_\gamma - g^\epsilon_{\gamma,h}), \nabla(u - u_h)\right)\\ &\leq \|g_\gamma - g^\epsilon_{\gamma,h}\|_V \|u - u_h\|_V \leq \frac{1}{2}\sqrt{h}\|u - u_h\|_V.\end{aligned}$$

In particular, if the continuous and the numerical Green's functions are equal, as is the case with matching mesh, u and u_h are equal at γ. In general, the pointwise error decreases at faster rate than the error measured in the energy norm. We note that the argument above was used in [20], see also [66, ch 3.4], to study pointwise (super-)convergence of finite element solutions.

To verify the results of the analysis we shall perform two numerical experiments. First, (11) is tested by considering a generalized eigenvalue problem

$$\begin{bmatrix}\mathsf{A} & \mathsf{T}^\top\\ \mathsf{T} & \end{bmatrix}\begin{bmatrix}\mathsf{u}\\ \mathsf{p}\end{bmatrix} = \lambda \mathsf{B}_i^{-1}\begin{bmatrix}\mathsf{u}\\ \mathsf{p}\end{bmatrix}, \quad i \in \{1, \gamma\} \tag{16}$$

where $\mathsf{B}_i^{-1} = \operatorname{diag}(\mathsf{A}, C)$ and, as in the continuous case, $C = 1$ or $C = \|g_\gamma\|^2_V$ for $i = 1$ or $i = \gamma$ respectively. Matrices A, T then represent the Galerkin approximations of operators A, B defined in (9) in the basis of V_h and Q_h.

Table 1 confirms the theoretical conclusions. In particular, the Schur complement preconditioner B_γ leads to a constant condition number [7] regardless of the position of the constrained point γ. As predicted, for B_1 the condition number increases as γ approaches the boundary.

In the second experiment the approximation properties and the convergence of the iterative method will be of interest. To this end (8) is considered with $f = 25\pi^2 \cos\frac{5\pi}{2}x$, $h = 2$, $\gamma = 0$ and discretized with (stable) continuous linear Lagrange elements using both matching and nonmatching meshes. The related preconditioned linear system

$$\tilde{\mathsf{B}}_1\begin{bmatrix}\mathsf{A} & \mathsf{T}^\top\\ \mathsf{T} & \end{bmatrix}\begin{bmatrix}\mathsf{u}\\ \mathsf{p}\end{bmatrix} = \tilde{\mathsf{B}}_1\begin{bmatrix}\mathsf{f}\\ \mathsf{h}\end{bmatrix}$$

is then solved with MINRES [8]. Here $\tilde{\mathsf{B}}_1$ is a spectrally equivalent and cost efficient approximation of the preconditioner B_1. The operator used a single sweep of algebraic multigrid to approximate A^{-1}.

The error convergence of the proposed discretization is shown in Figure 2. Using the matching meshes, linear and quadratic convergence is observed respectively for u_h and p_h.

[7] For each γ, the listed condition number is computed on a series of refined meshes until $|\varkappa_h - \varkappa_{2h}| < 10^{-8}$, where $\varkappa_h$ is the element size h. Then we set $\varkappa = \varkappa_h$.

[8] Random initial vector was used to start the iterations. The stopping criterion for the norm of the preconditioned residual was set to 10^{-13}.

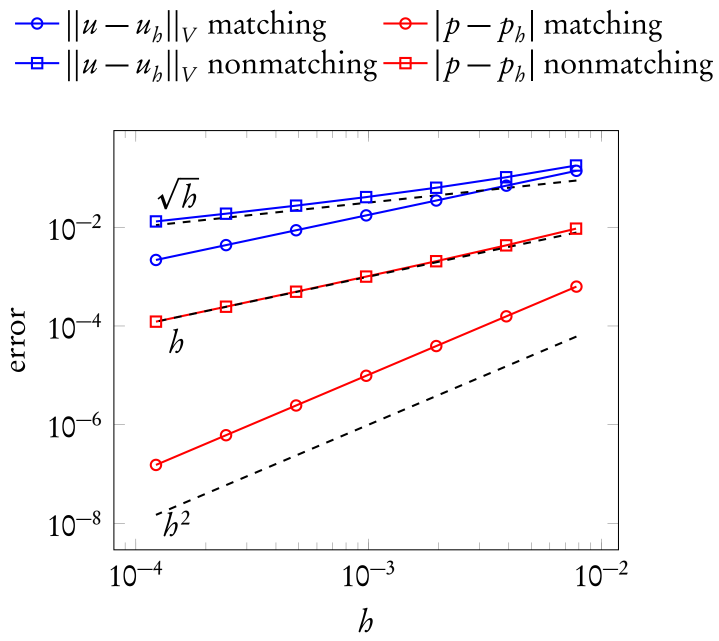

Figure 2: Error convergence of the finite element method for (8) using piecewise linear continuous Lagrange elements. On a matching discretization the convergence is optimal. Nonmatching discretization with constrained point inside an element yields suboptimal convergence.

With nonmatching meshes the orders are halved, cf. Remark 3. Regarding the efficiency of the constructed preconditioner, we note that approximately four MINRES iterations were required regardless of the mesh size h. Moreover, the number of iterations was not effected by varying γ. This observation is to be expected as $\dim Q_h = 1$ and thus the spectrum is dominated by the Poisson problem for which multigrid is an optimal preconditioner, see [68, 4].

Having successfully found an efficient preconditioner for the coupled $1d$-$0d$ problem, we shall comment on some of the upcoming challenges which cannot be appreciated on the simple example. First, the multiplier space here, $Q=\mathbb{R}$, is deceivingly simple; the Lagrange multiplier is a single number. In the $2d$-$1d$ and $3d$-$1d$ coupled problems studied in the papers Q becomes an infinite dimensional Hilbert space. Further, due to functions $H^1_0(\Omega)$, $\Omega=(-1,1)$, being continuous, the range of the trace operator here is simply $\mathbb{R}$. Characterizing the trace space in the later problems is far more involved. Finally, unlike here, there will be a PDE posed on the lower dimensional manifold.

Let us also briefly comment on the role of the abstract framework of operator preconditioning for the singular problems such as the Neumann problem of linear elasticity. A possible approach to transform a singular system into a well-posed one is to require orthogonality of the solution and the singular modes. This constraint is enforced by Lagrange multipliers and the new problem then takes the form of the saddle-point system (1). At this point the abstract framework can be applied to establish well-posedness and obtain a preconditioner. A simple example which could be used for illustration here is the singular Poisson problem with a one dimensional nullspace of constant functions. However, the problem is sufficiently well-known, see e.g. [8], and we therefore choose not to include it here.

We conclude the introduction by providing a summary of the papers that form the main body of the thesis.

Summary of the Papers

Paper I 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 Journal on Scientific Computing, 38 (2016), pp. B962–B987.

Paper II 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,
SIAM Journal on Scientific Computing, (2016). Submitted.

Paper III M. KUCHTA, K.-A. MARDAL, AND M. MORTENSEN,
Characterisation of the space of rigid motions in arbitrary domains,
in Proc. of 8th National Conference on Computational Mechanics, Barcelona, Spain, 2015, CIMNE.

Paper IV M. KUCHTA, K.-A. MARDAL, AND M. MORTENSEN,
On the singular Neumann problem in linear elasticity,
Numerical Linear Algebra with Applications, (2016). Submitted.

Paper I

The paper is concerned with preconditioning of a model coupled $2d$-$1d$ problem

$$\begin{aligned} -\Delta_\Omega w + \epsilon\delta_\Gamma p &= f \quad \text{in } \Omega,\\ -\Delta_\Gamma v - p &= g \quad \text{on } \Gamma,\\ \epsilon T_\Gamma w - v &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{17}$$

Here $\Omega \subset \mathbb{R}^2$ while $\Gamma \subset \Omega$ is a curve such that $\partial\Gamma \in \partial\Omega$. The system is considered with homogeneous Dirichlet boundary conditions. Recasting (17) as a saddle point system, two different preconditioners

$$\mathscr{B}_Q = \begin{bmatrix} -\Delta_\Omega & & \\ & -\Delta_\Gamma & \\ & & -\epsilon^2\Delta_\Gamma^{-\frac{1}{2}} - \Delta_\Gamma^{-1} \end{bmatrix}^{-1}, \mathscr{B}_W = \begin{bmatrix} -\Delta_\Omega + T'_\Gamma(-\epsilon^2\Delta_\Gamma)T_\Gamma & & \\ & -\Delta_\Gamma & \\ & & -\Delta_\Gamma^{-1} \end{bmatrix}^{-1}$$

are derived within the framework of operator preconditioning [44]. To this end existence and uniqueness of the weak solution are shown using the Brezzi theory [12] with the solution $(u, v, p) \in W \times V \times Q$ where the spaces are respectively

$$H_0^1(\Omega) \times H_0^1(\Gamma) \times \left(\epsilon H^{-\frac{1}{2}}(\Gamma) \cap H^{-1}(\Gamma)\right) \text{ and } \left(H_0^1(\Omega) \cap \epsilon H_0^1(\Gamma)\right) \times H_0^1(\Gamma) \times H^{-1}(\Gamma).$$

Consequently, finite element discretization of the problem is discussed. The discrete inf-sup condition is shown to be satisfied with the discrete spaces from continuous linear Lagrange elements and meshes of Γ and Ω such that the cells of the former are edges in the latter mesh (matching meshes). The discrete approximation of $\mathscr{B}_W$ then uses standard preconditioners (multigrid and LU factorization) for the individual blocks. In case of $\mathscr{B}_Q$ the approximation of the fractional Laplacian is constructed by spectral decomposition. Finally, numerical experiments are presented which show robustness of both preconditioners with respect to ϵ and the discretization parameter. Moreover, computational efficiency is assessed. For the considered examples, the costs of both preconditioners are similar, however, the generalized eigenvalue problem used to approximate $\mathscr{B}_Q$ can become a burden for large meshes of Γ. To address the issue mass lumping is shown to work as a simple trick that reduces the computational cost.

The main novelty and contribution of the paper is the mathematical analysis of the two proposed preconditioners. Altogether a complete and efficient algorithm for obtaining the numerical solution of (17) is presented.

In the future, it would be interesting to apply the proposed ideas to real-life applications. To this end the results might have to be extended to different equations, e.g. advection-diffusion or Stokes flow. Further, if the preconditioner $\mathscr{B}_Q$ (or similar operator using mappings in the fractional Sobolev spaces) is to be applied in large scale computations, e.g. $3d$-$2d$ coupling, a more efficient realization of the approximation needs to be used. Here, domain embedding preconditioners [59, 30] or Lanczos iterations as presented e.g. in [2, 3] are promising methods. Finally, a natural extension of this work is to consider preconditioning of $3d$-$1d$ coupled problems.

Paper II

This paper extends Paper I in several ways. In particular coupling between the domains $\Omega \subset \mathbb{R}^3$ and $\Gamma \subset \Omega$ a manifold of codimension two is discussed. Moreover, finite element discretization with nonmatching (cf. summary of Paper I) triangulations is considered. Finally, in addition to piecewise linear continuous Lagrange elements, the Lagrange multiplier space is also approximated by piecewise constant elements. The considered example of a $3d$-$1d$ problem is

$$\begin{aligned} -\Delta_\Omega w + w + \delta_\Gamma p &= f \quad \text{in } \Omega,\\ -\Delta_\Gamma v + v - p &= g \quad \text{on } \Gamma,\\ T_\Gamma w - v &= 0 \quad \text{on } \Gamma, \end{aligned} \tag{18}$$

equipped (for the ease of implementation) with homogeneous Neumann boundary conditions.

The main contribution of the paper is establishing a robust preconditioner for the model problem (18). The structure of the discrete preconditioner is motivated by ideas of Paper I

$$\mathscr{B}_h = \begin{bmatrix} -\Delta_{\Omega,h} + I_{\Omega,h} & & \\ & -\Delta_\Gamma + I_{\Gamma,h} & \\ & & -\Delta^s_{\Gamma,h} - \Delta^{-1}_{\Gamma,h} \end{bmatrix}^{-1}.$$

Here, however, the exponent of the fractional Laplacian is not derived from theory since establishing a well-defined continuous trace operator T_Γ requires higher than H^1 regularity. The discrete trace operator, on the other hand, is well defined as the considered finite element functions are continuous. The focus is therefore on numerical experiments which identify the exponent s that leads to stable numerical behavior. Using a simplified $3d$-$1d$ coupled problem a range of such exponents is identified in interval $(-0.2, -0.1)$. Taking $s = -0.14$, it is demonstrated that $\mathscr{B}_h$ is a suitable preconditioner for (18). The preconditioner uses standard multigrid for the two leading blocks while spectral decomposition is used for the trailing block. The preconditioner is further shown to work with nonmatching triangulations and the multiplier space constructed from continuous piecewise linear and discontinuous piecewise constant Lagrange elements. For inf-sup stability of both discretizations, the restriction on the mesh sizes of Ω, Γ inspired by $2d$-$1d$ problems, see e.g. [16, 54], is shown to be crucial.

A weakness of the presented work is the missing theoretical foundation as the preconditioner was established by reasoning about the discrete rather than the continuous problem. Putting the preconditioner on a proper mathematical footing is therefore a natural direction for the future work. We would further like to apply the preconditioner to practical applications in biomechanics. A challenge here might be the nature of domain Γ, which in the applications often resembles a space-filling curve, cf. [41, 22, 57].

Paper III

This conference paper compares two approaches for solving the singular Neumann problem of linear elasticity

$$\begin{aligned} -\nabla\cdot\sigma(u) &= f && \text{in } \Omega, \\ \sigma(u) &= 2\mu\epsilon(u)+\lambda(\nabla\cdot u)I && \text{in } \Omega, \\ \sigma(u)\cdot n &= h && \text{on } \partial\Omega. \end{aligned} \tag{19}$$

In the first case, the kernel of rigid motions is handled on a discrete level by solving the positive semi-definite linear system by (preconditioned) conjugate gradient method with a nullspace of the discrete operator passed to the solver. Alternatively, the singularity is dealt with on a continuous level and the orthonormal basis of the space of rigid motions is used to formulate a variational problem which upon discretization yields a positive definite linear system. In both cases finite element discretization is used. It is shown by numerical experiments that the solutions due to the first method may be wrong approximations in the H^1 norm if the mesh of Ω is nonuniform (or anisotropically refined). Similar observation was made in [8]. The second approach gives optimal convergence even in this case.

The novelty of this paper is the construction of the orthonormal basis of the space of rigid motions based on a tensorial quantity which describes rotational energy of the body with respect to its center of mass c. The tensor I with components

$$I_{ij}=\int_{\Omega}(x-c)\cdot(x-c)\delta_{ij}+(x-c)_i(x-c)_j\mathrm{d}x$$

is known in classical mechanics as inertia tensor of Ω, e.g. [27], however, its application in numerical methods for (19) is new.

As the focus of the paper is mainly on handling the rigid motions and numerical experiments, several aspects of the methods are omitted. Most notably, a thorough discussion of the preconditioning, in particular robustness with respect to Lamé constants, is missing. These aspects are left for future work.

Paper IV

The final paper discusses well-posedness of different variational formulations of the singular problem of linear elasticity with an aim to derive parameter robust preconditioners for the resulting equations. The formulations from Paper III are included and in this sense Paper IV is a completion and extension of the previous work. With $\mu\geq\lambda$ robust preconditioners are established for Lagrange multiplier formulation of (19) and the two formulations from Paper III. To derive robust preconditioner in case $\lambda\gg\mu$ a new variable, solid pressure $p=\lambda\nabla\cdot u$, is introduced leading to a singular system

$$\begin{aligned} \nabla\cdot(2\mu\epsilon(u))-\nabla p=f &\quad \text{in } \Omega,\\ \lambda\nabla\cdot u-p=0 &\quad \text{in } \Omega,\\ \sigma(u)\cdot n=h &\quad \text{on } \partial\Omega. \end{aligned} \tag{20}$$

Here the established preconditioner is similar to [33, 34].

The main contribution of the paper is that it presents, in a systematic way, the different approaches to solving singular systems with a finite dimensional kernel of which the Neumann problem in linear elasticity is an example.

Among the future extensions of this work is extending the ideas to different material models, e.g. poroelasticity or hyperelasticity. Another fruitful direction are the practical applications.

Additional Works

In addition to the summarized articles, three other papers have been written during the course of the project. Their brief summary is provided below.

Paper [45] presents a simple Python library for computing deformation of a plate-beam system governed by

$$\left.\begin{aligned} &\int_\Omega |\Delta U|^2(x)\,\mathrm{d}x-2\int_\Omega U(x)F(x)\,\mathrm{d}x\\ &\sum_{i=1}^{n}\int_{\Gamma_i}|\Delta u_i|^2(t)\,\mathrm{d}t-2\int_{\Gamma_i}u_i(t)f_i(t)\,\mathrm{d}t \end{aligned}\right\}\rightarrow \min$$

subject to n constraints

$$\epsilon T_{\Gamma_i}U-u_i=0.$$

Here U, u_i and F, f_i are respectively the deflections and loads of the plate occupying the domain $\Omega\subset\mathbb{R}^2$ and the beams $\Gamma_i\subset\Omega$ which are one dimensional curves. The above system is an example of a problem where equations on domains with different dimensionality are coupled. In fact, this problem motivated the later Papers I and II. The focus here is on the implementation which uses the Galerkin method with eigenfunctions of the related biharmonic operator as the basis functions and utilizes symbolic computations (Sympy [67]) to assemble the discrete problem. The paper also features a discussion of the conditioning of the assembled linear system for the case $n=1$ using H^{-2} norm as the preconditioner for the Schur complement. The preconditioner improved markedly the condition number of the system, however, in the light of the findings from Paper I, the preconditioner was not optimal and ϵ robust. In particular, the proposed norm ignores the intersection structure of the multiplier space and is therefore a good approximation only for the case $\epsilon\ll 1$.

In [28] the drug delivery via injection into the spinal chord is studied by simulating the flow of cerebrospinal fluid and the therapeutic agent using the Navier-Stokes equations coupled to the equation for advection of a passive scalar. Following [32] the algorithm used for the transport problem was implemented for FEniCS [42, 1] by Mikael Mortensen and the author. In the flow regimes dominated by convection, the method based on Lagrangian particle tracking showed more robust properties than the techniques used with the finite element discretization (and the Eulerian description of the flow) such as SUPG [13, 9] or stabilized discontinuous Galerkin method [19, ch 2., 3.]. The main outcome of the study are indications of injection locations and angles which lead to faster spread of the drug.

Finally, [36] is concerned with simulations of the free surface Stokes flow with high density and viscosity ratios for the purpose of investigating the evolution history of Saturn's moon Iapetus. In particular, the moon's large flattening and its connection with giant impact craters observed on the surface are of interest. The paper is an extension of the author's master thesis work where the foundations of the numerical code were laid. The study concludes that a collision with an external body is a plausible explanation for the current shape.

Bibliography

[1] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. Rognes, and G. Wells, *The FEniCS project version 1.5*, Archive of Numerical Software, 3 (2015).

[2] M. Arioli, D. Kourounis, and D. Loghin, *Discrete fractional Sobolev norms for domain decomposition preconditioning*, IMA Journal of Numerical Analysis, (2012), p. drr024.

[3] M. Arioli and D. Loghin, *Discrete interpolation norms with applications*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 2924–2951.

[4] S. F. Ashby and R. D. Falgout, *A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations*, Nuclear Science and Engineering, 124 (1996), pp. 145–159.

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

[6] ——, *The finite element method with Lagrangian multipliers*, Numerische Mathematik, 20 (1973), pp. 179–192.

[7] T. Belytschko, R. Gracie, and G. Ventura, *A review of extended/generalized finite element methods for material modeling*, Modelling and Simulation in Materials Science and Engineering, 17 (2009), p. 043001.

[8] P. Bochev and R. B. Lehoucq, *On the finite element solution of the pure Neumann problem*, SIAM review, 47 (2005), pp. 50–66.

[9] P. B. Bochev, M. D. Gunzburger, and J. N. Shadid, *Stability of the SUPG finite element method for transient advection–diffusion problems*, Computer methods in applied mechanics and engineering, 193 (2004), pp. 2301–2323.

[10] D. Braess, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, 2001.

[11] S. Brenner and R. Scott, *The Mathematical Theory of Finite Element Methods*, Texts in Applied Mathematics, Springer New York, 2007.

[12] F. Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique, 8 (1974), pp. 129–151.

[13] A. N. Brooks and T. J. R. Hughes, *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Computer Methods in Applied Mechanics and Engineering, 32 (1982), pp. 199 – 259.

[14] L. Cattaneo and P. Zunino, *A computational model of drug delivery through microcirculation to compare different tumor treatments*, International Journal for Numerical Methods in Biomedical Engineering, 30 (2014), pp. 1347–1371.

[15] ——, *Computational models for fluid exchange between microcirculation and tissue interstitium*, Networks and Heterogeneous Media, 9 (2014), pp. 135–159.

[16] W. Dahmen and A. Kunoth, *Appending boundary conditions by Lagrange multipliers: Analysis of the LBB condition*, Numerische Mathematik, 88 (2001), pp. 9–42.

[17] D. Daners, *A short elementary proof of* $\Sigma 1/k^2 = \pi^2/6$, Mathematics Magazine, 85 (2012), pp. 361–364.

[18] L. Demkowicz, *Babuška⟷Brezzi*.

[19] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Mathématiques et Applications, Springer Berlin Heidelberg, 2011.

[20] J. Douglas and T. Dupont, *Superconvergence for Galerkin methods for the two point boundary problem via local projections*, Numerische Mathematik, 21, pp. 270–278.

[21] T. Dutta-Roy, A. Wittek, and K. Miller, *Biomechanical modelling of normal pressure hydrocephalus*, Journal of Biomechanics, 41 (2008), pp. 2263 – 2271.

[22] Q. Fang, S. Sakadžić, L. Ruvinskaya, A. Devor, A. M. Dale, and D. A. Boas, *Oxygen advection and diffusion in a three-dimensional vascular anatomical network*, Optics express, 16 (2008), pp. 17530–17541.

[23] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics, Vol. I: The New Millennium Edition: Mainly Mechanics, Radiation, and Heat*, Basic Books, 2015.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Matrix Computations, Johns Hopkins University Press, 2012.

[25] L. Grinberg, E. Cheever, T. Anor, J. R. Madsen, and G. E. Karniadakis, *Modeling blood flow circulation in intracranial arterial networks: a comparative 3D/1D simulation study*, Annals of biomedical engineering, 39 (2011), pp. 297–309.

[26] 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*, Electronic Transactions on Numerical Analysis, 41 (2014), pp. 13–20.

[27] M. E. Gurtin, *An Introduction to Continuum Mechanics*, Mathematics in Science and Engineering, Elsevier Science, 1982.

[28] P. T. Haga, G. Pizzichelli, M. Mortensen, M. Kuchta, S. H. Pahlavian, E. Sinibaldi, B. Martin, and K.-A. Mardal, *A numerical investigation of intrathecal drug and gene vector dispersion within the cervical subarachnoid space*, PLOS ONE, (2016). Submitted.

[29] A. Halevy, P. Norvig, and F. Pereira, *The unreasonable effectiveness of data*, IEEE Intelligent Systems, 24 (2009), pp. 8–12.

[30] E. Haug and R. Winther, *A domain embedding preconditioner for the Lagrange multiplier system*, Mathematics of Computation of the American Mathematical Society, 69 (2000), pp. 65–82.

[31] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, vol. 49, 1952.

[32] S. Ianniello and A. Di Mascio, *A self-adaptive oriented particles level-set method for tracking interfaces*, J. Comput. Phys., 229 (2010), pp. 1353–1380.

[33] A. Klawonn, *Block-triangular preconditioners for saddle point problems with a penalty term*, SIAM Journal on Scientific Computing, 19 (1998), pp. 172–184.

[34] ——, *An optimal preconditioner for a class of saddle point problems with a penalty term*, SIAM Journal on Scientific Computing, 19 (1998), pp. 540–552.

[35] W. Krendl, V. Simoncini, and W. Zulehner, *Stability estimates and structural spectral properties of saddle point problems*, Numerische Mathematik, 124 (2013), pp. 183–213.

[36] M. Kuchta, G. Tobie, K. Miljković, M. Běhounková, O. Souček, G. Choblet, and O. Čadek, *Despinning and shape evolution of Saturn's moon Iapetus triggered by a giant impact*, Icarus, 252 (2015), pp. 454–465.

[37] C. Lanczos, *Linear Differential Operators*, Dover books on mathematics, Dover Publications, 1997.

[38] R. B. Larson, *Models for the formation of elliptical galaxies*, Monthly Notices of the Royal Astronomical Society, 173 (1975), pp. 671–699.

[39] X. S. Li and J. W. Demmel, *Making sparse Gaussian elimination scalable by static pivoting*, in Proceedings of the 1998 ACM/IEEE conference on Supercomputing, IEEE Computer Society, 1998, pp. 1–17.

[40] ——, *SuperLU_DIST: A scalable distributed-memory sparse direct solver for unsymmetric linear systems*, ACM Transactions on Mathematical Software (TOMS), 29 (2003), pp. 110–140.

[41] A. A. Linninger, I. G. Gould, T. Marinnan, C.-Y. Hsu, M. Chojecki, and A. Alaraj, *Cerebral microcirculation and oxygen tension in the human secondary cortex*, Annals of biomedical engineering, 41 (2013), pp. 2264–2284.

[42] A. Logg, K.-A. Mardal, and G. Wells, *Automated solution of differential equations by the finite element method: The FEniCS book*, vol. 84, Springer Science & Business Media, 2012.

[43] J. Málek and Z. Strakoš, *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

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

[45] M. Mortensen, M. Kuchta, J.C.G. Verschaeve, and K.-A. Mardal, *BENDPY: Python framework for computing bending of complex plate-beam systems*, in Proc. of 8th National Conference on Computational Mechanics, Barcelona, Spain, 2015, CIMNE.

[46] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 1969–1972.

[47] M. Nabil and P. Zunino, *A computational study of cancer hyperthermia based on vascular magnetic nanoconstructs*, Open Science, 3 (2016).

[48] J. Nečas, *Sur une méthode pour résoudre les équations aux dérivées partielles du type elliptique, voisine de la variationnelle*, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, 16 (1962), pp. 305–326.

[49] M. Ochs, J. R. Nyengaard, A. Jung, L. Knudsen, M. Voigt, T. Wahlers, J. Richter, and H. J. G. Gundersen, *The number of alveoli in the human lung*, American Journal of Respiratory and Critical Care Medicine, 169 (2004), pp. 120–124.

[50] SIAM Working Group on CSE Education, *Graduate education in computational science and engineering*, SIAM Review, 43 (2001), pp. 163–177.

[51] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM Journal on Numerical Analysis, 12 (1975), pp. 617–629.

[52] C. S. Peskin, *Numerical analysis of blood flow in the heart*, Journal of Computational Physics, 25 (1977), pp. 220 – 252.

[53] J. Pestana and A. J. Wathen, *Natural preconditioning and iterative methods for saddle point systems*, SIAM Review, 57 (2015), pp. 71–91.

[54] J. Pitkäranta, *Boundary subspaces for the finite element method with Lagrange multipliers*, Numerische Mathematik, 33 (1979), pp. 273–289.

[55] R. F. POTTER AND A. C. GROOM, *Capillary diameter and geometry in cardiac and skeletal muscle studied by means of corrosion casts*, Microvascular research, 25 (1983), pp. 68–84.

[56] A. QUARTERONI, *Numerical models for differential problems*, Springer Science & Business Media, 2010.

[57] J. REICHOLD, M. STAMPANONI, A. L. KELLER, A. BUCK, P. JENNY, AND B. WEBER, *Vascular graph model to simulate the cerebral blood flow in realistic vascular networks*, Journal of Cerebral Blood Flow & Metabolism, 29 (2009), pp. 1429–1443.

[58] Y. RICARD AND C. VIGNY, *Mantle dynamics with induced plate tectonics*, Journal of Geophysical Research: Solid Earth, 94 (1989), pp. 17543–17559.

[59] T. RUSTEN, P. A. VASSILEVSKI, AND R. WINTHER, *Domain embedding preconditioners for mixed systems*, Numerical Lin. Alg. with Applic., 5 (1998), pp. 321–345.

[60] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 856–869.

[61] M. B. SANDERUD, *Patient-specific modeling of normal pressure hydrocephalus*, master's thesis, University of Oslo, 2012.

[62] K. SCOTT, *On Proebsting's law*, tech. report, 2001.

[63] J. R. SHEWCHUK, *An introduction to the conjugate gradient method without the agonizing pain*, tech. report, Pittsburgh, PA, USA, 1994.

[64] D. SILVESTER, H. ELMAN, D. KAY, AND A. WATHEN, *Efficient preconditioning of the linearized Navier–Stokes equations for incompressible flow*, Journal of Computational and Applied Mathematics, 128 (2001), pp. 261–279.

[65] D. SILVESTER AND A. WATHEN, *Fast iterative solution of stabilised Stokes systems. Part II: Using general block preconditioners*, SIAM Journal on Numerical Analysis, 31 (1994), pp. 1352–1367.

[66] G. STRANG AND G. J. FIX, *An analysis of the finite element method*, Wellesley-Cambridge Press, 1988.

[67] SYMPY DEVELOPMENT TEAM, *SymPy: Python library for symbolic mathematics*, 2016.

[68] O. TATEBE, *The multigrid preconditioned conjugate gradient method*, (1993).

[69] G. TOBIE, O. ČADEK, AND C. SOTIN, *Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus*, Icarus, 196 (2008), pp. 642 – 652. Mars Polar Science IV.

[70] L. N. TREFETHEN AND D. BAU, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, 1997.

[71] A. WATHEN AND D. SILVESTER, *Fast iterative solution of stabilised Stokes systems. Part I: Using simple diagonal preconditioners*, SIAM Journal on Numerical Analysis, 30 (1993), pp. 630–649.

[72] E. P. WIGNER, *The unreasonable effectiveness of mathematics in the natural sciences. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959*, Communications on pure and applied mathematics, 13 (1960), pp. 1–14.

Paper I

Preconditioners for saddle point systems with trace constraints coupling 2d and 1d domains

M. Kuchta, M. Nordaas, J. C. G. Verschaeve, M. Mortensen, and K.-A. Mardal

SIAM Journal on Scientific Computing, 38 (2016), pp. B962–B987.

SIAM J. SCI. COMPUT.
Vol. 38, No. 6, pp. B962–B987

PRECONDITIONERS FOR SADDLE POINT SYSTEMS WITH TRACE CONSTRAINTS COUPLING 2D AND 1D DOMAINS[*]

MIROSLAV KUCHTA[†], MAGNE NORDAAS[‡], JORIS C. G. VERSCHAEVE[†], MIKAEL MORTENSEN[†‡], AND KENT-ANDRE MARDAL[†‡]

Abstract. We study preconditioners for a model problem describing the coupling of two elliptic subproblems posed over domains with different topological dimension by a parameter dependent constraint. A pair of parameter robust and efficient preconditioners is proposed and analyzed. Robustness and efficiency of the preconditioners is demonstrated by numerical experiments.

Key words. preconditioning, saddle-point problem, Lagrange multipliers

AMS subject classification. 65F08

DOI. 10.1137/15M1052822

1. Introduction.

This paper is concerned with preconditioning of multiphysics problems where two subproblems of different dimensionality are coupled. We assume that Γ is a submanifold contained within $\Omega \in \mathbb{R}^n$ and consider the following problem:

$$-\Delta u + \epsilon \delta_\Gamma p = f \quad \text{in } \Omega, \tag{1a}$$

$$-\Delta v - p = g \quad \text{on } \Gamma, \tag{1b}$$

$$\epsilon u - v = 0 \quad \text{on } \Gamma, \tag{1c}$$

where δ_Γ is a function with properties similar to the Dirac delta function, as will be discussed later. To allow for a unique solution (u, v, p), the system must be equipped with suitable boundary conditions, and we shall here, for simplicity, consider homogeneous Dirichlet boundary conditions for u and v on $\partial\Omega$ and $\partial\Gamma$, respectively. We note that the unknowns u, v are here the primary variables, while the unknown p should be interpreted as a Lagrange multiplier associated with the constraint (1c).

The two elliptic equations that are stated on two different domains, Ω and Γ, are coupled, and therefore the restriction of u to Γ and the extension of p to Ω are crucial. When the codimension of Γ is one, the restriction operator is a trace operator, and the extension operator is similar to the Dirac delta function. We note that $\epsilon \in (0, 1)$ and that the typical scenario will be that $\epsilon \ll 1$. We will therefore focus on methods that are robust in ϵ.

The problem (1a)–(1c) is relevant to biomedical applications [18, 15, 2, 17] where it models the coupling of the porous media flow inside tissue to the vascular bed through Starling's law. Further, problems involving coupling of the finite element method

[*] Submitted to the journal's Computational Methods in Science and Engineering section December 14, 2015; accepted for publication (in revised form) August 24, 2016; published electronically November 1, 2016.

http://www.siam.org/journals/sisc/38-6/M105282.html

Funding: The work of Magne Nordaas, Mikael Mortensen, and Kent-Andre Mardal has been supported by Center of Excellence grant 179578 from the Research Council of Norway to the Center for Biomedical Computing at Simula Research Laboratory. The work of Kent-Andre Mardal has also been supported by the Research Council of Norway through grant 209951.

[†] Department of Mathematics, Division of Mechanics, University of Oslo, Oslo, Norway (mirok@math.uio.no, joris@math.uio.no, mikaem@math.uio.no).

[‡] Center for Biomedical Computing, Simula Research Laboratory, 1325 Lysaker, Norway (magneano@simula.no, kent-and@simula.no).

and the boundary element method, e.g., [24, 26], are of the form (1). The system is also relevant for domain decomposition methods based on Lagrange multipliers [32]. Finally, in solid mechanics, the problem of plates reinforced with ribs (cf., for example, [44, Ch. 9.11]) can be recast into a related fourth order problem. We also note that the techniques developed here to address the constraint (1c) are applicable in preconditioning fluid-structure interaction problems involving interactions with thin structures, e.g., filaments [22].

One way of deriving equations (1) is to consider the following minimization problem:

$$\left.\begin{aligned} &\int_\Omega (\nabla u)^2 - 2uf \,\mathrm{d}x \\ &\int_\Gamma (\nabla v)^2 - 2vg \,\mathrm{d}s \end{aligned}\right\} \to \min \tag{2}$$

subject to the constraint

$$\epsilon u - v = 0 \quad \text{on } \Gamma. \tag{3}$$

Using the method of Lagrange multipliers, the constrained minimization problem will be recast as a saddle-point problem. The saddle-point problem is then analyzed in terms of the Brezzi conditions [13], and efficient solution algorithms are obtained using operator preconditioning [35]. A main challenge is the fact that the constraint (3) necessitates the use of trace operators, which leads to operators in fractional Sobolev spaces on Γ.

An outline of the paper is as follows: Section 2 presents the necessary notation and mathematical framework needed for the analysis. Then the mathemathical analysis as well as the numerical experiments of two different preconditioners are presented in sections 3 and 4, respectively. Section 5 discusses the computational efficiency of both methods.

2. Preliminaries. Let X be a Hilbert space of functions defined on a domain D, and let $\|\cdot\|_X$ denote its norm. The L^2 inner product on a domain D is denoted $(\cdot,\cdot)_D$ or $\int_D \cdot$, while $\langle\cdot,\cdot\rangle_D$ denotes the corresponding duality pairing between a Hilbert space X and its dual space X^*. We will use $H^m = H^m(D)$ to denote the Sobolev space of functions on D with m derivatives in $L^2 = L^2(D)$. The corresponding norm is denoted $\|\cdot\|_{m,D}$. In general, we will use H_0^m to denote the closure in H^m of the space of smooth functions with compact support in D, and the seminorm is denoted as $|\cdot|_{m,D}$.

The space of bounded linear operators mapping elements of X to Y is denoted $\mathcal{L}(X,Y)$, and if $Y = X$, we simply write $\mathcal{L}(X)$ instead of $\mathcal{L}(X,X)$. If X and Y are Hilbert spaces, both continuously contained in some larger Hilbert space, then the intersection $X \cap Y$ and the sum $X + Y$ are both Hilbert spaces with norms given by

$$\|x\|^2_{X\cap Y} = \|x\|^2_X + \|x\|^2_Y \quad \text{and} \quad \|z\|^2_{X+Y} = \inf_{\substack{x\in X, y\in Y\\ z=x+y}} (\|x\|^2_X + \|y\|^2_Y).$$

In the following $\Omega \subset \mathbb{R}^n$ is an open connected domain with Lipschitz boundary $\partial\Omega$. The trace operator T is defined by $Tu = u|_\Gamma$ for $u \in C(\overline{\Omega})$ and Γ a Lipschitz submanifold of codimension one in Ω. The trace operator extends to bounded and surjective linear operator $T : H^1(\Omega) \to H^{\frac{1}{2}}(\Gamma)$; see, e.g., [1, Ch. 7]. The fractional Sobolev space $H^{\frac{1}{2}}(\Gamma)$ can be equipped with the norm

$$\|u\|^2_{H^{\frac{1}{2}}(\Gamma)} = \|u\|^2_{L^2(\Gamma)} + \int_{\Gamma\times\Gamma} \frac{|u(x)-u(y)|^2}{|x-y|^{n+1}} \,\mathrm{d}x\mathrm{d}y. \tag{4}$$

However, the trace is not surjective as an operator from $H_0^1(\Omega)$ into $H^{\frac{1}{2}}(\Gamma)$; in particular, the constant function $1 \in H^{\frac{1}{2}}(\Gamma)$ is not in the image of the trace operator. Note that $H_0^{\frac{1}{2}}(\Gamma)$ does not characterize the trace space, since $H_0^{\frac{1}{2}}(\Gamma) = H^{\frac{1}{2}}(\Gamma)$; see [30, Ch. 2, Thm. 11.1]. Instead, the trace space can be identified as $H_{00}^{\frac{1}{2}}(\Gamma)$, defined as the subspace of $H^{\frac{1}{2}}(\Gamma)$ for which extension by zero into $H^{\frac{1}{2}}(\tilde{\Gamma})$ is continuous, for some suitable extension domain $\tilde{\Gamma}$ extending Γ (e.g., $\tilde{\Gamma} = \Gamma \cup \partial\Omega$). To be precise, the space $H_{00}^{\frac{1}{2}}(\Gamma)$ can be characterized with the norm

$$\|u\|_{H_{00}^{\frac{1}{2}}(\Gamma)} = \|\tilde{u}\|_{H^{\frac{1}{2}}(\tilde{\Gamma})}, \quad \tilde{u}(x) = \begin{cases} u(x), & x \in \Gamma, \\ 0, & x \notin \Gamma. \end{cases} \tag{5}$$

The space $H_{00}^{\frac{1}{2}}(\Gamma)$ does not depend on the extension domain $\tilde{\Gamma}$, since the norms induced by different choices of $\tilde{\Gamma}$ will be equivalent.

The above norms (4)–(5) for the fractional spaces are impractical from an implementation point of view, and we will therefore consider the alternative construction following [30, Ch. 2.1] and [16]. For $u, v \in H_0^1(\Gamma)$, set $L_u(v) = (u, v)_\Gamma$. Then L_u is a bounded linear functional on $H_0^1(\Gamma)$, and in accordance with the Riesz–Fréchet theorem there is an operator $S \in \mathcal{L}\big(H_0^1(\Gamma)\big)$ such that

$$(Su, w)_{H_0^1(\Omega)} = L_u(w) = (u, w)_\Gamma\,, \qquad u, w \in H_0^1(\Gamma). \tag{6}$$

The operator S is self-adjoint, positive definite, injective, and compact. Therefore, the spectrum of S consists of a nonincreasing sequence of positive eigenvalues $\{\lambda_k\}_{k=1}^\infty$ such that $0 < \lambda_{k+1} \leq \lambda_k$ and $\lambda_k \to 0$; see, e.g., [48, Ch. X.5, Thm. 2]. The eigenvectors $\{\phi_k\}_{k=1}^\infty$ of S satisfy the generalized eigenvalue problem

$$A\phi_k = \lambda_k^{-1} M \phi_k,$$

where operators A, M are such that $\langle Au, v\rangle_\Gamma = (\nabla u, \nabla v)_\Gamma$ and $\langle Mu, v\rangle_\Gamma = (u, v)_\Gamma$. The set of eigenvectors $\{\phi_k\}_{k=1}^\infty$ forms a basis of $H_0^1(\Gamma)$ orthogonal with respect to the inner product of $H_0^1(\Gamma)$ and orthonormal with respect to the inner product on $L^2(\Gamma)$. Then for $u = \sum_k c_k \phi_k \in \text{span}\ \{\phi_k\}_{k=1}^\infty$ and $s \in [-1, 1]$, we set

$$\|u\|_{H_s} = \sqrt{\sum_k c_k^2 \lambda_k^{-s}} \tag{7}$$

and define H_s to be the closure of span $\{\phi_k\}_{k=1}^\infty$ in the above norm. Then $H_0 = L^2(\Gamma)$ and $H_1 = H_0^1(\Gamma)$ with equality of norms. Moreover, we have $H_{\frac{1}{2}} = H_{00}^{\frac{1}{2}}(\Gamma)$ with equivalence of norms. This essentially follows from the fact that $H_{\frac{1}{2}}$ and $H_{00}^{\frac{1}{2}}(\Gamma)$ are closely related interpolation spaces; see [16, Thm. 3.4]. Note that we also have $H_{-1} = (H_0^1(\Gamma))^* = H^{-1}(\Gamma)$ and $H_{-\frac{1}{2}} = (H_{00}^{\frac{1}{2}}(\Gamma))^* = H^{-\frac{1}{2}}(\Gamma)$.

As the preceding paragraph suggests, we shall use the normal font to denote linear operators, e.g., A. To signify that the particular operator acts on a vector space with multiple components, we employ the calligraphic font, e.g., $\mathcal{A}$. Vectors and matrices are denoted by the sans serif font, e.g., A and x. In the case when the matrix has a block structure, it is typeset with the blackboard bold font, e.g., $\mathbb{A}$. Matrices and vectors are related to the discrete problems as follows (see also [35, Ch. 6]). Let $V_h \subset H_0^1(D)$, and let the discrete operator $A_h : V_h \to V_h^*$ be defined in terms of the Galerkin method:

$$\langle A_h u_h, v_h\rangle_D = \langle Au, v_h\rangle_D \quad \text{for } \ u_h, v_h \in V_h \text{ and } u \in H_0^1(D).$$

Let $\psi_j, j \in [1, m]$ be the basis functions of V_h. The matrix equation,

$$\mathsf{A}\mathsf{u} = \mathsf{f}, \quad \mathsf{u} \in \mathbb{R}^m \text{ and } \mathsf{f} \in \mathbb{R}^m,$$

is obtained as follows: Let $\pi_h : V_h \to \mathbb{R}^m$ and $\mu_h : V_h^* \to \mathbb{R}^m$ be given by

$$v_h = \sum_j (\pi_h v_h)_j \, \psi_j, \quad v_h \in V_h, \qquad \text{and} \qquad (\mu_h f_h)_j = \langle f_h, \psi_j \rangle_D, \quad f_h \in V_h^*.$$

Then

$$\mathsf{A} = \mu_h A_h \pi_h^{-1}, \quad \mathsf{v} = \pi_h v_h, \quad \mathsf{f} = \mu_h f_h.$$

A discrete equivalent to the H_s inner product (7) is constructed in the following manner, similarly to the continuous case. There exist a complete set of eigenvectors $\mathsf{u}_i \in \mathbb{R}^m$ with the property $\mathsf{u}_j{}^\top \mathsf{M} \mathsf{u}_i = \delta_{ij}$ and m positive definite (not necessarily distinct) eigenvalues λ_i of the generalized eigenvalue problem $\mathsf{A}\mathsf{u}_i = \lambda_i \mathsf{M}\mathsf{u}_i$. Equivalently, the matrix A can be decomposed as $\mathsf{A} = (\mathsf{MU})\,\Lambda(\mathsf{MU})^\top$ with $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ and $\operatorname{col}_i \mathsf{U} = \mathsf{u}_i$ so that $\mathsf{U}^\top\mathsf{MU} = \mathsf{I}$ and $\mathsf{U}^\top\mathsf{AU} = \Lambda$. We remark that A is the stiffness matrix, while M is the mass matrix.

Let now $\mathsf{H} : \mathbb{R} \to \mathsf{P}_{\text{sym}}$, where P_{sym} denotes the space of symmetric positive definite matrices, be defined as

$$\mathsf{H}(s) = (\mathsf{MU})\,\Lambda^s(\mathsf{MU})^\top. \tag{8}$$

Note that, due to M orthonormality of the eigenvectors, the inverse of $\mathsf{H}(s)$ is given as $\mathsf{H}(s)^{-1} = \mathsf{U}\Lambda^{-s}\mathsf{U}^\top$. To motivate the definition of the mapping, we shall in the following example consider several values $\mathsf{H}(s)$ and show the relation of the matrices to different Sobolev (semi)norms of functions in V_h.

Example 1 (L_2, H_0^1, and H^{-1} norms in terms of matrices). Let $V_h \subset H_0^1(\Gamma)$, $\dim V_h = m$, $v_h \in V_h$, and $\mathsf{v} \in \mathbb{R}^m$ be the representation of v_h in the basis of V_h, i.e., $\mathsf{v} = \pi_h v_h$. The L^2 norm of v_h is given through the mass matrix M as $\|v_h\|^2_{0,\Gamma} = \mathsf{v}^\top\mathsf{M}\mathsf{v}$ and $\mathsf{M} = \mathsf{H}(0)$. Similarly for the H_0^1 (semi)norm, it holds that $|v_h|^2_{1,\Gamma} = \mathsf{v}^\top\mathsf{A}\mathsf{v}$, where A is the stiffness matrix, and $\mathsf{A} = \mathsf{H}(1)$. Finally, for a less trivial example, let $f_h \in V_h$, and consider f_h as a bounded linear functional, $\langle f_h, v_h\rangle_\Gamma = (f_h, v_h)_\Gamma$ for $v_h \in V_h$. Then $\|f_h\|^2_{-1,\Gamma} = \mathsf{f}^\top\mathsf{H}(-1)\,\mathsf{f}$. By the Riesz representation theorem there exists a unique $u_h \in V_h$ such that $(\nabla u_h, \nabla v_h)_\Gamma = \langle f_h, v_h\rangle_\Gamma$ for all $v_h \in V_h$ and $\|f_h\|_{-1,\Gamma} = |u_h|_{1,\Gamma}$. The latter equality yields $\|f_h\|^2_{-1,\Gamma} = \mathsf{u}^\top\mathsf{A}\mathsf{u}$, but since $u_h \in V_h$ is given by the Riesz map, the coordinate vector comes as a unique solution of the system $\mathsf{A}\mathsf{u} = \mathsf{M}\mathsf{f}$, i.e., $\mathsf{u} = \mathsf{A}^{-1}\mathsf{M}\mathsf{f}$ (see, e.g., [33, Ch. 3]). Thus $\|f_h\|^2_{-1,\Gamma} = \mathsf{f}^\top\mathsf{M}\mathsf{A}^{-1}\mathsf{M}\mathsf{f}$. The matrix product in the expression is then $\mathsf{H}(-1)$.

In general, let c be the representation of vector $\mathsf{u} \in \mathbb{R}^m$ in the basis of eigenvectors u_i, $\mathsf{u} = \mathsf{U}\mathsf{c}$. Then

$$\mathsf{u}^\top\mathsf{H}(s)\,\mathsf{u} = \mathsf{c}^\top\Lambda^s\mathsf{c} = \sum_j c_j^2\lambda_j^s,$$

and so $\mathsf{u}^\top\mathsf{H}(s)\,\mathsf{u} = \|u_h\|^2_{H_s}$ for $u_h \in V_h$ such that $u_h = \pi_h^{-1}\mathsf{u}$. Similarly to the continuous case, the norm can be obtained in terms of powers of an operator

$$\mathsf{u}^\top\mathsf{H}(s)\,\mathsf{u} = \left[\mathsf{U}\Lambda^{\frac{s}{2}}(\mathsf{MU})^\top\mathsf{u}\right]^\top\mathsf{M}\left[\mathsf{U}\Lambda^{\frac{s}{2}}(\mathsf{MU})^\top\mathsf{u}\right] = \left[\mathsf{S}^{-\frac{s}{2}}\mathsf{u}\right]^\top\mathsf{M}\left[\mathsf{S}^{-\frac{s}{2}}\mathsf{u}\right],$$

where $\mathsf{S} = \mathsf{A}^{-1}\mathsf{M}$ is the matrix representation of the Riesz map $H^{-1}(\Gamma) \to H_0^1(\Gamma)$ in the basis of V_h.

Remark 2. The norms constructed above for the discrete space are equivalent to, but not identical to, the H_s-norm from the continuous case.

Before considering proper preconditioning of the weak formulation of problem (1), we illustrate the use of operator preconditioning with an example of a boundary value problem where operators in fractional spaces are utilized to weakly enforce the Dirichlet boundary conditions by Lagrange multipliers [6].

Example 3 (Dirichlet boundary conditions using the Lagrange multiplier). The problem considered in [6] reads as follows: Find u such that

$$\begin{aligned} -\Delta u + u &= f && \text{in } \Omega, \\ u &= g && \text{on } \Gamma \subset \partial\Omega, \\ \partial_n u &= 0 && \text{on } \partial\Omega \setminus \Gamma. \end{aligned} \tag{9}$$

Introducing a Lagrange multiplier p for the boundary value constraint and a trace operator $T : H^1(\Omega) \to H^{\frac{1}{2}}(\Gamma)$ leads to a variational problem for $(u,p) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ satisfying

$$\begin{aligned} (\nabla u, \nabla v)_\Omega + (u,v)_\Omega + \langle p, Tv\rangle_\Gamma &= (f,v)_\Omega\,, && v \in H^1(\Omega)\,, \\ \langle q, Tu\rangle_\Gamma &= \langle q, g\rangle_\Gamma, && q \in H^{-\frac{1}{2}}(\Gamma)\,. \end{aligned} \tag{10}$$

In terms of the framework of operator preconditioning, the variational problem (10) defines an equation

$$\mathcal{A}x = b, \quad \text{where} \quad \mathcal{A} = \begin{bmatrix} -\Delta_\Omega + I & T' \\ T & 0 \end{bmatrix}. \tag{11}$$

In [6] the problem is proved to be well-posed, and therefore $\mathcal{A} : V \to V^*$ is a symmetric isomorphism, where $V = H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ and $x \in V$, $b \in V^*$. A preconditioner is then $\mathcal{B} \in \mathcal{L}(V^*, V)$, constructed such that $\mathcal{B}$ is a positive, self-adjoint isomorphism. Then $\mathcal{B}\mathcal{A} \in \mathcal{L}(V)$ is an isomorphism.

To discretize (11) we shall here employ finite element spaces V_h consisting of linear continuous finite elements where Γ_h is formed by the facets of Ω_h; cf. Figure 1. The stability of discretizations of (10) (for the more general case where the discretization of Ω and Γ are independent) is studied, e.g., in [40] and [42, Ch. 11.3].

The linear system resulting from discretization leads to the following system of equations:

$$\mathbb{B}\mathbb{A}\mathsf{x} = \mathbb{B}\mathsf{b}, \tag{12}$$

where

$$\mathbb{B} = \begin{bmatrix} \mathsf{A}^{-1} & \\ & \mathsf{H}(-\frac{1}{2})^{-1} \end{bmatrix} \quad \text{and} \quad \mathbb{A} = \begin{bmatrix} \mathsf{A} & \mathsf{B}^\top \\ \mathsf{B} & \end{bmatrix}.$$

The last block of the matrix preconditioner $\mathbb{B}$ is the inverse of the matrix constructed by (8) (using discretization of an operator inducing the $H^1(\Gamma)$ norm on the second subspace of V_h), and matrix $\mathbb{B}\mathbb{A}$ has the same eigenvalues as operator $\mathcal{B}_h\mathcal{A}_h$.

Tables 1 and 2 consider the problem (10) with Ω the unit square and Γ its left edge. In Table 1 we show the spectral condition number of the matrix $\mathbb{B}\mathbb{A}$ as a function of the discretization parameter h. It is evident that the condition number is bounded by a constant.

TABLE 1
The smallest and the largest eigenvalues and the spectral condition number of matrix $\mathbb{B}\mathbb{A}$ from system (12).

h	$\lambda_{\min}$	$\lambda_{\max}$	κ
1.77×10^{-1}	0.311	1.750	5.622
8.84×10^{-2}	0.311	1.750	5.622
4.42×10^{-2}	0.311	1.750	5.622
2.21×10^{-2}	0.311	1.750	5.622
1.11×10^{-2}	0.311	1.750	5.622

TABLE 2
The number of iterations required for convergence of the minimal residual method for system (12) *with* $\mathbb{B}$ *replaced by the approximation* (13).

Size	n_{iters}	$\|u - u_h\|_{1,\Omega}$
4290	38	$6.76 \times 10^{-2}(1.00)$
16770	40	$3.38 \times 10^{-2}(1.00)$
66306	38	$1.69 \times 10^{-2}(1.00)$
263682	38	$8.45 \times 10^{-3}(1.00)$
1051650	39	$4.23 \times 10^{-3}(1.00)$

Table 2 then reports the number of iterations required for convergence of the minimal residual method [38] with the system (12) of different sizes. The iterations are started from a random initial vector, and for convergence it is required that r_k, the kth residuum, satisfy ${\mathsf{r}_k}^\top \overline{\mathbb{B}} \mathsf{r}_k < 10^{-10}$. The operator $\overline{\mathbb{B}}$ is the spectrally equivalent approximation of $\mathbb{B}$ given as[1]

$$\overline{\mathbb{B}} = \operatorname{diag}\left(\mathrm{AMG}\left(\mathsf{A}\right), \mathrm{LU}\left(\mathsf{H}\left(-\tfrac{1}{2}\right)\right)\right). \tag{13}$$

The iteration count appears to be bounded independently of the size of the linear system.

Together the presented results indicate that the constructed preconditioner whose discrete approximation utilizes matrices (8) is a good preconditioner for system (9).

Finally, with $\Omega \in \mathbb{R}^2$, $\Gamma \subset \Omega$ of codimension one, we consider problem (1). The weak formulation of (1a)–(1c), using the method of Lagrange multipliers, defines a variational problem for the triplet $(u, v, p) \in U \times V \times Q$,

$$\begin{aligned} (\nabla u, \nabla \phi)_\Omega + \langle p, \epsilon T_\Gamma \phi\rangle_\Gamma &= (f, \phi)_\Omega, & \phi &\in U,\\ (\nabla v, \nabla \psi)_\Gamma - \langle p, \psi\rangle_\Gamma &= (g, \psi)_\Gamma, & \psi &\in V,\\ \langle \chi, \epsilon T_\Gamma u - v\rangle_\Gamma &= 0, & \chi &\in Q, \end{aligned} \tag{14}$$

where U, V, Q are Hilbert spaces to be specified later. The well-posedness of (14) is guaranteed provided that the celebrated Brezzi conditions (see Appendix A) are fulfilled. We remark that

$$\langle p, T_\Gamma \phi\rangle_\Gamma = \langle \delta_\Gamma p, \phi\rangle_\Omega.$$

Hence δ_Γ is in our context the dual operator to the trace operator T_Γ. Since $T_\Gamma : H^1_0(\Omega) \rightarrow H^{\frac{1}{2}}_{00}(\Gamma)$, then $\delta_\Gamma : H^{-\frac{1}{2}}(\Gamma) \rightarrow H^{-1}(\Omega)$.

For our discussion of preconditioners it is suitable to recast (14) as an operator equation for the self-adjoint operator $\mathcal{A}$,

$$\mathcal{A}\begin{bmatrix} u\\ v\\ p\end{bmatrix} = \begin{bmatrix} A_U & & B_U^*\\ & A_V & B_V^*\\ B_U & B_V & \end{bmatrix}\begin{bmatrix} u\\ v\\ p\end{bmatrix} = \begin{bmatrix} f\\ g\\ \end{bmatrix}, \tag{15}$$

[1] Here and in the subsequent numerical experiments AMG is the algebraic multigrid BOOMERAMG from the Hypre library [23], and LU is the direct solver from the UMFPACK library [19]. The libraries were accessed through the interface provided by PETSc [7] version 3.5.3. To assemble the relevant matrices FEniCS library [31] version 1.6.0 and its extension for block-structured systems cbc.block [34] were used. The AMG preconditioner was used with the default options, except for coarsening, which was set to Ruge–Stueben algorithm.

with the operators A_i, B_i, $i \in \{U, V\}$, given by

$$\begin{aligned} \langle A_U u, \phi\rangle_\Omega &= (\nabla u, \nabla\phi)_\Omega\,, \quad &\langle A_V v, \psi\rangle_\Gamma &= (\nabla v, \nabla\psi)_\Gamma\,, \\ \langle B_U u, \chi\rangle_\Gamma &= \langle \chi, \epsilon T_\Gamma u\rangle_\Gamma, \quad &\langle B_V v, \chi\rangle_\Gamma &= -\langle \chi, v\rangle_\Gamma. \end{aligned}$$

Further, for discussion of mapping properties of $\mathcal{A}$ it will be advantageous to consider the operator as a map defined over space $W \times Q$, $W = U \times V$ as

$$\mathcal{A} = \begin{bmatrix} A & B^* \\ B & \end{bmatrix} \quad \text{with} \quad A = \begin{bmatrix} A_U & \\ & A_V \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} B_U & B_V \end{bmatrix}. \tag{16}$$

Considering two different choices of spaces U, V, and Q, we will propose two formulations that lead to different preconditioners:

$$\mathcal{B}_Q^{-1} = \begin{bmatrix} A_U & & \\ & A_V & \\ & & B_U A_U^{-1} B_U^* + B_V A_V^{-1} B_V^* \end{bmatrix} \tag{17}$$

and

$$\mathcal{B}_W^{-1} = \begin{bmatrix} A_U + B_U^* R B_U & & \\ & A_V & \\ & & B_V A_V^{-1} B_V^* \end{bmatrix}. \tag{18}$$

Here R is the Riesz map from Q^* to Q. Preconditioners of the form (17)–(18) will be referred to as the Q-cap and the W-cap preconditioners. This naming convention reflects the role intersection spaces play in the respected formulations. We remark that the definitions should be understood as templates identifying the correct structure of the preconditioner.

3. *Q*-cap preconditioner.

Consider operator $\mathcal{A}$ from problem (15) as a mapping $W \times Q \to W^* \times Q^*$,

$$\begin{aligned} W &= H_0^1(\Omega) \times H_0^1(\Gamma)\,, \\ Q &= \epsilon H^{-\frac{1}{2}}(\Gamma) \cap H^{-1}(\Gamma)\,. \end{aligned} \tag{19}$$

The spaces are equipped with norms

$$\|w\|_W^2 = |u|_{1,\Omega}^2 + |v|_{1,\Gamma}^2 \quad \text{and} \quad \|p\|_Q^2 = \epsilon^2\|p\|_{-\frac{1}{2},\Gamma}^2 + \|p\|_{-1,\Gamma}^2. \tag{20}$$

Since $H^{-\frac{1}{2}}(\Gamma)$ is continuously embedded in $H^{-1}(\Gamma)$, the space Q is the same topological vector space as $H^{-\frac{1}{2}}(\Gamma)$, but equipped with an equivalent, ϵ-dependent inner product. See also [9, Ch. 2]. The next theorem shows that this definition leads to a well-posed problem.

We will need a right inverse of the trace operator and employ the following harmonic extension. Let $q \in H_{00}^{\frac{1}{2}}(\Gamma)$, and let u be the solution of the problem

$$\begin{aligned} -\Delta u &= 0 \quad &&\text{in } \Omega \setminus \Gamma, \\ u &= 0 &&\text{on } \partial\Omega, \\ u &= q &&\text{on } \Gamma. \end{aligned} \tag{21}$$

Since the trace is surjective onto $H_{00}^{\frac{1}{2}}(\Gamma)$, (21) has a solution $u \in H_0^1(\Omega)$ and $|u|_{1,\Omega} \le C|q|_{\frac{1}{2},\Gamma}$ for some constant C. We denote the harmonic extension operator by E, i.e., $u = Eq$ with $\|E\| \le C$.

THEOREM 4. *Let W and Q be the spaces* (19). *The operator $\mathcal{A} : W \times Q \to W^* \times Q^*$, defined in* (15), *is an isomorphism, and the condition number of $\mathcal{A}$ is bounded independently of $\epsilon > 0$.*

Proof. The statement follows from the Brezzi theorem, Theorem 13, once its assumptions are verified. Since A induces the inner product on W, A is continuous and coercive, and the conditions (51a) and (51b) hold. Next, we see that B is bounded:

$$\begin{aligned}\langle Bw, q\rangle_\Gamma &= \langle q, \epsilon T_\Gamma u - v\rangle_\Gamma \\ &\le \|q\|_{-\frac{1}{2},\Gamma}\|\epsilon T_\Gamma u\|_{\frac{1}{2},\Gamma} + \|q\|_{-1,\Gamma}|v|_{1,\Gamma} \\ &\le \big(1 + \|T_\Gamma\|\big)\sqrt{\epsilon^2\|q\|^2_{-\frac{1}{2},\Gamma} + \|q\|^2_{-1,\Gamma}}\sqrt{|u|^2_{1,\Omega} + |v|^2_{1,\Gamma}} \\ &= \big(1 + \|T_\Gamma\|\big)\|q\|_Q\|w\|_W.\end{aligned}$$

It remains to show the inf-sup condition (51d). Since the trace is bounded and surjective, for all $\xi \in H^{\frac{1}{2}}_{00}(\Gamma)$ we let u be defined in terms of the harmonic extension (21) such that $u = \epsilon^{-1}E\xi$ and $|u|_{1,\Omega} \le \epsilon^{-1}\|E\|\|\xi\|_{\frac{1}{2},\Gamma}$. Hence,

$$\begin{aligned}\sup_{w\in W}\frac{\langle Bw, q\rangle_\Gamma}{\|w\|_W} &= \sup_{w\in W}\frac{\langle q, \epsilon T_\Gamma u - v\rangle_\Gamma}{\sqrt{|u|^2_{1,\Omega} + |v|^2_{1,\Gamma}}} \\ &\ge \big(1 + \|E\|\big)^{-1}\sup_{(\xi,v)\in H^{\frac{1}{2}}_{00}(\Gamma)\times H^1_0(\Gamma)}\frac{\langle q, \xi + v\rangle_\Gamma}{\sqrt{\epsilon^{-2}\|\xi\|^2_{\frac{1}{2},\Gamma} + \|v\|^2_{1,\Gamma}}}.\end{aligned}$$

Note that we have the identity

$$Q^* = \big(\epsilon H^{-\frac{1}{2}}(\Gamma) \cap H^{-1}(\Gamma)\big)^* = \epsilon^{-1}H^{\frac{1}{2}}_{00}(\Gamma) + H^1_0(\Gamma),$$

equipped with the norm

$$\|q^*\|_{Q^*} = \inf_{q^*=q^*_1+q^*_2}\epsilon^{-2}\|q^*_1\|^2_{\frac{1}{2},\Gamma} + |q^*_2|^2_{1,\Gamma}.$$

See also [9]. It follows that

$$\begin{aligned}\sup_{(\xi,v)\in H^{\frac{1}{2}}(\Gamma)\times H^1_0(\Gamma)}\frac{\langle q, \xi + v\rangle_\Gamma}{\sqrt{\epsilon^{-2}\|\xi\|^2_{\frac{1}{2},\Gamma} + |v|^2_{1,\Gamma}}} &= \sup_{\zeta\in Q^*}\sup_{\substack{\xi+v=\zeta\\ v\in H^1_0(\Gamma)}}\frac{\langle q, \xi + v\rangle_\Gamma}{\sqrt{\epsilon^{-2}\|\xi\|^2_{\frac{1}{2},\Gamma} + |v|^2_{1,\Gamma}}} \\ &= \sup_{\zeta\in Q^*}\frac{\langle q, \zeta\rangle_\Gamma}{\inf\limits_{\substack{\xi+v=\zeta\\ v\in H^1_0(\Gamma)}}\sqrt{\epsilon^{-2}\|\xi\|^2_{\frac{1}{2},\Gamma} + |v|^2_{1,\Gamma}}} \\ &= \|q\|_{Q^{**}} = \|q\|_Q.\end{aligned}$$

Consequently, condition (51d) holds with a constant independent of ϵ. ∎

Following Theorem 4 and [35], a preconditioner for the symmetric isomorphic operator $\mathcal{A}$ is the Riesz mapping $W^* \times Q^*$ to $W \times Q$:

$$\mathcal{B}_Q = \begin{bmatrix}-\Delta_\Omega & & \\ & -\Delta_\Gamma & \\ & & \epsilon^2\Delta_\Gamma^{-\frac{1}{2}} + \Delta_\Gamma^{-1}\end{bmatrix}^{-1}. \tag{22}$$

Here Δ_Γ^s is defined by $\langle \Delta_\Gamma^s v, w\rangle_\Gamma = (v, w)_{H_s}$, with the H_s-inner product defined by (7). Hence the norm induced on $W \times Q$ by the operator $\mathcal{B}_Q^{-1}$ is not (20) but an equivalent norm

$$\langle {\mathcal{B}_Q}^{-1} x, x\rangle = |u|^2_{1,\Omega} + |v|^2_{1,\Gamma} + \epsilon^2 \|p\|^2_{H_{-\frac{1}{2}}(\Gamma)} + \|p\|^2_{H_{-1}(\Gamma)}$$

for any $x = (u, v, p) \in W \times Q$. Note that $\mathcal{B}_Q$ fits the template defined in (17).

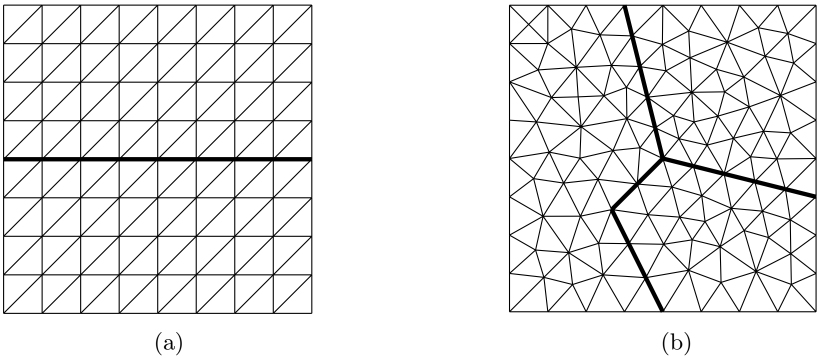

FIG. 1. *Geometrical configurations and their sample triangulations considered in the numerical experiments.*

3.1. Discrete Q-cap preconditioner. Following Theorem 4, the Q-cap preconditioner (22) is a good preconditioner for operator equation $\mathcal{A}x = b$ with the condition number independent of the material parameter ϵ. To translate the preconditioned operator equation $\mathcal{B}_Q \mathcal{A} x = \mathcal{B}_Q b$ into a stable linear system it is necessary to employ suitable discretization. In particular, the Brezzi conditions must hold on each approximation space $W_h \times Q_h$ with constants independent of the discretization parameter h. Such a suitable discretization will be referred to as stable.

Let us consider a stable discretization of operator $\mathcal{A}$ from Theorem 4 by finite dimensional spaces U_h, V_h, and Q_h defined as

$$U_h = \text{span}\ \{\phi_i\}_{i=1}^{n_U}, \quad V_h = \text{span}\ \{\psi_i\}_{i=1}^{n_V}, \quad Q_h = \text{span}\ \{\chi_i\}_{i=1}^{n_Q}.$$

Then the Galerkin method for problem (15) reads as follows: Find $(u_h, v_h, p_h) \in U_h \times V_h \times Q_h$ such that

$$\begin{aligned} (\nabla u_h, \nabla \phi)_\Omega + \langle p_h, \epsilon T_\Gamma \phi\rangle_\Gamma &= (f, \phi)_\Omega\,, & \phi &\in U_h,\\ (\nabla v_h, \nabla \psi)_\Gamma - \langle p_h, \psi\rangle_\Gamma &= (g, \psi)_\Gamma\,, & \psi &\in V_h,\\ \langle \chi, \epsilon T_\Gamma u_h - v_h\rangle_\Gamma &= 0, & \chi &\in Q_h. \end{aligned}$$

Further, we shall define matrices A_U, A_V and B_U, B_V in the following way:

$$\begin{aligned} &\mathsf{A}_U \in \mathbb{R}^{n_U \times n_U}, \quad (\mathsf{A}_U)_{i,j} = (\nabla \phi_j, \nabla \phi_i)_\Omega\,,\\ &\mathsf{A}_V \in \mathbb{R}^{n_V \times n_V}, \quad (\mathsf{A}_V)_{i,j} = (\nabla \psi_j, \nabla \psi_i)_\Gamma\,,\\ &\mathsf{B}_U \in \mathbb{R}^{n_Q \times n_U}, \quad (\mathsf{B}_U)_{i,j} = \langle \epsilon T_\Gamma \phi_j, \chi_i\rangle_\Gamma,\\ &\mathsf{B}_V \in \mathbb{R}^{n_Q \times n_V}, \quad (\mathsf{B}_V)_{i,j} = -\langle \psi_j, \chi_i\rangle_\Gamma. \end{aligned} \tag{23}$$

We note that B_V can be viewed as a representation of the negative identity mapping between spaces V_h and Q_h. Similarly, matrix B_U can be viewed as a composite, $\mathsf{B}_U = \mathsf{M}_{\overline{U}Q}\mathsf{T}$. Here $\mathsf{M}_{\overline{U}Q}$ is the representation of an identity map from space $\overline{U}_h$ to space Q_h. The space $\overline{U}_h$ is the image of U_h under the trace mapping T_Γ. We shall respectively denote the dimension of the space and its basis functions $n_{\overline{U}}$ and $\bar{\phi}_i$, $i \in [1, n_{\overline{U}}]$. Matrix $\mathsf{T} \in \mathbb{R}^{n_{\overline{U}} \times n_U}$ is then a representation of the trace mapping $T_\Gamma : U_h \to \overline{U}_h$.

We note that the rank of T is n_Q, and mirroring the continuous operator T_Γ, the matrix has a unique right inverse T^+. We refer the reader to [36] for the continuous case. The matrix T^+ can be computed as a pseudoinverse via the reduced singular value decomposition $\mathsf{TU} = \mathsf{Q\Sigma}$; see, e.g., [45, Ch. 11]. Then $\mathsf{T}^+ = \mathsf{U\Sigma}^{-1}\mathsf{Q}$. Here, the columns of U can be viewed as coordinates of functions $\bar{\phi}_i$ zero-extended to Ω such that they form the l^2 orthonormal basis of the subspace of $\mathbb{R}^{n_U}$ where the problem $\mathsf{Tu} = \bar{\mathsf{u}}$ is solvable. Further, the kernel of T is spanned by n_U-vectors representing those functions in U_h whose trace on Γ is zero.

For the space U_h constructed by the finite element method with the triangulation of Ω such that Γ is aligned with the element boundaries (cf. Figure 1), it is a consequence of the nodality of the basis that $\mathsf{T}^+ = \mathsf{T}^\top$.

With definitions (23) we use $\mathbb{A}$ to represent the operator $\mathcal{A}$ from (15) in the basis of $W_h \times Q_h$:

$$\mathbb{A} = \begin{bmatrix} \mathsf{A}_U & & \mathsf{B}_U{}^\top \\ & \mathsf{A}_V & \mathsf{B}_V{}^\top \\ \mathsf{B}_U & \mathsf{B}_V & \end{bmatrix}. \tag{24}$$

Finally, a discrete Q-cap preconditioner is defined as a matrix representation of (22) with respect to the basis of $W_h \times Q_h$:

$$\mathbb{B}_Q = \begin{bmatrix} \mathsf{A}_U & & \\ & \mathsf{A}_V & \\ & & \epsilon^2\mathsf{H}(-\tfrac{1}{2}) + \mathsf{H}(-1) \end{bmatrix}^{-1}. \tag{25}$$

The matrices A, M which are used to compute the values $\mathsf{H}(\cdot)$ through the definition (8) have the properties $|p|^2_{1,\Gamma} = \mathsf{p}^\top\mathsf{Ap}$ and $\|p\|^2_{0,\Gamma} = \mathsf{p}^\top\mathsf{Mp}$ for every $p \in Q_h$ and $\mathsf{p} \in \mathbb{R}^{n_Q}$ its coordinate vector. Note that due to properties of matrices $\mathsf{H}(\cdot)$, the matrix N_Q,

$$\mathsf{N}_Q = \left[\epsilon^2\mathsf{H}(-\tfrac{1}{2}) + \mathsf{H}(-1)\right]^{-1} = \mathsf{U}\left[\epsilon^2\Lambda^{-\frac{1}{2}} + \Lambda^{-1}\right]^{-1}\mathsf{U}^\top, \tag{26}$$

is the inverse of the final block of $\mathbb{B}_Q$.

By Theorem 4 and the assumption on spaces $W_h \times Q_h$ being stable, the matrix $\mathbb{B}_Q\mathbb{A}$ has a spectrum bounded independently of the parameter ϵ and the size of the system or equivalently discretization parameter h. In turn, $\mathbb{B}_Q$ is a good preconditioner for matrix $\mathbb{A}$. To demonstrate this property we shall now construct a stable discretization of the space $W \times Q$ using the finite element method.

3.2. Stable subspaces for *Q*-cap preconditioner. For $h > 0$ fixed, let Ω_h be the polygonal approximation of Ω. For the set $\bar{\Omega}_h$, we construct a shape-regular triangulation consisting of closed triangles K_i such that $\Gamma \cap K_i$ is an edge e_i of the triangle. Let Γ_h be a union of such edges. The discrete spaces $W_h \subset W$ and $Q_h \subset Q$ shall be defined in the following way. Let

$$\begin{aligned} U_h &= \{v \in C\left(\overline{\Omega}_h\right) : v|_K = \mathbb{P}_1(K)\}, \\ V_h &= \{v \in C\left(\overline{\Gamma}_h\right) : v|_e = \mathbb{P}_1(e)\}, \end{aligned} \tag{27}$$

where $\mathbb{P}_1(D)$ are linear polynomials on the simplex D. Then we set

$$\begin{aligned} W_h &= \left(U_h \cap H_0^1(\Omega)\right) \times \left(V_h \cap H_0^1(\Gamma)\right), \\ Q_h &= V_h \cap H_0^1(\Gamma). \end{aligned} \tag{28}$$

Let A_h, B_h be the finite dimensional operators defined on the approximation spaces (28) in terms of the Galerkin method for operators A, B in (16). Since the constructed spaces are conforming, the operators A_h, B_h are continuous with respect to the norms (20). Further, A_h is W-elliptic on W_h since the operator defines an inner product on the discrete space. Thus, to show that the spaces $W_h \times Q_h$ are stable, it remains to show that the discrete inf-sup condition holds.

LEMMA 5. *Let $W_h \subset W$, $Q_h \subset Q$ be the spaces* (28). *Further, let $\|\cdot\|_W$, $\|\cdot\|_Q$ be the norms* (20). *Finally, let B_h be such that $\langle B_h w_h, q_h\rangle_\Gamma = \langle Bw, q_h\rangle_\Gamma$, $w \in W$. There exists a constant $\beta > 0$ such that*

$$\inf_{q_h\in Q_h} \sup_{w_h\in W_h} \frac{\langle B_h w_h, q_h\rangle_\Gamma}{\|w_h\|_W \|q_h\|_Q} \geq \beta. \tag{29}$$

Proof. Recall that $Q = \epsilon H^{-\frac{1}{2}}(\Gamma) \cap H^{-1}(\Gamma)$. We follow the steps of the continuous inf-sup condition in reverse order. By definition,

$$\begin{aligned} \|q_h\|_Q &= \sup_{p\in\epsilon H_{00}^{\frac{1}{2}}(\Gamma)+H_0^1(\Gamma)} \frac{\langle q_h, p\rangle_\Gamma}{\inf\limits_{p=p_1+p_2} \sqrt{\epsilon^{-2}\|p_1\|^2_{\frac{1}{2},\Gamma} + |p_2|^2_{1,\Gamma}}} \\ &= \sup_{p} \sup_{p=p_1+p_2} \frac{\langle q_h, p_1\rangle_\Gamma + \langle q_h, p_2\rangle_\Gamma}{\sqrt{\epsilon^{-2}\|p_1\|^2_{\frac{1}{2},\Gamma} + |p_2|^2_{1,\Gamma}}}. \end{aligned} \tag{30}$$

For each $p_1 \in H_{00}^{\frac{1}{2}}(\Gamma)$, let $u_h \in U_h$ be the weak solution of the boundary value problem

$$\begin{aligned} -\Delta u &= 0 && \text{in } \Omega, \\ \epsilon u &= p_1 && \text{on } \Gamma, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Then $\epsilon T_\Gamma u_h = p_1$ in $H_{00}^{\frac{1}{2}}(\Gamma)$ and $\epsilon|u_h|_{1,\Omega} \leq C\|p_1\|_{\frac{1}{2},\Gamma}$ for some constant C depending only on Ω and Γ. For each $p_2 \in H_0^1(\Gamma)$, let $v_h \in V_h$ be the L^2 projection of p_2 onto the space V_h:

$$\langle v_h - p_2, z\rangle_\Gamma = 0, \quad z \in V_h. \tag{31}$$

By construction we then have $\langle q_h, p_2 - v_h\rangle_\Gamma = 0$ for all $q_h \in Q_h$ and $\|v_h\|_{0,\Gamma} \leq \|p_2\|_{0,\Gamma}$. Moreover, for shape-regular triangulation, the projection $\Pi : H_0^1(\Gamma) \to V_h$, $v_h = \Pi p_2$ is bounded in the H_0^1 norm:

$$|v_h|_{1,\Gamma} \leq |p_2|_{1,\Gamma}. \tag{32}$$

We refer the reader to [10, Ch. 7] for this result. For constructed u_h, v_h it follows from (30) that

$$\begin{aligned} \|q_h\|_Q &\lesssim \sup_{w_h\in U_h+V_h} \sup_{w_h=u_h+v_h} \frac{\langle q_h, \epsilon T_\Gamma u_h + v_h\rangle_\Gamma}{\sqrt{|u_h|^2_{1,\Omega} + |v_h|^2_{1,\Gamma}}} \\ &= \sup_{(u_h,v_h)\in U_h\times V_h} \frac{\langle q_h, \epsilon T_\Gamma u_h + v_h\rangle_\Gamma}{\|(u_h, v_h)\|_W} = \sup_{w_h\in W_h} \frac{\langle B_h w_h, q_h\rangle_\Gamma}{\|w_h\|_W}. \end{aligned}$$

□

The constructed stable discretizations (28) are a special case of conforming spaces built from $U_{h;k} \subset H^1(\Omega)$ and $V_{h;l} \subset H^1(\Gamma)$ defined as

$$\begin{aligned} U_{h;k} &= \{v \in C\left(\bar{\Omega}_h\right) \,:\, v|_K = \mathbb{P}_k(K)\}, \\ V_{h;l} &= \{v \in C\left(\bar{\Gamma}_h\right) \,:\, v|_e = \mathbb{P}_l(e)\}. \end{aligned} \tag{33}$$

The following corollary gives a necessary compatibility condition on polynomial degrees in order to build inf-sup stable spaces from components (33).

COROLLARY 6. *Let* $W_{h;k,l} = \left(U_{h;k} \cap H^1_0(\Omega)\right) \times \left(V_{h;l} \cap H^1_0(\Gamma)\right)$ *and* $Q_{h;m} = V_{h;m} \cap H^1_0(\Gamma)$. *The necessary condition for* (29) *to hold with space* $W_{h;k,l} \times Q_{h;m}$ *is that* $m \leq \max(k,l)$.

Proof. Note that $T_\Gamma u_h - v_h$ is a piecewise polynomial of degree $\max(k,l)$. Suppose $m > \max(k,l)$. Then for each $(u_h, v_h) \in W_{h;k,l}$ we can find an orthogonal polynomial $0 \neq q_h \in Q_{h;m}$ such that

$$\langle q_h, T_\Gamma u_h - v_h \rangle_\Gamma = 0.$$

In turn, $\beta = 0$ in (29), and the discrete inf-sup condition cannot hold. □

3.3. Numerical experiments. Let now $\mathbb{A}$, $\mathbb{B}_Q$ be the matrices (24), (25) assembled over the constructed stable spaces (28). We demonstrate the robustness of the Q-cap preconditioner (22) through a pair of numerical experiments. First, the *exact* preconditioner represented by the matrix $\mathbb{B}_Q$ is considered, and we are interested in the condition number of $\mathbb{B}_Q\mathbb{A}$ for different values of the parameter ϵ. The spectral condition number is computed from the smallest and largest (in magnitude) eigenvalues of the generalized eigenvalue problem $\mathbb{A}\mathsf{x} = \lambda \mathbb{B}_Q^{-1}\mathsf{x}$, which is here solved by SLEPc [27].[2] The obtained results are reported in Table 3. In general, the condition numbers are well behaved, indicating that $\mathbb{B}_Q$ defines a parameter robust preconditioner. We note that for $\epsilon \ll 1$ the spectral condition number is close to $\left(1+\sqrt{5}\right)/\left(\sqrt{5}-1\right) \approx 2.618$. In section 3.4 this observation is explained by the relation of the proposed preconditioner $\mathbb{B}_Q$ and the matrix preconditioner of Murphy, Golub, and Wathen [37].

TABLE 3
Spectral condition numbers of matrices $\mathbb{B}_Q\mathbb{A}$ *for the system assembled on geometry* (a) *in Figure* 1.

Size	n_Q	$\log_{10}\epsilon$						
		-3	-2	-1	0	1	2	3
99	9	2.655	2.969	4.786	6.979	7.328	7.357	7.360
323	17	2.698	3.323	5.966	7.597	7.697	7.715	7.717
1155	33	2.778	3.905	7.031	7.882	7.818	7.816	7.816
4355	65	2.932	4.769	7.830	8.016	7.855	7.843	7.843
16899	129	3.217	5.857	8.343	8.081	7.868	7.854	7.852
66563	257	3.710	6.964	8.637	8.113	7.872	7.856	7.855

In the second experiment, we monitor the number of iterations required for convergence of the MinRes method [38] (the implementation is provided by cbc.block [34]) applied to the preconditioned equation $\bar{\mathbb{B}}_Q\mathbb{A}\mathsf{x} = \bar{\mathbb{B}}_Q\mathsf{b}$. The operator $\bar{\mathbb{B}}_Q$ is an

[2] We use the generalized Davidson method with the Cholesky preconditioner and convergence tolerance 10^{-8}.

efficient and spectrally equivalent approximation of $\mathbb{B}_Q$,

$$\bar{\mathbb{B}}_Q = \begin{bmatrix} \mathrm{AMG}\,(\mathsf{A}_U) & & \\ & \mathrm{LU}(\mathsf{A}_V) & \\ & & \mathsf{N}_Q \end{bmatrix}, \tag{34}$$

with N_Q defined in (26). The iterations are started from a random initial vector, and as a stopping criterion a condition on the magnitude of the kth preconditioned residual r_k, ${\mathsf{r}_k}^\top \bar{\mathbb{B}}_Q \mathsf{r}_k < 10^{-12}$ is used. The observed number of iterations is shown in Table 4. Robustness with respect to size of the system and the material parameter is evident as the iteration count is bounded for all the considered discretizations and values of ϵ.

Table 4

Iteration count for convergence of $\bar{\mathbb{B}}_Q \mathbb{A}\mathsf{x} = \bar{\mathbb{B}}_Q \mathsf{b}$ solved with the minimal residual method. The problem is assembled on geometry (a) *from Figure* 1.

Size	n_Q	$\log_{10}\epsilon$						
		−3	−2	−1	0	1	2	3
66563	257	20	34	37	32	28	24	21
264195	513	22	34	34	30	26	24	20
1052675	1025	24	33	32	28	26	22	18
4202499	2049	26	32	30	26	24	20	17
8398403	2897	26	30	30	26	22	19	15
11075583	3327	26	30	30	26	22	19	15

Comparing Tables 3 and 4, we observe that the ϵ-behavior of the condition number and the iteration counts are different. In particular, fewer iterations are required for $\epsilon = 10^3$ than for $\epsilon = 10^{-3}$, while the condition number in the former case is larger. Moreover, the condition numbers for $\epsilon > 1$ are almost identical, whereas the iteration counts decrease as the parameter grows. We note that these observations should be viewed in light of the fact that the convergence of the minimal residual method in general does not depend solely on the condition number (e.g., [29]), and a more detailed knowledge of the eigenvalues is required to understand the behavior.

Having proved and numerically verified the properties of the Q-cap preconditioner, we shall in the next section link $\mathbb{B}_Q$ to a block diagonal matrix preconditioner suggested by Murphy, Golub, and Wathen [37]. Both matrices are assumed to be assembled on the spaces (28), and the main objective of the section is to prove spectral equivalence of the two preconditioners.

3.4. Relation to Schur complement preconditioner. Consider a linear system $\mathbb{A}\mathsf{x} = \mathsf{b}$ with an indefinite matrix (24) which shall be preconditioned by a block diagonal matrix

$$\mathbb{B} = \mathrm{diag}\,(\mathsf{A}_U, \mathsf{A}_V, \mathsf{S})^{-1}, \quad \mathsf{S} = \mathsf{B}_U {\mathsf{A}_U}^{-1} {\mathsf{B}_U}^\top + \mathsf{B}_V {\mathsf{A}_V}^{-1} {\mathsf{B}_V}^\top, \tag{35}$$

where S is the negative Schur complement of $\mathbb{A}$. Following [37], the spectrum of $\mathbb{B}\mathbb{A}$ consists of three distinct eigenvalues. In fact, $\rho\,(\mathbb{B}\mathbb{A}) = \{1, \frac{1}{2} \pm \frac{1}{2}\sqrt{5}\}$. A suitable Krylov method is thus expected to converge in no more than three iterations. However, in its presented form, $\mathbb{B}$ does not define an efficient preconditioner. In particular, the cost of setting up the Schur complement comes close to inverting the system matrix $\mathbb{A}$. Therefore, a cheaply computable approximation of S is needed to make the preconditioner practical (see, e.g., [8, Ch. 10.1] for an overview of generic methods for constructing the approximation). We proceed to show that if spaces (28) are used for discretization, the Schur complement is more efficiently approximated with the inverse of the matrix N_Q defined in (26).

Let W_h, Q_h be the spaces (28). Then the mass matrix $\mathsf{M}_{\overline{U}Q} = \mathsf{M}_{VQ}$ (cf. the discussion prior to (23)), and the matrix will be referred to as M. Moreover, let us set $\mathsf{A}_V = \mathsf{A}$. With these definitions the Schur complement of $\mathbb{A}$ reads

$$\mathsf{S} = \epsilon^2 \mathsf{M}\mathsf{T}{\mathsf{A}_U}^{-1}\mathsf{T}^\top\mathsf{M} + \mathsf{M}\mathsf{A}^{-1}\mathsf{M}. \tag{36}$$

Further, note that such matrices A, M are suitable for constructing the approximation of the H_s norm on the space Q_h by the mapping (8). In particular, A is such that $|p|^2_{1,\Gamma} = \mathsf{p}^\top\mathsf{A}\mathsf{p}$ with $p \in Q_h$ and $\mathsf{p} \in \mathbb{R}^{n_Q}$ its coordinate vector. In turn, the inverse of the matrix N_Q reads

$$\mathsf{N}_Q{}^{-1} = (\mathsf{M}\mathsf{U})\left(\epsilon^2\Lambda^{-\frac{1}{2}} + \Lambda^{-1}\right)(\mathsf{M}\mathsf{U})^\top = \epsilon^2\mathsf{H}\left(-\tfrac{1}{2}\right) + \mathsf{H}(-1)\,. \tag{37}$$

Recalling that $\mathsf{H}(-1) = \mathsf{M}\mathsf{A}^{-1}\mathsf{M}$ and contrasting (36) with (37), we see that the matrices differ only in the first terms. We shall first show that if the terms are spectrally equivalent, then so are S and $\mathsf{N}_Q{}^{-1}$.

THEOREM 7. *Let* S, $\mathsf{N}_Q{}^{-1}$ *be the matrices defined, respectively, in* (36) *and* (37), *and let* n_Q *be their size. Assume that there exist positive constants* c_1, c_2 *dependent only on* Ω *and* Γ *such that for every* $n_Q > 0$ *and any* $\mathsf{p} \in \mathbb{R}^{n_Q}$

$$c_1\mathsf{p}^\top\mathsf{H}\left(-\tfrac{1}{2}\right)\mathsf{p} \le \mathsf{p}^\top\mathsf{M}\mathsf{T}{\mathsf{A}_U}^{-1}\mathsf{T}^\top\mathsf{M}\mathsf{p} \le c_2\mathsf{p}^\top\mathsf{H}\left(-\tfrac{1}{2}\right)\mathsf{p}.$$

Then, for each $n_Q > 0$, *matrix* S *is spectrally equivalent with* $\mathsf{N}_Q{}^{-1}$.

Proof. By direct calculation we have

$$\begin{aligned}\mathsf{p}^\top\mathsf{S}\mathsf{p} &= \epsilon^2\mathsf{p}^\top\mathsf{M}\mathsf{T}{\mathsf{A}_U}^{-1}\mathsf{T}^\top\mathsf{M}\mathsf{p} + \mathsf{p}^\top\mathsf{H}(-1)\,\mathsf{p}\\ &\le c_2\epsilon^2\mathsf{p}^\top\mathsf{H}\left(-\tfrac{1}{2}\right)\mathsf{p} + \mathsf{p}^\top\mathsf{H}(-1)\,\mathsf{p}\\ &\le C_2\mathsf{p}^\top\mathsf{N}_Q{}^{-1}\mathsf{p}\end{aligned}$$

for $C_2 = \sqrt{1+c_2^2}$. The existence of the lower bound follows from the estimate

$$\mathsf{p}^\top\mathsf{S}\mathsf{p} \ge c_1\epsilon^2\mathsf{p}^\top\mathsf{H}\left(-\tfrac{1}{2}\right)\mathsf{p} + \mathsf{p}^\top\mathsf{H}(-1)\,\mathsf{p} \ge C_1\mathsf{p}^\top\mathsf{N}_Q{}^{-1}\mathsf{p}$$

with $C_1 = \min(1, c_1)$. □

The spectral equivalence of preconditioners $\mathbb{B}_Q$ and $\mathbb{B}$ now follows immediately from Theorem 7. Note that for $\epsilon \ll 1$ the term $\mathsf{H}(-1)$ dominates both S and $\mathsf{N}_Q{}^{-1}$. In turn, the spectrum of $\mathbb{B}\mathbb{A}$ is expected to approximate well the eigenvalues of $\mathbb{B}_Q\mathbb{A}$. This is then a qualitative explanation of why the spectral condition numbers of $\mathbb{B}_Q\mathbb{A}$ observed for $\epsilon = 10^{-3}$ in Table 3 are close to $\left(1+\sqrt{5}\right)/\left(\sqrt{5}-1\right)$. It remains to prove that the assumption of Theorem 7 holds.

LEMMA 8. *There exist constants* $c_1, c_2 > 0$ *depending only on* Ω, Γ *such that for all* $n_Q > 0$ *and* $p \in \mathbb{R}^{n_Q}$

$$c_1\mathsf{p}^\top\mathsf{H}\left(-\tfrac{1}{2}\right)\mathsf{p} \le \mathsf{p}^\top\mathsf{M}\mathsf{T}{\mathsf{A}_U}^{-1}\mathsf{T}^\top\mathsf{M}\mathsf{p} \le c_2\mathsf{p}^\top\mathsf{H}\left(-\tfrac{1}{2}\right)\mathsf{p}.$$

Proof. For the sake of readability let $n = n_Q$ and $m = n_U$. Since M is symmetric and invertible, $\mathsf{H}\left(-\frac{1}{2}\right) = \mathsf{M}\mathsf{U}\Lambda^{-\frac{1}{2}}\mathsf{U}^\top\mathsf{M}$ and $\mathsf{U}\Lambda^{-\frac{1}{2}}\mathsf{U}^\top = \mathsf{H}\left(\frac{1}{2}\right)^{-1}$, the statement is equivalent to

$$c_1\mathsf{y}^\top\mathsf{H}\left(\tfrac{1}{2}\right)^{-1}\mathsf{y} \leq \mathsf{y}^\top\mathsf{T}\mathsf{A}_U{}^{-1}\mathsf{T}^\top\mathsf{y} \leq c_2\mathsf{y}^\top\mathsf{H}\left(\tfrac{1}{2}\right)^{-1}\mathsf{y} \quad \text{for all } y \in \mathbb{R}^m. \tag{38}$$

The proof is based on properties of the continuous trace operator T_Γ. Recall the trace inequality: There exists a positive constant $K_2 = K_2(\Omega, \Gamma)$ such that $\|T_\Gamma u\|_{\frac{1}{2},\Gamma} \leq K_2|u|_{1,\Omega}$ for all $u \in H_0^1(\Omega)$. From here it follows that the sequence $\{\lambda_m^{\max}\}$, where for each m value $\lambda_m^{\max}$ is the largest eigenvalue of the eigenvalue problem

$$\mathsf{T}^\top\mathsf{H}\left(\tfrac{1}{2}\right)\mathsf{T}\mathsf{u} = \lambda\mathsf{A}_U\mathsf{u}, \tag{39}$$

is bounded from above by K_2. Note that the eigenvalue problem can be solved with a nontrivial eigenvalue only for $\mathsf{u} \in \mathbb{R}^n$ for which there exists some $\mathsf{q} \in \mathbb{R}^m$ such that $\mathsf{u} = \mathsf{T}^\top\mathsf{q}$. Consequently, the eigenvalue problem becomes $\mathsf{T}^\top\mathsf{H}\left(\frac{1}{2}\right)\mathsf{q} = \lambda\mathsf{A}_U\mathsf{T}^\top\mathsf{q}$. Next, applying the inverse of A_U and the trace matrix yields $\mathsf{T}\mathsf{A}_U{}^{-1}\mathsf{T}^\top\mathsf{H}\left(\frac{1}{2}\right)\mathsf{q} = \lambda\mathsf{q}$. Finally, setting $\mathsf{q} = \mathsf{H}\left(\frac{1}{2}\right)^{-1}\mathsf{p}$ yields

$$\mathsf{T}\mathsf{A}_U{}^{-1}\mathsf{T}^\top\mathsf{p} = \lambda\mathsf{H}\left(\tfrac{1}{2}\right)^{-1}\mathsf{p}. \tag{40}$$

Thus the largest eigenvalues of (39) and (40) coincide, and, in turn, $C_2 = K_2$. Further, (40) has only positive eigenvalues, and the smallest nonzero eigenvalue of (39) is the smallest eigenvalue $\lambda_m^{\min}$ of (40). Therefore, for all $\mathsf{y} \in \mathbb{R}^m$ it holds that $\lambda_m^{\min}\mathsf{y}^\top\mathsf{H}\left(\frac{1}{2}\right)^{-1}\mathsf{y} \leq \mathsf{y}^\top\mathsf{T}\mathsf{A}_U{}^{-1}\mathsf{T}^\top\mathsf{y}$. But the sequence $\{\lambda_m^{\min}\}$ is bounded from below since the right-inverse of the trace operator is bounded [36]. □

The proof of Lemma 8 suggests that the constants c_1, c_2 for spectral equivalence are computable as the limit of convergent sequences $\{\lambda_m^{\min}\}$, $\{\lambda_m^{\max}\}$ consisting of the smallest and largest eigenvalues of the generalized eigenvalue problem (40). Convergence of such sequences for the two geometries in Figure 1 is shown in Figure 2. For the simple geometry (a), the sequences converge rather fast, and the equivalence constants c_1, c_2 are clearly visible in the figure. Convergence on the more complex geometry (b) is slower.

So far we have by Theorem 4 and Lemma 5 that the condition numbers of matrices $\mathbb{B}_Q\mathbb{A}$ assembled over spaces (28) are bounded by constants independent of $\{h, \epsilon\}$. A more detailed characterization of the spectrum of the system preconditioned by the Q-cap preconditioner is given next. In particular, we relate the spectrum to computable bounds C_1, C_2 and characterize the distribution of eigenvalues. Further, the effect of varying ϵ (cf. Tables 3–4) is illustrated by numerical experiment.

3.5. Spectrum of the *Q*-cap preconditioned system.

In the following, the left-right preconditioning of $\mathbb{A}$ based on $\mathbb{B}_Q$ is considered, and we are interested in the spectrum of

$$\mathbb{B}_Q^{\frac{1}{2}}\mathbb{A}\mathbb{B}_Q^{\frac{1}{2}} = \begin{bmatrix} \mathsf{I}_U & & \mathsf{A}_U^{-\frac{1}{2}}\mathsf{B}_U{}^\top\mathsf{N}_Q{}^{\frac{1}{2}} \\ & \mathsf{I}_V & \mathsf{A}_V^{-\frac{1}{2}}\mathsf{B}_V{}^\top\mathsf{N}_Q{}^{\frac{1}{2}} \\ \mathsf{N}_Q^{\frac{1}{2}}\mathsf{B}_U\mathsf{A}_U^{-\frac{1}{2}} & \mathsf{N}_Q^{\frac{1}{2}}\mathsf{B}_V\mathsf{A}_V^{-\frac{1}{2}} & \end{bmatrix}. \tag{41}$$

The spectra of the left preconditioner system $\mathbb{B}_Q\mathbb{A}$ and the left-right preconditioned system $\mathbb{B}_Q^{\frac{1}{2}}\mathbb{A}\mathbb{B}_Q^{\frac{1}{2}}$ are identical. Using results of [41] the spectrum ρ of (41) is such that

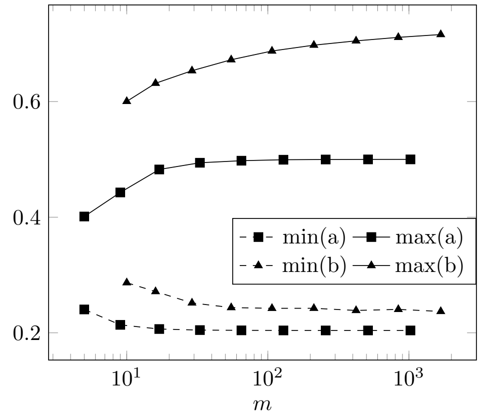

FIG. 2. *Convergence of sequences* $\{\lambda_m^{max}\}$ $\{\lambda_m^{min}\}$ *from Lemma* 8 *for geometries in Figure* 1*. For all sequences but max*(*b*) *the constant bound is reached within the considered range of discretization parameter* $m = n_Q$*.*

$\rho = I^- \cup I^+$ with

$$I^- = \left[\frac{1-\sqrt{1+4\sigma_{\max}^2}}{2}, \frac{1-\sqrt{1+4\sigma_{\min}^2}}{2}\right], \qquad I^+ = \left[1, \frac{1+\sqrt{1+4\sigma_{\max}^2}}{2}\right], \tag{42}$$

and $\sigma_{\min}, \sigma_{\max}$ the smallest and largest singular values of the block matrix formed by the first two row blocks in the last column of $\mathbb{B}_Q^{\frac{1}{2}}\mathbb{A}\mathbb{B}_Q^{\frac{1}{2}}$. We shall denote the matrix as $\mathbb{D}$:

$$\mathbb{D} = \begin{bmatrix} \mathsf{A}_U^{-\frac{1}{2}}\mathsf{B}_U{}^\top\mathsf{N}_Q{}^{\frac{1}{2}} \\ \mathsf{A}_V^{-\frac{1}{2}}\mathsf{B}_V{}^\top\mathsf{N}_Q{}^{\frac{1}{2}} \end{bmatrix}.$$

PROPOSITION 9. *The condition number* $\kappa(\mathbb{B}_Q\mathbb{A})$ *is bounded such that*

$$\kappa(\mathbb{B}_Q\mathbb{A}) \leq \frac{1+\sqrt{1+4C_2}}{1-\sqrt{1+4C_1}},$$

where C_1, C_2 *are the spectral equivalence bounds from Theorem* 7*.*

Proof. Note that the singular values of matrix $\mathbb{D}$ and the eigenvalues of matrix $\mathsf{N}_Q{}^{\frac{1}{2}}\mathsf{S}\mathsf{N}_Q{}^{\frac{1}{2}}$ are identical. Further, using Theorem 7 with $\mathsf{p} = \mathsf{N}_Q{}^{\frac{1}{2}}\mathsf{q}$, $\mathsf{q} \in \mathbb{R}^{n_Q}$ yields

$$C_1\mathsf{q}^\top\mathsf{q} \leq \mathsf{q}^\top\mathsf{N}_Q{}^{\frac{1}{2}}\mathsf{S}\mathsf{N}_Q{}^{\frac{1}{2}}\mathsf{q} \leq C_2\mathsf{q}^\top\mathsf{q} \quad \text{for all } q \in \mathbb{R}^{n_Q}.$$

In turn, the spectrum of matrices $\mathsf{N}_Q{}^{\frac{1}{2}}\mathsf{S}\mathsf{N}_Q^{\frac{1}{2}}$ is contained in the interval $[C_1, C_2]$. The statement now follows from (42). ∎

From numerical experiments we observe that the bound due to Proposition 9 slightly overestimates the condition number of the system. For example, using numerical trace bounds (cf. Figure 2) of geometry (a) in Figure 1, $c_1 = 0.204, c_2 = 0.499$, and Theorem 7, the formula yields 9.607 as the upper bound on the condition number. On the other hand, condition numbers reported in Table 3 do not exceed 8.637. Similarly, using estimated bounds for geometry (b), $c_1 = 0.237, c_2 = 0.716$, the formula gives the upper bound 8.676. The largest condition number in our experiments (not reported here) was 7.404.

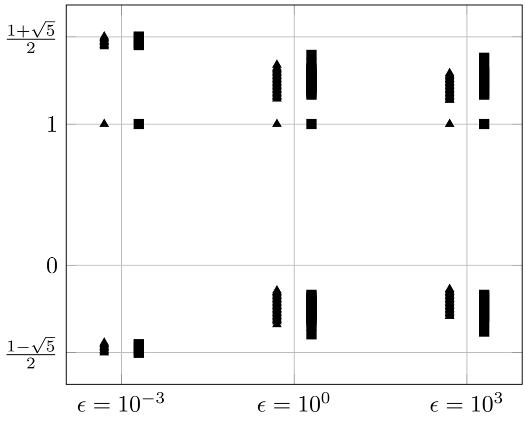

FIG. 3. *Eigenvalues of matrices $\mathbb{B}_Q\mathbb{A}$ assembled on geometries from Figure* 1 *for three different values of ϵ. The value of ϵ is indicated by gray vertical lines. On the left side of the lines is the spectrum for configuration* (a). *The spectrum for geometry* (b) *is then plotted on the right side. For $\epsilon \ll 1$ the eigenvalues cluster near $\lambda = 1$ and $\lambda = \frac{1}{2} \pm \frac{1}{2}\sqrt{5}$ (indicated by gray horizontal lines), which form the spectrum of $\mathbb{B}\mathbb{A}$.*

It is clear that (42) could be used to analyze the effect of the parameter ϵ on the spectrum provided that the singular values $\sigma_{\min}$, $\sigma_{\max}$ were given as functions of ϵ. We do not attempt to give this characterization here. Instead the effect of ϵ is illustrated by a numerical experiment. Figure 3 considers the spectrum of $\mathbb{B}_Q\mathbb{A}$ assembled on geometries from Figure 1 and three different values of the parameter. The systems from the two geometrical configurations are similar in size: 4355 for (a) and 4493 for (b). Note that for $\epsilon \ll 1$ the eigenvalues for both configurations cluster near $\lambda = 1$ and $\lambda = \frac{1}{2} \pm \frac{1}{2}\sqrt{5}$, that is, near the eigenvalues of $\mathbb{B}\mathbb{A}$. This observation is expected in light of the discussion following Theorem 7. With ϵ increasing, the difference between $\mathbb{B}_Q$ and $\mathbb{B}$ caused by $\mathsf{H}\left(-\frac{1}{2}\right)$ becomes visible as the eigenvalues are no longer clustered. Observe that in these cases the lengths of intervals I^-, I^+ are greater for geometry (b). This observation can be qualitatively understood via Proposition 9, Theorem 7, and Figure 2, where the trace map constants c_1, c_2 of configuration (a) are more widely spread than those of (b).

4. *W*-cap preconditioner.

To circumvent the need for mappings involving fractional Sobolev spaces, we shall next study a different preconditioner for (14). As will be seen, the new W-cap preconditioner (18) is still robust with respect to the material and discretization parameters.

Consider operator $\mathcal{A}$ from problem (15) as a mapping $W \times Q \to W^* \times Q^*$, with spaces W, Q defined as

$$\begin{aligned} W &= \left(H^1_0(\Omega) \cap \epsilon H^1_0(\Gamma)\right) \times H^1_0(\Gamma), \\ Q &= H^{-1}(\Gamma). \end{aligned} \tag{43}$$

The spaces are equipped with norms

$$\|w\|^2_W = |u|^2_{1,\Omega} + \epsilon^2 |T_\Gamma u|^2_{1,\Gamma} + |v|^2_{1,\Gamma} \quad \text{and} \quad \|p\|^2_Q = \|p\|^2_{-1,\Gamma}. \tag{44}$$

Note that the trace of functions from space U is here controlled in the norm $|\cdot|_{1,\Gamma}$ and not the fractional norm $\|\cdot\|_{\frac{1}{2},\Gamma}$, as was the case in section 3. Also note that the space

W now is dependent on ϵ while Q is not. The following result establishes the well-posedness of (14) with the above spaces.

THEOREM 10. *Let* W *and* Q *be the spaces* (43)*. The operator* $\mathcal{A} : W \times Q \to W^* \times Q^*$*, defined in* (15)*, is an isomorphism, and the condition number of* $\mathcal{A}$ *is bounded independently of* $\epsilon > 0$*.*

Proof. The proof proceeds by verifying the Brezzi conditions in Theorem 13. With $w = (u, v)$, $\omega = (\phi, \psi)$, application of the Cauchy–Schwarz inequality yields

$$\begin{aligned}\langle Aw, \omega\rangle_\Omega &= (\nabla u, \nabla\phi)_\Omega + (\nabla v, \nabla\psi)_\Gamma \\ &\leq |u|_{1,\Omega}|\phi|_{1,\Omega} + |v|_{1,\Gamma}|\psi|_{1,\Gamma} \\ &\leq |u|_{1,\Omega}|\phi|_{1,\Omega} + \epsilon^2|T_\Gamma u|_{1,\Gamma}|\phi|_{1,\Gamma} + |v|_{1,\Gamma}|\psi|_{1,\Gamma} \\ &\leq \|w\|_W\|\omega\|_W.\end{aligned}$$

Therefore, A is bounded with $\|A\| = 1$, and (51a) holds. The coercivity of A on $\ker B$ for (51b) is obtained from

$$\begin{aligned}\inf_{w\in\ker B}\frac{\langle Aw, w\rangle_\Omega}{\|w\|_W^2} &= \inf_{w\in\ker B}\frac{|u|_{1,\Omega}^2 + |v|_{1,\Gamma}^2}{|u|_{1,\Omega}^2 + \epsilon^2|T_\Gamma u|_{1,\Gamma}^2 + |v|_{1,\Gamma}^2} \\ &= \inf_{w\in\ker B}\frac{|u|_{1,\Omega}^2 + |v|_{1,\Gamma}^2}{|u|_{1,\Omega}^2 + 2|v|_{1,\Gamma}^2} \geq \frac{1}{2},\end{aligned}$$

where we used that $\epsilon T_\Gamma u = v$ a.e. on the kernel. Consequently, $\alpha = \frac{1}{2}$. Boundedness of B in (51c) with a constant $\|B\| = \sqrt{2}$ follows from the Cauchy–Schwarz inequality:

$$\begin{aligned}\langle Bw, q\rangle_\Gamma &\leq \|q\|_{-1,\Gamma}\epsilon|T_\Gamma u|_{1,\Gamma} + \|q\|_{-1,\Gamma}|v|_{1,\Gamma} \\ &\leq \sqrt{2}\|q\|_Q\sqrt{\epsilon^2|T_\Gamma u|_{1,\Gamma}^2 + |v|_{1,\Gamma}^2} \\ &\leq \sqrt{2}\|q\|_Q\sqrt{|u|_{1,\Omega}^2 + \epsilon^2|T_\Gamma u|_{1,\Gamma}^2 + |v|_{1,\Gamma}^2} \\ &\leq \sqrt{2}\|q\|_Q\|w\|_W.\end{aligned}$$

To show that the inf-sup condition holds, compute

$$\begin{aligned}\sup_{w\in W}\frac{\langle Bw, q\rangle_\Gamma}{\|w\|_W} &= \sup_{w\in W}\frac{\langle q, \epsilon T_\Gamma u - v\rangle_\Gamma}{\sqrt{|u|_{1,\Omega}^2 + \epsilon^2|T_\Gamma u|_{1,\Gamma}^2 + |v|_{1,\Gamma}^2}} \\ &\overset{u=0}{\geq} \sup_{v\in V}\frac{\langle q, v\rangle_\Gamma}{|v|_{1,\Gamma}} = \|q\|_Q.\end{aligned}$$

Thus $\beta = 1$ in condition (51d). □

Following Theorem 10, the operator $\mathcal{A}$ is a symmetric isomorphism between spaces $W \times Q$ and $W^* \times Q^*$. As a preconditioner we shall consider a symmetric positive-definite isomorphism $W^* \times Q^* \to W \times Q$:

$$\mathcal{B}_W = \begin{bmatrix}\left(-\Delta_\Omega + T_\Gamma^*\left(-\epsilon^2\Delta_\Gamma\right)T_\Gamma\right)^{-1} & & \\ & \left(-\Delta_\Gamma\right)^{-1} & \\ & & -\Delta_\Gamma\end{bmatrix}. \tag{45}$$

4.1. Discrete preconditioner. Similar to section 3.1, we shall construct discretizations $W_h \times Q_h$ of space $W \times Q$ (43) such that the finite dimensional operator $\mathcal{A}_h$ defined by considering $\mathcal{A}$ from (15) on the constructed spaces satisfies the Brezzi conditions in Theorem 13.

Let $W_h \subset W$ and $Q_h \subset Q$ be the spaces (28) of continuous piecewise linear polynomials. Then A_h, B_h are continuous with respect to norms (44), and it remains to verify conditions (51a) and (51d). First, coercivity of A_h is considered.

LEMMA 11. *Let* W_h, Q_h *be the spaces* (28)*, and let* A_h, B_h *be such that* $\langle Aw, \omega_h\rangle_\Omega = \langle A_h w_h, \omega_h\rangle_\Omega$, $\langle Bw, q_h\rangle_\Gamma = \langle B_h w_h, q_h\rangle_\Gamma$ *for* $\omega_h, w_h \in W_h$, $w \in W$, *and* $q_h \in Q_h$. *Then there exists a constant* $\alpha > 0$ *such that, for all* $z_h \in \ker B_h$,

$$\langle A_h z_h, z_h\rangle \geq \alpha \|z_h\|_W,$$

where $\|\cdot\|_W$ *is defined in* (44).

Proof. The claim follows from coercivity of A over $\ker B$ (cf. Theorem 10) and the property $\ker B_h \subset \ker B$. To see that the inclusion holds, let $z_h \in \ker B_h$. Since z_h is continuous on Γ, we have from definition $\langle z_h, q_h\rangle_\Gamma = 0$ for all $q_h \in Q_h$ that $z_h|_\Gamma = 0$. But then $\langle z_h, q\rangle = 0$ for all $q \in Q$, and therefore $z_h \in \ker B$. □

Finally, to show that the discretization $W_h \times Q_h$ is stable, we show that the inf-sup condition for B_h holds.

LEMMA 12. *Let spaces* W_h, Q_h *and operator* B_h *from Lemma* 11 *be given. Then there exists* $\beta > 0$ *such that*

$$\inf_{q_h \in Q_h} \sup_{w_h \in W_h} \frac{\langle B_h w_h, q_h\rangle_\Gamma}{\|w_h\|_W \|q_h\|_Q} \geq \beta, \tag{46}$$

where $\|\cdot\|_Q$ *is defined in* (44).

Proof. We first proceed as in the proof of Theorem 10 and compute

$$\sup_{w_h \in W_h} \frac{\langle q_h, \epsilon T_\Gamma u_h - v_h\rangle_\Gamma}{\|w_h\|_W} \overset{u_h=0}{\geq} \sup_{v_h \in V_h} \frac{\langle v_h, q_h\rangle_\Gamma}{|v_h|_{1,\Gamma}}. \tag{47}$$

Next, for each $p \in H^1_0(\Gamma)$, let $v_h = \Pi p$ be the element of V_h defined in the proof of Lemma 5. In particular, it holds that

$$\langle p - v_h, q_h\rangle_\Gamma = 0, \quad q_h \in Q_h,$$

and $|v_h|_{1,\Gamma} \leq C|p|_{1,\Gamma}$ for some constant C depending only on Ω and Γ. Then

$$\|q_h\|_{-1,\Gamma} = \sup_{p \in H^1_0(\Gamma)} \frac{\langle q_h, p\rangle_\Gamma}{|p|_{1,\Gamma}} \leq C \sup_{v_h \in V_h} \frac{\langle q_h, v_h\rangle_\Gamma}{|v_h|_{1,\Gamma}}.$$

The estimate together with (47) proves the claim of the lemma. □

Let now $\mathsf{A}_U, \mathsf{A}_V$ and $\mathsf{B}_U, \mathsf{B}_V$ be the matrices defined in (23) as representations of the corresponding finite dimensional operators in the basis of the stable spaces W_h and Q_h. We shall represent the preconditioner $\mathcal{B}_W$ by a matrix

$$\mathbb{B}_W = \begin{bmatrix} \left(\mathsf{A}_U + \epsilon^2 \mathsf{T}^\top \mathsf{A} \mathsf{T}\right)^{-1} & & \\ & \left(\mathsf{A}_V\right)^{-1} & \\ & & \mathsf{H}(-1)^{-1} \end{bmatrix}, \tag{48}$$

where $\mathsf{H}(-1)^{-1} = \mathsf{M}^{-1}\mathsf{A}\mathsf{M}^{-1}$ (cf. (8)) and M, A are the matrices inducing L^2 and H_0^1 inner products on Q_h. Let us point out that there is an obvious correspondence between the matrix preconditioner $\mathbb{B}_W$ and the operator $\mathcal{B}_W$ defined in (18). On the other hand, it is not entirely straightforward that the matrix $\mathbb{B}_W$ represents the W-cap preconditioner defined in (45). In particular, since the isomorphism from $Q^* = H_0^1(\Gamma)$ to $Q = H^{-1}(\Gamma)$ is realized by the Laplacian, a case could be made for using the stiffness matrix A as a suitable representation of the operator.

Let us first argue for A not being a suitable representation for preconditioning. Note that the role of matrix $\mathbb{A} \in \mathbb{R}^{m\times n}$ in a linear system $\mathbb{A}\mathsf{x} = \mathsf{b}$ is to transform vectors from the solution space $\mathbb{R}^n$ to the residual space $\mathbb{R}^m$. In the case when the matrix is invertible, the spaces coincide. However, to emphasize the conceptual difference between the spaces, let us write $\mathbb{A} : \mathbb{R}^n \to \mathbb{R}^{n*}$. Then a preconditioner matrix is a mapping $\mathbb{B} : \mathbb{R}^{n*} \to \mathbb{R}^n$. The stiffness matrix A, however, is such that $\mathsf{A} : \mathbb{R}^{n_Q} \to \mathbb{R}^{n_Q*}$.

It remains to show that $\mathsf{M}^{-1}\mathsf{A}\mathsf{M}^{-1}$ is the correct representation of $A = -\Delta_\Gamma$. Recall that $Q_h \subset Q^*$ and A is the matrix representation of operator $A_h : Q_h \to Q_h^*$. Further, mappings $\pi_h : Q_h \to \mathbb{R}^{n_Q}$, $\mu_h : Q_h^* \to \mathbb{R}^{n_Q*}$,

$$p_h = \sum_j (\pi_h p_h)_j \chi_j, \quad p_h \in Q_h, \qquad \text{and} \qquad (\mu_h f_h)_j = \langle f_j, \chi_j\rangle, \quad f_h \in Q_h^*,$$

define isomorphisms between[3] spaces Q_h, $\mathbb{R}^{n_Q}$ and Q_h^*, $\mathbb{R}^{n_Q*}$, respectively. We can uniquely associate each $p_h \in Q_h$ with a functional in Q_h^* via the Riesz map $I_h : Q_h \to Q_h^*$ defined as $\langle I_h p_h, q_h\rangle_\Gamma = (p_h, q_h)_\Gamma$. Since

$$\left(\mu_h I_h p_h\right)_j = \left(I_h p_h, \chi_j\right)_\Gamma = \sum_i \left(\pi_h p_h\right)_i \left(\chi_i, \chi_j\right)_\Gamma,$$

the operator I_h is represented as the mass matrix M. The matrix then provides a natural isomorphism from $\mathbb{R}^{n_Q}$ to $\mathbb{R}^{n_Q*}$. In turn, $\mathsf{M}^{-1}\mathsf{A}\mathsf{M}^{-1} : \mathbb{R}^{n_Q*} \to \mathbb{R}^{n_Q}$ has the desired mapping properties. In conclusion, the inverse of the mass matrix was used in (48) as a natural adapter to obtain a matrix operating between spaces suitable for preconditioning.

Finally, we make a few observations about the matrix preconditioner $\mathbb{B}_W$. Recall that the Q-cap preconditioner $\mathbb{B}_Q$ could be related to the Schur complement based preconditioner (35) obtained by factorizing $\mathbb{A}$ in (24). The relation of $\mathbb{A}$ to the W-cap preconditioner matrix (48) is revealed in the following calculation:

$$\mathbb{U}\mathbb{L}\mathbb{A} = \begin{bmatrix} \mathsf{A}_V + \epsilon^2\mathsf{T}^\top\mathsf{A}\mathsf{T} & & \\ & \tau^2\mathsf{A} & -\mathsf{M} \\ -\epsilon\mathsf{M}\mathsf{T} & & \mathsf{M}\mathsf{A}^{-1}\mathsf{M} \end{bmatrix}, \tag{49}$$

where

$$\mathbb{U} = \begin{bmatrix} \mathsf{I} & & -\mathsf{T}^\top\epsilon\mathsf{A}\mathsf{M}^{-1} \\ & \mathsf{I} & \\ & & \mathsf{I} \end{bmatrix} \quad \text{and} \quad \mathbb{L} = \begin{bmatrix} \mathsf{I} & & \\ & \mathsf{I} & \\ & -\mathsf{M}\mathsf{A}^{-1} & -\mathsf{I} \end{bmatrix}.$$

[3] Note that in section 1 the mapping μ_h was considered as $\mu_h : Q_h^* \to \mathbb{R}^{n_Q}$. The definition used here reflects the conceptual distinction between spaces $\mathbb{R}^{n_Q}$ and $\mathbb{R}^{n_Q*}$. That is, μ_h is viewed as a map from the space of right-hand sides of the operator equation $A_h p_h = L_h$ to the space of right-hand sides of the corresponding matrix equation $\mathsf{A}\mathsf{p} = \mathsf{b}$.

Here the matrix $\mathbb{L}$ introduces a Schur complement of a submatrix of $\mathbb{A}$ corresponding to spaces V_h, Q_h.The matrix $\mathbb{U}$ then eliminates the constraint on the space U_h. Preconditioner $\mathbb{B}_W$ could now be interpreted as coming from the diagonal of the resulting matrix in (49). Further, note that the action of the Q_h-block can be computed cheaply by Jacobi iterations with a diagonally preconditioned mass matrix (cf. [47]).

Table 5

Spectral condition numbers of matrices $\mathbb{B}_W\mathbb{A}$ for the system assembled on geometry (a) *in Figure* 1.

Size	$\log_{10}\epsilon$						
	-3	-2	-1	0	1	2	3
99	2.619	2.627	2.546	3.615	3.998	4.044	4.048
323	2.623	2.653	2.780	3.813	4.023	4.046	4.049
1155	2.631	2.692	3.194	3.925	4.036	4.048	4.049
4355	2.644	2.740	3.533	3.986	4.042	4.048	4.049
16899	2.668	2.788	3.761	4.017	4.046	4.049	4.049
66563	2.703	3.066	3.896	4.033	4.047	4.049	4.049

4.2. Numerical experiments. Parameter robust properties of the W-cap preconditioner are demonstrated by the two numerical experiments used to validate the Q-cap preconditioner in section 3.3. Both experiments use discretization of domain (a) from Figure 1. First, using the *exact* preconditioner, we consider the spectral condition numbers of matrices $\mathbb{B}_W\mathbb{A}$. Next, using an approximation of $\mathbb{B}_W$, the linear system $\bar{\mathbb{B}}_W\mathbb{A}\mathsf{x} = \bar{\mathbb{B}}_W\mathsf{f}$ is solved with the minimal residual method. The operator $\bar{\mathbb{B}}_W$ is defined as

$$\bar{\mathbb{B}}_W = \begin{bmatrix} \mathrm{AMG}\big(\mathsf{A}_U + \epsilon^2\mathsf{T}^\top\mathsf{A}\mathsf{T}\big) & & \\ & \mathrm{LU}(\mathsf{A}) & \\ & & \mathrm{LU}\,(\mathsf{M})\;\mathsf{A}\,\mathrm{LU}\,(\mathsf{M}) \end{bmatrix}. \tag{50}$$

The spectral condition numbers of matrices $\mathbb{B}_W\mathbb{A}$ for different values of material parameter ϵ are listed in Table 5. For all the considered discretizations, the condition numbers are bounded with respect to ϵ. We note that the mesh convergence of the condition numbers appears to be faster and the obtained values are in general smaller than in case of the Q-cap preconditioner (cf. Table 3).

Table 6 reports the number of iterations required for convergence of the minimal residual method for the linear system $\bar{\mathbb{B}}_W\mathbb{A}\mathsf{x} = \bar{\mathbb{B}}_W\mathsf{f}$. Like for the Q-cap preconditioner, the method is started from a random initial vector, and the condition $\mathsf{r}_k{}^\top\bar{\mathbb{B}}_W\mathsf{r}_k < 10^{-12}$ is used as a stopping criterion. We find that the iteration counts with the W-cap preconditioner are again bounded for all the values of the parameter ϵ. Consistent with the observations about the spectral condition number, the iteration count is in general smaller than for the system preconditioned with the Q-cap preconditioner.

We note that the observations from section 3.3 about the difference in ϵ-dependence of condition numbers and iteration counts of the Q-cap preconditioner apply to the W-cap preconditioner as well.

Before addressing the question of computational costs of the proposed preconditioners, let us remark that the Q-cap preconditioner and the W-cap preconditioner are not spectrally equivalent. Further, both preconditioners yield numerical solutions with linearly (optimally) converging error; see Appendix B.

5. Computational costs. We conclude by assessing computational efficiency of the proposed preconditioners. In particular, the setup cost and its relation to the

Table 6

Iteration count for system $\bar{\mathbb{B}}_W \mathbb{A}\mathsf{x} = \bar{\mathbb{B}}_W \mathsf{f}$ solved with the minimal residual method. The problem is assembled on geometry (a) *from Figure* 1. *A comparison to the number of iterations with the Q-cap preconditioned system is shown in the brackets (cf. also Table* 4*).*

Size	$\log_{10}\epsilon$						
	-3	-2	-1	0	1	2	3
66563	17(-3)	33(-1)	40(3)	30(-2)	20(-8)	14(-10)	12(-9)
264195	19(-3)	35(1)	39(5)	28(-2)	19(-7)	14(-10)	11(-9)
1052675	22(-2)	34(1)	37(5)	27(-1)	19(-7)	14(-8)	11(-7)
4202499	24(-2)	34(2)	34(4)	25(-1)	17(-7)	12(-8)	9(-8)
8398403	25(-1)	32(2)	32(2)	24(-2)	16(-6)	11(-8)	8(-7)
11075583	25(-1)	32(2)	32(2)	25(-1)	16(-6)	13(-6)	11(-4)

aggregate solution time of the Krylov method is of interest. For simplicity we let $\epsilon = 1$.

In case of the Q-cap preconditioner discretized as (34) the setup cost is determined by the construction of algebraic multigrid (AMG) and the solution of the generalized eigenvalue problem $\mathsf{A}\mathsf{x} = \lambda\mathsf{M}\mathsf{x}$ (GEVP). The problem is here solved by calling the OpenBLAS [46] implementation of the LAPACK [3] routine DSYGVD. The setup cost of the W-cap preconditioner is dominated by the construction of multigrid for operator $\mathsf{A}_U + \mathsf{T}^\top\mathsf{A}\mathsf{T}$. We found that the operator can be assembled with negligible costs and therefore do not report the timings of this operation.

The setup costs of the preconditioners obtained on a Linux machine with 16GB RAM and a single Intel Core i5-2500 CPU clocking at 3.3 GHz are reported in Table 7. We remark that the timings on the finest discretization deviate from the trend set by the predecessors. This is due to SWAP memory being required to complete the operations and the case should therefore be omitted from the discussion. On the remaining discretizations the following observations can be made: (i) the solution time always dominates the construction time by a factor 5.5 for W-cap and 3.5 for Q-cap; (ii) W-cap preconditioner is close to two times cheaper to construct than the Q-cap preconditioner in the form (34); (iii) the eigenvalue problem always takes fewer seconds to solve than the construction of multigrid.

For our problems of about 11 million nodes in the $2d$ domain, the strategy of solving the generalized eigenvalue problem using a standard LAPACK routine provided an adequate solution. However, the DSYGVD routine appears to be nearly cubic in complexity ($\mathcal{O}(n_Q^{2.70})$ or $\mathcal{O}(n_U^{1.35})$; cf. Table 7), which may represent a bottleneck for larger problems. However, the transformation $\mathsf{M}_l^{-\frac{1}{2}}\mathsf{A}\mathsf{M}_l^{-\frac{1}{2}}$ with M_l the lumped mass matrix presents a simple trick providing significant speed-up. In fact, the resulting eigenvalue problem is symmetric and tridiagonal and can be solved with fast algorithms of nearly quadratic complexity [20, 21]. We note that due to the spectral equivalence of M and M_l (e.g., [47]), the trick leads to a preconditioner spectrally equivalent to (25). In particular, the iteration count with lumping is expected to remain bounded. In our experiments (not reported here) the lumped preconditioner leads to convergence in 3–10 fewer iterations than (34). However, the savings should be interpreted in light of the fact that convergence in the two cases is measured with respect to different norms. Note also that the tridiagonal property holds under the assumption of Γ having no bifurcations and that the elements are linear. To illustrate the potential gains with mass lumping, using the transformation and applying the dedicated LAPACK routine DSTEGR, we were able to compute eigenpairs for systems of order 16,000 in about 50 seconds. This presents more than a factor 10 speed-up relative to the original generalized eigenvalue problem. The value should also be viewed in light of the fact that the relevant space U_h has in this case about a quarter billion degrees of freedom. We remark that [28] presents a method for computing all the eigenpairs of the generalized symmetric tridiagonal eigenvalue problem with an estimated quadratic complexity.

Let us briefly mention a few alternative methods for realizing the mapping between fractional Sobolev spaces needed by the Q-cap preconditioner. The methods have a common feature of computing the action of operators rather than constructing the operators themselves. Taking advantage of the fact that $\mathsf{H}(s) = \mathsf{MS}^{-s}$, $\mathsf{S} = \mathsf{A}^{-1}\mathsf{M}$, the action of the powers of the matrix S is efficiently computable by contour integrals [25], by the symmetric Lanczos process [4, 5], or, in cases when the matrices A, M are structured, by fast Fourier transform [39]. Alternatively, the mapping can be realized by the BPX preconditioner [12, 11] or integral operator based preconditioners (e.g., [43]). The above-mentioned techniques are all less than $\mathcal{O}(n_Q^2)$ in complexity.

In summary, for linear elements and geometrical configurations where Γ is free of bifurcations, the eigenvalue problem required for (8) lends itself to solution methods with complexity nearing that of the multigrid construction. In such cases the Q-cap preconditioner (34) is feasible whenever the methods deliver acceptable performance ($n_Q \sim 10^4$). For larger spaces Q_h, a practical realization of the Q-cap preconditioner could be achieved by one of the listed alternatives.

Table 7

Timings of elements of construction of the Q, W-cap for $\epsilon = 1$ and discretizations from Tables 4 and 6. Estimated complexity of computing quantity v at the ith row, $r_i = {}^{\log v_i - \log v_{i-1}}/_{\log m_i - \log m_{i-1}}$, is shown in the brackets. Fitted complexity of computing v, $\mathcal{O}(n_Q^r)$ is obtained by least-squares. All fits but GEVP ignore the SWAP-affected final discretization.

n_U	n_Q	Q-cap			W-cap	
		AMG[s]	GEVP[s]	MinRes[s]	AMG[s]	MinRes[s]
66049	257	0.075(1.98)	0.014(1.81)	0.579(1.69)	0.078(1.94)	0.514(1.73)
263169	513	0.299(2.01)	0.066(2.27)	2.286(1.99)	0.309(1.99)	2.019(1.98)
1050625	1025	1.201(2.01)	0.477(2.87)	8.032(1.82)	1.228(1.99)	7.909(1.97)
4198401	2049	4.983(2.05)	3.311(2.80)	30.81(1.94)	4.930(2.01)	30.31(1.94)
8392609	2897	9.686(1.92)	8.384(2.68)	62.67(2.05)	10.64(2.22)	59.13(1.93)
11068929	3327	15.94(3.60)	12.25(2.74)	84.43(2.15)	15.65(2.79)	82.13(2.37)
Fitted complexity		(2.02)	(2.70)	(1.92)	(2.02)	(1.96)

6. Conclusions.

We have studied preconditioning of model multiphysics problem (1) with Γ being the subdomain of Ω having codimension one. Using operator preconditioning [35], two robust preconditioners were proposed and analyzed. Theoretical findings obtained in the present treatise about robustness of preconditioners with respect to material and discretization parameter were demonstrated by numerical experiments using a stable finite element approximation for the related saddle-point problem developed herein. Computational efficiency of the preconditioners was assessed revealing that the W-cap preconditioner is more practical. The Q-cap preconditioner with discretization based on eigenvalue factorization is efficient for smaller problems, and its application to large scale computing possibly requires different means of realizing the mapping between the fractional Sobolev spaces.

Possible future work based on the presented ideas includes extending the preconditioners to problems coupling three-dimensional and one-dimensional domains, problems with multiple disjoint subdomains, and problems describing different physics on the coupled domains. In addition, a finite element discretization of the problem which avoids the constraint of Γ_h being aligned with facets of Ω_h is of general interest.

Appendix A. Brezzi theory.

THEOREM 13 (Brezzi). *The operator* $\mathcal{A} : V \times Q \to V^* \times Q^*$ *in* (16) *is an isomorphism if the following conditions are satisfied:*
(a) *A is bounded,*

$$\sup_{u\in V}\sup_{v\in V}\frac{\langle Au, v\rangle}{\|u\|_V\|v\|_V} = c_A \equiv \|A\| < \infty; \tag{51a}$$

(b) *A is invertible on* $\ker B$*, with*

$$\inf_{u\in\ker B}\frac{\langle Au, u\rangle}{\|u\|_V^2} \geq \alpha > 0; \tag{51b}$$

(c) *B is bounded,*

$$\sup_{q\in Q}\sup_{v\in V}\frac{\langle Bv, q\rangle}{\|v\|_V\|q\|_Q} = c_B \equiv \|B\| < \infty; \tag{51c}$$

(d) *B is surjective (this is also the inf-sup or LBB condition), with*

$$\inf_{q\in Q}\sup_{v\in V}\frac{\langle Bv, q\rangle}{\|v\|_V\|q\|_Q} \geq \beta > 0. \tag{51d}$$

The operator norms $\|\mathcal{A}\|$ and $\|\mathcal{A}^{-1}\|$ are bounded in terms of the constants appearing in (a)–(d).

Proof. See, for example, [14]. □

Appendix B. Estimated order of convergence. Refinements of a uniform discretization of geometry (a) in Figure 1 are used to establish order of convergence of numerical solutions of a manufactured problem obtained using Q-cap and W-cap preconditioners. The error of discrete solutions u_h and v_h is interpolated by discontinuous piecewise cubic polynomials and measured in the H_0^1 norm. The observed convergence rate is linear (optimal).

Size	Q-cap		W-cap	
	$\lvert u-u_h\rvert_{1,\Omega}$	$\lvert v-v_h\rvert_{1,\Gamma}$	$\lvert u-u_h\rvert_{1,\Omega}$	$\lvert v-v_h\rvert_{1,\Gamma}$
16899	$3.76\times10^{-2}(1.00)$	$1.32\times10^{-2}(1.00)$	$3.76\times10^{-2}(1.00)$	$1.32\times10^{-2}(1.00)$
66563	$1.88\times10^{-2}(1.00)$	$6.58\times10^{-3}(1.00)$	$1.88\times10^{-2}(1.00)$	$6.58\times10^{-3}(1.00)$
264195	$9.39\times10^{-3}(1.00)$	$3.29\times10^{-3}(1.00)$	$9.39\times10^{-3}(1.00)$	$3.29\times10^{-3}(1.00)$
1052675	$4.70\times10^{-3}(1.00)$	$1.64\times10^{-3}(1.00)$	$4.70\times10^{-3}(1.00)$	$1.64\times10^{-3}(1.00)$
4202499	$2.35\times10^{-3}(1.00)$	$8.22\times10^{-4}(1.00)$	$2.35\times10^{-3}(1.00)$	$8.22\times10^{-4}(1.00)$

Acknowledgment. We would like to thank the anonymous referees for their valuable and constructive comments, which improved the presentation of this paper.

REFERENCES

[1] R. A. Adams and J. F. Fournier, *Sobolev Spaces*, 2nd ed., Pure Appl. Math. 140, Elsevier, Academic Press, Amsterdam, 2003.

[2] I. Ambartsumyan, E. Khattatov, I. Yotov, and P. Zunino, *Simulation of flow in fractured poroelastic media: A comparison of different discretization approaches*, in Finite Difference Methods, Theory and Applications, I. Dimov, I. Faragó, and L. Vulkov, eds., Lecture Notes in Comput. Sci. 9045, Springer, Berlin, 2015, pp. 3–14.

[3] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed., SIAM, Philadelphia, 1999.

[4] M. Arioli, D. Kourounis, and D. Loghin, *Discrete fractional Sobolev norms for domain decomposition preconditioning*, IMA J. Numer. Anal., 33 (2012), pp. 318–342.

[5] M. Arioli and D. Loghin, *Discrete interpolation norms with applications*, SIAM J. Numer. Anal., 47 (2009), pp. 2924–2951, https://doi.org/10.1137/080729360.

[6] I. Babuška, *The finite element method with Lagrangian multipliers*, Numer. Math., 20 (1973), pp. 179–192.

[7] S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, *PETSc Users' Manual*, Tech. Report ANL-95/11, Revision 3.4, Argonne National Laboratory, Lemont, IL, 2013.

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

[9] J. Bergh and J. Löfström, *Interpolation Spaces. An Introduction*, Grundlehren Math. Wiss. 223, Springer, Berlin, 1976.

[10] D. Braess, *Finite Elements*, 3rd ed., Cambridge University Press, Cambridge, UK, 2007.

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

[12] J. H. Bramble, J. E. Pasciak, and J. Xu, *Parallel multilevel preconditioners*, Math. Comp., 55 (1990), pp. 1–22.

[13] F. Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, Rev. Française Automat. Informat. Recherche Opérationnelle Ser. Rouge, 8 (1974), pp. 129–151.

[14] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Ser. Comput. Math. 15, Springer, New York, 1991.

[15] L. Cattaneo and P. Zunino, *Computational models for fluid exchange between microcirculation and tissue interstitium*, Netw. Heterog. Media, 9 (2014), pp. 135–159.

[16] S. N. Chandler-Wilde, D. P. Hewett, and A. Moiola, *Interpolation of Hilbert and Sobolev spaces: Quantitative estimates and counterexamples*, Mathematika, 61 (2015), pp. 414–443.

[17] C. D'Angelo, *Finite element approximation of elliptic problems with Dirac measure terms in weighted spaces: Applications to one- and three-dimensional coupled problems*, SIAM J. Numer. Anal., 50 (2012), pp. 194–215, https://doi.org/10.1137/100813853.

[18] 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.

[19] T. A. Davis, *Algorithm* 832: *Umfpack V*4.3*—an unsymmetric-pattern multifrontal method*, ACM Trans. Math. Software, 30 (2004), pp. 196–199.

[20] J. W. Demmel, O. A. Marques, B. N. Parlett, and C. Vömel, *Performance and accuracy of LAPACK's symmetric tridiagonal eigensolvers*, SIAM J. Sci. Comput., 30 (2008), pp. 1508–1526, https://doi.org/10.1137/070688778.

[21] S. I. Dhillon and B. N. Parlett, *Multiple representations to compute orthogonal eigenvectors of symmetric tridiagonal matrices*, Linear Algebra Appl., 387 (2004), pp. 1–28.

[22] J. Etienne, J. Lohéac, and P. Saramito, *A Lagrange-Multiplier Approach for the Numerical Simulation of an Inextensible Membrane or Thread Immersed in a Fluid*, preprint, https://hal.inria.fr/inria-00449805, 2010.

[23] R. D. Falgout and U. Meier Yang, *Hypre: A library of high performance preconditioners*, in Computational Science, ICCS 2002, P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, and J. J. Dongarra, eds., Lecture Notes in Comput. Sci. 2331, Springer, Berlin, Heidelberg, 2002, pp. 632–641.

[24] S. A. Funken and E. P. Stephan, *Hierarchical basis preconditioners for coupled FEM-BEM equations*, in Boundary Elements: Implementation and Analysis of Advanced Algorithms, W. Hackbusch and G. Wittum, eds., Notes Numer. Fluid Mech. 50, Vieweg+Teubner Verlag, Berlin, 1996, pp. 92–101.

[25] N. Hale, N. J. Higham, and L. N. Trefethen, *Computing* A^{α}, $\log(A)$*, and related matrix functions by contour integrals*, SIAM J. Numer. Anal., 46 (2008), pp. 2505–2523, https://doi.org/10.1137/070700607.

[26] H. Harbrecht, F. Paiva, C. Pérez, and R. Schneider, *Multiscale preconditioning for the coupling of FEM-BEM*, Numer. Linear Algebra Appl., 10 (2003), pp. 197–222.

[27] V. Hernandez, J. S. Roman, and V. Vidal, *SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems*, ACM Trans. Math. Software, 31 (2005), pp. 351–362.

[28] K. Li, T.-Y. Li, and Z. Zeng, *An algorithm for the generalized symmetric tridiagonal eigenvalue problem*, Numer. Algorithms, 8 (1994), pp. 269–291.

[29] J. Liesen and . Tichý, *Convergence analysis of Krylov subspace methods*, GAMM Mitt. Ges. Angew. Math. Mech., 27 (2004), pp. 153–173.

[30] J. L. Lions and E. Magenes, *Non-Homogeneous Boundary Value Problems and Applications, Vol.* 1, Grundlehren Math. Wiss. 181, Springer, Berlin, 1972.

[31] A. Logg, K.-A. Mardal, and G. N. Wells, eds., *Automated Solution of Differential Equations by the Finite Element Method*, Lect. Notes Comput. Sci. Eng. 84, Springer, Berlin, 2012.

[32] F. Magouls and F. X. Roux, *Lagrangian formulation of domain decomposition methods: A unified theory*, Appl. Math. Model., 30 (2006), pp. 593–615.

[33] J. Málek and Z. Strakoš, *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs*, SIAM Spotlights 1, SIAM, Philadelphia, 2015.

[34] K.-A. Mardal and J. B. Haga, *Block preconditioning of systems of PDEs*, in Automated Solution of Differential Equations by the Finite Element Method, A. Logg, K.-A. Mardal, and G. N. Wells, eds., Lect. Notes Comput. Sci. Eng. 84, Springer, Berlin, 2012, pp. 643–655.

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

[36] J. Marschall, *The trace of Sobolev-Slobodeckij spaces on Lipschitz domains*, Manuscripta Math., 58 (1987), pp. 47–65.

[37] 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, https://doi.org/10.1137/S1064827599355153.

[38] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629, https://doi.org/10.1137/0712047.

[39] P. Peisker, *On the numerical solution of the first biharmonic equation*, RAIRO Modél. Math. Anal. Numér., 22 (1988), pp. 655–676.

[40] J. Pitkäranta, *Boundary subspaces for the finite element method with Lagrange multipliers*, Numer. Math., 33 (1979), pp. 273–289.

[41] T. Rusten and R. Winther, *A preconditioned iterative method for saddlepoint problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887–904, https://doi.org/10.1137/0613054.

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

[43] O. Steinbach and W. L. Wendland, *The construction of some efficient preconditioners in the boundary element method*, Adv. Comput. Math., 9 (1998), pp. 191–216.

[44] S. Timoshenko, *Theory of Elastic Stability*, 2nd ed., Engineering Societies Monographs, McGraw–Hill, New York, 1961.

[45] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[46] Q. Wang, X. Zhang, Y. Zhang, and Q. Yi, *AUGEM: Automatically generate high performance dense linear algebra kernels on* x86 *CPUs*, in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC '13, ACM, New York, 2013, 25, https://doi.org/10.1145/2503210.2503219.

[47] A. J. Wathen, *Realistic eigenvalue bounds for the Galerkin mass matrix*, IMA J. Numer. Anal., 7 (1987), pp. 449–457.

[48] K. Yosida, *Functional Analysis*, 6th ed., Springer, New York, 1980.

Paper II

On preconditioning saddle point systems with trace constraints coupling 3d and 1d domains – applications to matching and nonmatching FEM discretizations

M. Kuchta, K.-A. Mardal, and M. Mortensen

SIAM Journal on Scientific Computing, (2016). Submitted.

ON PRECONDITIONING SADDLE POINT SYSTEMS WITH TRACE CONSTRAINTS COUPLING 3D AND 1D DOMAINS – APPLICATIONS TO MATCHING AND NONMATCHING FEM DISCRETIZATIONS [*]

MIROSLAV KUCHTA [†], KENT-ANDRE MARDAL [†‡], AND MIKAEL MORTENSEN [†]

Abstract. Multiscale or multiphysics problems often involve the coupling of partial differential equations posed on domains of different dimensionality. In this work we consider a simplified model problem of a 3d-1d coupling and the main motivation is to construct algorithms that may utilize standard multilevel algorithms for the 3d domain, which has the dominating computational complexity. Preconditioning for a system of two elliptic problems posed, respectively, in a three dimensional domain and an embedded one dimensional curve and coupled by the trace constraint is discussed. Investigating numerically the properties of the well-defined discrete trace operator, it is found that negative fractional Sobolev norms are suitable preconditioners for the Schur complement. The norms are employed to construct a robust block diagonal preconditioner for the coupled problem.

Key words. preconditioning, saddle-point problem, Lagrange multipliers, trace

AMS subject classifications. 65F08

1. Introduction.

Let Ω be a bounded domain in 3d, while Γ represents a 1d structure inside Ω, and consider the following coupled problem

$$-\Delta u + u + p\delta_\Gamma = f \qquad \text{in } \Omega, \tag{1.1a}$$
$$-\Delta v + v - p = g \qquad \text{on } \Gamma, \tag{1.1b}$$
$$Tu - v = h \qquad \text{on } \Gamma. \tag{1.1c}$$

Here the term $p\delta_\Gamma$ is to be understood as a Dirac measure such that $\int_\Omega p(x)\delta_\Gamma v(x)\,\mathrm{d}x = \int_\Gamma p(t)v(t)\,\mathrm{d}t$ for a continuous function v. We remark that from a mathematical point of view the trace T of u required in (1.1c) is in the continuous case not well-defined unless the functions are sufficiently regular. For simplicity, the system shall be considered with homogeneous Neumann boundary conditions.

The system (1.1) is relevant in numerous biological applications where the embedded (three dimensional) structure is such that order reduction techniques can be used to capture its response by a one dimensional model. Equation (1.1a) then models processes in the bulk, while (1.1c) is the coupling between the domains. A typical example of such a system is a vascular network surrounded by a tissue and the order reduction is due to assumption of radii of the arteries being negligible in comparison to their lengths. To list a few concrete applications, the 3d-1d models have been used, e.g., in [17, 21, 16, 31] to study blood and oxygen transport in the brain or in [11] to describe fluid exchange between microcirculation and tissue interstitium. Efficiency of cancer therapies delivered through microcirculation was studied in [10], and hyperthermia as a cancer treatment in [27]. We note that the employed models are more involved than (1.1), but that the system still qualifies as a relevant model problem.

[†] Department of Mathematics, Division of Mechanics, University of Oslo {mirok, mikaem}@math.uio.no

[‡] Center for Biomedical Computing, Simula Research Laboratory, kent-and@simula.no

[*] The work of Kent-Andre Mardal and Mikael Mortensen has been supported by a Center of Excellence grant no. 179578 from the Research Council of Norway to the Center for Biomedical Computing at Simula Research Laboratory. The work of Kent-Andre Mardal has also been supported by the Research Council of Norway through grant no. 209951.

Due to the measure term and the three-to-one dimensional trace operator, the problem (1.1) is not standard and establishing its well-posedness is a delicate issue. In fact, considering (1.1a) with a known p and homogeneous Dirichlet boundary conditions, the equation is not solvable in $H^1_0(\Omega)$, as ∇u may be unbounded in the neighborhood of Γ. A similar problem was studied in [14], where two elliptic problems were coupled via a measure source term, and a unique weak solution was found using weighted Sobolev spaces. In particular, the weighted spaces ensured that the trace could be defined as a bounded operator. A corresponding finite element method (FEM) for the problem was discussed in [13], where optimal convergence in the weighted Sobolev norm was shown using graded meshes. For an elliptic problem with measure data, it was shown in [19] that FEM with regular meshes yields optimal convergence in the L^2 norm outside of the fixed neighborhood of the singularity. We note that the more application oriented works [11, 10, 27], that build on the analysis in [14, 13], relied on incomplete LU preconditioning.

In the current paper we shall *assume* that (1.1) is well posed and the focus is then on the construction of optimal preconditioners for the linear system due to (1.1) and FEM. Because the computational complexity of the $3d$ problem dominates the $1d$ problem, we put focus on preconditioners that are composed of standard multilevel algorithms for the $3d$ problem. This means that the weighted Sobolev spaces, or extra regularity in the equation in the $3d$ domain (1.1a), is disregarded, and that we rather add an extra requirement to (1.1c). We note that u shall be approximated within H^1 conforming finite element spaces and as such the approximation has a well defined trace.

The current paper is an extension of [20], where a system similar to (1.1a)–(1.1c) was analyzed for the case Ω a bounded domain in $2d$ and Γ a structure of codimension one. Therein, robust preconditioners were established, based on the operator preconditioning framework [26], in which preconditioners are constructed as approximate Riesz mappings in properly chosen Hilbert spaces. The framework often allows for construction of order-optimal preconditioners, with convergence independent of material and discretization parameters, directly from the analysis of the continuous system of equations. In particular, in [20] it was shown that the proper preconditioning relied on a nonstandard fractional $H^{\frac{1}{2}}$ inner product. Crucial for the analysis was the fact that the trace operator T is a well-defined mapping between $H^1(\Omega)$ and $H^{\frac{1}{2}}(\Gamma)$, when Γ is of codimension one with respect to Ω. Furthermore, for the finite element approximation in [20], it was assumed that the discrete meshes representing Γ and Ω *matched* in the sense that the cells of the mesh of Γ were edges in the mesh of Ω. Finally, only continuous linear Lagrange elements were used.

This paper utilizes ideas presented in [20] for the construction of the preconditioner, but here we go beyond what was theoretically established. In particular, we consider the case where Γ is of codimension two with respect to Ω. As the trace operator T is not well-defined in the continuous case, we do not attempt to provide mathematically rigorous proofs, as this would require additional regularity of the solution. Instead, we study for which s the H^s inner product provides numerically stable behaviour. In addition, we consider the case where the discretizations of Γ and Ω do not match. Finally, an approximation by discontinuous elements is discussed.

Our work is structured as follows. In §2 the theoretical background is presented. Section 3 discusses numerical experiments using spectral and finite element discretizations that identify suitable norms for the discrete $3d$-$1d$ trace operator. In §4 the identified norms are employed to construct optimal preconditioners for coupled model

$3d$-$1d$ problems discretized with FEM and matched discretizations of Ω and Γ. In §5 this restriction is lifted, the corresponding inf-sup condition is discussed, and we present numerical experiments that suggest the identified norms lead to good preconditioners. Finally, conclusions are drawn in §6.

2. Notation and preliminaries. Let X be a Hilbert space of functions defined on a domain $D \subset \mathbb{R}^d$, $d = 1, 2, 3$. The norm of the space is denoted by $\|\cdot\|_X$, while $\langle\cdot,\cdot\rangle_{X',X}$ is the duality pairing between X and its dual space X'. We let $(\cdot,\cdot)_X$ denote the inner product of X, while, to simplify the notation, $(\cdot,\cdot)_D$ is the L^2 inner product. The Sobolev space of functions with m square integrable derivatives is $H^m(D)$. Finally, $H^m_0(D)$ denotes the closure of the space of smooth functions with compact support in D in the $H^m(D)$ norm.

We use normal capital font to denote operators over infinite dimensional spaces, e.g. $A : X \rightarrow X'$. For a discrete subspace $X_h \subset X$, $\dim X_h = n$, the subscript h is used to distinguish the finite dimensional operator due to the Galerkin method, e.g., $A_h : X_h \rightarrow X'_h$ defined by

$$\langle A_h u_h, v_h\rangle_{X',X} = \langle Au, v_h\rangle_{X',X} \quad u_h, v_h \in X_h \text{ and } u \in X.$$

For a given basis, $\{\phi_i\}_{i=1}^n$ of X_h, the matrix representation of the operator is denoted by sans serif font. Thus A_h is represented by $\mathsf{A} \in \mathbb{R}^{n\times n}$ with entries

$$\mathsf{A}_{i,j} = \langle A_h\phi_j, \phi_i\rangle_{X',X}.$$

Finally, the function $u_h \in X_h$ is represented in the basis by a coefficient vector $\mathsf{u} \in \mathbb{R}^n$, where $u_h = \mathsf{u}_i\phi_i$ (summation convention invoked).

2.1. Properties of the trace operator. We consider $\Omega \subset \mathbb{R}^d$ an open connected domain with Lipschitz boundary $\partial\Omega$ and Γ a Lipschitz submanifold of codimension one or two in Ω. The trace operator T is defined by $Tu = u|_\Gamma$ for $u \in C(\overline{\Omega})$.

In case the codimension of Γ is one, the properties of the trace operator are well known. In particular, $T : H^1(\Omega) \rightarrow H^{\frac{1}{2}}(\Gamma)$ is bounded and surjective, see, e.g., [1, ch. 7], where $H^{\frac{1}{2}}(\Gamma)$ is a fractional Sobolev space equipped with the norm

$$\|u\|^2_{H^{\frac{1}{2}}(\Gamma)} = \|u\|^2_{L^2(\Gamma)} + \int_{\Gamma\times\Gamma} \frac{|u(x) - u(y)|^2}{|x - y|^{d+1}}\,\mathrm{d}x\mathrm{d}y.$$

Moreover, $T : H^1_0(\Omega) \rightarrow H^{\frac{1}{2}}(\Gamma)$ is bounded, but not surjective. To define the trace over $H^1_0(\Omega)$ as a surjective operator, the range is given as $H^{\frac{1}{2}}_{00}(\Gamma)$,

$$H^{\frac{1}{2}}_{00}(\Gamma) = \{u \in H^{\frac{1}{2}}(\Gamma); \tilde{u} \in H^{\frac{1}{2}}(\tilde{\Gamma})\} \text{ where } \tilde{u}(x) = \begin{cases} u(x) & x \in \Gamma \\ 0 & x \in \tilde{\Gamma}\setminus\Gamma \end{cases}$$

and $\tilde{\Gamma}$ is some suitable extension of Γ, e.g., $\tilde{\Gamma} = \Gamma \cup \partial\Omega$, in which case $\|u\|_{H^{\frac{1}{2}}_{00}(\Gamma)} = \|\tilde{u}\|_{H^{\frac{1}{2}}(\tilde{\Gamma})}$. We refer to [20] for these results.

The integral norms of $H^{\frac{1}{2}}(\Gamma)$ and $H^{\frac{1}{2}}_{00}(\Gamma)$ can be expensive to compute. For construction of efficient numerical algorithms, it is therefore more suitable to relate the spaces to interpolation spaces, see [22, 5] or [20]. For the sake of completeness, we review here the presentation from [20]. Let $u, v \in X = H^1(\Gamma)$. For u fixed $v \mapsto (u, v)_\Gamma$ is in X' and by the Riesz-Fréchet theorem there is a unique $w \in X$ such that $(w, v)_X = (u, v)_\Gamma$ for any $v \in X$. The operator $S : u \rightarrow w$ is injective and compact and thus the eigenvalue problem $S\phi_i = \lambda_i\phi_i$ (no summation implied) is well-defined. In addition, S is self-adjoint and positive-definite such that the eigenvalues form a nonincreasing sequence $0 < \lambda_{k+1} \leq \lambda_k$ and $\lambda_k \to 0$. By definition, the eigenvectors satisfy

$$(\phi_i, v)_X = \lambda_i^{-1}(\phi_i, v)_\Gamma \quad v \in X,$$

or equivalently

$$A\phi_i = \lambda_i^{-1}M\phi_i \text{ with } \langle Au, v\rangle_{X',X} = (u,v)_X \text{ and } \langle Mu, v\rangle_{X',X} = (u,v)_\Gamma. \tag{2.1}$$

Further, the set of eigenvectors $\{\phi_k\}_{k=1}^{\infty}$ forms a basis of X, which is orthogonal in the inner product of X and orthonormal in the $L^2(\Gamma)$ inner product. Finally, for $s \in [-1, 1]$ we define the s-norm of $u = c_k\phi_k \in \operatorname{span}\{\phi_k\}_{k=1}^{\infty}$ as

$$\|u\|_{H_s(\Gamma)} = \sqrt{c_k^2\lambda_k^{-s}}. \tag{2.2}$$

The space $H_s(\Gamma)$ is defined as the closure of the span$\{\phi_k\}_{k=1}^{\infty}$ in the s-norm, while $H_{s,0}(\Gamma)$ is then defined analogically to $H^s(\Gamma)$ with $X = H_0^1(\Gamma)$ in the construction. We remark that $H_0(\Gamma) = L^2(\Gamma)$, $H_{1,0} = H_0^1(\Gamma)$, $H_1 = H^1(\Gamma)$ and the norms of the spaces are equal. Moreover $H_{\frac{1}{2}}(\Gamma) = H^{\frac{1}{2}}(\Gamma)$ and $H_{\frac{1}{2},0}(\Gamma) = H_{00}^{\frac{1}{2}}(\Gamma)$ with the equivalence of norms.

Following the approach in [20], a weak formulation of the homogeneous Dirichlet problem for (1.1)–(1.1c) with $\Omega \in \mathbb{R}^3$, $\Gamma \subset \Omega$ of codimension one, using the method of Lagrange multipliers, reads: Find $(u, v, p) \in H_0^1(\Omega) \times H_0^1(\Gamma) \times Q$ such that

$$\begin{aligned}
(\nabla u, \nabla\phi)_\Omega + (u, \phi)_\Omega + (p, T\phi)_\Gamma &= (f, \phi)_\Omega & \phi &\in H_0^1(\Omega),\\
(\nabla v, \nabla\psi)_\Gamma + (v, \psi)_\Gamma - (p, \psi)_\Gamma &= (g, \psi)_\Gamma & \psi &\in H_0^1(\Gamma),\\
(\chi, Tu - v)_\Gamma &= (h, \chi)_\Gamma & \chi &\in Q.
\end{aligned} \tag{2.3}$$

Letting $Q = H_{s,0}$, the well-posedness of (2.3) is guaranteed as the Brezzi conditions are satisfied with $s = -\frac{1}{2}$, see [20] for the proof in $2d$-$1d$ setting, which immediately generalizes to $3d$-$2d$. Crucial for the well-posedness is the fact that $T : H_0^1(\Omega) \to H_{\frac{1}{2},0}(\Gamma)$ is an isomorphism. Consequently, [26] is invoked to yield a block diagonal preconditioner for the discretized problem where individual blocks are conceived as approximations of the corresponding Riesz mappings.

Let now $V_h \subset H^1(\Omega)$. Considering (1.1) on the finite dimensional spaces, we obtain a variational problem: Find $(u_h, v_h, p_h) \in V_h \times W_h \times Q_h$ such that

$$\begin{aligned}
(\nabla u_h, \nabla\phi_h)_\Omega + (u_h, \phi_h)_\Omega + (p_h, T_h\phi_h)_\Gamma &= (f, \phi_h)_\Omega & \phi_h &\in V_h,\\
(\nabla v_h, \nabla\psi_h)_\Gamma + (v_h, \psi_h)_\Gamma - (p_h, \psi_h)_\Gamma &= (g, \psi_h)_\Gamma & \psi_h &\in W_h,\\
\langle\chi_h, T_hu_h - v_h\rangle_\Gamma &= (h, \chi_h)_\Gamma & \chi_h &\in Q_h.
\end{aligned} \tag{2.4}$$

Here the discrete trace operator T_h is well-defined as the functions in V_h are continuous. The continuous problem, on the other hand, is not well defined since H^1 does not to permit a bounded trace in $L^2(\Gamma)$, see [14]. In turn we cannot directly follow the steps of [20] and employ operator preconditioning [26] to construct an optimal preconditioner. Instead, we shall reason about the properties of the discrete system.

From a linear algebra point of view, the problem (2.4) is a saddle-point system

$$\begin{bmatrix} \mathsf{A} & \mathsf{B}^\top \\ \mathsf{B} & \end{bmatrix}\begin{bmatrix} \mathsf{x} \\ \mathsf{y} \end{bmatrix} = \begin{bmatrix} \mathsf{b} \\ \mathsf{c} \end{bmatrix}.$$

Block diagonal preconditioners for such problems can be constructed as an approximate inverse of the matrix $\mathrm{diag}(\mathsf{K}, \mathsf{L})$, where K should be spectrally equivalent with A and L should be spectrally equivalent with the Schur complement $\mathsf{B}\mathsf{A}^{-1}\mathsf{B}^{\top}$, see, e.g., [32, 33]. Considering (2.4), the key question is thus whether it is possible (in an efficient and systematic manner) to construct an operator that is spectrally equivalent with the Schur complement. Motivated by the $2d$-$1d$, the operator shall be based on the fractional s-norm (2.2).

Following [20], the discrete approximation of the s-norm shall be constructed by mirroring the continuous eigenvalue problem (2.1). More specifically, let $X_h \subset X$ and matrices A, M be the representations of A_h, M_h; the Galerkin approximations of operators A, M from (2.1). Then there exists an invertible matrix U and diagonal, positive-definite matrix Λ satisfying $\mathsf{AU} = \mathsf{MU}\Lambda$. Moreover, the product $\mathsf{U}^{\top}\mathsf{MU}$ is an identity such that the columns of U form an A orthogonal and M orthonormal basis of $\mathbb{R}^n$. In order to define the discrete norm, we let H_s be a symmetric, positive-definite matrix

$$\mathsf{H}_s = (\mathsf{MU})^{\top}\Lambda^s\,(\mathsf{MU})\,. \tag{2.5}$$

The matrices $\mathsf{H}_{s,0}$ are defined analogically to (2.5), using the eigenvalue problem for the Laplace operator with homogeneous Dirichlet boundary conditions. For $u_h \in X_h$ represented in the basis of the space by a coefficient vector u, let c be the representation of u in the basis of eigenvectors, that is, $\mathsf{u} = \mathsf{Uc}$. We then set

$$\|u_h\|_{H_s(\Gamma)} = \sqrt{\mathsf{u}^{\top}\mathsf{H}_s\mathsf{u}} = \sqrt{\mathsf{c}^{\top}\Lambda^s\mathsf{c}}. \tag{2.6}$$

The generalized eigenvalue problem required for evaluating the discrete s-norm (2.6) becomes trivial if the approximation space V_n is such that $V_n = \mathrm{span}\{\phi_i\}_{i=1}^n$, i.e. the basis is formed by the eigenvectors of the continuous problem (2.1). Such a discretization is practically limited to Cartesian domains, however, it will prove useful in studying the trace operator when the codimension of Γ in Ω is two. The technique is introduced in Example 2.1.

Example 2.1 (Spectral method for $2d$-$1d$ coupled problem). *Let $\Omega = [0,1]^2$, $\Gamma = \{(t, \frac{1}{2}); t \in [0,1]\}$ and consider the task of finding $u \in H^1(\Omega)$ which minimizes $v \mapsto (\nabla v, \nabla v)_\Omega - 2(f, v)_\Omega$ subject to $Tu = g$ and $u = 0$ on the boundary (in the sense of traces). Introducing $V = H_0^1(\Omega)$, $Q = H_{-\frac{1}{2},0}(\Gamma)$ the problem is formulated as a saddle point system for $u \in V$, $p \in Q$ satisfying*

$$\begin{aligned}(\nabla u, \nabla v)_\Omega + \langle p, Tv\rangle_{Q',Q} &= (f, v)_\Omega \qquad && v \in V,\\ \langle q, Tu\rangle_{Q',Q} &= \langle q, g\rangle_{Q',Q} && q \in Q.\end{aligned} \tag{2.7}$$

Well-posedness of (2.7) is readily established by verifying the Brezzi conditions [9]. In particular, the inf-sup condition can be shown to hold, see, e.g., Appendix A. By operator preconditioning [26] the canonical preconditioner for (2.7) is the Riesz map with respect to the inner product inducing the norm of $V \times Q$.

The Galerkin approximation of (2.7) is defined with spaces Q_m, V_n such that $Q_m = \mathrm{span}\{\phi_k(t)\}_{k=1}^m$ and $V_n = \mathrm{span}\{\phi_i(x)\phi_j(y)\}_{i,j=1}^n$. Recall that functions $\phi_k(t) = \sqrt{2}\sin k\pi t$ are the eigenfunctions of (2.1) satisfying $-\Delta\phi_k = (k\pi)^2\phi_k$ on Γ. For greater readability, let us introduce $N = n^2$. In the basis of eigenfunctions, the discrete trace operator is represented by a trace matrix $\mathsf{T} \in \mathbb{R}^{m\times N}$ with entries

$$\mathsf{T}_{k,(i,j)} = \begin{cases} 0 & j \text{ even} \\ (-1)^{j+1}\sqrt{2}\delta_{ik} & j \text{ odd} \end{cases}\,.$$

Here (i,j) is a column index $n(i-1)+j$. With matrix $\mathsf{A}\in\mathbb{R}^{N\times N}$ and vectors $\mathsf{f}\in\mathbb{R}^N$, $\mathsf{g}\in\mathbb{R}^m$ defined in a natural way, the preconditioned linear system from (2.7) *is*

$$\begin{bmatrix}\mathsf{A} & \\ & \mathsf{H}_{-\frac{1}{2},0}\end{bmatrix}^{-1}\begin{bmatrix}\mathsf{A} & \mathsf{T}^{\top}\\ \mathsf{T} & \end{bmatrix}\begin{bmatrix}\mathsf{u}\\ \mathsf{p}\end{bmatrix}=\begin{bmatrix}\mathsf{A} & \\ & \mathsf{H}_{-\frac{1}{2},0}\end{bmatrix}^{-1}\begin{bmatrix}\mathsf{f}\\ \mathsf{g}\end{bmatrix}, \tag{2.8}$$

where the first matrix is the discretization of the canonical preconditioner. We note that matrices $\mathsf{H}_{s,0}$ are diagonal. Consequently ${\mathsf{H}_{s,0}}^{-1}=\mathsf{H}_{-s,0}$.

For stability of the solution obtained by solving (2.8)*, it is required that the discrete inf-sup condition holds. Note that for $m>n$ the trace matrix does not have a full row rank and thus $m\leq n$ is necessary. We shall set $m=n$ and show that for this choice the inf-sup condition is satisfied.*

By, e.g., [24] or [7], the constant of the discrete inf-sup condition is the smallest eigenvalue of the generalized eigenvalue problem for the negative Schur complement of the system matrix, i.e.,

$$\mathsf{T}\mathsf{A}^{-1}\mathsf{T}^{\top}\mathsf{e}=\lambda\mathsf{H}_{-\frac{1}{2},0}\mathsf{e},$$

where $(\lambda\in\mathbb{R},\mathsf{e}\in\mathbb{R}^m)$ is the sought eigenpair. The simple structure of the involved matrices allows us to compute all the eigenvalues of the problem analytically. In fact,

$$\left(\mathsf{H}_{\frac{1}{2},0}\mathsf{T}\mathsf{A}^{-1}\mathsf{T}^{\top}\right)_{ij}=S_j\delta_{ij}\ \textit{where}\ S_j=\frac{2}{\pi}\sum_{l\ odd}^{n}\frac{j}{j^2+l^2}.$$

Then $S_j\geq S_n$ and the lower bound can be evaluated. Using Mathematica [36], we have obtained $\lim_{n\to\infty}S_n=\frac{1}{8}$ *as the discrete inf-sup constant. For the largest eigenvalue, the following estimate can be established*

$$S_j\leq\frac{2}{\pi}\sum_{l\ odd}^{\infty}\frac{j}{j^2+l^2}=\frac{1}{4}\tanh\tfrac{j\pi}{2}\leq\frac{1}{4}$$

and in turn the theoretical spectral condition number for the preconditioned Schur complement is $\kappa=2$. We remark that the remaining Brezzi constants are both equal to one and the condition number of (2.8) *can be determined from spectral bounds presented in [32].*

The theoretical findings about the condition number of the preconditioned Schur complement are confirmed by numerical experiments, with results shown in Figure 2.1 and Table 2.1. The figure shows that there is a range of exponents $s\in[-0.52,-0.5]$ for which the condition numbers are stable. Interestingly, in this range $s=-0.5$ gives the largest condition number while for $s=-0.52$ a slightly smaller value is observed, cf. Table 2.1.

3. Norms for the discrete $3d$-$1d$ trace.

In Example 2.1 a priori knowledge of the trace space lead to an optimal preconditioner for the model problem (2.7). In particular, the norm of the trace space was used to construct a spectrally equivalent operator to the Schur complement of (2.8). For $\Omega\subset\mathbb{R}^3$ and Γ a one dimensional curve, the trace space is not a priori known and we shall therefore attempt to characterize it numerically. To this end, we shall at first use the spectral discretization and search for the s-norm (2.6) for which the condition number of the preconditioned Schur complement is bounded in the discretization parameter. We note that the condition is motivated by the fact that convergence of the preconditioned conjugate gradient

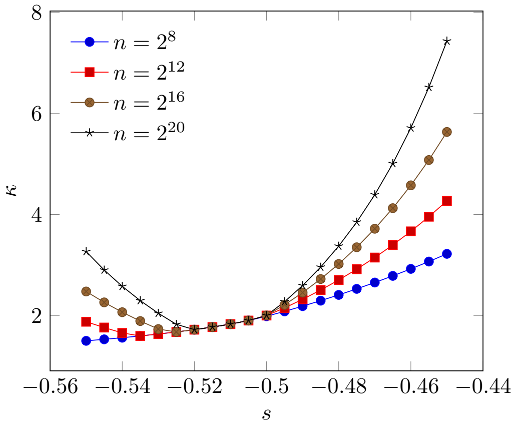

Fig. 2.1: Spectral condition numbers of the generalized eigenvalue problem for matrices $\mathsf{T}\mathsf{A}^{-1}\mathsf{T}^{\top}$ and $\mathsf{H}_{s,0}$, (cf. (2.8)), and different values of the discretization parameter n. The exponent $s = -\frac{1}{2}$ yields condition number 2, independent of n. Only the exponents $s \in [-0.52, 0.5]$ yield stable condition numbers.

Table 2.1: Spectral condition numbers κ of the preconditioned Schur complement of (2.8) with two preconditioners $\mathsf{H}_{-0.5,0}$ and $\mathsf{H}_{-0.52,0}$. In agreement with analysis, $s = -0.5$ yields bounded κ. The value $s = -0.52$, determined from observations, cf. Figure 2.1, yields a smaller condition number.

$\log_2 n$	8	10	12	14	16	18	20
$s = -0.5$	1.9848	1.9959	1.9987	1.9997	1.9999	2.0000	2.0000
$s = -0.52$	1.7189	1.7190	1.7190	1.7190	1.7190	1.7190	1.7190

method is estimated in terms of the condition number, see, e.g., [35]. For suitable s the linear system with the Schur complement could thus be solved efficiently. We also note that the condition is weaker than spectral equivalence. In fact, if such s exists, the matrix $\mathsf{H}_{s,0}$ is spectrally equivalent with the Schur complement if and only if one of the extremal eigenvalues is bounded by a constant.

3.1. Trace operator with spectral discretization. Let $\Omega = [0,1]^3$, $Q_m = \operatorname{span}\{\phi_j(t)\}_{j=1}^m$ and $V_n = \operatorname{span}\{\phi_i(x)\phi_k(y)\phi_l(z)\}_{i,k,l=1}^n$, cf. Example 2.1. We consider the problem of minimizing $v \mapsto (\nabla v, \nabla v)_\Omega - 2(f, v)_\Omega$, $v \in V_n$, subject to $v = 0$ on the boundary and the constraint $Tv = g$ on Γ, where the trace operator restricts v either to $\Gamma_1 = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0,1]\}$ or $\Gamma_2 = \{(t,t,t); t \in [0,1]\}$ respectively. The weak formulation of the problem reads

$$\begin{aligned} (\nabla u, \nabla v)_\Omega + (p, Tv)_\Gamma &= (f, v)_\Omega \qquad && v \in V_n, \\ (q, Tu)_\Gamma &= (q, g)_\Gamma \qquad && q \in Q_m \end{aligned} \tag{3.1}$$

and is equivalent with a linear system

$$\begin{bmatrix} \mathsf{A} & \mathsf{T}^{\top} \\ \mathsf{T} & \end{bmatrix} \begin{bmatrix} \mathsf{u} \\ \mathsf{p} \end{bmatrix} = \begin{bmatrix} \mathsf{f} \\ \mathsf{g} \end{bmatrix}. \tag{3.2}$$

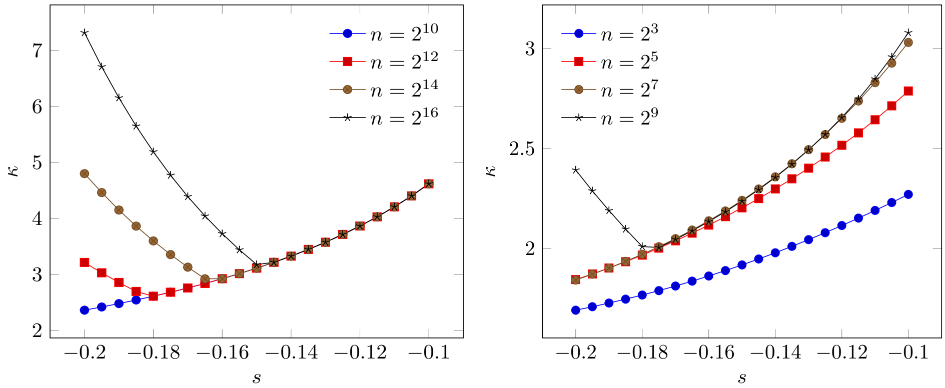

Fig. 3.1: Spectral condition numbers computed from the generalized eigenvalue problem for Schur complement of (3.2) and matrices $\mathsf{H}_{s,0}$, see (2.6). (Left) The constraint is considered on $\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0, 1]\}$. (Right) $\Gamma = \{(t, t, t); t \in [0, 1]\}$ is considered. With both configurations, values s close to -0.14 yield bounded κ.

In (3.2) the trace matrix $\mathsf{T} \in \mathbb{R}^{m \times N}$ for curve Γ_1 is sparse with entries

$$\mathsf{T}_{j,(i,k,l)} = \begin{cases} 0 & k \text{ or } l \text{ even} \\ (-1)^{k+1}(-1)^{l+1} 2\delta_{ij} & \text{otherwise} \end{cases}.$$

Here $N = n^3$ was introduced for readability. Note that for $m > n$ the matrix does not have a full row rank and the system is singular. We therefore set $m = n$. For Γ_2 the trace matrix is sparse with a more involved sparsity pattern and at most four nonzero entries per row

$$\mathsf{T}_{j,(i,k,l)} = 4\sqrt{3} \int_0^1 \sin j\pi t \sin i\pi t \sin k\pi t \sin l\pi t \, \mathrm{d}t.$$

Finally, we consider the generalized eigenvalue problem for the Schur complement of (3.2) and matrices $\mathsf{H}_{s,0}$ where such exponents are of interest, for which the spectral condition number $\kappa = \lambda_{\text{max}}/\lambda_{\text{min}}$ is bounded in the discretization parameter. Note that with Γ_1 the Schur complement is a diagonal matrix $S_j \delta_{ij}$,

$$S_j = \frac{4}{\pi^2} \sum_{l,m \text{ odd}}^{n} \frac{1}{j^2 + l^2 + m^2}. \tag{3.3}$$

For Γ_2 the matrix is dense and shall be computed from assembled terms. As such a smaller n is explored in this configuration.

The results of the numerical experiments with $s \in [-0.2, -0.1]$ are summarized in Figure 3.1. We observe that values $s \in [-0.145, -0.1]$ yield bounded condition numbers for Γ_1. The condition numbers are not quite converged for the other configuration, however, it is possible to identify unstable exponents $s < -0.18$. Moreover, the values close to $s = -0.14$ appear to be stable also in this configuration. This fact is easier to appreciate in Table 3.1, which shows λ_{min}, λ_{max} and κ as functions of the discretization parameter for $s = -0.14$. With Γ_1 the condition number is evidently

Table 3.1: Smallest and largest eigenvalues $\lambda_{\min}$, $\lambda_{\max}$ and the spectral condition numbers κ of the preconditioned Schur complement of (3.2). (Top) The preconditioner is $\mathsf{H}_{-0.14,0}$. While the eigenvalues are unbounded the condition number is bounded in n. (Bottom) Matrix $\mathsf{H}_{0,0}$ (identity matrix) is used as the preconditioner. In agreement with the analysis in Remark 3.1, constant $\lambda_{\min}$ and $\lambda_{\max}$ with a logarithmic growth are observed.

$\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0,1]\}$				$\Gamma = \{(t,t,t); t \in [0,1]\}$			
$\log_2 n$	$\lambda_{\min}$	$\lambda_{\max}$	κ	$\log_2 n$	$\lambda_{\min}$	$\lambda_{\max}$	κ
10	0.6218	2.0696	3.3285	6	0.8476	1.9916	2.3496
12	0.9167	3.0511	3.3285	7	1.0298	2.4283	2.3581
14	1.3514	4.4982	3.3285	8	1.2513	2.9491	2.3569
16	1.9923	6.6315	3.3285	9	1.5201	3.5804	2.3553
11	0.0648	1.2167	18.7767	6	0.1939	1.2180	6.2807
12	0.0648	1.3270	20.4792	7	0.1938	1.4080	7.2655
13	0.0648	1.4373	22.1818	8	0.1938	1.5985	8.2487
14	0.0648	1.5476	23.8843	9	0.1938	1.7893	9.2312

constant, while for Γ_2 the number appears to be bounded. Note that with both configurations the smallest and largest eigenvalues are not bounded and thus $\mathsf{H}_{-0.14,0}$ is not spectrally equivalent with the Schur complement with the bounds independent of n. However, any of $\lambda_{\min}(n)$, $\lambda_{\max}(n)$ (or their linear combinations) define a mesh-depenedent scale $\tau(n)\mathsf{H}_{-0.14,0}$ that yields spectral equivalence. Such scale, however, is not easily computable in general.

In the numerical experiment the range of exponents was limited to $s \in [-0.2, -0.1]$ and the upper bound yielded condition numbers independent of the discretization parameter, cf. Figure 3.1. The observation raises a question about the suitablity of $s = 0$, i.e. considering the multiplier space Q_m with the L^2 norm. It is shown in Remark 3.1 that the choice leads to a condition number with logarithmic growth.

Remark 3.1. *We consider* (3.2) *with* Γ_1. *Since* $\mathsf{H}_{0,0}$ *is (due to the employed discretization) an identity, the values* S_j *in* (3.3) *are the eigenvalues of the preconditioned Schur complement, where* $\mathsf{H}_{0,0}$ *is the preconditioner. We have* $S_j \geq S_n$ *and observe that the lower bound sums* $\mathcal{O}(n^2)$ *terms that are at most* n^{-2} *in magnitude. Thus* S_n *is bounded from below by a constant. On the other hand the upper bound* $S_j \leq S_1$ *grows as* $\log n$.

Note that for the 2d-1d trace and $s = 0$ *we have, cf. Example 2.1,*

$$\frac{2}{\pi}\sum_{\iota\ odd}^{n} \frac{1}{n^2 + l^2} \leq S_j = \frac{2}{\pi}\sum_{\iota\ odd}^{n} \frac{1}{j^2 + l^2} \leq \frac{2}{\pi}\sum_{\iota\ odd}^{n} \frac{1}{1 + l^2} \leq C,$$

while the lower bound as a sum of $\mathcal{O}(n)$ *terms with* n^{-2} *magnitude decays as* n^{-1}. *Thus* $s = 0$ *leads to a linearly growing condition number.*

The estimates for $\Omega \subset \mathbb{R}^3$ *are confirmed by numerical experiment summarized in Table 3.1. In particular, the constant lower bound and the upper bound growing as* $\log n$, *are visible for both configurations.*

Experiments with the spectral discretization suggest that there exists an exponent s, independent of Γ, such that the *discrete* trace operator T_h defined over V_h can be controlled by the s-norm (2.6). However, the space V_h considered thus far consisted of infinitely smooth functions. We proceed to show that a similar conjecture holds if the discrete spaces are obtained by FEM. In particular, the space V_h shall be constructed using the H^1 conforming continuous linear Lagrage elements.

3.2. Trace operator with FEM discretizaton. Let $V_h \subset H^1(\Omega)$. Further, let $\{\psi_k\}_{k=1}^m$ and $\{L_j\}_{j=1}^m$ be, respectively, the basis and degrees of freedom/dual basis nodal with respect to $\{\psi_k\}_{k=1}^m$ of the finite element space Q_h over Γ. The trace mapping $T_h : V_h \rightarrow Q_h$ shall be defined by interpolation so that $p_h = T_h u_h$ is represented in the basis by vector $\mathsf{p} \in \mathbb{R}^m$,

$$\mathsf{p}_j = \langle L_j, u_h|_\Gamma \rangle. \tag{3.4}$$

Equivalently we have $\mathsf{p} = \mathsf{T}\mathsf{u}$ where $\mathsf{u} \in \mathbb{R}^n$ and the matrix representing the trace operator has entries

$$\mathsf{T}_{i,j} = \langle L_i, \phi_j|_\Gamma \rangle,$$

where $\{\phi_j\}_{j=1}^n$ are the basis functions of V_h.

Lemma 3.1 (Discrete trace operator by projection). *Let $u_h \in V_h$ be given and $\tilde{p}_h \in Q_h$ be the L^2 projection*

$$(\tilde{p}_h, q)_\Gamma = (u_h|_\Gamma, q)_\Gamma, \quad q \in Q_h.$$

Further let $p_h \in Q_h$ be defined via (3.4). Then $V_h|_\Gamma \subseteq Q_h$ is necessary and sufficient for $p_h = \tilde{p}_h$.

Proof. To verify the assertion let $q_k \in Q_h$ be the Riesz representation of L_k, i.e. $(q_k, v)_\Gamma = \langle L_k, v \rangle$, $v \in Q_h$, and $u_h \in V_h$ arbitrary. Then by definition $(p_h, q_k)_\Gamma = \langle L_i, u_h|_\Gamma \rangle (\psi_i, q_k)_\Gamma$ and

$$\langle L_i, u_h|_\Gamma \rangle (\psi_i, q_k)_\Gamma = (q_i, u_h|_\Gamma)_\Gamma \langle L_k, \psi_i \rangle = (q_k, u_h|_\Gamma)_\Gamma = (q_k, \tilde{p}_h)_\Gamma$$

by the property of the Riesz basis $\{q_k\}_{k=1}^m$, nodality of the basis $\{\psi_i\}_{i=1}^m$ and definition of $\tilde{p}_h$. It follows that $(p_h - \tilde{p}_h, q_k)_\Gamma = 0$. Note that $u_h|_\Gamma \in Q_h$ was required to apply the Riesz theorem. □

Definition 3.2 (Γ-matching spaces). *Let Γ be a manifold in Ω and Q_h, V_h the finite element spaces over the respected domains. The spaces are called Γ-matching if (i) V_h and Q_h are constructed from the same elements and (ii) meshes of Ω and Γ are matched.*

Remark 3.2 (Equivalence of interpolation and projection trace). *The condition from Lemma 3.1 is satisfied with $V_h|_\Gamma = Q_h$ if V_h and Q_h are Γ-matching.* Finally, note that the interpolation trace is in general cheaper to construct than the trace due to projection. We shall employ (3.4) throughout the rest of the paper.

Let now V_h, Q_h be a pair of Γ-matching spaces constructed from continuous linear Lagrange elements. Further, the discretization of the geometry shall be such that the mesh of Ω is *finer* at/near Γ than in the rest of the domain, cf. Table B.1 in Appendix B and Figure 4.1. This way the dimensionality of Q_h is increased. Finally, we consider the Schur complement[1] of (3.1) preconditioned by different matrices $\mathsf{H}_{s,0}$. Recall that previously global trigonometric polynomial basis functions were used with (3.1) and $-0.2 < s \leq -0.1$ yielded condition numbers bounded in the discretization parameter. Figure 3.2 and Table 3.2 show that the same conclusions hold also if the finite element discratization is employed.

[1] The Schur comeplement is computed from its definition, where the components T, A are assembled using FEniCS [23, 2] and PETSc [8] libraries. The Laplacian matrix is then inverted by conjugate gradient method with algebraic multigrid (AMG) preconditioner from Hypre library [15]. Relative tolerance 10^{-15} was set as a convergence criterion.

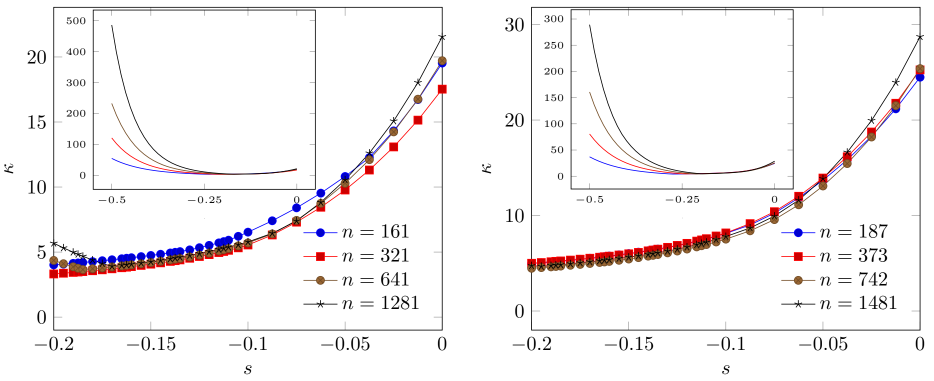

Fig. 3.2: Condition numbers of the Schur complement of (3.1) with finite element discretization, $n = \dim Q_h$, and different preconditioners $\mathsf{H}_{s,0}$. (Left) the curve is Γ_1. (Right) the curve is Γ_2. The zoomed out plot shows that $s < -0.25$ yields unbounded κ. For both configurations exponents from the interval around $s = -0.1$ yield bounded condition numbers.

Table 3.2: Condition numbers of $\mathsf{H}_{s,0}$ preconditioned Schur complement of (3.1) for selected values of s. The finite element discretization is considered on a sequence of uniformly refined meshes, see Table B.1. For each discretization the mesh is finer near Γ than in the rest of the domain. Exponent $s = -0.14$ observed in the spectral discretization, cf. Table 3.1, yields bounded κ also with discrization by FEM. Note that similar to the spectral discretization there is a slight growth of κ for $s = 0$.

L\\s	$\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0, 1]\}$					$\Gamma = \{(t, t, t); t \in [0, 1]\}$				
	-0.16	-0.14	-0.12	-0.1	0	-0.16	-0.14	-0.12	-0.1	0
1	4.568	4.932	5.517	6.531	19.530	5.760	6.316	7.064	8.129	24.484
2	3.883	4.282	4.804	5.545	17.525	5.743	6.300	7.085	8.175	25.253
3	4.023	4.400	4.927	5.710	19.713	5.192	5.744	6.488	7.525	25.386
4	4.062	4.477	5.045	5.781	21.561	5.381	5.926	6.698	7.798	28.731

Figure 3.2 explores the condition numbers for $s \in [-0.5, 0]$. It is evident, cf. the zoom-out plot, that for $s < -0.25$, $\mathsf{H}_{s,0}$ is not a good preconditioner for the Schur complement. For both configurations there are exponents in $(-0.2, 0)$ that lead to bounded condition numbers. For several values of s in this interval, the condition numbers observed on a sequence of uniformly refined meshes are reported in Table 3.2. Therein $s \leq -0.1$ can be observed to lead to bounded κ. Exponent $s = 0$, i.e. the L^2 norm, leads to a slight growth in κ with both Γ_1 and Γ_2.

We note that in both configurations the behaviour of the eigenvalues is similar to the spectral case. In particular, $\lambda_{\max}$ and $\lambda_{\min}$ grow for $s \leq -0.1$, whereas for $s = 0$ only $\lambda_{\max}$ grows while $\lambda_{\min}$ is bounded by a constant, see Table 3.3. Since the extremal eigenvalues are in general not bounded by a constant, $\mathsf{H}_{s,0}$ is not a discretization of an operator spectrally equivalent to the Schur complement with constants independent of the discretization parameter. However, the relation observed in the experiments

$$0 < \lambda_{\min}(h) \leq \frac{\mathsf{x}^\top \mathsf{T}\mathsf{A}^{-1}\mathsf{T}^\top \mathsf{x}}{\mathsf{x}^\top \mathsf{H}_{s,0}\mathsf{x}} \leq \lambda_{\max}(h) \quad \mathsf{x} \in \mathbb{R}^m \tag{3.5}$$

Table 3.3: Smallest and largest eigenvalues of the $\mathsf{H}_{s,0}$ preconditioned Schur complement considered in Table 3.2. Similar to spectral discretization both the extremal eigenvalues grow for $s=-0.14$ while the lower bound is constant and the upper one grows for $s=0$.

L	$\Gamma=\{(t,\frac{1}{2},\frac{1}{2}); t\in[0,1]\}$		$\Gamma=\{(t,t,t); t\in[0,1]\}$	
	$s=-0.14$	$s=0$	$s=-0.14$	$s=0$
1	(0.290, 1.433)	(0.051, 1.000)	(0.207, 1.310)	(0.041, 1.000)
2	(0.420, 1.799)	(0.059, 1.040)	(0.256, 1.610)	(0.041, 1.026)
3	(0.502, 2.208)	(0.059, 1.161)	(0.342, 1.965)	(0.045, 1.145)
4	(0.603, 2.701)	(0.059, 1.276)	(0.401, 2.379)	(0.044, 1.265)

suggests existence of a mesh dependent scale in which spectral equivalence can be achieved. In particular, rescaling the s-norm matrix as $\lambda_{\min}(h)\mathsf{H}_{s,0}$ leads to constant bounds, cf. observed constant spectral condition number. We remark that $\lambda_{\min}$ is bounded away from zero for all h, in fact the eigenvalue increases with h^{-1}, and in this sense the discrete inf-sup constant never approaches zero.

Based on the mesh-dependent s-norm a block-diagonal preconditioner $\text{diag}(\mathsf{A}, \lambda_{\min}(h)\mathsf{H}_{s,0})^{-1}$ could be analysed and shown to be optimal using the results of [32, 33] (see also the review paper [29]). However, obtaining the scale is computationally expensive. We shall therefore proceed with (2.6) only. In particular, the exponents s identified previously shall be used to construct preconditioners for several $3d$-$1d$ constrained problems. We note that the bounds (3.5) enter estimates for convergence of iterative solvers, see, e.g., [33], and since the bounds here are not constant, the proposed preconditioners are theoretically suboptimal. Nevertheless, the number of iterations in the studied examples will be bounded. We remark that the smallest and largest eigenvalues are never far from unity in our examples.

4. Trace coupled problems.

The previous experiments revealed a range of exponents s for which matrices H_s behaved similarly to the Schur complement, in terms of stability of the condition number, of the related generalized eigenvalue problem. To simplify the discussion, we shall in the following employ $s=-0.14$. The exponent shall be used to construct preconditioners for two model $3d$-$1d$ coupled problems.

4.1. Babuška's problem.

Let V_h, Q_h be a pair of Γ-matching spaces constructed by continuous linear Lagrange elements and consider the problem: Find $u\in V_h\subset H^1(\Omega)$, $p\in Q_h$ such that

$$\begin{aligned}(\nabla u,\nabla v)_\Omega+(u,v)_\Omega+(p,Tv)_\Gamma&=(f,v)_\Omega \qquad v\in V_h,\\ (q,Tu)_\Gamma&=(q,g)_\Gamma \qquad q\in Q_h.\end{aligned}\tag{4.1}$$

The system (4.1) is a Lagrange multiplier formulation of the minimization problem for $v\mapsto\|v\|^2_{H^1(\Omega)}-2(f,v)_\Omega$, and the constraint $Tv-g=0$ on Γ. We note that the problem is considered with homogeneous Neumann boundary conditions. A similar problem with $\Omega\subset\mathbb{R}^2$ and $\Gamma\subset\partial\Omega$ was first studied in [6] to introduce Lagrange multipliers as means of prescribing boundary data.

Similar to the Schur complement study in Section 3.2, the problem shall be considered with two different curves Γ. Moreover, for each configuration we consider three different sequences of uniformly refined meshes, to investigate numerically whether the construction of the preconditioner relies on a quasi-uniform mesh, or if shape-regular elements are sufficient. In a *uniform* discretization the characteristic mesh

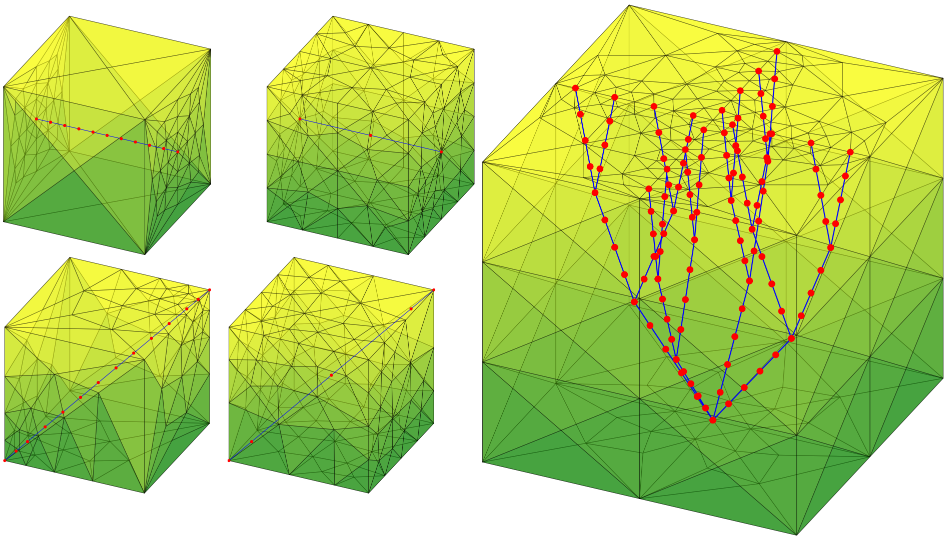

Fig. 4.1: Domains used in experiments with matching discretization. The one dimensional curve Γ is drawn in blue with element boundaries signified by red dots. (Left) The curve is, respectively, a horizontal or diagonal segment. The triangulation of Ω is either refined or coarsened at Γ. (Right) The curve contains branches and bifurcations, thus capturing some of the features of complex vascular systems.

size of Ω and Γ are identical and the tessellation of Ω is structured. In *finer* and *coarser* discretizations the mesh is unstructured and is either finer or coarser near Γ than in the rest of the domain. The example meshes are pictured in Figure 4.1. Information about the parameters of the discretizations and sizes of the corresponding finite element spaces are then summarized in Table B.1.

Since (4.1) is considered with Neumann boundary conditions, the block diagonal preconditioner for the system shall have the multiplier block based on H_s (not $\mathsf{H}_{s,0}$). We propose the following preconditioned linear system

$$\begin{bmatrix} \mathsf{A}+\mathsf{M} & \\ & \mathsf{H}_{-0.14} \end{bmatrix}^{-1} \begin{bmatrix} \mathsf{A}+\mathsf{M} & (\mathsf{M}_\Gamma \mathsf{T})^\top \\ (\mathsf{M}_\Gamma \mathsf{T}) & \end{bmatrix} \begin{bmatrix} \mathsf{u} \\ \mathsf{p} \end{bmatrix} = \begin{bmatrix} \mathsf{A}+\mathsf{M} & \\ & \mathsf{H}_{-0.14} \end{bmatrix}^{-1} \begin{bmatrix} \mathsf{f} \\ \mathsf{g} \end{bmatrix}, \tag{4.2}$$

where M and M_Γ are, respectively, the mass matrices of V_h and Q_h. We remark that the proposed preconditioner is not theoretically optimal because of the estimate (3.5).

In our implementation the leading block of the preconditioner is inverted by algebraic multigrid from the Hypre[2] library [15]. The system is then solved iteratively with the minimal residual method (MINRES) implemented in cbc.block [25] and requiring a preconditioned residual norm smaller than 10^{-12} for convergence. The initial vectors were random.

The recorded iterations counts are reported in Table 4.1. It can be seen that the proposed preconditioner results in a bounded number of iterations for all the considered geometrical configurations and their discretizations. In the table we also report iteration counts for the preconditioner that employs $\mathsf{H}_0 = \mathsf{M}_\Gamma$ for the multiplier

[2] We have used default values of all the parameters.

Table 4.1: Iteration counts for preconditioned Babuška's problem (4.1) with preconditioners based on (2.5) and $s = -0.14$ or $s = 0$ (discrete L^2 norm). Two geometric configurations and their different discretizations (L denotes the refinement level) are considered cf. Figure 4.1 and Table B.1. Both preconditioners yield bounded number of iterations. The L^2 norm leads to a less efficient preconditioner.

L	$\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0,1]\}$			$\Gamma = \{(t, t, t); t \in [0,1]\}$		
	uniform	finer	coarser	uniform	finer	coarser
2	(28, 59)	(53, 81)	(44, 46)	(29, 57)	(73, 107)	(62, 71)
3	(27, 68)	(52, 82)	(49, 58)	(27, 59)	(69, 103)	(64, 81)
4	(25, 70)	(52, 83)	(47, 62)	(25, 61)	(69, 105)	(67, 88)
5	(23, 70)	(53, 83)	(51, 71)	(25, 62)	(70, 105)	(67, 91)

block. Recall that with $s = 0$ and spectral discretization, the spectral condition number of the preconditioned Schur complement showed a logarithmic growth, cf. Table 3.1. Using FEM, the growth was less evident (see Table 3.2), however, the condition number was significantly larger than for $s = -0.14$. The iteration counts agreee with this observation; the L^2 norm leads to at least 20 more iterations. We remark that the norms in which the convergence criterion is measured differ between the two cases.

4.2. Model multiphysics problem. Building upon the Babuška problem we next consider a model multiphysics problem (1.1). A similar problem with $\Omega \subset \mathbb{R}^2$ and Γ a manifold of codimension one was previously studied by the authors in [20]. Therein it was found that the problem is well posed with the Lagrange multiplier in the intersection space $H^{-\frac{1}{2}}(\Gamma) \cap H^{-1}(\Gamma)$. The structure of the space was mirrored by the preconditioner, which used $(\mathsf{H}_{-0.5} + \mathsf{H}_{-1})^{-1}$ in the corresponding block.

We note that the exponent $-\frac{1}{2}$ was dictated by the properties of the continuous trace operator. In the $3d$-$1d$ case, which is of interest here, we shall instead base the exponent/preconditioner on the previous numerical experiments. More specifically, the linear system obtained by considering (2.4) on finite dimensional finite element subspaces

$$\begin{bmatrix} \mathsf{A}_\Omega + \mathsf{M}_\Omega & & (\mathsf{M}_\Gamma \mathsf{T})^\top \\ & \mathsf{A}_\Gamma + \mathsf{M}_\Gamma & \mathsf{M}_\Gamma \\ (\mathsf{M}_\Gamma \mathsf{T}) & \mathsf{M}_\Gamma & \end{bmatrix} \begin{bmatrix} \mathsf{u} \\ \mathsf{w} \\ \mathsf{p} \end{bmatrix} = \begin{bmatrix} \mathsf{f} \\ \mathsf{g} \\ \mathsf{h} \end{bmatrix} \tag{4.3}$$

shall be considered with the preconditioner

$$\begin{bmatrix} \mathsf{A}_\Omega + \mathsf{M}_\Omega & & \\ & \mathsf{A}_\Gamma + \mathsf{M}_\Gamma & \\ & & \mathsf{H}_{-0.14} + \mathsf{H}_{-1} \end{bmatrix}^{-1}. \tag{4.4}$$

Note that in (4.4) the structure of the trailing block mimics the related $2d$-$1d$ problem. We remark that in the implementation, the remaining two blocks are inverted by AMG. Moreover the discrete spaces are such that $W_h = Q_h$ and V_h, Q_h are Γ-matching. As in the previous example, continuous linear Lagrange elements are used. To demonstrate the performance of the preconditioner, (2.4) is considered on the same geometrical configurations and their discretizations as (4.1). The preconditioned system is then solved by MINRES, starting from a random initial vector and terminating if the preconditioned residuum is less than 10^{-12} in magnitude. As can be seen in

Table 4.2: Iteration counts for the model problem (4.3) with preconditioner (4.4). Spatial configurations and disretizations from Table 4.1 are considered. In all the cases the number of iterations is bounded.

L	$\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0, 1]\}$			$\Gamma = \{(t, t, t); t \in [0, 1]\}$		
	uniform	finer	coarser	uniform	finer	coarser
2	51	45	42	44	62	62
3	49	45	48	43	59	62
4	47	43	47	43	59	64
5	46	43	49	42	59	66

Table 4.2, the preconditioner yields bounded iteration counts. Interestingly, the convergence is faster on the *finer* discretization than on the *coarser* one. We note that the systems on the latter discretization are in general of smaller size and have more than a factor 10 fewer degrees of freedom in Q_h.

In the examples presented thus far, Γ was always a straight segment. To show that the preconditioner (4.4) (or the general idea of H_s based preconditioners for $3d$-$1d$ problems) is not limited to such simple curves, we shall in the final example consider (2.4) with Γ having a more complicated stucture. The considered domain, pictured in the right pane of Figure 4.1, is inspired by biomechnical applications and is intended to mimic some of the features of the vasculature. In particular, the domain consists of numerous branches and contains multiple bifurcations.

Repeating the setup of the previous experiment, Table 4.3 reports the iteration counts for the (4.4) preconditioned linear system (4.3), obtained by considering (2.4) on the complex Γ. The number of iterations is clearly bounded. In fact, the number decreases with refinement.

The good performance of the proposed preconditioner in all the considered examples brings in the question of practicability of its construction. Here, the question shall be addressed by considering the setup costs of the preconditioner for the domain with complex Γ. The choice is motivated by the fact that (i) the domain is potentially relevant for practical applications and (ii) the large (releative to $\dim V_h$) number of degrees of freedom of Q_h puts the emphasis on the construction of (2.5). We note that the costs are expected to be determined by the multigrid setup and the solution time of the generalized eigenvalue problem (2.5). As in [20] the eigenvalue problem is solved by the DSYGVD routine from LAPACK [3].

The timings obtained on a Linux machine with a single Intel Xeon E5-2680 CPU with 2.5GHz and 32GB of RAM are reported in Table 4.3. The observed costs of the eigenvalue solve are 3-4 times smaller than that of the multigrid setup, and thus the spectral construction does not present a bottleneck. Morover, both AMG and GEVP are expected to scale roughly as $\dim {Q_h}^3$. However, due to the cubic scaling, the system/preconditioner is unlikely to be assembled/setup in serial. For such a case, a scalable parallel implementation, for the construction of (2.5), remains an issue, and approaches that provide the approximate action of H_s matrices may offer better performance. Examples of such approaches are the Lanczos method [5, 4], contour integrals [18] or fast Fourier transforms [28]. We refer to [20] for a more thorough discussion of the subject.

5. Nonmatching discrete trace. The numerical examples presented thus far have always employed Γ-matching finite element spaces. We note that in [20] this construction is shown to imply that the discrete inf-sup condition holds for problems

Table 4.3: Iteration counts and setup costs (in seconds) for system (4.3) and preconditioner (4.4). Both operators are assembled for the complex Γ pictured in Figure 4.1. The number of iterations is bounded in the discretization parameter. In the considered example, the eigenvalue (GEVP) based construction (2.5) does not present a bottleneck as it is 3-4 times cheaper than setting up the algebraic multigrid (AMG).

$\dim V_h$	$\dim Q_h$	#	AMG [s]	GEVP [s]
18K	817	86	0.2	0.1
100K	1605	81	1.9	0.6
634K	3193	76	15.0	4.2
4.8M	6381	68	141.6	36.4

(4.1) and (2.4) considered with $\Omega \subset \mathbb{R}^2$ and Γ a one dimensional curve. However, the assumption of matched discretizations of Ω and Γ can be too limiting, e.g, if fine resolution is requested on the curve. In this section we present numerical examples using the Babuška problem (4.1), which demonstrate that the assumption is not necessary and to the extent given by the new inf-sup condition the discretizations can be independent. For stable discretizations, preconditioners based on characterization of the trace space will remain optimal. We note that from the point of view of Lemma 3.1 the spaces shall be such that $V_h|_\Gamma \supset Q_h$.

5.1. Codimension 1. Consider (4.1) with $\Omega \subset \mathbb{R}^2$. For $\Gamma \subset \partial\Omega$, the finite element discretization of the problem requires that the spaces V_h, Q_H (we use different subscripts to indicate the difference in underlying triangulations) are such that $h \leq cH$ for some $c < 1$. Here h is understood as a mesh size of V_h on Γ. The inequality ensures that the discrete inf-sup condition is satisfied, see, e.g., [34, 12]. We note that [30] shows that the inequality is not necessary.

Let now Γ be a curve, contained in Ω, where the domains are discretized such that the condition from the previous paragraph is met. Further, the space V_h shall be discretized by continuous linear Lagrange elements, while, for the construction of Q_H, either the same elements or piecewise constant Lagrange elements are employed. We note that with the latter choice the eigenvalue problem for the discrete s-norm simplifies, since the mass matrix is diagonal in this case.

Table 5.1 reports the number of MINRES iterations on the system (4.1), using $\text{diag}(\text{AMG}(\mathsf{A}+\mathsf{M}), {\mathsf{H}_{-0.5}}^{-1})$ as the preconditioner. The iterations are started from a random vector using 10^{-12} as the stopping tolerance for the magnitude of the preconditioned residuum. With both considered finite element discretizations of the multiplier space the number of iterations is bounded indicating (i) that the inf-sup condition is satisfied and (ii) the optimality of the preconditioner. We note that for $h > H$, the iterations are unbounded (not reported here) and thus the inf-sup condition is clearly not satisfied. An example of a pair of inf-sup stable and unstable discretizations is shown in Figure 5.1.

5.2. Codimension 2. For Γ a manifold of codimension two a condition guaranteeing the discrete inf-sup condition and stability of the discretization of (4.1) is not available. However, we shall assume that the inequality $h \leq cH$, $c < 1$ plays a role also in the $3d$-$1d$ case and discretize the domains accordingly.

The problem (4.1) is considered with two carefully constructed curves Γ, see Figure 5.1, and Ω a unit cube discretized such that the inequality is ensured. As before, the spaces Q_H are constructed from continuous piecewise linear or discon-

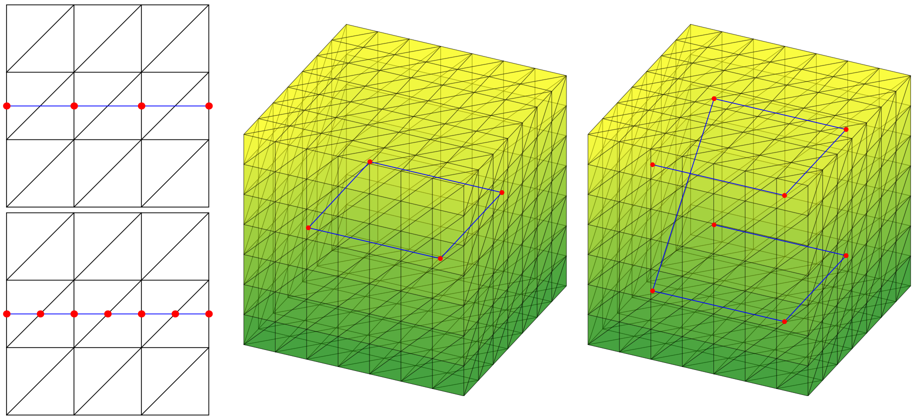

Fig. 5.1: Domains used in experiments with nonmatching discretization. (Left) The spaces V_h and Q_H are inf-sup stable for (4.1) if $h \leq cH$, $c < 1$. The condition is satisfied/violated in the top/bottom configurations. (Right) The $3d$-$1d$ experiments use two curves Γ. The mesh of Ω is obtained by first subdividing the domain into odd number of cubes in each direction. Thus degrees of freedom of V_h, Q_H are not associated with identical spatial points. Moreover $h \ll H$ is ensured in the refinement.

Table 5.1: Iteration counts and error convergence for (4.1) and Ω a unit square and Γ a circle. The spaces V_h and Q_H are formed either by continuous linear Lagrange elements or Q_H uses discontinuous piecewise constant Lagrange elements. Note that Γ is closed and thus Q_H has the same dimension with either of the elements. The inequality $h \leq cH$, $c < 1$ is respected ensuring that the inf-sup condition is satisfied. Consequently the iteration count is bounded. Both pairs yield optimal, order 1, convergence in $H^1(\Omega)$ norm of the error $u - u_h$. We note that the exact solution is smooth. The error of the Lagrange multiplier measured in the $s = -\frac{1}{2}$ norm (2.6) (on the same mesh) norm decays with order 1.5.

$\dim V_h$	$\dim Q_H$	Q_H continuous			Q_H discontinuous		
		#	$\lVert u - u_h \rVert_V$	$\lVert p - p_h \rVert_Q$	#	$\lVert u - u_h \rVert_V$	$\lVert p - p_h \rVert_Q$
22K	136	52	9.54E-02	5.28E-03	47	9.54E-02	3.68E-03
87K	272	52	4.78E-02	1.71E-03	48	4.78E-02	1.15E-03
348K	544	51	2.39E-02	5.77E-04	49	2.39E-02	4.18E-04
1.4M	1088	51	1.19E-02	1.87E-04	50	1.19E-02	1.49E-04

tinuous piecewise constant Lagrange elements. We note that the $\dim Q_h \ll \dim V_h$. Further, the MINRES iterations use the same initial and convergence conditions as in §5.1, while $\mathrm{diag}(\mathrm{AMG}(\mathsf{A}+\mathsf{M}), \mathsf{H}_{-0.14}{}^{-1})$ is used as the preconditioner. In Table 5.2 we observe that the discretization and the preconditioner lead to bounded iteration counts. We note that too fine a discretization of Γ, i.e., violating the inequality, leads to unbounded iterations.

6. Conclusions. We have discussed preconditioning of a model multiphysics problem (1.1), where two elliptic subproblems were coupled by a trace constraint, bridging the dimensionality gap of size two. The design of preconditioners for the problem followed our previous work [20]. In particular, the discrete fractional Sobolev

Table 5.2: Iteration counts for (4.2) posed on $\Omega \subset \mathbb{R}^3$ and the two curves pictured in Figure 5.1. For each domain, Q_H from continuous linear (first column) or discontinuous constant (second column) Lagrange elements is considered. The domains are discretized such that $h \leq cH$, $c < 1$. In all the cases, the number of iterations is bounded.

$\dim V_h$	Square				Spiral			
	$\dim Q_H$	#	$\dim Q_H$	#	$\dim Q_H$	#	$\dim Q_H$	#
33K	16	36	16	24	29	48	28	36
262K	32	38	32	24	57	48	56	35
2.1M	64	36	64	23	113	46	112	35
6.0M	128	38	128	24	225	48	224	36

norm was employed in order to facilitate the re-use of standard multilevel preconditioners for the 3*d* domain. Previously, the Sobolev index was dictated by properties of the *continuous* 2*d*-1*d* trace operator. Due to the difficulty of establishing a well-posed 3*d*-1*d* trace operator for functions in H^1, we relied on properties of the *discrete* trace. Using spectral and finite element discretizations, a range of suitable (negative) Sobolev indices was found. Consequently, preconditioners for coupled problems were built and their performance was demonstrated by a series of numerical experiments. The proposed preconditioners were robust with respect to the discretization parameter. Finally, the work in [20] was extended by considering independent discretizations of the bulk and embedded domains and by using discontinuous elements for the Lagrange multiplier space.

An obvious weakness of the presented work is the lack of a theoretical foundation, as the well-posedness of (1.1) was assumed and not established. Thus, referring to the idea of operator preconditioning, the continuous picture behind the preconditioner is missing.

Appendix A. Inf-sup condition for Example 2.1. Let $\Omega^- = [0,1] \times \left[0, \frac{1}{2}\right]$, $\Omega^+ = [0,1] \times \left[\frac{1}{2}, 1\right]$. Further, let $g \in H_{\frac{1}{2},0}(\Gamma)$ be given. The functions u_g^i, $i \in \{-,+\}$ are the unique weak solutions of $-\Delta u_g^i = 0$ in Ω^i with homogeneous boundary conditions on $\partial\Omega^i \setminus \Gamma$ and $Tu_g = g$ on Γ. Then

$$\|u_g^i\|_{H_0^1(\Omega^i)} \leq C_i \|g\|_{H_{\frac{1}{2},0}(\Gamma)}$$

for some constants independent of the domain. Moreover,

$$u_g(x) = \begin{cases} u_g^-(x) & x \in \Omega^- \\ u_g^+(x) & x \in \Omega^+ \end{cases}$$

is such that ∇u_g (defined piecewise) is in $L^2(\Omega)$ and thus $u_g \in H_0^1(\Omega)$. Finally, the estimate $\|u_g\|_{H_0^1(\Omega)} \leq C\|g\|_{H_{\frac{1}{2},0}(\Gamma)}$ holds. Then, by surjectivity of the trace, we get the estimate

$$\begin{aligned} \|q\|_{H_{-\frac{1}{2},0}(\Gamma)} = \sup_{g \in H_{\frac{1}{2},0}(\Gamma)} \frac{\langle q, g\rangle_{H_{-\frac{1}{2},0}(\Gamma), H_{\frac{1}{2},0}(\Gamma)}}{\|g\|_{H_{\frac{1}{2},0}(\Gamma)}} &\leq \frac{1}{C} \sup_{u_g \in H_0^1(\Omega)} \frac{\langle q, Tu_g\rangle_{H_{-\frac{1}{2},0}(\Gamma), H_{\frac{1}{2},0}(\Gamma)}}{\|u_g\|_{H_0^1(\Gamma)}} \\ &\leq \frac{1}{C} \sup_{v \in H_0^1(\Omega)} \frac{\langle q, Tv\rangle_{H_{-\frac{1}{2},0}(\Gamma), H_{\frac{1}{2},0}(\Gamma)}}{\|v\|_{H_0^1(\Gamma)}} \end{aligned}$$

and the inf-sup condition for (2.7) is satisfied.

Appendix B. Geometrical configurations and their discretization. Numerical experiments with the Schur complement in §3.2 and the coupled problem in §4 are considered on sequences of uniformly refined meshes, discretizing the geometrical configurations shown in Figure 4.1. The Schur complement experiment is considered with straight segments $\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0, 1]\}$ or $\Gamma = \{(t, t, t); t \in [0, 1]\}$. For each case the domains are discretized in three ways: (*uniform*) the meshes for Ω, Γ have the same characteristic size, (*finer*) the mesh of Ω is finer at Γ than in the rest of the domain, (*coarser*) the mesh of Ω is coarser at Γ than in the rest of the domain. Parameters of the meshes for each refinement level are summarized in Table B.1.

Table B.1: Sizes of FEM spaces and mesh parameters for different levels of refinements (L). The length of the largest cell in the mesh of Γ is denoted by H. For readability the reported value is $H \times 10^3$. Lengths of smallest/largest edges of cells of the mesh for $\Omega \backslash \Gamma$ are respectively $h_{\min}$ and $h_{\max}$. (Top) In *uniform* discretization the characteristic mesh size of Ω and Γ triangulations are identical. (Middle) *Finer* discretization uses finer mesh near Γ. (Bottom) In the *coarser* cases the mesh of Γ is coarser near the curve.

L	$\Gamma = \{(t, \frac{1}{2}, \frac{1}{2}); t \in [0, 1]\}$					$\Gamma = \{(t, t, t); t \in [0, 1]\}$				
	$\dim V_h$	$\dim Q_H$	$\frac{h_{\min}}{H}$	$\frac{h_{\max}}{H}$	H	$\dim V_h$	$\dim Q_H$	$\frac{h_{\min}}{H}$	$\frac{h_{\max}}{H}$	H
1	5K	17	1.7	1.7	62.5	5K	17	1.0	1.0	108.3
2	36K	33	1.7	1.7	31.2	36K	33	1.0	1.0	54.1
3	275K	65	1.7	1.7	15.6	275K	65	1.0	1.0	27.1
4	2.1M	129	1.7	1.7	7.8	2.1M	129	1.0	1.0	13.5
5	6.1M	183	1.7	1.7	5.5	6.1M	183	1.0	1.0	9.5
1	12K	161	1.1	32.9	6.2	9K	187	1.0	22.5	9.4
2	72K	321	1.0	35.3	3.1	46K	373	0.9	24.7	4.7
3	476K	641	0.9	39.0	1.6	308K	742	0.8	27.3	2.3
4	3.7M	1281	0.8	40.6	0.8	2.2M	1481	0.8	27.0	1.2
5	6.8M	1601	0.7	40.8	0.6	7.4M	2220	0.8	27.0	0.8
1	11K	9	0.2	1.7	125.0	5K	16	0.2	1.7	122.5
2	59K	17	0.2	1.9	62.5	30K	31	0.2	2.0	61.2
3	375K	33	0.2	2.1	31.2	194K	59	0.2	2.2	30.6
4	2.7M	65	0.2	2.1	15.6	1.4M	114	0.2	2.3	15.5
5	8.5M	97	0.2	2.5	10.4	4.4M	169	0.2	3.2	10.4

REFERENCES

[1] R. A. Adams and J. F. Fournier, *Sobolev spaces*, vol. 140, Academic press, 2003.

[2] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. Rognes, and G. Wells, *The FEniCS project version 1.5*, Archive of Numerical Software, 3 (2015).

[3] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, third ed., 1999.

[4] M. Arioli, D. Kourounis, and D. Loghin, *Discrete fractional Sobolev norms for domain decomposition preconditioning*, IMA Journal of Numerical Analysis, (2012), p. drr024.

[5] M. Arioli and D. Loghin, *Discrete interpolation norms with applications*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 2924–2951.

[6] I. Babuška, *The finite element method with Lagrangian multipliers*, Numerische Mathematik, 20 (1973), pp. 179–192.

[7] C. Bacuta, *A unified approach for Uzawa algorithms*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 2633–2649.

[8] S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, *PETSc users manual*, Tech. Report ANL-95/11 - Revision 3.4, Argonne National Laboratory, 2013.

[9] F. Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique, 8 (1974), pp. 129–151.

[10] L. Cattaneo and P. Zunino, *A computational model of drug delivery through microcirculation to compare different tumor treatments*, International Journal for Numerical Methods in Biomedical Engineering, 30 (2014), pp. 1347–1371.

[11] ———, *Computational models for fluid exchange between microcirculation and tissue interstitium*, Networks and Heterogeneous Media, 9 (2014), pp. 135–159.

[12] W. Dahmen and A. Kunoth, *Appending boundary conditions by Lagrange multipliers: Analysis of the lbb condition*, Numerische Mathematik, 88 (2001), pp. 9–42.

[13] C. D'Angelo, *Finite element approximation of elliptic problems with Dirac measure terms in weighted spaces: applications to one-and three-dimensional coupled problems*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 194–215.

[14] C. D'Angelo and A. Quarteroni, *On the coupling of 1D and 3D diffusion-reaction equations: Application to tissue perfusion problems*, Mathematical Models and Methods in Applied Sciences, 18 (2008), pp. 1481–1504.

[15] R. D. Falgout and U. Meier Yang, *hypre: A library of high performance preconditioners*, in Computational Science ICCS 2002, P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, and J. J. Dongarra, eds., vol. 2331 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2002, pp. 632–641.

[16] Q. Fang, S. Sakadžić, L. Ruvinskaya, A. Devor, A. M. Dale, and D. A. Boas, *Oxygen advection and diffusion in a three-dimensional vascular anatomical network*, Optics express, 16 (2008), pp. 17530–17541.

[17] L. Grinberg, E. Cheever, T. Anor, J. R. Madsen, and G. E. Karniadakis, *Modeling blood flow circulation in intracranial arterial networks: a comparative 3D/1D simulation study*, Annals of biomedical engineering, 39 (2011), pp. 297–309.

[18] N. Hale, N. J. Higham, and L. N. Trefethen, *Computing a^{α}, $\log A$, and related matrix functions by contour integrals*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2505–2523.

[19] T. Koppl and B. Wohlmuth, *Optimal a priori error estimates for an elliptic problem with Dirac right-hand side*, SIAM Journal on Numerical Analysis, 52 (2014), pp. 1753–1769.

[20] 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 Journal on Scientific Computing, 38 (2016), pp. B962–B987.

[21] A. A. Linninger, I. G. Gould, T. Marinnan, C.-Y. Hsu, M. Chojecki, and A. Alaraj, *Cerebral microcirculation and oxygen tension in the human secondary cortex*, Annals of biomedical engineering, 41 (2013), pp. 2264–2284.

[22] J. L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications*, vol. 1, Springer Science & Business Media, 2012.

[23] A. Logg, K.-A. Mardal, and G. Wells, *Automated solution of differential equations by the finite element method: The FEniCS book*, vol. 84, Springer Science & Business Media, 2012.

[24] D.S. Malkus, *Eigenproblems associated with the discrete LBB condition for incompressible finite elements*, International Journal of Engineering Science, 19 (1981), pp. 1299–1310.

[25] K.-A. Mardal and J. B. Haga, *Block preconditioning of systems of PDEs*, in Automated Solution of Differential Equations by the Finite Element Method, G. N. Wells et al. A. Logg, K.-A. Mardal, ed., Springer, 2012.

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

[27] M. Nabil and P. Zunino, *A computational study of cancer hyperthermia based on vascular magnetic nanoconstructs*, Open Science, 3 (2016).

[28] P. Peisker, *On the numerical solution of the first biharmonic equation*, ESAIM: Mathematical Modelling and Numerical Analysis - Modlisation Mathmatique et Analyse Numrique, 22 (1988), pp. 655–676.

[29] J. Pestana and A. J. Wathen, *Natural preconditioning and iterative methods for saddle point systems*, SIAM Review, 57 (2015), pp. 71–91.

[30] J. Pitkäranta, *Boundary subspaces for the finite element method with Lagrange multipliers*,

Numerische Mathematik, 33 (1979), pp. 273–289.

[31] J. Reichold, M. Stampanoni, A. L. Keller, A. Buck, P. Jenny, and B. Weber, *Vascular graph model to simulate the cerebral blood flow in realistic vascular networks*, Journal of Cerebral Blood Flow & Metabolism, 29 (2009), pp. 1429–1443.

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

[33] D. Silvester and A. Wathen, *Fast iterative solution of stabilised Stokes systems part ii: using general block preconditioners*, SIAM Journal on Numerical Analysis, 31 (1994), pp. 1352–1367.

[34] O. Steinbach, *Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements*, Texts in applied mathematics, Springer New York, 2007.

[35] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, 1997.

[36] Wolfram Research Inc., *Mathematica 8.0*, 2010.

Paper III

Characterisation of the space of rigid motions in arbitrary domains

M. Kuchta, K.-A. Mardal, and M. Mortensen

Proc. of 8th National Conference on Computational Mechanics, Barcelona, Spain, 2015, CIMNE.

8. National Conference on Computational Mechanics
MekIT'15
B. Skallerud and H I Andersson (Eds)

CHARACTERISATION OF THE SPACE OF RIGID MOTIONS IN ARBITRARY DOMAINS

MIROSLAV KUCHTA[1], KENT-ANDRE MARDAL[1,2] AND MIKAEL MORTENSEN[1,2]

[1] Department of Mathematics, Division of Mechanics, University of Oslo,
0316 Oslo, Norway

[2] Center for Biomedical Computing, Simula Research Laboratory,
P.O. Box 134, No-134 Lysaker, Norway

Key words: rigid motions, linear elasticity, singular problems, conjugate gradients

Abstract. The Neumann problem for linear elasticity is singular with a kernel formed by rigid motions. Unless removed, the kernel causes problems for both direct and iterative methods. In this paper we shall first discuss how the basis of the nullspace may be used to formulate a well posed problem. Afterwards we present a simple and efficient technique for computing an orthonormal basis of the kernel. We give examples where the technique is used to characterize the space of rigid motions in complex domains. Finally, we show how iterative methods may exploit this basis for efficiency.

1 Introduction

In this paper we shall study efficient iterative methods for solving the linear elasticity equation subject to Neumann boundary conditions. The problem reads

$$\begin{aligned} -\mathrm{div}\cdot\left(\sigma\left(u\right)\right) &= f \;\text{ in } \Omega, \\ \sigma\left(u\right)\cdot n &= h \;\text{ on } \partial\Omega, \end{aligned} \tag{1}$$

where u is the unknown displacement vector while σ is the (symmetric) stress tensor defined by the constitutive law $\sigma\left(u\right) = 2\mu\epsilon\left(u\right) + \mathrm{tr}\left(\epsilon\left(u\right)\right)\mathbb{I}$ with μ, λ denoting the material parameters and $\epsilon\left(u\right)$ the symmetric part of the displacement gradient. Moreover $\mathbb{I}$ is the identity tensor while tr denotes a trace. Finally n is the outer normal vector of the body Ω and f, h are respectively the volume and surface forces acting on the body.

Variants of the Neumann problem describe a range of physical phenomena. For example in Tobie et. al. [1] the solution provides insight into internal processes of the Saturn's moon Enceladus, while Sanderud [2] uses the Neumann problem to study a condition of the brain known as hydrocephalus. In order to solve (1) one must recognize the fact that the problem is not well posed. First of all, a solution exists if and only if the net force and the net torque on the body are zero (see, e.g., [4]). This condition places a restriction on the forces f, h and it is also referred to as a compatibility condition. Secondly, even if u solves (1) for some compatible data, the solution is not unique. In fact, for $\Omega \subset \mathbb{R}^d, d = 2, 3$ there exist, respectively, three and six linearly independent functions z^i such that $\sigma\left(z^i\right) = 0$. Consequently function $u + v$ for v an arbitrary combination of z^i is also a solution of (1). Such functions v are called rigid motions. Due to the ambiguity of solution a careless discretization of the problem results in a singular system.

To circumvent the singular system a displacement can be fixed in d nodes or different boundary conditions can be used for the continuous problem. For example, Dirichlet boundary condition may be prescribed on some part of the boundary or Robin boundary conditions may be used with a small weight factor controlling the displacement. Both of these methods modify the matrix of the system such that it is no longer singular. Bochev and Lehoucq [3] report that the preferred method amongst practitioners is to fix the solution datum. The latter approach is applied in [2] where it is also found that the solution close to the boundary is affected by the modified boundary conditions. If modifying the system matrix is not desirable, the solution can be obtained by Krylov subspace iterations. In fact, a symmetric singular linear system $\mathbb{A}u = b$ can be solved by the Krylov method provided that the vector b is l^2-orthogonal to the nullspace of $\mathbb{A}$ and the approximate solution remains orthogonal to the nullspace throughout the iterations. The approach is supported by linear algebra packages such as Trilinos [6], PETSc [5], or PyAMG [13] where the nullspace is provided in the form of an l^2-orthonormal basis. As we shall demonstrate in the later sections the numerical solution obtained by the iterative method is not always a good approximation to the true solution. Moreover the convergence rate of the error in the L^2-norm is suboptimal.

Broadly speaking the methods for solving (1) listed thus far treat the singularity on a discrete/algebraic level. Either the invertible system was obtained by modification of the matrix via artificial boundary conditions or the kernel of the matrix was considered. On the other hand there exists a class of methods where the singularity is treated on a continuous level. A well known method of this type is the Lagrange multipliers which yields a saddle point formulation of (1) and after discretization a symmetric indefinite linear system. The main focus of the current paper is a symmetric positive definite formulation of (1) and its connection to the saddle point formulation. In Bochev and Lehoucq [3] similar idea is explored in the context of augmented/stabilized Lagrangian formulation of the singular Poisson problem. Here we shall follow a different line of reasoning.

The symmetric positive definite formulation is derived from an abstract framework for solving singular problems with a known kernel which is specified as an orthogonal basis. In the framework the relation of the formulation to the saddle point problem becomes evident. The framework is developed in the next section. Then in Section 3 the basis for the space of rigid motions over arbitrary body is constructed. Section 4 begins with a discussion of preconditioners that are used to solve efficiently the linear system stemming from the proposed formulation of the Neumann problem (1). Afterwards the numerical results are shown to verify the properties of the method. Finally the singular Poisson problem is used to compare the presented method with Krylov method for solving symmetric singular linear systems.

2 Abstract framework for singular problems with known kernel

For completeness and clarity we list here some basic properties of the singular variational problems. The results can be found in books on mathematical theory of the finite element method, e.g., [4, 11].

We let V denote a Hilbert space over body Ω and $(\cdot,\cdot)$ the L^2-inner product. We assume that $a : V \times V \mapsto \mathbb{R}$ is a symmetric continuous bilinear form and that there exist finite $k = \dim(Z)$, where Z is the nullspace of the bilinear form. For $z \in Z$, we have $a\left(z, v\right) = 0$ for any $v \in V$. Finally we assume that the form is coercive on the orthogonal complement of Z in V which we denote as $Z^{\perp}$. Under these assumptions on a we want to solve a variational problem:

$$\text{Find } u \in V \text{ such that } a\left(u, v\right) = L(v) \quad \forall v \in V, \tag{2}$$

for some linear continuous functional L. The problem (2) is not solvable for all L. In fact the solution exists if and only if for all $z \in Z$ we have $L(z) = 0$. For such (compatible) L the solution is not unique as for u some solution of (2) the function $u+z$ is also a solution. The space V is therefore too large to find a unique solution. However a unique solution of (2) with compatible L exists in $Z^{\perp}$. The unique solution can be found by considering a constrained saddle point problem: Find $[u, \alpha] \in V \times \mathbb{R}^k$ such that

$$a\left(u, v\right) + \alpha_j\left(v, z^j\right) + \beta_j\left(u, z^j\right) = L(v) \quad \forall\, [v, \beta] \in V \times \mathbb{R}^k. \tag{3}$$

Here $z^j, j = 1, \cdots, k$ are the orthonormal basis functions of Z (note that the basis functions here do not need to be orthonormal). Problem (3) is uniquely solvable since the requirements of the Brezzi theory [10] on the form a are satisfied by assumption while choosing $w = z^1 + z^2 + \cdots + z^k$ yields

$$\sup_{v \in V} \frac{\alpha_j\left(v, z^j\right)}{\|v\|_V} \geq \frac{\alpha_j\left(w, z^j\right)}{\|v\|_V} = \frac{\alpha^T 1}{\|v\|_V}.$$

The inf-sup condition is thus satisfied with constant $\gamma = \frac{1}{\|w\|_V} > 0$. Here $\alpha^{\mathrm{T}}1$ denotes an Euclidean inner product between vectors α and 1. We remark that in (2) no compatibility condition is required from L.

In the saddle point formulation k new unknowns have been introduced. However, an informal calculation shows that the vector of Lagrange multipliers $\alpha \in \mathbb{R}^k$ is not a true unknown. Indeed, testing (3) with functions $[z^i, 0]\,, i = 1, 2, \cdots k$ it follows that $\alpha_i = L(z^i)$. The variational problem can now be rearranged yielding: Find $u \in V$ such that

$$a\left(u, v\right) + \beta_j \left(u, z^j\right) = L\left(v - \left(v, z^i\right) z^i\right) \quad \forall\left[v, \beta\right] \in V \times \mathbb{R}^k. \tag{4}$$

In (4) observe that $L\left(v - \left(v, z^i\right) z^i\right) = 0$ for all $v \in Z$ and thus the role of α is to ensure that the functional L is compatible. We remark that by testing (4) and (3) with $[0, \beta]$ it follows that the solution belongs to the orthogonal complement $Z^\perp$.

The observation about the role of multipliers suggests that there exists a formulation where the only unknown is the primary unknown u. The formulation can be obtained based on the following reasoning. Let operator $\mathcal{A}$ be defined by

$$\left\langle \mathcal{A}\left(\left[u, \alpha\right]\right), \left[v, \beta\right]\right\rangle = a\left(u, v\right) + \alpha_j \left(v, z^j\right) + \beta_j \left(u, z^j\right),$$

where $\left[u, \alpha\right], \left[v, \beta\right] \in V \times \mathbb{R}^k$. Further let $\mathcal{B}$,

$$\left\langle \mathcal{B}\left(\left[u, \alpha\right]\right), \left[v, \beta\right]\right\rangle = a\left(u, v\right) + \left(u, z_j\right)\left(v, z_j\right) + \alpha^{\mathrm{T}}\beta.$$

Operator $\mathcal{B}$ defines an inner product over $V \times \mathbb{R}^k$. The norm it induces controls the multiplier in its natural space l^2. Further letting $u = u_Z + u_{Z^\perp}$, where $u_Z = \left(u, z^j\right) z^j$, $u_{Z^\perp} = u - \left(u, z^j\right) z^j$, we have $\left\langle \mathcal{B}\left(\left[u, \alpha\right]\right), \left[u, \alpha\right]\right\rangle = a\left(u_{Z^\perp}, u_{Z^\perp}\right) + \left(u_Z, z_j\right)\left(u_Z, z_j\right) + \alpha^{\mathrm{T}}\alpha$. The norm induced by $\mathcal{B}$ thus controls the part of u in the orthogonal complement only by the energy norm and in this sense it is optimal.

Next consider a generalized eigenvalue problem $\mathcal{A}\left(\left[u, \alpha\right]\right) = \lambda\mathcal{B}\left(\left[u, \alpha\right]\right)$. Seeing how the norm induced by $\mathcal{B}$ separates the kernel from the complement we propose functions $\left[z^j, -t^j\right], j = 1, 2, \cdots, k$ and $\left[w, 0\right]$ as solution candidates. Here t^j is the j-th standard unit vector in $\mathbb{R}^k$ and w is arbitrary element of $Z^\perp$. Indeed we get by a direct calculations that for all $\left[v, \beta\right] \in V \times \mathbb{R}^k$

$$a\left(w, v\right) = \lambda a\left(w, v\right) \qquad \text{and} \qquad -\left(v, z^j\right) + \beta_j = \lambda\left(\left(v, z^j\right) - \beta_j\right). \tag{5}$$

Thus $(1, [w, 0])$ and $(-1, [z^j, -t^j])$ are solutions of the eigenvalue problem. As a corollary we have that a spectrum of the generalized eigenvalue problem is formed only by eigenvalues $\lambda = 1$ and $\lambda = -1$. Moreover, the number of negative eigenvalues matches the dimension of the kernel Z. A computational verification of the claims can be found in Figure 1 which shows computed spectrum of the generalized eigenvalue problem for (1) posed over a three dimensional domain. As a further observation note that the second equation in (5) is also solved by $(1, [z^j, t^j])$. The first set of solutions, $(-1, [z^j, -t^j])$, highlights the indefiniteness of $\mathcal{A}$ but with the second one $\lambda = 1$ is the only eigenvalue. Therefore for any $\left[u, \alpha\right] \in V \times \mathbb{R}^k$ it holds that $\mathcal{A}\left(\left[u, \alpha\right]\right) = \mathcal{B}\left(\left[u, \alpha\right]\right)$. Thus especially for $\alpha = 0$ we have

$$\left\langle \mathcal{A}\left(\left[u, 0\right]\right), \left[v, \beta\right]\right\rangle = a\left(u, v\right) + \left(u, z^j\right)\left(v, z^j\right).$$

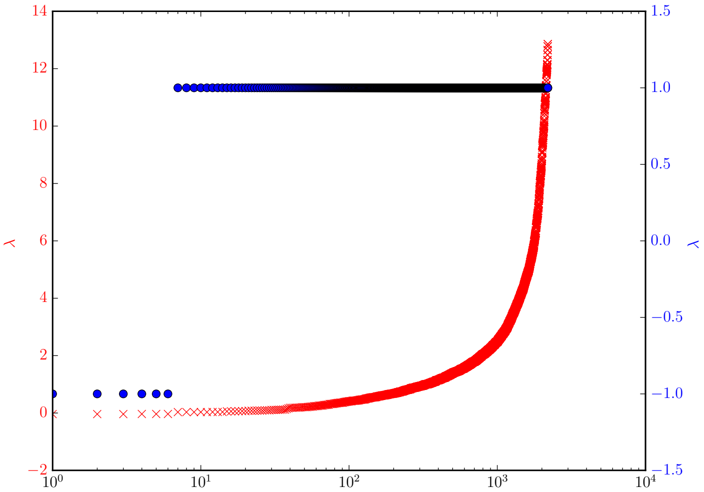

Figure 1: Spectrum of the (generalized) eigenvalue problem for linear elasticity equation with Neumann boundary conditions. Spectrum of the eigenvalue problem for operator $\mathcal{A}$ is plotted in red against the left axis while the spectrum of the generalized eigenvalue problem is shown in the blue color against the right axis. The body was $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^3$ and piecewise-linear continuous elements were used to construct space V_h. In agreement with our analysis the computed spectrum consists only of eigenvalues $\lambda = 1$ and $\lambda = -1$, where the number of the negative eigenvalues is six. Recall that for a three dimensional body the space of rigid motions is six dimensional.

Finally comparing the left hand side with (4) we observe that for *compatible* L the solution of saddle point problem (3) can be instead obtained as a solution to the symmetric positive-definite problem: Find $u \in V$ such that

$$a\left(u, v\right) + \left(u, z^j\right)\left(v, z^j\right) = L\left(v\right) \quad v \in V. \tag{6}$$

Note that by compatibility of L the solution of (6) belongs to $Z^{\perp}$. Also, observe that for $u \in Z^{\perp}$ it holds that $(u, z^j)\,(v, z^j) = 0$ for all $v \in V$. The additional term is therefore consistent.

While (6) was derived from reasoning about spectral properties of operators $\mathcal{A}, \mathcal{B}$ we note that the resulting formulation can be also viewed as a special case of a variational problem: Find $u \in V$ such that

$$a\left(u, v\right) + \gamma\left(u, z^j\right)\left(v, z^j\right) = L\left(v\right) \quad v \in V, \tag{7}$$

which defines extreme points of the augmented Lagrangian

$$\mathcal{L}\left(u\right) = \frac{1}{2}a\left(u, u\right) + \gamma\frac{\left(u, z^j\right)\left(u, z^j\right)}{2} - L\left(u\right),$$

for some $\gamma \neq 0$. Application of the augmented Lagrangian in the context of finite element method can be found in the pioneering work of Babuška [16] where it is referred to as a variational principle with penalty. We remark that therein $\gamma = \gamma\left(h\right)$ where h is a parameter of the discretization. In [3] the formulation (7) with constant stabilization/penalty parameter is used to solve the singular Poisson problem.

We remark that the augmented Lagrangian formulation (7) can be included into the presented spectral considerations by replacing $\mathcal{B}$ in the eigenvalue problem $\mathcal{A}\left(\left[u,\alpha\right]\right) = \lambda\mathcal{B}\left(\left[u,\alpha\right]\right)$ by operator $\mathcal{B}_\gamma$

$$\left\langle \mathcal{B}_\gamma\left(\left[u,\alpha\right]\right),\left[v,\beta\right]\right\rangle = a\left(u,v\right) + \gamma\left(u,z_j\right)\left(v,z_j\right) + \gamma\alpha^{\mathrm{T}}\beta.$$

With the substitution the eigenvalues of the problem are $\lambda = 1$ and $\lambda = -\gamma^{-1}$.

2.1 Finite element method for the symmetric positive-definite formulation

To solve Eq. (6) we let V_h denote an N-dimensional finite element space subspace of V. The variational problem posed over the constructed space then reads: Find $u_h \in V_h$ such that

$$a\left(u_h,\phi^i\right) + \left(u_h,z^j\right)\left(\phi^i,z^j\right) = L\left(\phi^i\right) \quad i = 1,2,\cdots N, \tag{8}$$

where ϕ^i are the basis functions of V_h. We let $\mathbb{A} \in \mathbb{R}^{N\times N}$ denote a matrix corresponding to the bilinear form a. The components of the matrix are $\mathbb{A}_{ij} = a\left(\phi^i,\phi^j\right)$. For every function $z^j \in Z$ we let z_h^j denote its interpolant in V_h and $\pi(z_h^j) \in R^N$ is then the primal representation(see Mardal and Winther [8]) of the interpolant. Further we define matrix $\mathbb{Z} \in \mathbb{R}^{N\times k}$ whose j-column is $\pi(z_h^j)$ and the mass matrix $\mathbb{M} \in \mathbb{R}^{N\times N}$ with components $\mathbb{M}_{ij} = \left(\phi^i,\phi^j\right)$. Finally let $b \in \mathbb{R}^N$ be the vector representing the right-hand side of (6), i.e., $b_j = L\left(\phi^j\right)$ and $u \in \mathbb{R}^N$ be the unknown vector of expansion coefficients of the solution u_h, i.e., $u_h = u_j\phi^j$. With this notation the discretization of (6) yields a linear system

$$\mathbb{A}u + \left(\mathbb{M}\mathbb{Z}\right)\left(\mathbb{M}\mathbb{Z}\right)^{\mathrm{T}}u = b. \tag{9}$$

We note that while matrix $\mathbb{A}$ is sparse, the second matrix in (9) is dense and thus storing the linear system requires $\mathcal{O}\left(N^2\right)$ storage. However, if the vectors $\mathbb{M}\pi\left(z_h^j\right)$ are stored ($\mathcal{O}\left(N\right)$ additional storage requirement) the matrix-vector product which is needed for iterative methods can be obtained efficiently in $\mathcal{O}\left(N\right)$ operations.

Recall that vector b should represent the discretization of a compatible functional L. For any L a compatible functional is obtained by $L\left(v\right) \leftarrow L\left(v - \left(v,z^j\right)z^j\right), v \in V$. For $v \in V_h$ this transformation becomes

$$b \leftarrow \left(\mathbb{I} - \left(\mathbb{M}\mathbb{Z}\right)\mathbb{Z}^{\mathrm{T}}\right)b. \tag{10}$$

and can again be represented in $\mathcal{O}\left(N\right)$ storage and operations. We remark that orthogonalization of *function* $v \in V_h$ with respect to Z is provided by a matrix $\mathbb{I} - \mathbb{Z}\left(\mathbb{M}\mathbb{Z}\right)^{\mathrm{T}}$ which is a transpose of the transformation which orthogonalized functionals.

2.2 Neumann problem of linear elasticity in the general framework

To apply the presented framework to the Neumann problem (1), the first Korn's lemma (see, e.g., [12]) is used to show coercivity of the bilinear

$$a\left(u,v\right)=2\mu\left(\epsilon\left(u\right),\epsilon\left(u\right)\right)+\lambda\left(\mathrm{div}u,\mathrm{div}v\right), \tag{11}$$

over the orthogonal complement of the rigid motions. Here the space V is the Sobolev space $H^1\left(\Omega\right)$. For completeness we state here also the linear form used in the variational problem

$$L\left(v\right)=\int_{\Omega}f^{\mathrm{T}}v\,\mathrm{d}x+\int_{\partial\Omega}h^{\mathrm{T}}v\,\mathrm{d}s.$$

Further ingredient of the framework, the orthonormal basis of the space of rigid motions of arbitrary body Ω is needed. Construction of such basis is the subject of the following section.

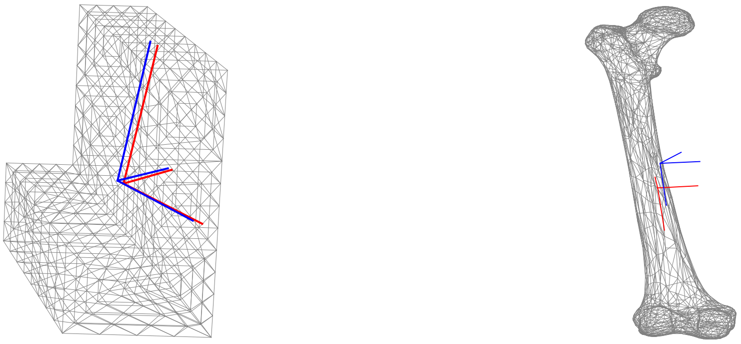

Figure 2: Eigenvectors of the matrix $\mathbb{G}$ which define the basis of the space of rigid motions. In red and blue the vectors used to define respectively the L^2-orthonormal and l^2-orthonormal basis are shown.

3 Constructing an orthonormal basis for the space of rigid motions

Let $t^i, i=1,2,3$ denote the Cartesian unit vectors. With r the position vector in $\mathbb{R}^3$ we further define $r^i=r\times t^i, i=1,2,3$. For any $L>0$ the set $RM_0=\{t^i,r^i\}_{i=1}^{i=3}$ is then an orthogonal basis of rigid motions for body $\Omega_0=\left[-L,L\right]^3$, where functions t^i,r^i are, respectively, the basis of translations and rotations. It is obvious that for general domain Ω an orthogonal basis of rigid motions can be obtained by using Gramm-Schmidt orthogonalization properly on RM_0. However it is the author's opinion that such a process leaves out several physical insights. Therefore we advocate a constructive method which is closely connected to rotational motions.

Let $\mathbb{D} \in \mathbb{R}^{6\times 6}$ be the matrix of mutual L^2-inner products between the functions from RM_0

$$\mathbb{D} = \begin{bmatrix} (t^i, t^j)\,, (t^i, r^j) \\ (r^i, t^j)\,, (r^i, r^j) \end{bmatrix}.$$

Clearly orthonormality of RM_0 is equivalent to $\mathbb{D} = \mathbb{I}$. If the body Ω_0 is shifted so that the geometrical center of the domain is no longer in the origin, then the basis seizes to be orthogonal. However, using identity $(r \times t^i)^{\mathrm{T}}\, t^j = (t^i \times t^j)^{\mathrm{T}}\, r$ we get that a transformation $r \leftarrow r - c$ where $c = \frac{(r,1)}{(1,1)}$ restores orthogonality. Physically this transformation means that the rotations are described relative to the geometrical center c. In fact, this transformation restores nullity of the translation-rotation block of $\mathbb{D}$ for any deformation of the body. However it is not sufficient to yield a diagonal block measuring inner products between rotations. Let then $\mathbb{G} \in \mathbb{R}^{3\times 3}, \mathbb{G}_{ij} = ((r-c) \times t^i, (r-c) \times t^j)$. The matrix $\mathbb{G}$ being symmetric, positive-definite has three orthonormal eigenvectors e^i and three positive (not necessarily distinct) eigenvalues λ_i. We now set $t^i = e^i$ while the rotations are defined with respect to the eigenvectors. By properties of the eigenvectors, the matrix $\mathbb{D}$ assembled in this new basis is diagonal with $|\Omega| = (1, 1)$ the diagonal values for the translation-translation block, while the rotation-rotation entries are the eigenvalues. The L^2-orthonormal basis for arbitrary body is therefore

$$RM = \left\{ \frac{e^1}{\sqrt{|\Omega|}}, \frac{e^2}{\sqrt{|\Omega|}}, \frac{e^3}{\sqrt{|\Omega|}}, \frac{(r-c) \times e^1}{\sqrt{\lambda_1}}, \frac{(r-c) \times e^2}{\sqrt{\lambda_2}}, \frac{(r-c) \times e^3}{\sqrt{\lambda_3}} \right\}$$

To uncover the physical importance of the constructed basis note that $\mathbb{G}$ can be equivalently written as

$$\int_\Omega (r-c)^{\mathrm{T}} (r-c)\, \mathbb{I} - (r-c) \otimes (r-c)\; \mathrm{d}r.$$

Gurtin in [7] gives the last expression as the inertia tensor relative to c, while the eigenvalues λ_i are termed moments of inertia and the eigenvectors e^i form a coordinate system that simplifies considerations about the kinetic energy of rotation. In this sense the constructed basis RM is a natural one. We remark that in the basis the translations are "along“ the eigenvectors while the rotations are "around“ the eigenvectors.

The L^2-orthonormal basis of the rigid motions, represented in some finite element function space V_h, is obtained simply by interpolating the functions in RM. Further the six vectors in $\mathbb{R}^N$ that form an l^2-orthonormal basis of the rigid motions in the Euclidean space can be obtained by the constructive method with two modifications: (i) the interpolated functions are replaced by their primal representation, (ii) the L^2-inner products in the definition of $\mathbb{G}$ and c are replaced by the discrete l^2-inner product. This basis can be used by the Krylov methods mentioned in the introduction. In Figure 2 we present some examples of the constructed basis for two complex domains. Rather than by plotting the vector fields, the difference between the basis is illustrated by showing the eigenvectors. In general, the defining axes for l^2 and L^2-orthonormal basis are rotated

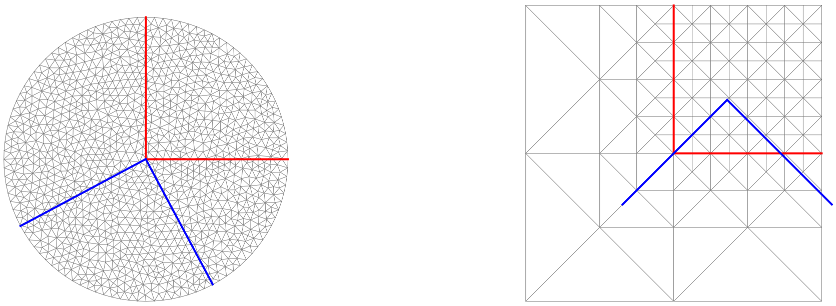

Figure 3: Eigenvectors of the matrix $\mathbb{G}$ which define the basis of the space of rigid motions. In red and blue the vectors used to define respectively the L^2-orthonormal and l^2-orthonormal basis are shown. On the right figure note that only the L^2-inner product yields c as the true center of gravity.

and/or translated with respect to each other. However, this difference should not be interpreted as a difference between spaces. The vectors from the l^2-orthonormal basis and the primal vector representation of the L^2-orthonormal basis span the same subspace of $\mathbb{R}^N$.

Finally we note that the method can be equally well used to construct an orthonormal basis for rigid motions of $\Omega \subset \mathbb{R}^2$. With c still denoting the center of mass of the body, we define $\mathbb{G}_{ij} = \Big((r-c)_i\,, (r-c)_j\Big)$. We denote e^0, e^1 the two orthonormal eigenvectors of $\mathbb{G}$ and λ_1, λ_2 the two corresponding eigenvalues. Then

$$RM = \left\{ \frac{e^1}{\sqrt{|\Omega|}}, \frac{e^2}{\sqrt{|\Omega|}}, \frac{(r-c)\times(e^1\times e^2)}{\sqrt{\lambda_1+\lambda_2}} \right\}$$

is the L^2-orthonormal basis of the rigid motions for body Ω. Two examples of the basis constructed with the method are shown in Figure 3. Note that for the left body the center points c_l, c_L computed respectively using the l^2 and L^2-inner products are practically identical. On the other hand the center points for the right body are clearly distinct. The relative distance between the computed centers c_l, c_L is directly linked to the approximation properties of the discrete inner product. On a homogeneous triangulation the l^2-inner product provides a good(convergent) approximation of the L^2-inner product whereas on an inhomogeneous triangulation the two inner products diverge in general.

4 Validation of method

In this section the formulation (8) presented in Section 2 and the basis for the space of rigid motions presented in Section 3 are used to solve the Neumann problem (1). The numerical experiments presented here have been conducted using the software frameworks

FEniCS[15] and cbc.block[14]. The problem has been discretized using piecewise linear continuous finite elements and the resulting linear system has been solved by the preconditioned conjugate gradient(CG) method. As as a preconditioner we have employed the multigrid method(ML) with a linear system $\mathbb{A}$+$\mathbb{M}$. We used the multigrid implementation from the PETSc library[5].

Before discussing the numerical results we shall comment on the choice of the preconditioner. Let $\mathcal{A}$ be defined as

$$\langle \mathcal{A}(u), v\rangle = a(u,v) + \left(u, z^j\right)\left(v, z^j\right),$$

with the bilinear form (11) and z^j the orthonormal basis functions of RM over Ω. Using Korn's second inequality (see, e.g., Marsden [12]) the operator $\mathcal{A}$ is an isomorphism between $V = H^1(\Omega)$ and its dual space. Following Mardal and Winther [8] a canonical preconditioner for the problem (6) is then based on an operator $\langle \mathcal{B}_1(u), v\rangle = (\nabla u, \nabla v) + (u, v)$ which defines and inner product over V. However, the bounds provided by the Korn's inequality are not sufficiently sharp and the condition number of the preconditioned system is about three hundred. To obtain a better conditioned system we consider an operator

$$\langle \mathcal{B}(u), v\rangle = a(u,v) + (u,v).$$

We do not discuss spectral equivalence of operators $\mathcal{B}, \mathcal{B}_1$. Instead, the relevance of the proposed operator is illustrated numerically by computing the spectrum of a generalized eigenvalue problem $\mathcal{A}u = \lambda\mathcal{B}u$. The calculated spectrum of two and three dimensional rectangular bodies is shown in Figure 4. For both domains the spectrum lies within a narrow interval $\lambda \in (0.97, 1]$. The upper bound can be obtained analytically. Indeed a simple calculation shows that $\lambda = 1, u = z^j$ solves the eigenvalue problem. Moreover $\lambda = 1$ if and only if $u \in RM$. The condition number estimated from the calculated spectrum is about 1.02. This establishes $\mathcal{B}^{-1}$ as a suitable preconditioner for problem (8). We note that the matrix $\mathbb{A} + \mathbb{M}$ is a discretization of operator $\mathcal{B}$.

The results of a convergence study for a two dimensional domain $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$ are presented in Table 1. In the tests we set traction force h equal to zero while as the volume force we set $f = (f_x, f_y)$, where $f_x = 4\lambda\sin 2x + 8\mu\sin 2x + 2\sqrt{6}y + 1$ and $f_y = 9\lambda\sin 3y + 18\mu\sin 3y - 2\sqrt{6}x$. The volume force was purposely chosen to render the functional L incompatible. With the compatible functional the problem (6) has a solution $u = (\sin 2x, \sin 3y)$. We observed optimal convergence rate of the computed solution in both the H^1 and the L^2-norm.

To evaluate the performance of the method on a three dimensional problem we used domain $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^3$. The surface force was identically zero while for the volume force $f = (f_x, f_y, f_z)$, $f_x = 4\lambda\sin 2x + 8\mu\sin 2x + \sqrt{6}y + 1$, $f_y = 9\lambda\sin 3y + 18\mu\sin 3y - \sqrt{6}x$, $f_z = \lambda\sin z + 2\mu\sin z - 2$ was used. Again the linear form with this data was not orthogonal to the rigid motions. After orthogonalization the problem allows solution $u = (\sin 2x, \sin 3y, \sin z)$. The results of the convergence study are presented in Table 2. Similar to the two dimensional case we observed optimal convergence rates in both norms.

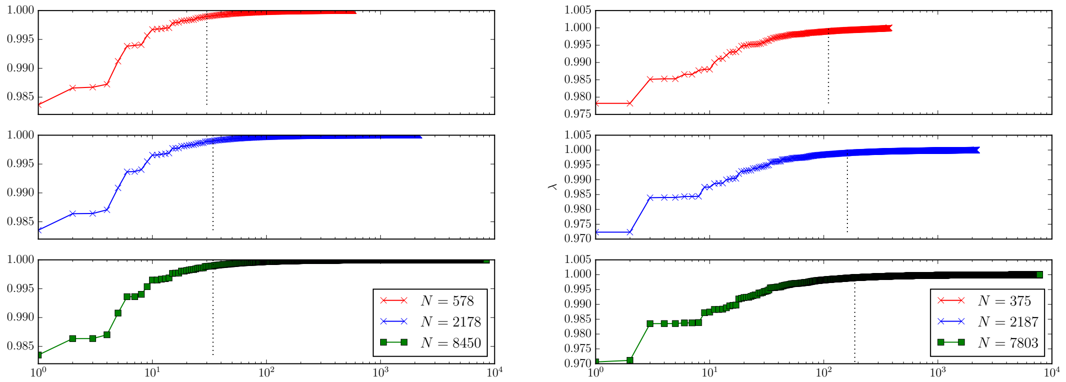

Figure 4: Spectrum of the generalized eigenvalue problem for the preconditioned formulation (6) of the Neumann problem (1). (Left) Domain $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$. (Right) $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^3$. The spectrum lies in a narrow interval bounded from above by $\lambda = 1$. The dashed line indicates the first eigenvalue with magnitude greater than $1 - 1.10^{-3}$.

Table 1: Convergence study of (6) used to solve problem (1) with body $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$. The first three columns list respectively the mesh size and errors together with convergence rates in L^2 and H^1-norms. The norms were computed by interpolating the error into the space of discontinuous polynomials of degree four. Due to memory requirement the last norm was computed from error represented in the same space as the solution. This change is indicated by a horizontal line. The fourth and fifth columns indicate if the solution is perpendicular to the kernel in the L^2 and l^2-inner products. Finally the last two columns list the size of the linear system and number of iterations needed to reduce the error to required tolerance 10^{-10}.

h	$\|e\|_0$	$\|e\|_1$	(u_h, z)	$Z^{\mathrm{T}}U_h$	N	n
4.42E-02	8.21E-04(1.94)	5.94E-02(1.01)	1.78E-12	5.61E-07	2178	32
2.21E-02	2.08E-04(1.98)	2.96E-02(1.00)	8.41E-13	3.84E-08	8450	38
1.10E-02	5.23E-05(1.99)	1.48E-02(1.00)	2.70E-12	2.49E-09	33282	42
5.52E-03	1.31E-05(2.00)	7.39E-03(1.00)	1.32E-11	1.58E-10	132098	48
2.76E-03	3.28E-06(2.00)	3.70E-03(1.00)	5.55E-11	9.98E-12	526338	53
6.91E-04	2.05E-07(2.00)	9.24E-04(1.00)	8.62E-10	2.49E-13	8396802	65

We remark that in both test cases the number of iterations required for convergence of the preconditioned conjugate gradient method grows logarithmically with the size of the linear system. For the two dimensional case the iteration count obeys $n \propto 4 \ln N$. In three dimensions the count grows as $n \propto 5 \ln N$. We observe similar growth when using ML as a preconditioner for CG solution of the vector Poisson equation or linear elasticity equation both considered with the homogeneous Dirichlet boundary conditions. Specifically, with vector Poisson equation and two dimensional domain the iterations grow as $n \propto 3 \ln N$, while in three dimensions $n \propto 2 \ln N$ is observed. For linear elasticity equation the

Table 2: Convergence study of Eq. (6) used to solve problem (1) with body $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^3$. Notation from Table 1 is reused.

h	$\|e\|_0$	$\|e\|_1$	(u_h, z)	$Z^{\mathrm{T}}U_h$	N	n
2.17E-01	1.45E-02(1.63)	2.47E-01(0.95)	8.82E-13	2.80E-05	2187	34
1.08E-01	4.22E-03(1.78)	1.22E-01(1.01)	9.26E-13	2.01E-06	14739	43
5.41E-02	1.12E-03(1.91)	6.00E-02(1.02)	3.81E-12	1.13E-07	107811	53
2.71E-02	2.87E-04(1.97)	2.98E-02(1.01)	1.07E-11	5.55E-09	823875	63
1.35E-02	7.22E-05(1.99)	1.48E-02(1.01)	2.92E-12	2.56E-10	6440067	75

observed growth was $n \propto 8 \ln N$ and $n \propto 5 \ln N$ for two and three dimensional body respectively.

4.1 Comparison with the conjugate gradient method for solving singular linear systems

The presented results demonstrate that the studied numerical method performs optimally when applied to the Neumann problem (1). In this section the method is compared against the iterative method for solving singular linear systems from Sec. 1. For comparison both methods are applied to the singular Poisson problem. We note that this test case is less complex than problem (1). Nevertheless it sufficiently illustrates the important differences between the methods.

In the singular Poisson problem the unknown scalar u is obtained from the equations

$$\begin{aligned} -\Delta u &= f \ \text{ in } \Omega, \\ \mathrm{grad} u \cdot n &= h \ \text{ on } \partial\Omega, \end{aligned} \tag{12}$$

where f, h are given scalar fields representing the volume and surface forces. The problem has a one dimensional kernel spanned by the function $z = \frac{1}{(1,1)}$. Furthermore, the problem is solvable only for data satisfying the compatibility condition $(f, 1) + (h, 1)_{\partial\Omega} = 0$. Here we used $(\cdot,\cdot)_{\partial\Omega}$ to indicate the L^2-inner product over the boundary $\partial\Omega$.

The bilinear form a and the linear form L needed to apply (8) to (12) are defined as

$$a(u, v) = (\mathrm{grad} u, \mathrm{grad} v) \qquad \text{and} \qquad L(v) = (f, v) + (h, v)_{\partial\Omega} \tag{13}$$

for $u, v \in V = H^1(\Omega)$. Note that the defined L is not necessarily compatible and its orthogonalization is to be handled by the numerical method. Finally, the space $V_h \subset V, \dim(V_h) = N$ is constructed from the continuous piecewise linear finite elements.

Considering the problem (2) on the space V_h yields a singular linear system $\mathbb{A}u = b$ with $\mathbb{A} \in \mathbb{R}^{N\times N}$ and $b \in \mathbb{R}^N$ having respectively the entries $\mathbb{A}_{ij} = (\mathrm{grad}\phi^i, \mathrm{grad}\phi^j)$ and $b_j = L(\phi^j)$. The matrix $\mathbb{A}$ has a one dimensional nullspace whose l^2-orthonormal basis is formed by a single vector $y = \frac{1}{\sqrt{N}}$. The linear system is therefore solvable if and only if $y^{\mathrm{T}}b = 0$ (see Lanczos [9]). We note that for any b such that $y^{\mathrm{T}}b \neq 0$, the transformation

Table 3: Convergence rates of methods (6) and (2) for the singular Poisson test case with a homogeneous triangulation of the domain. Notation from Table 1 is used. The results of method (6) are shown in the top half of the table. In the bottom half the results of method (2) are listed. Method (6) gives optimal convergence rates and enforces orthogonality in the L^2-inner product. The convergence rate of method (2) in the L^2-norm is suboptimal. The orthogonality with respect to kernel is enforced in the l^2-inner product.

h	$\|e\|_0$	$\|e\|_1$	(u_h, z)	$Z^{\mathrm{T}}U_h$	N	n
4.42E-02	4.08E-04(1.98)	4.35E-02(0.99)	5.63E-13	3.26E-05	1089	13
2.21E-02	1.02E-04(1.99)	2.18E-02(1.00)	1.65E-13	8.45E-06	4225	14
1.10E-02	2.57E-05(2.00)	1.09E-02(1.00)	2.35E-14	2.15E-06	16441	14
5.52E-03	6.42E-06(2.00)	5.45E-03(1.00)	2.40E-13	5.41E-07	66049	16
4.42E-02	2.48E-02(0.95)	2.01E-01(0.96)	1.33E-03	5.73E-20	1089	12
2.21E-02	1.27E-02(0.97)	1.02E-01(0.98)	7.09E-05	2.58E-20	4225	12
1.10E-02	6.40E-03(0.98)	5.13E-02(0.99)	1.20E-04	1.51E-19	16441	12
5.52E-03	3.22E-03(0.99)	2.57E-02(0.99)	9.95E-05	1.69E-19	66049	13

$b \leftarrow b - yy^{\mathrm{T}}b$ yields a vector orthogonal to y. With the orthogonalized right hand side the singular linear system can be solved by the Krylov subspace method provided that the solver is aware of the kernel. This ensures orthogonality of the constructed Krylov subspaces with respect to the nullspace. We shall refer to this method as method (2).

Methods (2) and (6) are compared on a singular Poisson problem (12) with domain $\Omega = [0,1]^2$ and forces $f = \pi^2 \sin \pi x + \pi^2 \sin \pi y$, $h = \frac{\pi}{4}$. Note that the chosen data is not compatible. The unique solution of the test case is then $u = \sin \pi x + \sin \pi y + 5\pi \frac{x(x-1)+y(y-1)}{4} - \frac{4}{\pi} + \frac{5\pi}{12}$. The resulting linear systems were solved by the preconditioned conjugate gradient method with the multigrid method on a linear system $\mathbb{A} + \mathbb{M}$ used as a preconditioner. The stopping criterion was that the absolute magnitude of the error be less than 10^{-10}. For method (6) the CG and ML implementations were taken respectively from cbc.block[14] and PETSc[13]. In method (2) both CG and ML from PyAMG[13] were used. Due to differences in ML implementations the methods are not expected to converge in the same number of iterations. However, the difference in iteration counts should be (order one) small.

Table 3 shows results of a convergence test performed on a uniformly discretized mesh. Both methods perform similarly in terms of iteration counts. Method (6) yields optimal convergence rates in both H^1 and L^2-norms. Moreover, orthogonality of the solution to the kernel is enforced in the L^2-inner product. Only the convergence rate in the energy norm is optimal with method (2). The orthonormality constraint is enforced in the l^2-inner product.

Table 4 shows results of a convergence test performed on successive uniform refinements of the mesh pictured in the right pane of Figure 3. Convergence rates by method (6) remain optimal. Method (2) fails to converge to the true solution if vector b is modified

Table 4: Convergence rates of methods (6) and (2) for the singular Poisson test case with a domain triangulated as in Figure 3. Notation from Table 1 is used. The results of method (6) are shown above the double horizontal line. The convergence rate in both norms is optimal. Below the line, the results of method (2) are listed. The results for the right hand side made compatible with the matrix problem are listed first. The Krylov method converges to the wrong solution. If the right hand side compatible with the variational problem is used the solution converges linearly in the energy norm. There is no convergence in the L^2-norm.

h	$\|e\|_0$	$\|e\|_1$	(u_h, z)	$Z^{\mathrm{T}}U_h$	N	n
2.21E-02	1.61E-03(1.78)	7.54E-02(0.91)	2.36E-13	1.88E-04	1621	14
1.10E-02	4.24E-04(1.93)	3.85E-02(0.97)	9.30E-14	9.61E-05	6377	16
5.52E-03	1.08E-04(1.98)	1.94E-02(0.99)	8.91E-14	4.81E-05	25297	18
2.21E-02	2.01E+00(-0.01)	5.12E+00(-0.00)	1.40E+00	2.04E-17	1621	12
1.10E-02	2.01E+00(-0.00)	5.12E+00(-0.00)	1.40E+00	5.67E-17	6377	16
5.52E-03	2.01E+00(-0.00)	5.12E+00(-0.00)	1.40E+00	1.23E-18	25297	15
2.21E-02	7.38E-03(0.25)	7.58E-02(0.91)	7.21E-03	9.09E-06	1621	12
1.10E-02	7.46E-03(-0.02)	3.92E-02(0.95)	7.45E-03	2.79E-06	6377	17
5.52E-03	7.53E-03(-0.01)	2.08E-02(0.92)	7.53E-03	7.72E-07	25297	16

by the l^2-projection used on the uniform mesh. Note that this is not signaled by an increase in the iteration count. Convergence properties of the method are restored by mapping b with (10) which represents orthogonalization of the functional L.

5 Conclusions

In this paper we have presented a method for solving the singular Neumann problem of linear elasticity based on the symmetric positive-definite formulation of the problem. The method takes advantage of the orthonormal basis of the nullspace which is herein obtained by a constructive algorithm closely related to the physical nature of rigid motions. The method has been verified by a series of numerical tests. It has been found that while having superior convergence properties the method's performance is similar in terms of efficiency to the Krylov method for symmetric singular linear systems.

REFERENCES

[1] Tobie G., Čadek O. and Sotin. C, Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus, *Icarus* (2008) **196**: 642–652.

[2] Sanderud M.B., Patient-specific modeling of normal pressure hydrocephalus, *Master thesis, University of Oslo* (2012).

[3] Bochev P. and Lehoucq R.B, On the finite element solution of the pure Neumann problem, *SIAM review* (2005) **47**: 50–66.

[4] Reddy B.D., *Introductory functional analysis: with applications to boundary value problems and finite elements*, Springer, (1998).

[5] Balay S., Abhyankar S., Adams M.F., Brown J., Brune P., Buschelman K., Eijkhout V., Gropp W.D., Kaushik D., Knepley M.G., McInnes L.C., Rupp K, Smith B.F. and Zhang H., PETSc users manual, *Technical report ANL-95/11 - Revision 3.5, Argonne National Laboratory*, (2014).

[6] Heroux M.A. and Willenbring J.M., Trilinos users guide, *Technical report SAND2003-2952, Sandia National Laboratories*, (2003).

[7] Gurtin, M.E., *An Introduction to Continuum Mechanics*, Elsevier Science, (1982).

[8] Mardal K.-A. and Winther R., *Preconditioning discretizations of systems of partial differential equations*, (2009).

[9] Lanczos, C., *Linear Differential Operators*, Dover Publications, (1997).

[10] Brezzi F., On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers, *Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, (1974), **8**: 129–151.

[11] Brenner S.C. and Scott R., *The Mathematical Theory of Finite Element Methods*, Springer, (2008).

[12] Marsden J.E. and Hughes T.J.R., *Mathematical Foundations of Elasticity*, Dover, (1994).

[13] Bell W. N., Olson L. N. and Schroder J. B., *PyAMG: Algebraic Multigrid Solvers in Python v2.0*, (2011), http://www.pyamg.org.

[14] Mardal K.-A. and Haga J.B., *Block preconditioning of systems of PDEs* In A. Logg, K.-A. Mardal, G. N. Wells et al. (ed) Automated Solution of Differential Equations by the Finite Element Method, Springer, (2012).

[15] Logg A., Mardal K.-A., Wells G.N. et al., *Automated Solution of Differential Equations by the Finite Element Method*, Springer, (2012).

[16] Babuška I., The finite element method with penalty, *Math. Comp.*, (1973), **27**:221–228.

Paper IV

On the singular Neumann problem in linear elasticity

M. Kuchta, K.-A. Mardal, and M. Mortensen

Numerical Linear Algebra with Applications, (2016). Submitted.

NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2010; **00**:1–22
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla

On the Singular Neumann Problem in Linear Elasticity

Miroslav Kuchta[1]* and Kent-Andre Mardal[12] and Mikael Mortensen[1]

[1]*Department of Mathematics, Division of Mechanics, University of Oslo*
[2]*Center for Biomedical Computing, Simula Research Laboratory*

SUMMARY

The Neumann problem of linear elasticity is singular with a kernel formed by the rigid motions of the body. There are several tricks that are commonly used to obtain a non-singular linear system. However, they often cause reduced accuracy or lead to poor convergence of the iterative solvers. In this paper, four different well-posed formulations of the problem are studied through discretization by the finite element method, and preconditioning strategies based on operator preconditioning are discussed. For each problem we derive preconditioners that are independent of the discretization parameter. Preconditoners that are robust with respect to the first Lamé constant are constructed for the pure displacement formulations, while a preconditioner that is robust in both Lamé constants is constructed for the mixed formulation. It is shown that, for convergence in the first Sobolev norm, it is crucial to respect the orthogonality constraint derived from the continuous problem. Based on this observation a modification to the conjugate gradient method is proposed that achieves optimal error convergence of the computed solution. Copyright © 2010 John Wiley & Sons, Ltd.

Received . . .

KEY WORDS: linear elasticity; rigid motions; singular problems; preconditioning; conjugate gradient

1. INTRODUCTION

The presented paper discusses numerical techniques for solving the singular problem of linear elasticity. Let $\Omega \subset \mathbb{R}^3$ be the body subjected to volume forces $f : \Omega \to \mathbb{R}^3$ and surface forces $h : \partial\Omega \to \mathbb{R}^3$. The body's displacement $u : \Omega \to \mathbb{R}^3$ is then found as a solution to

$$\begin{aligned} -\nabla \cdot \sigma(u) &= f && \text{in } \Omega, \\ \sigma(u) &= 2\mu\epsilon(u) + \lambda(\nabla \cdot u)I && \text{in } \Omega, \\ \sigma(u) \cdot n &= h && \text{on } \partial\Omega, \end{aligned} \tag{1}$$

with $\mu > 0$, $\lambda \geq 0$ the Lamé constants of the material, I the identity matrix, $\epsilon(u) = \frac{1}{2}\left(\nabla u + (\nabla u)^{\top}\right)$ the strain and n the outward-pointing surface normal, see [1]. The system is used extensively in structural analysis [2], and is relevant in numerous applications for, e.g., marine engineering [3], biomechanics of brain [4], spine [5] or the mechanics of planetary bodies [6].

Due to the absence of a Dirichlet boundary condition that can anchor the body (coordinate system) in space, the problem can be solved if and only if the net force and the net torque on Ω are zero, i.e.,

*Correspondence to: E-mail: mirok@math.uio.no

Contract/grant sponsor: The work of Kent-Andre Mardal and Mikael Mortensen has been supported by a Center of Excellence grant no. 179578 from the Research Council of Norway to the Center for Biomedical Computing at Simula Research Laboratory. The work of Kent-Andre Mardal has also been supported by the Research Council of Norway through grant no. 209951.

the forces f, h satisfy the compatibility conditions

$$\begin{aligned} \int_\Omega f \,\mathrm{d}x + \int_{\partial\Omega} h \,\mathrm{d}s &= 0, \\ \int_\Omega f \times x \,\mathrm{d}x + \int_{\partial\Omega} h \times x \,\mathrm{d}s &= 0. \end{aligned} \tag{2}$$

With such compatible data the now solvable (1) is singular as any rigid motion can be added to the solution. We note that the space of rigid motions $z : \Omega \to \mathbb{R}^3$ such that $\epsilon(z) = 0$, consists of translations and rigid rotations and for a body in $3d$ the space is six-dimensional.

The ambiguity of the solution of (1) can be removed by adding constraints by means of Lagrange multipliers which enforce that the solution is free of rigid motions. When discretized, this approach yields an invertible saddle point system. Alternatively, discretizing (1) directly leads to a symmetric, positive semi-definite matrix with a six dimensional kernel. If (2) holds, such a system can be solved by the conjugate gradient (CG) method [7]. Finally, a common approach (here termed *pinpointing*) in engineering literature, e.g. [3], is to remove the nullspace by prescribing the displacement in selected points of $\partial\Omega$.

In this work we focus on analysis of the Lagrange multiplier method and the conjugate gradient method for the singular problem (1). Well-posedness of both the methods is discussed and efficient preconditioners are established based on operator preconditioning [8]. Further, connections between the two methods and the question of whether they yield identically converging numerical solutions are elucidated.

The manuscript is structured as follows. In §2 the necessary notation is introduced and shortcomings of pinpointing and CG are illustrated by numerical examples. Section 3 discusses Lagrange multiplier formulation and two preconditioners for the method. Section 4 deals with the preconditioned CG method and two preconditioners are proposed. Further, it is revealed that if the variational origin of the discretized problem is ignored, the method, in general, will not yield convergent solutions. A variational setting is introduced to modify the CG to yield a convergent method. Section §5 discusses well-posedness and preconditioning of an alternative formulation of (1). The proposed formulation leads to a symmetric, positive definite linear system. In §3-5 we assume that λ and μ are of comparable magnitude in order to put the focus on proper handling of the rigid motions. In §6 we consider the case where $\lambda \gg \mu$. The focus here is on a well-known and simple technique to remove the problems of locking, namely the mixed formulation of linear elasticity. This formulation yields robust approximation and preconditioning in λ when care is taken of proper handling of the rigid motions. Finally, conclusions are drawn in §7.

2. PRELIMINARIES

Let V be the Sobolev space of vector (or scalar or tensor) valued functions, which, together with their weak derivatives of order one, are in space $L^2(\Omega)$. We denote $(\cdot,\cdot)$ the L^2 inner product of functions in V while $\|\cdot\|$ is the corresponding norm. The standard inner product of V is $(u,v)_1 = (u,v) + (\nabla u, \nabla v)$, $u, v \in V$ and $\|\cdot\|_1$ shall be the induced norm. For any Hilbert space V its dual space is denoted as V' and we use capital or calligraphy letters to denote operators, e.g. $A : V \to V'$ or $\mathcal{A} : (V \times V) \to (V \times V)'$. Finally, $\langle\cdot,\cdot\rangle$ is the duality pairing between V' and V.

The space $\mathbb{R}^n$ is considered with the l^2 inner product $\mathbf{x}^\top\mathbf{y} = x_i y_i$ (invoking the summation convention), $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and the norm $|\mathbf{x}| = \sqrt{\mathbf{x}^\top\mathbf{x}}$. For clarity of notation bold fonts are used to denote vectors and operators(matrices) in $\mathbb{R}^n$ that represent functions and operators from finite dimensional finite element approximation space $V_h \subset V$ with respect to its nodal basis $\{\phi_i\}_{i=1}^n$. The representations are obtained by mappings $\pi_h : V_h \to \mathbb{R}^n$ (the nodal interpolant) and $\mu_h : V_h' \to \mathbb{R}^n$ such that for $v \in V_h$, $f \in V_h'$

$$v = (\pi_h v)_i \phi_i \quad \text{and} \quad (\mu_h f)_i = \langle f, \phi_i \rangle. \tag{3}$$

We refer to [8, ch 6.] for a detailed discussion of the properties of the mappings, e.g. invertibility, and note here that $M : V_h \to V_h'$ is represented by a matrix $\mathbf{M} = \mu_h M {\pi_h}^{-1}$. In particular, the

DOI: 10.1002/nla

mass matrix $\mathbf{M}$, $M_{ij} = (\phi_j, \phi_i)$ represents the Riesz map with respect to the L^2-inner product, $\langle Mu, v\rangle = (u, v)$, $u \in V_h$. On the other hand the duality pairing between V_h' and V_h is represented by the l^2 inner product $\langle f, v\rangle = \mathbf{f}^\top\mathbf{v}$, $\mathbf{f} = \mu_h f$. We remark that for V_h set up on a sequence of non-uniformly refined triangulations of Ω, the l^2 inner product $\mathbf{u}^\top\mathbf{v}$ may not provide a converging approximation of (u, v) and the distinction between the two becomes crucial for the construction of converging methods.

Finally, Korn's inequalities on $V = \left[H^1(\Omega)\right]^3$ and $Z^\perp = \{v \in V; (v, z) = 0, z \in Z\}$, $Z = \{v \in V; \epsilon(v) = 0\}$ are invoked, see [9, thm 2.1] and [9, thm 2.3]. There exist a positive constant $C = C(\Omega)$ such that

$$C\|u\|_1^2 \leq \|\epsilon(u)\|^2 + \|u\|^2 \quad u \in V. \tag{4}$$

and

$$C\|u\|_1^2 \leq \|\epsilon(u)\| \quad u \in Z^\perp. \tag{5}$$

To motivate our investigations, we present three numerical examples which discuss performance of CG and pinpointing for solving (1). That the pinpointing can be a suitable method for treating a singular problem is shown in the first example which considers the Poisson problem with Neumann boundary conditions. However, the method in not a cure-all as the second example shows that it does not work well for (1). In the third example, the singular elasticity problem is therefore solved with preconditioned CG. The employed preconditioner ignores the rigid motions leading to lack of convergence and unbounded iterations.

Bochev and Lehoucq [10] report an increase in iteration count due to pinpointing for a non-preconditioned CG in the context of singular Poisson problem. However, Krylov methods are in practice rarely applied without a preconditioner. For this reason, Example 2.1 solves the singular Poisson problem in two and three dimensions by means of pinpointing and a preconditioned CG, where algebraic multigrid (AMG) from the Hypre library [11] is used as a preconditioner. We will see that the preconditioned method yields convergent numerical solutions without increasing the iteration count.

Example 2.1
We consider $\Omega = [0, 1]^d$, $d = 2, 3$ and the singular Poisson equation

$$\begin{aligned} -\Delta u = f \qquad & \text{in } \Omega, \\ \nabla u \cdot n = 0 \quad & \text{on } \partial\Omega, \end{aligned}$$

with unique exact solution obtained by subtracting its mean value $|\Omega|^{-1}\int_\Omega u\,\mathrm{d}x$ from a manufactured u. The value of the exact solution is prescribed as a constraint for the degree of freedom at the (bottom) lower left corner of the domain, which is triangulated such that the computational mesh is refined towards the origin.

To discretize the system continuous linear Lagrange elements[†] from the FEniCS library [12] were used. The resulting linear system was solved by the preconditioned CG method implemented in the PETSc library [13], taking HypreAMG with default settings as a preconditioner. The iterations were started from a random initial guess and a relative preconditioned residual magnitude of 10^{-11} was required for convergence.

The number of iterations together with error and convergence rate based on the H^1 norm are reported in Table I. Pinpointing yields numerical solutions u_h that converge with optimal rate. The number of iterations is bounded.

Following the performance of pinpointing in the singular Poisson problem, the same approach is now applied to (1) in Example 2.2. Here, we will observe that fixing the solution datum in vertices of the mesh leads to slightly increased iteration counts. More importantly, we will see that the method in general does not yield converging solutions.

[†] Unless stated otherwise continuous linear Lagrange elements (P_1) are used to discretize all the presented numerical examples.

DOI: 10.1002/nla

Table I. Convergence of the pinpointing approach for the singular Poisson problem.

$d=2$			$d=3$		
size	$\|u-u_h\|_1$	#	size	$\|u-u_h\|_1$	#
40849	2.49E-01 (1.00)	11	12347	2.72E+00 (1.22)	10
162593	1.25E-01 (1.00)	11	92685	1.36E+00 (1.01)	11
648769	6.23E-02 (1.00)	11	718649	6.78E-01 (1.00)	12
2591873	3.11E-02 (1.00)	12	5660913	3.39E-01 (1.00)	12

Example 2.2

We consider the singular elasticity problem (1) with $\mu=384$, $\lambda=577$ and Ω obtained by rigid deformation of the box $\left[-\frac{1}{4},\frac{1}{4}\right]\times\left[-\frac{1}{2},\frac{1}{2}\right]\times\left[-\frac{1}{8},\frac{1}{8}\right]$. The box was first rotated around x, y and z axes by angles $\frac{\pi}{2}$, $\frac{\pi}{4}$ and $\frac{\pi}{5}$ respectively. Afterwards it was translated by the vector $(0.1, 0.2, 0.3)$. The unique exact solution is obtained by orthogonalizing $u=\frac{1}{4}(\sin\frac{\pi}{4}x, z^3, -y)$ with respect to the rigid motions of Ω where the orthogonality is enforced in the L^2 inner product. The solution is pictured in Figure 1. We note that in this example a uniform triangulation is used.

To obtain from (1) an invertible linear system, the exact displacement was prescribed in four different ways, cf. Table II below. (3○) constrains six degrees of freedom in three corners of the body such that in i-th corner there are i components prescribed. This choice is motivated by the dimensionality of the space of rigid motions, cf. [3]. The fact that fixing three points in space is sufficient to prevent the body from rigid motions motivates (1▷) where all three components of displacement are prescribed on vertices of a single triangular element on $\partial\Omega$. However, with mesh size decreasing this constraint effectively becomes a constraint for a single (mid)point. Thus in (3▷) the displacement in three arbitrary triangles is fixed. Finally in (3●) the displacement is prescribed in three corners of the body.

The iterative solver used the same tolerances and parameters as in Example 2.1. In particular, default settings of the multigrid preconditioner were utilized and the iterations were started from random initial vectors.

The number of iterations together with error and convergence rates based on the H^1 norm are reported in Table II. Note that all the considered pinpointing strategies lead to moderately increased iteration counts. The increase is most notable for (1▷), which effectively constrains a single point as the mesh is refined. On the other hand, strategies (3▷) and (3●), that always constrain all three components of the displacement in at least three points, yield the slowest growth rates. However, neither strategy yields convergent numerical solutions. In fact, the numerical error can often be seen to increase with resolution.

Table II. Convergence of the pinpointing approach for the singular Poisson problem.

size	3○		1▷		3▷		3●	
	$\|u-u_h\|_1$	#	$\|u-u_h\|_1$	#	$\|u-u_h\|_1$	#	$\|u-u_h\|_1$	#
2187	6.69E-02 (-0.02)	30	1.01E-01 (-0.70)	32	2.82E-02 (0.88)	24	2.89E-02 (0.99)	25
14739	1.27E-01 (-0.92)	35	9.61E-01 (-3.25)	40	1.08E-02 (1.38)	28	1.35E-02 (1.10)	29
107811	2.57E-01 (-1.02)	36	7.89E+00 (-3.04)	48	1.72E-02 (-0.66)	31	1.08E-02 (0.31)	32
823875	5.17E-01 (-1.01)	41	6.36E+01 (-3.01)	54	3.96E-02 (-1.21)	33	1.82E-02 (-0.75)	35

In the final example a preconditioned CG method will be applied to solve the singular elasticity problem with data such that the compatibility conditions (2) are met. Based on whether or not the components of the kernel are removed from the converged vector, we will see that the method yields convergent/divergent numerical solutions. We will also see that the iteration counts are not bounded.

Example 2.3

We consider again the problem from Example 2.2. As the data satisfy (2), the discrete linear system is solvable and amiable to solution by the preconditioned CG method. The mass matrix is added to the singular system matrix in order to obtain a positive definite matrix in the construction of the preconditioner based on AMG. We consider two cases where the converged vector is either

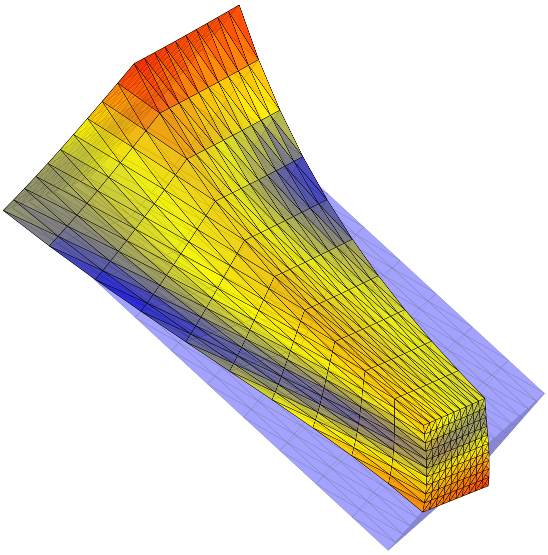

Figure 1. Computational domain (blue) deformed by exaggerated(4x) analytical displacement used in the numerical examples. The deformed body is colored by the magnitude of the displacement.

postprocessed by removing from it the components of the nullspace or no postprocessing is applied. We note that in this example the iterations are started from a zero initial vector and the relative tolerance of 10^{-10} is used as a convergence criterion.

The number of iterations together with error and convergence rates based on the H^1 norm are reported in Table III. We observe that the method yields convergent solutions only if postprocessing is applied. This is expected as the current preconditioner introduces components of the nullspace into the solution even if the iterations are started from a right hand side and initial vector (here zero) that are orthogonal to the kernel. We note that in exact arithmetic, the CG method without preconditioner will maintain orthogonality. We further observe that the choice of preconditioner leads to unbounded iteration counts.

Table III. Convergence of the preconditioned CG method with positive definite preconditioner for the singular elasticity problem.

size	kernel removed		kernel not removed	
	$\|u - u_h\|_1$	#	$\|u - u_h\|_1$	#
14739	1.14E-02 (1.09)	34	4.01E-02 (-0.12)	22
107811	5.49E-03 (1.06)	22	5.02E-02 (-0.32)	34
823875	2.71E-03 (1.02)	78	5.51E-02 (-0.13)	53
6440067	1.35E-03 (1.00)	150	5.18E-02 (0.09)	150

Examples 2.1–2.3 have illustrated some of the issues that might be encountered when solving the singular problem (1) with the finite element method. In particular, the following questions may be posed: (i) What is the cause of the poor convergence properties of pinpointing? (ii) What should be the optimal preconditioner for CG? (iii) What should be the optimal preconditioner for the Lagrange multiplier formulation?

With questions (ii) and (iii) answered in detail in the remainder of the text let us briefly comment on the first question. As will become apparent, the singular problem with a known kernel, such as (1), possesses all the information necessary to formulate a well-posed problem and a convergent numerical method. In this sense, coming up with a datum to be prescribed in the pinpointed nodes is theoretically redundant, but usually required for implementation. Further, as pointed out in [10] there are stability issues with prescribing point values of H^1 functions for $d \geq 2$. However, we note that we have not explored settings of HypreAMG that could potentially improve convergence properties of the method in Example 2.2.

3. LAGRANGE MULTIPLIER FORMULATION

Let $Z \subset V = \left[H^1(\Omega)\right]^3$ denote the space of rigid motions of Ω, $m = \dim Z$. For compatible data a unique solution u of (1) is required to be linearly independent of functions in Z. To this end a Lagrange multiplier $p \in Q$, $Q = \mathbb{R}^m$ is introduced which enforces orthogonality of u with respect to Z. The constrained variational formulation of (1) seeks $u \in V, p \in Q$ such that[‡]

$$\begin{aligned} 2\mu(\epsilon(u),\epsilon(v)) + \lambda(\nabla\cdot u, \nabla\cdot v) - p_k(v, z_k) &= (f,v) + (h,v) \qquad && v \in V,\\ -q_k(u,q_k) &= 0 && q \in Q, \end{aligned} \tag{6}$$

for some basis vectors $z_k \in V$, $Z = \operatorname{span}\{z_k\}_{k=1}^m$. Equation (6) defines a saddle point problem for $(u,p) \in W$, $W = V \times Q$ satisfying

$$\mathcal{A}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} A & B \\ B' & \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} b \\ \end{pmatrix}, \tag{7}$$

where $b \in V'$ such that $\langle b, v\rangle = (f,v) + (h,v)$ and operators $A : V \rightarrow V'$, $B : Q \rightarrow V'$ are defined in terms of bilinear forms

$$a(u,v) = 2\mu(\epsilon(u),\epsilon(v)) + \lambda(\nabla\cdot u,\nabla\cdot v) \quad \text{and} \quad b(u,q) = q_k(u,q_k) \tag{8}$$

as $\langle Au, v\rangle = a(u,v)$ and $\langle Bq, u\rangle = -b(u,q)$. We note that in (7) operator B' is the adjoint of B.

Existence and uniqueness of the solution to (7) follows from the Brezzi theory [14], see also [15, ch 3.4]. The proof shall utilize the inequalities given in Lemma 3.1.

Lemma 3.1
Let $u \in V$ and $\omega(u)$ be the skew symmetric part of the displacement gradient ∇u and $u_x, v_y \in V$ the rigid rotations around vectors $x, y \in \mathbb{R}^3$. Then

$$\|\epsilon(u)\| \leq \|\nabla u\| \quad \text{and} \quad \|\omega(u)\| \leq \|\nabla u\|, \tag{9a}$$

$$\|\nabla\cdot u\| \leq \sqrt{3}\|\nabla u\|, \tag{9b}$$

$$(\omega(u),\omega(v)) = \tfrac{1}{2}(\nabla\times u, \nabla\times v), \tag{9c}$$

$$(\nabla u_x, \nabla v_y) = 2|\Omega| x^\top y. \tag{9d}$$

Proof
Inequalities (9a) follow from the orthogonal decomposition $\nabla u = \epsilon(u) + \omega(u)$. Inequalities (9b) and (9c) follow from the definitions of the terms and the Young's inequality. Finally (9d) is a special case of (9a) and (9c). □

Theorem 3.1
Let f, h such that $b \in V'$. Then there exists a unique solution $u \in V$, $p \in Q$ of (7).

Proof
We proceed by establishing the Brezzi constants. First, the bilinear form a is shown to be bounded with respect to the $\|\cdot\|_1$. Indeed, by Cauchy-Schwarz inequality and inequalities (9a), (9b)

$$\begin{aligned} a(u,v) = 2\mu(\epsilon(u),\epsilon(v)) + \lambda(\nabla\cdot u,\nabla\cdot v) &\leq 2\mu\|\epsilon(u)\|\|\epsilon(u)\| + \lambda\|\nabla\cdot u\|\|\nabla\cdot v\| \\ &\leq (2\mu + 3\lambda)\|\nabla v\|\|\nabla u\| \leq \alpha^*\|u\|_1\|v\|_1. \end{aligned}$$

[‡] Note that (h,v) is to be understood as the L^2 inner product over $\partial\Omega$

DOI: 10.1002/nla

with $\alpha^* = 2\lambda + 3\mu$. Ellipticity of a on $Z^\perp = \{v \in V; (v, z) = 0, z \in Z\} = \{v \in V; b(v, p), p \in Q\}$ follows from Korn's inequality (5). Since $\lambda \geq 0$ by assumption

$$a(u, u) = 2\mu\|\epsilon(u)\|^2 + \lambda\|\nabla \cdot u\|^2 \geq 2\mu\|\epsilon(u)\|^2 \geq \alpha_*\|u\|_1^2,$$

with $\alpha_* = 2\mu C$ and $C = C(\Omega)$ the constant from (5). To verify boundedness of b let $G \in \mathbb{R}^{m\times m}$ be the Gram matrix of the basis of Z with entries $G_{ij} = (z_i, z_j)$. By assumption of complete basis of Z, G is a positive definite matrix. Further $G = G^\top$ and we let $0 < \lambda_* \leq \lambda^*$ be, respectively, the smallest and largest eigenvalues of G. Then

$$b(v, p) = (v, p_k z_k) \leq \|v\|\|p_k z_k\| \leq \sqrt{\lambda^*}\|v\|_1|p|$$

and b is bounded with constant $\beta^* = \sqrt{\lambda^*}$. Finally, we show that the inf-sup property of b is satisfied. By (9d)

$$\sup_{v\in V} \frac{b(v,p)}{\|v\|_1} \geq \frac{(p_k z_k, p_i z_i)}{\|p_i z_i\|_1} = \frac{p^\top G p}{\sqrt{p^\top G p + 2|\Omega| p^\top D p}},$$

with $D \in \mathbb{R}^{m\times m}$ a block diagonal matrix $D = \text{diag}(I, R)$ and $R \in \mathbb{R}^{3\times 3}$ such that $R_{ij} = {e_i}^\top e_j$ for axes of rigid rotations e_i. Denoting C the largest eigenvalue of the symmetric positive definite generalized eigenvalue problem for matrices D and G we have

$$\sup_{v\in V} \frac{b(v,p)}{\|v\|_1} \geq \frac{\sqrt{p^\top G p}}{\sqrt{1 + 2|\Omega| C}} \geq \sqrt{\frac{\lambda_*}{1 + 2|\Omega| C}}|p| = \beta_*|p|.$$

□

We remark that Theorem 3.1 implies that the operator $\mathcal{A} : W \to W'$ from (7) is an isomorphism. In particular, conditions (2) need not to hold for there to exist a unique solution of (6).

In order to find the solution of the well-posed (7) numerically, conditions from Theorem 3.1 must hold with discrete spaces $V_h \subset V$, $Q_h \subset Q$, see [16] or [15, ch 3.4]. Note that $Q_h = Q$ in the case studied here. Typically, satisfying the discrete inf-sup condition presents an issue and requires choice of compatible finite element discretization of the involved spaces, e.g. Taylor-Hood or MINI elements [17] for the Stokes equations. For the conforming discretization $V_h \subset V$ the following result shows that the discrete inf-sup condition holds.

Theorem 3.2
Let $V_h \subset V$ and b the bilinear form defined in (8). Then there is a constant β_* independent of h such that $\sup_{v\in V_h} \frac{b(v,p)}{\|v\|_1} \geq \beta_*|p|$.

Proof
Since the continuous inf-sup condition holds the statement follows from Fortin's criterion [16] and we shall construct Fortin's projector $\Pi : V \to V_h$ such that $\|\Pi u\|_1 \leq C\|u\|_1$ with C independent of h and $b(u - \Pi u, q) = 0$ for any $q \in Q_h$. For given $u \in V$ we consider $u_h = \Pi u \in V_h$ the satisfies

$$2\mu(\epsilon(u_h), \epsilon(v)) + \lambda(\nabla \cdot u_h, \nabla \cdot v) + (u_h, v) = 2\mu(\epsilon(u), \epsilon(v)) + \lambda(\nabla \cdot u, \nabla \cdot v) + (u, v) \quad v \in V_h.$$

Then, testing the equation with $z_h \in V_h$, an interpolant of $z \in Z$ in V_h, gives $(u - u_h, z_h) = 0$ and in turn $b(u - \Pi u, q) = 0$. Moreover by Korn's inequality (4)

$$\begin{aligned} 2\mu(\epsilon(u_h), \epsilon(u_h)) + \lambda(\nabla \cdot u_h, \nabla \cdot u_h) + (u_h, u_h) &\geq 2\mu(\epsilon(u_h), \epsilon(u_h)) + (u_h, u_h) \\ &\geq C\min(2\mu, 1)\|u_h\|_1^2 = c\|u_h\|_1^2, \end{aligned}$$

while the estimate

$$2\mu(\epsilon(u), \epsilon(u_h)) + \lambda(\nabla \cdot u, \nabla \cdot u_h) + (u, u_h) \leq \max(2\mu + 3\lambda, 1)\|u\|_1\|u_h\|_1 = C\|u\|_1\|u_h\|_1$$

follows from (9a), (9b). Thus $\|u_h\|_1 \leq \frac{C}{c}\|u\|_1$ and the projector is bounded. □

DOI: 10.1002/nla

Following Theorems 3.1, 3.2 and operator preconditioning [8, 18] the Riesz map $\mathcal{B}_1 : W' \to W$ with respect to the W inner product $(u, v)_1 + p^\top q$ with $(u, p), (v, q) \in W$

$$\mathcal{B}_1 = \begin{pmatrix} H & \\ & I \end{pmatrix}^{-1}, \quad H : V \to V', \langle Hu, v\rangle = (u, v)_1 \quad \text{and} \quad I : Q \to Q, I_{ij} = \delta_{ij}, \tag{10}$$

defines a preconditioner for discretized (7) whose condition number is independent of h. This follows from Brezzi constants in Theorems 3.1, 3.2 being free of the discretization parameter. However the constants depend on the skewness of the basis of Z and material parameters. To remove the former dependency, an orthonormal basis of the space of rigid motions shall be constructed.

3.1. Construction for orthonormal basis of rigid motions

Consider a unit cube $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^3$ centered at the origin. Denoting e_i, $i = 1, 2, 3$ the canonical unit vectors the set

$$Z_{\square} = \{e_1, e_2, e_3, x \wedge e_1, x \wedge e_2, x \wedge e_3\}$$

constitutes an orthonormal basis of the rigid motions of Ω with respect to the L^2 inner product. Clearly, the basis for an arbitrary body can be obtained from $Z_{\square}$ by a Gram-Schmidt process. However, we shall advocate here a construction derived from physical considerations. The construction was originally presented by the authors in [19].

Lemma 3.2
Let $c = |\Omega|^{-1}(x, 1)$ be the center of mass of Ω, I_Ω the tensor of inertia [20, ch 4.] of Ω with respect to c

$$I_\Omega = \int_\Omega I(x - c)^\top(x - c) + (x - c) \otimes (x - c)\,\mathrm{d}x$$

and (λ_i, v_i), $i = 1, 2, 3$ the eigenpairs of the tensor. Then the set

$$Z_\Omega = \{|\Omega|^{-\frac{1}{2}}v_1, |\Omega|^{-\frac{1}{2}}v_2, |\Omega|^{-\frac{1}{2}}v_3, \lambda_1^{-\frac{1}{2}}(x - c) \wedge v_1, \lambda_2^{-\frac{1}{2}}(x - c) \wedge v_2, \lambda_3^{-\frac{1}{2}}(x - c) \wedge v_3\} \tag{11}$$

is the L^2 orthonormal basis of rigid motions of Ω.

Proof
Note that by construction I_Ω is a symmetric positive definite tensor. Thus $\lambda_i > 0$ and there exists a complete set of eigenvectors ${v_i}^\top v_j = \delta_{ij}$. We proceed to show that the Gram matrix of the proposed basis is an identity. First $(v_i, v_j) = |\Omega|\delta_{ij}$ by orthonormality of the eigenvectors. Further, for $((x - c) \wedge v_i, v_j) = (v_i \wedge v_j, (x - c))$ and in the nontrivial case $i \neq j$ the product is zero since c is the center of mass. Finally $((x - c) \wedge v_i, (x - c) \wedge v_j) = {v_i}^\top I_\Omega v_j = \lambda_i\delta_{ij}$. □

We remark that the rigid motions of the body are in the constructed basis given in terms of translations along and rotations around the principal axes of the tensor that describes its rotational kinetic energy.

Note also, that the construction can be generalized to yield an orthonormal basis with respect to different inner products. In particular, let $Z_h = \operatorname{span}\{z_k\}_{k=1}^m \subset V_h$ be functions approximating Z. For $u_h \in V_h$ let $\mathbf{u} = \pi_h u$ be a coefficient vector in the nodal basis of V_h. The l^2 orthonormal basis of Z_h can be created using Lemma 3.2 by replacing (u, v) with $\mathbf{u}^\top\mathbf{v}$. The differences between the bases are shown in Figure 2 where the defining principal axes of the L^2 and l^2 orthonormal basis of rigid motions are drawn. If Ω is uniformly triangulated the bases are practically identical. However, the l^2 basis changes in the presence of a non-uniform mesh refinement.

The assumption of orthonormal basis in (6) modifies the Brezzi constants in Theorem 3.1 (and Theorem 3.2). More specifically the Gram matrix of Z becomes identity and $\beta^* = 1$ while β_* newly depends only on the domain size. In turn, if (6) is considered with the orthonormal basis of the space of rigid motions, then $\mathcal{B}_1$ (see (10)) is a suitable preconditioner for (7) with a condition number dependent only on the geometry and material parameters.

To address the dependence on material parameters, we shall at first assume that μ and λ are comparable in magnitude. The case $\lambda \gg \mu$ is postponed until §6.

DOI: 10.1002/nla

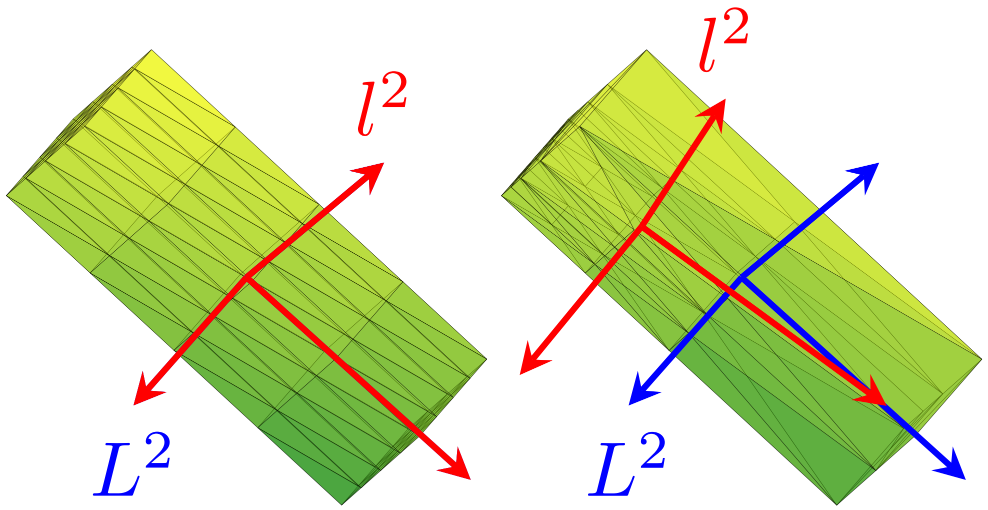

Figure 2. Computational domains considered in the numerical examples for linear elasticity are obtained by uniformly refining the parent mesh. (Left) Parent is close to uniformly triangulated. (Right) The parent mesh is refined near a single edge of the domain. The blue and red arrows indicate the principal axes of the tensor I_Ω, cf. Lemma 3.2, defined using the L^2 and l^2 inner products. Axes are drawn from the center of mass computed using the respected inner products. Only the L^2 basis is stable upon change of triangulation from uniform (left) to nonuniform (right).

3.2. *Robust preconditioning of the singular problem*

Parameter robust preconditioners for the Lagrange multiplier formulation of the singular elasticity problem (6) can be analyzed by the operator preconditioning framework of [8]. The preconditioners are constructed by considering (7) in parameter dependent spaces, e.g. [21], which are equivalent with V as a set, but the topology of the spaces are given by different, parameter dependent, norms. Two such norms leading to two different preconditioners are constructed next.

Let $\{z_k\}_{k=1}^m$ be the L^2 orthonormal basis of the space of rigid motions of Ω. Bilinear forms $(\cdot,\cdot)_E$, $(\cdot,\cdot)_M$ over V are defined in terms of A from (7) and operators $Y : V \to V'$, $M : V \to V'$ as

$$\begin{aligned} \langle Yu, v\rangle &= (u, z_k)(v, z_k), & (u,v)_E &= \langle Au, v\rangle + \langle Yu, v\rangle,\\ \langle Mu, v\rangle &= (u, v), & (u,v)_M &= \langle Au, v\rangle + \langle Mu, v\rangle. \end{aligned} \tag{12}$$

The forms (12) define functionals $\|\cdot\|_E$ and $\|\cdot\|_M$ over V such that

$$\|u\|_E = \sqrt{(u,u)_E} \quad \text{and} \quad \|u\|_M = \sqrt{(u,u)_M}. \tag{13}$$

Lemma 3.3
Let $\|\cdot\|_E$ and $\|\cdot\|_M$ be the functionals (13). Then $\|\cdot\|_E$ and $\|\cdot\|_M$ define norms on V equivalent with the H^1 norm.

Proof
From an orthogonal decomposition of $u \in V$, $u = u - (u, z_k)z_k + (u, z_k)z_k$ into $u_{Z^\perp} = u - (u, z_k)z_k \in Z^\perp$ and $u_Z = (u, z_k)z_k \in Z$ it follows that $\|u\|_M^2 = \|u\|_E^2 + \|u_{Z^\perp}\|^2$. Together with Lemma 3.1 we thus establish

$$\|u\|_E^2 \le \|u\|_M^2 \le (2\mu + 3\lambda + 1)\|u\|_1^2.$$

To complete the equivalence, let $C = C(\Omega)$ be the constant from Korn's inequality (4). Then

$$\|u\|_M^2 \ge 2\mu\|\epsilon(u)\|^2 + \|u\|^2 \ge c\|u\|_1^2,$$

DOI: 10.1002/nla

with $c = C$ for $2\mu > 1$ and $c = 2\mu C$ otherwise. Finally, for equivalence of the E-norm, the Korn's inequality for $u \in Z^{\perp}$, see (5) also Theorem 3.1, yields

$$\|u\|_E^2 = 2\mu\|\epsilon(u)\|^2 + \lambda\|\nabla \cdot u\|^2 \geq 2\mu C\|u\|_1^2$$

with $C = C(\Omega)$, while from (9d) in Lemma 3.1

$$\|u\|_E^2 = \|u\|^2 = \frac{1}{1+2|\Omega|}\|u\|_1^2$$

for $u \in Z$. Thus E and H^1 norms are equivalent on $Z^{\perp}$ and Z respectively. The proof is completed by observing that u_Z and $u_{Z^{\perp}}$ are orthogonal in the E inner product

$$\begin{aligned}\|u\|_E^2 = 2\mu\|\epsilon(u_{Z^{\perp}})\|^2 + \lambda\|\nabla \cdot u_{Z^{\perp}}\|^2 + \|u_Z\|^2 &\geq 2\mu C\|u_{Z^{\perp}}\|_1^2 + \frac{1}{1+2|\Omega|}\|u_Z\|_1^2\\ &\geq c(\|u_{Z^{\perp}}\|_1^2 + \|u_Z\|_1^2),\end{aligned}$$

$c = \min(2\mu C, (1+2|\Omega|)^{-1})$, while for the H^1 inner product $\|u\|_1^2 \leq 2(\|u_{Z^{\perp}}\|_1^2 + \|u_Z\|_1^2)$ holds. Thus $\|u\|_E^2 \geq \frac{c}{2}\|u\|_1^2$, for $u \in V$. □

Using equivalent norms of V from Lemma 3.3 we readily establish equivalent norms for the product space $W = V \times Q$

$$\|w\|_E = \|(u,p)\|_E = \sqrt{\|u\|_E^2 + p^{\top}q} \quad \text{and} \quad \|w\|_M = \|(u,p)\|_M = \sqrt{\|u\|_M^2 + p^{\top}q} \tag{14}$$

and consider as preconditioners for (7) the operators $\mathcal{B}_E : W' \rightarrow W$ and $\mathcal{B}_M : W' \rightarrow W$

$$\mathcal{B}_E = \begin{pmatrix} A+Y & \\ & I \end{pmatrix}^{-1} \quad \text{and} \quad \mathcal{B}_M = \begin{pmatrix} A+M & \\ & I \end{pmatrix}^{-1}. \tag{15}$$

Note that the mappings (15) are the Riesz maps with respect to the inner products which induce norms (14). We proceed with analysis of the properties of $\mathcal{B}_E$.

Theorem 3.3
Let $\{z_k\}_{k=1}^{m}$ be the L^2 orthonormal basis of the space of rigid motions of Ω, $\mathcal{A} : W \rightarrow W'$ from (7) and W_E the space W considered with $\|\cdot\|_E$ norm (14). Then $\mathcal{A} : W_E \rightarrow W_E'$ is an isomorphism. Moreover the Riesz map $\mathcal{B}_E : W_E' \rightarrow W_E$ in (15) defines the canonical preconditioner for (7).

Proof
We shall show that the first assertion holds by establishing the Brezzi constants. Recall the definition of the bilinear form a given in (8). Then, by the Cauchy-Schwarz inequality and (9a) in Lemma 3.1, the inequality $a(u,v) \leq \sqrt{a(u,u)}\sqrt{a(v,v)}$ holds. In turn

$$a(u,v) \leq \sqrt{a(u,u)}\sqrt{a(v,v)} \leq \sqrt{a(u,u) + (u,z_k)(u,z_k)}\sqrt{a(v,v) + (v,z_k)(v,z_k)} = \|u\|_E\|v\|_E$$

and a is bounded with respect to E norm with a constant $\alpha^* = 1$. Further $(u, z_k) = 0$ for $u \in Z^{\perp}$. Hence $a(u,u) = a(u,u) + (u,z_k)(u,z_k) = \|u\|_E^2$ and the form is E elliptic on $Z^{\perp}$ with constant $\alpha^* = 1$. To compute the boundedness constant of the form b, the orthogonal decomposition $u = u_Z + u_{Z^{\perp}}$ is used together with equality $(p_i z_i, p_j z_j) = |p|^2$ which is due to orthonormality of the basis of the space of rigid motions. In turn

$$b(u,q) = q_k(u,z_k) = (u_Z, q_k z_k) \leq \|u_Z\|\|q_k z_k\| = \sqrt{a(u,u) + (u,z_k)(u,z_k)}|p| = \|u\|_E|p|.$$

We have $\beta^* = 1$. Finally, $\beta_* = 1$ in the inf-sup property

$$\sup_{u\in V}\frac{b(u,q)}{\|u\|_E} \geq \frac{(p_k z_k, p_i z_i)}{\sqrt{a(p_k z_k, p_i z_i) + (p_k z_k, p_i z_i)}} \geq \frac{|p|^2}{\sqrt{0+|p|^2}} = |p|.$$

As all the constants are independent of material parameters, the second assertion follows from the first one by operator preconditioning [8, ch 5.]. □

Using Theorem 3.3 it is readily established that the condition number of the composed operator $\mathcal{B}_E\mathcal{A} : W \mapsto W$ is equal to one. We further note that discretizing operator $\mathcal{B}_E$ leads to discrete nullspace preconditioners of [22, ch 6.].

While the spectral properties of $\mathcal{B}_E$ are appealing, the preconditioner is impractical. Consider $\mathbf{B}_E$ as a matrix representation of the Galerkin approximation of $\mathcal{B}_E$ in $W_h \subset W$. Then $\mathbf{B}_E = \text{diag}(\mathbf{A} + \mathbf{Y}\mathbf{Y}^\top, I)^{-1}$ where $\mathbf{Y} = \mathbb{R}^{n\times m}$, $\mathbf{y}_i = \text{col}_k\mathbf{Y} = \pi_h z_k$ and $z_k \in V_h$ is the basis function of the space of rigid motions. Due to the second (nonlocal) term the matrix $\mathbf{A} + \mathbf{Y}\mathbf{Y}^\top$ is dense. Further, as shall be discussed in §4, inverting the operator requires computing (the action of) the pseudoinverse of the singular matrix $\mathbf{A}$. The mapping $\mathcal{B}_M$, on the other hand, leads to a more practical preconditioner.

Theorem 3.4
Let $2\mu \geq 1$, $\{z_k\}_{k=1}^m$ be the L^2 orthonormal basis of the space of rigid motions of Ω, $\mathcal{A} : W \to W'$ defined in (7) and W_M defined analogically to Theorem 3.3. Then $\mathcal{A} : W_M \to W_M'$ is an isomorphism. Moreover the Riesz map $\mathcal{B}_M : W_M' \to W_M$ in (15) defines a parameter robust preconditioner for (7).

Proof
As in the proof of Theorem 3.3 we establish that $a(u, v) \leq \|u\|_M\|v\|_E$ and $b(v, p) \leq \|v\|\|p_k z_k\| \leq \|v\|_M|p|$. Setting $v = p_k z_k$ orthonormality of the basis yields $\inf_{p\in Q}\sup_{v\in V}\frac{b(v,p)}{\|v\|_M} \geq 1$. For M ellipticity of a on $Z^\perp$, assume existence of $C = C(\Omega)$ such that $\|u\|^2 \leq C\|\epsilon(u)\|^2$ for $u \in Z^\perp$. Then on $Z^\perp$

$$\|u\|^2 \leq C\|\epsilon(u)\|^2 \leq C\mu\|\epsilon(u)\|^2 \leq C(2\mu\|\epsilon(u)\|^2 + \lambda\|\nabla\cdot u\|^2) = C\|u\|_E^2$$

and

$$\|u\|_M^2 = \|u\|_E^2 + \|u\|^2 \leq (C+1)\|u\|_E^2$$

so that $a(u, u) = \|u\|_E^2 \geq (1 + C)^{-1}\|u\|_M^2$. Finally we comment on the assumption of existence of the constant C. Assume the contrary. Then there is $u \in Z^\perp$ such that $\|e(u)\| = 1$, $\|w(u)\| = 0$ and the $\|u\|$ unbounded. However, such u violates Korn's inequality (5). □

We remark that Theorem 3.4 required an additional assumption $2\mu \geq 1$. The assumption is not restrictive as it can be always achieved by scaling the equations such that the inequality is satisfied. Note also that the discrete preconditioner based on $\mathcal{B}_M$ is such that $\mathbf{B}_M{}^{-1} = \text{diag}(\mathbf{A} + \mathbf{M}, I)$, with $\mathbf{M}$ the mass matrix. The system to be assembled is therefore sparse.

Following Theorem 3.4 the condition number of the preconditioned operator $\mathcal{B}_M\mathcal{A} : W \to W$ depends solely on the constant C from Korn's inequality (5). An approximation for the constant is provided by the smallest positive eigenvalue λ_{min}^+ of the problem

$$\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^\top & \end{pmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{p}\end{pmatrix} = \lambda\begin{pmatrix}\mathbf{A}+\mathbf{M} & \\ & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{p}\end{pmatrix}.$$

In Table IX, Appendix A, the constant has been computed for two different domains; a cube from Example 2.2 and a hollow cylinder. In both cases $C \approx 1$ can be observed.

In order to demonstrate h robust properties of $\mathcal{B}_M$, the problem from Example 2.2 is discretized on $V_h \subset V$ and the resulting preconditioned linear system is solved by the minimal residual (MinRes) method [23]. Here the approximation of the preconditioner is provided by an algebraic multigrid (AMG), leading to the discrete operator

$$\mathbf{B}_M = \begin{pmatrix}\text{AMG}(\mathbf{A}+\mathbf{M}) & \\ & \mathbf{I}\end{pmatrix}.$$

The saddle point system was assembled and inverted using *cbc.block*, the FEniCS library for block matrices [24]. The results of the experiment are presented in Table V. Clearly, the number of iterations required for convergence is independent of the discretization. Moreover, the method yields

DOI: 10.1002/nla

numerical solutions which converge in the H^1 norm with the optimal rate[§] on both the uniform and nonuniform meshes, cf. Figure 2.

A drawback of the Lagrange multiplier formulation is the cost of solving the resulting indefinite linear system. Let us denote κ the condition number of an indefinite matrix $\mathbf{A}$. Under simplifying assumptions on the spectrum [25, ch 3.2] gives the following bound on the relative error in residual r_n at step n

$$\frac{|r_n|}{|r_0|} \le 2\left(\frac{\kappa-1}{\kappa+1}\right)^{\lfloor n/2 \rfloor}.$$

The result should be contrasted with a similar one for the error e_n at the n-th step of CG method on positive definite matrix $\mathbf{A}$, e.g. [26, thm 38.5],

$$\frac{{e_n}^\top \mathbf{A} e_n}{{e_0}^\top \mathbf{A} e_0} \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{n}.$$

While the above estimates are known to give the worst case behaviour of the two methods, the faster rate of convergence of CG motivates investigating formulations of (1) to which the conjugate gradient method can be applied.

4. CONJUGATE GRADIENT METHOD FOR DISCRETE SINGULAR PROBLEMS

We consider a variational formulation of (1) for $u \in V = \left[H^1(\Omega)\right]^3$ such that

$$2\mu(\epsilon(u), \epsilon(v)) + \lambda(\nabla \cdot u, \nabla \cdot v) = (f, v) + (h, v) \quad v \in V. \tag{16}$$

Denoting $a : V \times V \to \mathbb{R}, l : V' \to \mathbb{R}$ the bilinear and linear forms defined by (16), we note that the problem is not well-posed in V. Indeed, the compatibility conditions (2) restrict the functionals for which the solution can be found to $l \in Z^0 = \{f \in V'; \langle f, z\rangle = 0, z \in Z\}$. Moreover, only the part of u in $Z^\perp$ is uniquely determined by (16). More precisely we have the following result.

Theorem 4.1
Let $l \in Z^0$. Then there exists a unique solution of the problem

$$\text{Find } u \in Z^\perp \text{ such that for any } v \in Z^\perp \text{ it holds that } a(u, v) = \langle l, v\rangle. \tag{17}$$

Proof
The complete proof can be found as Theorem 11.2.30 in [27]. Note that boundedness and ellipticity of a on $Z^\perp$ with $\|\cdot\|_1$ are proven as part of Theorem 3.1. □

We remark that if (2) holds then $u \in Z^\perp$ solves (17) if and only if $(u, 0)$ solves the Lagrange multiplier problem (7). Further, the well-posed variational problem (17) is not suitable for discretization by the finite element method as the approximation leads to a dense linear system. A sparse discrete problem to which the conjugate gradient method shall be applied is therefore derived from (16).

Recall $m = \dim Z$, $n = \dim V_h$ and let $V_h = \operatorname{span}\{\phi_i\}_{i=1}^n$. Discretizing the variational problem (16) leads to a linear system

$$\mathbf{A}\mathbf{u} = \mathbf{b}, \tag{18}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ such that $A_{ij} = a(\phi_j, \phi_i)$ and vector $\mathbf{b} \in \mathbb{R}^n$, $b_i = \langle l, \phi_i\rangle$. Note that we shall consider (18) for a general right hand side, that is, not necessarily a discretization of $l \in Z^0$. We proceed by reviewing properties of the discrete system.

Due to symmetry and ellipticity of the bilinear form a on $Z^\perp$ there exists respectively m vectors $\mathbf{z}_k$ and $n - m$ eigenpairs $(\gamma_i, \mathbf{u}_i)$, $\gamma_i > 0$ such that $\mathbf{A}\mathbf{z}_k = 0$, ${\mathbf{z}_k}^\top \mathbf{u}_i = 0$, $\mathbf{A}\mathbf{u}_i = \gamma_i \mathbf{u}_i$ and

[§] We recall that V_h is constructed from continuous linear Lagrange elements

DOI: 10.1002/nla

Table IV. Preconditioned CG iterations on (18) obtained by discretization of (16) with problem parameters as in Example 2.2 and two preconditioners. Both systems are solved with relative tolerance of 10^{-10}. Uniform mesh is used.

size	$\mathbf{P}_z$AMG$(\mathbf{A}+\mathbf{M})$			AMG$(\mathbf{A}\vert\mathbf{Z})$		
	$\Vert u-u_h\Vert_1$	#	time $[s]$	$\Vert u-u_h\Vert_1$	#	time $[s]$
14739	1.14E-02 (1.09)	22	0.491	1.14E-02 (1.09)	21	0.537
107811	5.49E-03 (1.06)	23	10.17	5.49E-03 (1.06)	23	10.96
823875	2.71E-03 (1.02)	24	103.5	2.71E-03 (1.02)	25	86.51
6440067	1.35E-03 (1.00)	26	1580	1.35E-03 (1.00)	26	911.9

${\mathbf{u}_i}^\top\mathbf{u}_j=\delta_{ij}$. From the decomposition of $\mathbf{A}$ it follows that the system (18) is solvable if and only if ${\mathbf{z}_k}^\top\mathbf{b}=0$ for any k and the unique solution of the system is $\mathbf{u}\in\operatorname{span}\{\mathbf{u}_i\}_{i=1}^{n-m}$. We note that the last statement is the Fredholm alternative for (18). As a further consequence of the decomposition it is readily verified that given compatible vector $\mathbf{b}$, the solution of (18) is $\mathbf{u}=\mathbf{B}_A\mathbf{b}$ with $\mathbf{B}_A$ such that $\mathbf{B}_A\mathbf{y}=\sum_i{\gamma_i}^{-1}\left({\mathbf{u}_i}^\top\mathbf{y}\right)\mathbf{u}_i$. The matrix $\mathbf{B}_A$ is the pseudoinverse [28] or natural inverse [29, ch 3.] of $\mathbf{A}$.

We note that any vector from $\mathbb{R}^n$ can be orthogonalized with respect to the kernel of $\mathbf{A}$ by a projector $\mathbf{P}_Z=\mathbf{I}-\mathbf{Z}\mathbf{Z}^\top$, where $\mathbf{Z}\in\mathbb{R}^{n\times m}$ is the matrix consisting of l^2 orthonormal basis vectors of the kernel.

With $\mathbf{b}$ such that $\mathbf{Z}^\top\mathbf{b}=0$ the solution $\mathbf{u}$ of linear system (18) can be computed by the conjugate gradient method, e.g. [30]. Let $\mathbf{u}^0$ be the starting vector for the iterations. Then, assuming exact arithmetic and no preconditioner, the method preserves the component of $\mathbf{u}^0$ in $\mathbf{Z}$, i.e. $\mathbf{Z}^\top\mathbf{u}^0=\mathbf{Z}^\top\mathbf{u}$. In particular, $\mathbf{Z}^\top\mathbf{u}^0=0$ is required to obtain a solution orthogonal to the kernel. On the other hand, let $\mathbf{B}$ be the CG preconditioner. Then the iterations introduce components of the kernel to the solution even if $\mathbf{Z}^\top\mathbf{u}^0=0$, unless the range of $\mathbf{B}$ is orthogonal to $\mathbf{Z}$.

4.1. *Preconditioned CG for singular elasticity problem*

A suitable preconditioner for (18) is obtained by a composition with the $\mathbf{P}_Z$ projector and we shall consider $\mathbf{B}_M=\mathbf{P}_Z(\mathbf{A}+\mathbf{M})^{-1}$ where $\mathbf{M}$ is the mass matrix. That the preconditioner leads to bounded iteration count (and converging numerical solutions) is demonstrated in Table IV, cf. left pane. The preconditioner is also compared with a different preconditioner based on the approximation of the pseudoinverse $\mathbf{B}_A$. The approximation can be constructed by passing a kernel of the operator to the multigrid preconditioner, in the form of the l^2 orthonormal basis vectors, see [13]. Note that the preconditioners perform similarly in terms of iteration count, however, for large systems the pseudoinverse is cheaper.

We remark that in terms of operator preconditioning, the preconditioner based on the pseudoinverse can be interpreted as a Riesz map $Z^0\rightarrow Z^\perp$ defined with respect to the inner product induced by the bilinear form a. Recall that a is symmetric and elliptic on $Z^\perp$. On the other hand $\mathbf{B}_M$ approximates a mapping $Z^0\rightarrow V\rightarrow Z^\perp$.

Having established preconditioners for the indefinite system stemming from the Lagrange multiplier formulation (7) and the positive semi-definite problem stemming from (16), we shall finally discuss approximation properties of the computed solutions. To this end the problem from Example 2.2 is considered with f perturbed by rigid motions. Note that with the new functional l the problem (7) is well-posed while in (18) a compatible right hand side $\mathbf{b}$ will be obtained by projector $\mathbf{P}_Z$.

Results of the experiment are listed in Table V. The Lagrange multiplier method converges with an optimal rate on both the uniformly and non-uniformly discretized mesh, cf. Figure 2. On the other hand, solutions to (18) converge to the true solution *only* on the uniform mesh while there is no convergence with nonuniform discretization. Note that this is not signaled by growth of the iterations - for both methods the iteration counts are bounded. Note also that MinRes takes about twice as many iterations as CG.

Table V. (top) Convergence properties of the Lagrange multiplier formulation (7) and (bottom) the singular formulation (16) utilizing l^2 orthogonal basis of the nullspace to invert the system (18). Only the multiplier formulation yields solutions converging on uniform and nonuniform meshes. Relative tolerances of 10^{-11} and 10^{-10} are used for MinRes and CG respectively.

uniform				refined			
size	$\lVert u-u_h\rVert_1$	#	$\max_Z \lvert(u_h,z)\rvert$	size	$\lVert u-u_h\rVert_1$	#	$\max_Z \lvert(u_h,z)\rvert$
14745	1.03E-02 (1.14)	44	3.54E-07	13080	3.11E-02 (0.99)	50	1.68E-07
107817	4.84E-03 (1.09)	45	2.77E-06	98052	1.41E-02 (1.14)	53	6.73E-08
823881	2.36E-03 (1.03)	45	1.38E-06	759546	6.53E-03 (1.11)	54	8.11E-07
6440073	1.18E-03 (1.00)	44	1.75E-05	5978835	3.20E-03 (1.03)	55	2.94E-06
14739	1.14E-02 (1.09)	21	1.30E-03	13074	5.51E-02 (0.45)	26	6.06E-03
107811	5.49E-03 (1.06)	23	6.66E-04	98046	5.05E-02 (0.12)	27	6.32E-03
823875	2.71E-03 (1.02)	25	3.36E-04	759540	5.00E-02 (0.02)	29	6.43E-03
6440067	1.35E-03 (1.00)	26	1.69E-04	5978829	4.98E-02 (0.01)	31	6.49E-03

From the experiment we conclude that the conjugate gradient method for (18), as applied so far, in general does not yield converging numerical solutions of (16). It is next shown that the issue is due projector $\mathbf{P}_Z = \mathbf{I} - \mathbf{Z}\mathbf{Z}^\top$ which the method uses and which is derived from the discrete problem. In particular, we show that $\mathbf{P}_Z$ is not a correct discretization of a projector used in the continuous problem (17) (and (7)). Following the continuous problem, a modification to CG is proposed, which leads to a converging method.

4.2. *Conjugate gradient method with Z^0, $Z^\perp$ projectors*

Consider the variational problem (17) which was proven well-posed in Theorem 4.1 under the assumptions $l \in Z^0 \subset V'$ and $u \in Z^\perp \subset V$. In this respect, there are two subspaces associated with (17) and we shall define two projectors $P : V \rightarrow Z^\perp$, $P' : V' \rightarrow Z^0$ such that for $v \in V$

$$\begin{aligned} (Pu, v) &= (u, v) - (u, z_k)(v, z_k), \\ \langle P'f, v\rangle &= \langle f, v\rangle - \langle f, z_k\rangle(v, z_k), \end{aligned} \tag{19}$$

where $Z = \operatorname{span}\{z_k\}_{k=1}^m$ is the L^2 orthonormal basis of the space of rigid motions (e.g. constructed by Lemma 3.2). Similar projectors were discussed in [10] for the singular Poisson problem. We note that $\langle f, Pu\rangle = \langle P'f, u\rangle$ and thus P' is the adjoint of P. Note also that the two projectors are present in the multiplier formulation (7).

Lemma 4.1
Let $f \in V'$ and P, P' be the projectors (19). Then $(u, p) \in V \times Q$ solves (7) with the right hand side $(v, q) \mapsto \langle f, v\rangle + \langle 0, q\rangle$ if and only if $u \in Z^\perp$ and u solves (17) with the right hand side $P'(f)$.

Proof
It suffices to establish the relation between the right hand sides. Using orthogonality of the basis it follows from testing (7) with $(z_k, 0)$ that $p_k = \langle f, z_k\rangle$. Substituting the obtained Lagrange multiplier, the new right hand side of (7) is $(v, q) \mapsto \langle f, v\rangle - \langle f, z_k\rangle(v, z_k) + \langle 0, q\rangle = \langle P'(f), v\rangle + \langle 0, q\rangle$. □

To derive a matrix representation of the projectors with respect to nodal basis $V_h = \operatorname{span}\{\phi_i\}_{i=1}^n$, the mappings $\pi_h : V_h \rightarrow \mathbb{R}^n$ (the nodal interpolant) and $\mu_h : V_h' \rightarrow \mathbb{R}^n$ from (3) are used. We recall that $(u, v) = \mathbf{v}^\top\mathbf{M}\mathbf{u}$ for $\mathbf{u} = \pi_h u$, $\mathbf{v} = \pi_h v$ and $\mathbf{M}$, $M_{ij} = (\phi_j, \phi_i)$ the mass matrix while $\langle f, v\rangle = \mathbf{f}^\top\mathbf{v}$ with $\mathbf{f} = \mu_h f$. Finally, matrix $\mathbf{Y} = \mathbb{R}^{n\times m}$ is such that $\mathbf{y}_i = \operatorname{col}_k\mathbf{Y} = \pi_h z_k$ where $z_k \in V_h$ belongs to the L^2 orthogonal basis of the space of rigid motions. Then

$$\begin{aligned} \mathbf{v}^\top\mathbf{M}\mathbf{P}\mathbf{u} &= (Pu, v) = (u, v) - (u, z_k)(v, z_k) = \mathbf{V}^\top\mathbf{M}\left(\mathbf{I} - \mathbf{Y}\mathbf{Y}^\top\mathbf{M}\right)\mathbf{u}, \\ \mathbf{f}^\top{\mathbf{P}'}^\top\mathbf{v} &= \langle f, Pv\rangle = \langle f, v\rangle - \langle f, z_k\rangle(v, z_k) = \mathbf{f}^\top\left(\mathbf{I} - \mathbf{Y}\mathbf{Y}^\top\mathbf{M}\right)\mathbf{v} \end{aligned} \tag{20}$$

and $\mathbf{P} = (\mathbf{I} - \mathbf{Y}\mathbf{Y}^\top\mathbf{M})$ is the representation of P while P' is represented by $\mathbf{P}^\top$. We remark that in addition to $\mathbf{Y}$, the rigid motions $Z_h = \operatorname{span}\{z_k\}_{k=1}^m$ can be represented in $\mathbb{R}^n$ by an additional

Table VI. Convergence of conjugate gradient solutions for (18) with different combinations of right hand (horizontal) side and left hand side (vertical) projectors. The problem from Example 2.2 is considered. Preprocessing the right hand side and postprocessing the solution by projectors $(\mathbf{P}^\top, \mathbf{P})$ yields solutions converging with optimal rate.

	size	$\mathbf{P}_Z$			$\mathbf{P}^\top$		
		$\|u-u_h\|_1$	#	$\max_Z \|(u_h, z)\|$	$\|u-u_h\|_1$	#	$\max_Z \|(u_h, z)\|$
$\mathbf{P}_Z$	13074	5.51E-02 (0.45)	26	6.06E-03	5.53E-02 (0.44)	27	6.05E-03
	98046	5.05E-02 (0.12)	27	6.32E-03	5.11E-02 (0.12)	28	6.31E-03
	759540	5.00E-02 (0.02)	29	6.43E-03	5.06E-02 (0.01)	29	6.42E-03
	5978829	4.98E-02 (0.01)	31	6.49E-03	5.05E-02 (0.00)	31	6.48E-03
$\mathbf{P}$	13074	3.13E-02 (0.98)	27	6.84E-16	3.11E-02 (0.99)	25	6.15E-16
	98046	1.45E-02 (1.11)	28	2.94E-14	1.41E-02 (1.14)	27	2.92E-14
	759540	6.92E-03 (1.07)	29	6.39E-14	6.53E-03 (1.11)	29	6.40E-14
	5978829	3.63E-03 (0.93)	31	2.89E-13	3.20E-03 (1.03)	31	2.86E-13

matrix $\mathbf{W} = \mathbf{M}\mathbf{Y}$, which is μ_h applied to functionals $v \mapsto (z_k, v)$. Following [8] the matrices $\mathbf{Y}$, $\mathbf{W}$ are termed respectively the primal and dual representation of Z_h. Observe that in (20) matrix $\mathbf{P}$ uses the primal representation for $\mathbf{u}$ while the vector is expanded in the dual representation by $\mathbf{P}'$. Moreover, L^2 orthogonality of Z_h yields ${\mathbf{y}_i}^\top \mathbf{w}_j = \delta_{ij}$. Finally note that the projectors $\mathbf{P}^\top$, $\mathbf{P}$ are implicitly present in the linear system which is the discretization of the multiplier problem (7) with the orthogonal basis of rigid motions

$$\begin{pmatrix} \mathbf{A} & \mathbf{W} \\ \mathbf{W}^\top & \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}. \tag{21}$$

Indeed, $\mathbf{p} = \mathbf{Y}^\top \mathbf{b}$ from premultiplying the first equation by $\mathbf{Y}^\top$. Upon substitution the equation reads $\mathbf{A}\mathbf{u} = \mathbf{b} - \mathbf{W}\mathbf{Y}^\top \mathbf{b} = \mathbf{P}^\top \mathbf{b}$. Further the solution is such that $\mathbf{P}\mathbf{u} = 0$.

The situation where the continuous problems (7), (17) and the discrete problem (21) use different projectors for the left and right hand sides contrasts with (18) which utilizes $\mathbf{P}_Z$ to obtain consistent right hand side and the solution is such that $\mathbf{P}_Z \mathbf{u} = 0$ as well. This observation together with the lack of convergence of the CG method, cf. Table V, motivate that the CG method on (18) is used with the following two modifications: (i) the iterations are started from vector $\mathbf{P}^\top \mathbf{b}$, (ii) $\mathbf{P}$ is applied to the final solution.

The effect of the proposed modifications is shown in Table VI. The problem from Example 2.2 is considered on a non-uniform mesh and CG on (18) is applied with different combinations of projectors used to obtain the right hand side from incompatible vector $\mathbf{b}$ and to orthogonalize the converged solution. We observe that only the case $(\mathbf{P}^\top, \mathbf{P})$[¶] yields optimal convergence. With $(\mathbf{P}_Z, \mathbf{P})$ the rate is slightly smaller than one. In the remaining two cases the solution do not converge suggesting that for convergence $\mathbf{P}$ must be applied to the solution.

The results shown in Table VI are satisfactory in a sense that preprocessing the right hand side with $\mathbf{P}^\top$ and postprocessing the solution with $\mathbf{P}$ improved the convergence properties of the CG method for (18). However, the modifications alter the original discrete problem and thus the properties of the new problem should be discussed. We note that in the discussion $\mathbf{Z}$, $\mathbf{Y}$ are respectively $\mathbf{I}$ and $\mathbf{M}$ orthogonal basis of the nullspace of $\mathbf{A}$. Further, the transformation matrix between the basis is $c \in \mathbb{R}^{m \times m}$ such that $\mathbf{Z} = \mathbf{Y}c$ and we have $\mathbf{Y}^\top \mathbf{M}\mathbf{Z} = c$.

First, admissibility of the modified right hand side $\mathbf{P}^\top \mathbf{b}$ is considered. Using the transformation matrix it holds that $\mathbf{Z}^\top \mathbf{P}^\top \mathbf{b} = 0$ and thus $\mathbf{P}^\top \mathbf{b}$ is compatible and the solution can be obtained by a pseudoinverse (or equivalently by CG). The computed solution of the new linear system then satisfies $\mathbf{Z}^\top \mathbf{u} = 0$. However, the continuous problem requires orthogonality $\mathbf{Y}^\top \mathbf{M}\mathbf{u} = Ch$. As the two conditions are related through $|\mathbf{Y}^\top \mathbf{M}\mathbf{u}|^2 = \mathbf{u}^\top \mathbf{M}\mathbf{Z}(c^\top c)^{-1} \mathbf{Z}^\top \mathbf{M}\mathbf{u} = \mathbf{u}^\top \mathbf{M}\mathbf{Z}(\mathbf{Z}^\top \mathbf{M}\mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{M}\mathbf{u}$, and $\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}$, orthogonality in the L^2 inner product depends on similarity of the mass matrix with

[¶] Elements of the tuple denote respectively the projector for the right hand side and the left hand side.

DOI: 10.1002/nla

identity. This is essentially a condition on the mesh and $|\mathbf{Y}^\top\mathbf{MZ}| \geq C$ is possible (as observed in Table VI).

To enforce orthogonality constraint $\mathbf{Y}^\top\mathbf{Mu} = 0$ without postprocessing we shall finally consider linear system $\mathbf{Au} = \mathbf{P}^\top\mathbf{b}$ and require $\mathbf{Pu} = 0$ for uniqueness. In this case the solution is not provided by pseudoinverse $\mathbf{B}_A$. However, a similar construction based on the generalized eigenvalue problem can be used instead.

Lemma 4.2
Let $\mathbf{u}$ be a unique solution of $\mathbf{Au} = \mathbf{P}^\top\mathbf{b}$, satisfying $\mathbf{Pu} = 0$ and $\mathbf{\Gamma} \in \mathbb{R}^{n\times n}$, $\mathbf{U} \in \mathbb{R}^{n\times n-m}$ such that $\mathbf{AU} = \mathbf{MU\Gamma}$, $\mathbf{U}^\top\mathbf{MU} = \mathbf{I}$. Then $\mathbf{u} = \mathbf{BP}^\top\mathbf{b}$ where $\mathbf{B} = \mathbf{U\Gamma}^{-1}\mathbf{U}^\top$.

Proof
First, note that the existence of matrices $\mathbf{U}, \mathbf{\Gamma}$ follows from positive semi-definiteness of $\mathbf{A}$. Further, by $\mathbf{M}$ orthogonality of the eigenvectors $\mathbf{MUx} = \mathbf{P}^\top\mathbf{b}$ holds with $\mathbf{x} = \mathbf{U}^\top\mathbf{b}$. As $\mathbf{Y}^\top\mathbf{MU} = 0$ any vector $\mathbf{Bb}$ is $\mathbf{M}$ orthogonal with $\mathbf{Y}$ and thus $\mathbf{PBb} = 0$. It remains to show that the composition $\mathbf{AB}$ is the identity on the subspace spanned by columns of $\mathbf{MU}$

$$\mathbf{ABMU} = \mathbf{AU\Gamma}^{-1}\mathbf{U}^\top\mathbf{MU} = \mathbf{AU\Gamma}^{-1} = \mathbf{MU\Gamma\Gamma}^{-1} = \mathbf{MU}.$$

□

5. NATURAL NORM FORMULATION

An attractive feature of the variational problem (16) is the fact that the resulting linear system is amiable to solution by the CG method, which when modified following §4 yields converging solutions. However, the projectors P', P are only applied as pre and postprocessor and the CG loop (Lanczos process) is in this respect detached from the continuous problem. Moreover the method requires a special preconditioner that handles the nullspace of matrix $\mathbf{A}$. A formulation which leads to a positive definite linear system requiring only a regular (not nullspace aware) preconditioner shall be studied next.

Theorem 5.1
Let $a : V \times V \to \mathbb{R}$, $a(u, v) = 2\mu(\epsilon(u), \epsilon(v)) + \lambda(\nabla \cdot u, \nabla \cdot v)$ and $Z = \operatorname{span}\{z_k\}_{k=1}^m$ the L^2 orthogonal basis of the space of rigid motions. Futher let $l \in Z^0$. There exists a unique $u \in V$ such

$$a(u, v) + (u, z_k)(v, z_k) = \langle l, v\rangle \quad v \in V. \tag{22}$$

Moreover $u \in Z^\perp$.

Proof
Recall that the bilinear form above is the inner product $(u, v)_E$ from (12) which induces an equivalent norm on V, cf. Lemma 3.3. The existence and uniqueness of the solution now follow from the Lax-Milgram lemma. Testing the equation with $v = z_i$ yields $(u, z_i) = 0$ and in turn $u \in Z^\perp$. □

We remark that the solution of (22) and (17) are equivalent because $l \in Z^0$. Note also that Theorem 3.4 gives equivalence bounds $(1 + C)^{-1}\|u\|^2_M \leq \|u\|^2_E \leq \|u\|^2_M$, $C = C(\Omega)$ and in turn the Riesz map with respect to the inner product $(u, v)_M = a(u, v) + (u, v)$ defines a suitable h robust preconditioner for (22). Finally, observe that the L^2 orthogonality of decomposition $u = u_Z + u_{Z^\perp}$, $u_Z = (u, z_k)z_k$ is respected by the inner product $(\cdot, \cdot)_E$, see (12). The norm $\|u\|_E$, see (13), thus considers Z and $Z^\perp$ with L^2 and a induced norms which are the natural norms for the subspaces.

Using (20) the natural norm formulation (22) leads to a positive definite linear system

$$\left[\mathbf{A} + \mathbf{MY}(\mathbf{MY})^\top\right]\mathbf{u} = \mathbf{P}^\top\mathbf{b}.$$

where we recognize a dense matrix from the discretization of $\mathcal{B}_E$ preconditioner of the Lagrange multiplier formulation, cf. Theorem 3.3. Therein the inverse of the matrix was of interest. However,

Table VII. Convergence study of the natural norm formulation (22) for the singular elasticity problem from Example 2.2. The system is solved with relative tolerance 10^{-11}. The CG iterations use a preconditioner AMG($\mathbf{A}+\mathbf{M}$). The iteration count remains bounded and the solutions converge with the optimal rate.

uniform				refined			
size	$\|u-u_h\|_1$	#	$\max_Z \|(u_h, z)\|$	size	$\|u-u_h\|_1$	#	$\max_Z \|(u_h, z)\|$
14739	1.03E-02 (1.14)	33	2.57E-08	13074	3.11E-02 (0.99)	39	3.70E-08
107811	4.84E-03 (1.09)	29	1.80E-05	98046	1.41E-02 (1.14)	41	3.46E-08
823875	2.36E-03 (1.03)	37	9.23E-09	759540	6.53E-03 (1.11)	43	8.90E-08
6440067	1.18E-03 (1.00)	33	2.38E-05	5978829	3.20E-03 (1.03)	46	3.53E-08

relevant for the CG method here is only the matrix vector product, which can be computed efficiently by storing separately $\mathbf{A}$ and $\mathbf{MY}$, the dual representation of rigid motions in V_h.

With (22) we finally revisit the test problem from Example 2.2. Results of the method are summarized in Table VII. Optimal convergence rate is observed with both uniform and nonuniform meshes. Moreover, the CG iteration count with the proposed Riesz map preconditioner approximated by AMG($\mathbf{A}+\mathbf{M}$) remains bounded. An interesting observation is the fact that the error in the orthogonality constraint is smaller in comparison to the Lagrange multiplier formulation, cf. Table V.

6. NEARLY INCOMPRESSIBLE MATERIALS

So far we have assumed that μ and λ are comparable in magnitude. In this section we handle the case where $\lambda \gg \mu$ and the material is nearly incompressible. The variational problems (6), (16), (22) studied thus far were based on the pure displacement formulation of linear elasticity (1) and H^1 conforming finite element spaces were used for their discretization. Due to the *locking* phenomenon the approximation properties of their respected solutions are known to degrade for nearly incompressible materials with $\lambda \gg \mu$, (equivalently Poisson ratio close to 1/2), see e.g. [15, ch 6.3]. Moreover, the incompressible limit presents a difficulty for convergence of iterative methods in the standard form.

Methods robust with respect to increasing λ can be formulated using a discretization with nonconforming elements, [27, ch 11.4]. However, this method fails to satisfy the Korn's inequality. To the authors' knowledge the only finite element method that is both robust in λ and satisfies Korn's inequality is [31, 32]. In addition to problems with the discretization, standard multigrid algorithms do not work well for large λ and special purpose algorithms must be used [33]. For this reason we resort to a more straightforward solution of the mixed formulation where an additional variable, the *solid pressure* p, is introduced. Let the solid pressure be defined as $p = \lambda\nabla\cdot u$ so that (6) is reformulated as

$$\begin{aligned} \nabla\cdot(2\mu\epsilon(u)) - \nabla p &= f && \text{in } \Omega,\\ \lambda\nabla\cdot u - p &= 0 && \text{in } \Omega,\\ \sigma(u)\cdot n &= h && \text{on } \partial\Omega. \end{aligned} \tag{23}$$

Note that the problem is singular, since any pair $u \in Z$, $p = 0$ can be added to the solution. In fact such pairs constitute the kernel of (23). To obtain a unique solution we shall as in §3, require that u is orthogonal to the rigid motions Z. We assume that the basis of Z is orthonormal.

Setting $Q = L^2(\Omega)$, $Y = \mathbb{R}^6$ we shall consider a variational problem for triplet $u \in V$, $p \in Q$, $x \in Y$ such that

$$\begin{aligned} &2\mu(\epsilon(u), \epsilon(v)) + (p, \nabla\cdot v) + x_k(v, z_k) = (f, v) + (h, v) && v \in V,\\ &(q, \nabla\cdot u) - \lambda^{-1}(p, q) = 0 && q \in Q,\\ &y_k(u, z_k) = 0 && y \in Y. \end{aligned} \tag{24}$$

Equation (24) is a double saddle point problem

$$\mathcal{A}\begin{pmatrix}u\\p\\x\end{pmatrix}=\begin{pmatrix}A & B & D\\ B' & -\lambda^{-1}C & \\ D' & & \end{pmatrix}\begin{pmatrix}u\\p\\x\end{pmatrix}=\begin{pmatrix}b\\ \\ \end{pmatrix},$$

with operators $A : V \to V'$, $B : Q \to V'$, $C : Q \to Q'$ and $D : X \to V'$ defined as

$$\begin{aligned}\langle Au, v\rangle &= 2\mu(\epsilon(u), \epsilon(v)), & \langle Bp, v\rangle &= (p, \nabla\cdot v),\\ \langle Cp, q\rangle &= (p, q), & \langle Dx, v\rangle &= x_k(v, z_k).\end{aligned}$$

To show well-posedness of the constrained mixed formulation (24) the abstract theory for saddle points problems with small (note that that $\lambda \gg 1$) penalty terms [15, ch 3.4] is applied. To this end we introduce the bilinear forms $a(u, v) = \langle Au, v\rangle$,

$$b(v, (p, x)) = \langle Bp, v\rangle + \langle Dx, v\rangle, \tag{25}$$

$c((p, y), (q, x)) = \langle Cp, q\rangle$ so that (24) is recast as

$$\begin{aligned}a(u, v) + b(v, (p, x)) &= (f, v) + (h, v) & v &\in V,\\ b(u, (q, y)) - \lambda^{-1}(p, q) &= 0 & (q, y) &\in Q \times Y.\end{aligned}\tag{26}$$

The space $Q \times Y$ will be considered with the norm $\|(p, x)\| = \sqrt{\|p\|^2 + |x|^2}$, while V is considered with the H^1 norm. Following [15, thm 4.11] the problem (26) is well-posed provided that the assumptions of Brezzi theory hold and in addition c is continuous and c and a are positive

$$a(u, u) \geq 0, \quad u \in V \qquad \text{and} \qquad c((p, x), (p, x)) \geq 0, \quad (p, x) \in Q \times Y.$$

We review that continuity and V-ellipticity of a on $Z^\perp$ was shown in Theorem 3.1 and as $a(z, z) = 0$, $z \in Z$, the form is positive on V. Moreover, by Lemma 3.1, Cauchy-Schwarz inequality and orthonormality of basis

$$\begin{aligned}b(v, (p, x)) = (p, \nabla\cdot v) + x_k(v, z_k) \leq \sqrt{3}\|p\|\|\nabla v\| + \|v\||x| &\leq \sqrt{3}\sqrt{\|v\|^2 + \|\nabla v\|^2}\sqrt{\|p\|^2 + |x|^2}\\ &\leq \beta^*\|v\|_1\|(p, x)\|.\end{aligned}$$

It is easy to observe that continuity and positivity of the bilinear form c hold and thus (26) is well-posed provided that the inf-sup condition is satisfied. We note that the proof requires extra regularity of the boundary.

Lemma 6.1
Let Ω with a smooth boundary and b be the bilinear form over $V \times (Q \times Y)$ defined in (25). There exists $\beta_* = \beta_*(\Omega)$ such that

$$\sup_{v\in V}\frac{b(v, (p, x))}{\|v\|_1} \geq \beta_*\|(p, x)\|.$$

Proof
Let $p \in Q$ and $x \in Y$ given. Following [27, thm 11.2.3] there exists for every p, $v^* \in V$ such that

$$p = \nabla\cdot v^*, \tag{27a}$$

$$\|v^*\|_1 \leq C(\Omega)\|p\|. \tag{27b}$$

The element v^* is constructed from the unique solution of the Poisson problem

$$\begin{aligned}-\Delta w &= p & &\text{in } \Omega,\\ w &= 0 & &\text{on } \partial\Omega,\end{aligned}\tag{28}$$

taking $v^* = -\nabla w$. Observe that the computed $v^* \in Z^{\perp}$

$$-\,(z, v^*) = \int_{\Omega} z\nabla w = \int_{\partial\Omega} wz\cdot n - \int_{\Omega} w\nabla\cdot z = 0 \quad z \in Z. \tag{29}$$

Orthogonality of v^* and (27a) yields that $b(v^* + x_k z_k, (p, x)) = (p, \nabla\cdot v^*) + (x_k z_k, x_l z_l) = \|p\|^2 + |x|^2$. Further, by Cauchy-Schwarz and Young's inequalities

$$\begin{aligned}\|v^* + x_k z_k\|_1^2 &= \|v^* + x_k z_k\|^2 + \|\nabla(v^* + x_k z_k)\|^2\\ &= \|v^*\|^2 + \|x_k z_k\|^2 + \|\nabla v^*\|^2 + 2(\nabla v^*, \nabla x_k z_k) + \|\nabla x_k z_k\|^2\\ &\leq 2\|v^*\|_1^2 + 2(\|v^*\|^2 + \|\nabla x_k z_k\|^2).\end{aligned}$$

Using (27b) and Lemma 3.1 gives $\|v^* + x_k z_k\|_1^2 \leq 2C(\Omega)\|p^2\| + (1 + 2|\Omega|)|x|^2 \leq c(\Omega)\|(p, x)\|^2$. Combining the observations

$$\sup_{v\in V} \frac{b(v, (p, x))}{\|v\|_1} \geq \frac{b(v^* + x_k z_k, (p, x))}{\|v^* + p_k z_k\|_1} = \frac{\|p\|^2 + |x|^2}{\|v^* + p_k z_k\|_1} \geq \frac{1}{c}\sqrt{\|p\|^2 + |x|^2} = \frac{1}{c}\|(p, x)\|.$$

□

We remark that none of the constants of the problem (26) depends on λ despite the norm of $Q \times Y$ being free of the parameter, cf. also [34, 35]. Observe also that with H^1 norm on V the boundedness constant of a depends on μ, cf. Theorem 3.1, and thus the parameter shall be included in the norm to get a μ independent preconditioner. Finally, note that tighter bounds (e.g. in the proof of Lemma 6.1) can be obtained if the space V is considered with the norm $u \mapsto \sqrt{\mu\|\epsilon(u)\|^2 + \|u\|^2}$.

Motivated by the above, we shall consider as the preconditioner for the well-posed problem (26) a Riesz map $\mathcal{B} : (V \times Q \times Y)' \to (V \times Q \times Y)$ with respect to the inner product inducing the norm $(u, p, x) \mapsto \sqrt{\mu\|\epsilon(u)\|^2 + \|u\|^2 + \|p\|^2 + |x|^2}$

$$\mathcal{B} = \begin{pmatrix} A + M & & \\ & C & \\ & & I \end{pmatrix}^{-1}, \tag{30}$$

where M was defined in (12). Similar preconditioners for the Dirichlet problem has been discussed in [35].

Remark 6.1 (Lemma 6.1 in the discrete case)
The continuous inf-sup condition can be extended to Taylor-Hood discretizations in the following way. We consider $V_h \subset V$, $Q_h \subset Q$ approximated with the lowest order Taylor-Hood element. Given $p_h \in Q_h$ both the element $v_h^* \in V_h$ and $w_h \in Q_h$ from Lemma 6.1 are found as the solution to the mixed Poisson problem

$$\begin{aligned}(v_h^*, v) + (\nabla_h w_h, v) &= 0 \qquad && v \in V_h,\\ (\nabla_h q, v_h^*) &= -(p_h, q) && q \in Q_h.\end{aligned}$$

The problem is well-posed due to the weak inf-sup condition

$$\sup_{v\in V_h} \frac{(v_h, \nabla_h q_h)}{\|v_h\|} \geq C\|q_h\|_1.$$

Since $z \in V_h$ a direct calculation shows that the orthogonality condition (29) is satisfied.

Both in the above and in the construction of the proof of Lemma 6.1 we relied on a well-posed mixed Poisson problem to obtain orthogonality with respect to the kernel. We note that stable Stokes element $P_2 - P_0$ does not allow for such a construction and does not give h uniform bounds.

To show that the preconditioner (30) is robust with respect to λ, the problem from Example 2.2 is considered with $\mu = 1$ and $\lambda \in \left[1, 10^8\right]$. The spaces V and Q are approximated by lowest order Taylor-Hood elements for which the discrete inf-sup condition from Lemma 6.1 holds following Remark 6.1. The non-trivial blocks of the preconditioner are inverted using algebraic multigrid and the system is solved using the MinRes method requiring reduction of the preconditioned residual norm by a factor of 10^{10} for convergence.

From the results of the experiment, summarized in Table VIII, it is evident that the iteration count is bounded in λ as well as in the discretization parameter.

Table VIII. Iteration counts of the preconditioned MinRes method for mixed linear elasticity problem (24) and different values of Lamé constant λ. The iteration counts remain bounded for the considered values of the parameter.

$\dim(V)$	$\dim(Q)$	λ				
		10^0	10^2	10^4	10^6	10^8
14739	729	109	113	100	70	36
107811	4913	107	109	103	69	36
823875	35937	109	109	107	72	36
6440067	274625	109	108	113	75	37

7. CONCLUSIONS

We have studied the singular Neumann problem of linear elasticity. Four different formulations of the problem have been analyzed and mesh independent preconditioners established for the resulting linear systems within the framework of operator preconditioning. We have proposed a preconditioner for the (singular) mixed formulation of linear elasticity, that is robust with respect to the material parameters. Using an orthonormal basis of the space of rigid motions, discrete projection operators have been derived and employed in a modification to the conjugate gradients method to ensure optimal error convergence of the solution.

A. EIGENVALUE BOUNDS FOR LAGRANGE MULTIPLIER PRECONDITIONERS

Bounds for the eigenvalues of operators $\mathcal{B}_E\mathcal{A}$ and $\mathcal{B}_M\mathcal{A}$ from (7) and (15) are approximated by considering the eigenvalue problems

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \lambda {\mathbf{B_i}}^{-1} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} \tag{31}$$

with the left hand side the discretization of (7) and $\mathbf{B_i}$, $i \in \{E, M\}$ discretizations of preconditioners $\mathcal{B}_i$ from (15). The spectrum of the symmetric, indefinite problem (31) is a union of negative and positive intervals $[\lambda^-_{\text{min}}, \lambda^-_{\text{max}}]$, $[\lambda^+_{\text{min}}, \lambda^+_{\text{max}}]$. Following the analysis in Theorems 3.3 and 3.4 negative bounds equal to -1 are expected for both preconditioners. Further, the positive eigenvalues are bounded from above by 1. Finally, $\lambda^+_{\text{min}} = -1$ for $\mathcal{B}_E$ while the constant $C = C(\Omega)$ from the Korn's inequality determines the bound for $\mathcal{B}_M$.

In the experiment, Ω as a cube from Example 2.2 and a hollow cylinder with inner and outer radii $\frac{1}{2}$, 1 and height 2 are considered. Lamé constants $\mu = 384$, $\lambda = 577$ are used. For both bodies $C \approx 1$ is observed, cf. Table IX. The remaining bounds agree well with the analysis.

REFERENCES

1. Marsden J, Hughes T. *Mathematical Foundations of Elasticity*. Dover Civil and Mechanical Engineering Series, Dover, 1994.

Table IX. Spectral bounds for eigenvalue problems (31). (Top) The body is cube. (Bottom) The body is a cylinder.

	size	κ	$\lambda^-_{\min}+1$	$\lambda^-_{\max}+1$	$\lambda^+_{\min}-1$	$\lambda^+_{\max}-1$
$\mathbf{B}_E$	87	1.0000	-6.83E-11	2.92E-11	-4.36E-11	5.89E-12
	381	1.0000	-1.38E-10	7.00E-12	-1.61E-10	5.55E-15
	2193	1.0000	-5.88E-10	1.65E-11	-6.23E-10	9.55E-15
	14745	1.0000	-1.10E-08	-4.27E-09	-2.00E-08	1.73E-14
$\mathbf{B}_M$	87	1.0001	-6.64E-11	4.46E-12	-1.10E-04	1.03E-11
	381	1.0002	-1.35E-10	-1.06E-11	-2.33E-04	-5.33E-12
	2193	1.0004	-5.73E-10	-1.12E-11	-4.00E-04	5.91E-12
	14745	1.0005	-2.37E-09	-7.73E-11	-4.97E-04	-4.47E-11
$\mathbf{B}_E$	210	1.0000	-3.91E-12	-4.46E-13	-4.58E-12	9.33E-15
	462	1.0000	-3.82E-12	-8.91E-13	-4.55E-12	5.77E-15
	1764	1.0000	-9.32E-12	-4.40E-12	-1.08E-11	1.31E-14
	8292	1.0000	-3.71E-11	-1.74E-11	-4.06E-11	6.26E-14
$\mathbf{B}_M$	210	1.0752	1.84E-02	7.00E-02	-7.00E-02	-2.57E-06
	462	1.0219	1.94E-03	2.14E-02	-2.14E-02	-2.21E-06
	1764	1.0069	1.14E-03	6.82E-03	-6.82E-03	-4.57E-07
	8292	1.0022	1.60E-04	1.66E-03	-2.17E-03	-2.10E-08

2. Bauchau OA, Craig JI. *Basic equations of linear elasticity*. Springer Netherlands: Dordrecht, 2009; 3–51, doi: 10.1007/978-90-481-2516-6_1.
3. Finite element analysis, DNVGL–CG–0127. *Technical Report*, Det Norske Veritas GL 10 2015.
4. Dutta-Roy T, Wittek A, Miller K. Biomechanical modelling of normal pressure hydrocephalus. *Journal of Biomechanics* 2008; **41**(10):2263 – 2271.
5. Støverud K, Alnæs M, Langtangen H, Haughton V, Mardal KA. 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* 2016; **19**(6):686–698.
6. Tobie G, Čadek O, Sotin C. Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus. *Icarus* 2008; **196**(2):642 – 652. Mars Polar Science IV.
7. Hestenes MR, Stiefel E. *Methods of conjugate gradients for solving linear systems*, vol. 49. NBS, 1952.
8. Mardal KA, Winther R. Preconditioning discretizations of systems of partial differential equations. *Numerical Linear Algebra with Applications* 2011; **18**(1):1–40.
9. Ciarlet PG. On Korn's inequality. *Chinese Annals of Mathematics, Series B* 2010; **31**(5):607–618.
10. Bochev P, Lehoucq RB. On the finite element solution of the pure Neumann problem. *SIAM review* 2005; **47**(1):50–66.
11. Falgout RD, Meier Yang U. hypre: A library of high performance preconditioners. *Computational Science ICCS 2002*, *Lecture Notes in Computer Science*, vol. 2331, Sloot PMA, Hoekstra AG, Tan CJK, Dongarra JJ (eds.). Springer Berlin Heidelberg, 2002; 632–641.
12. Alnæs M, Blechta J, Hake J, Johansson A, Kehlet B, Logg A, Richardson C, Ring J, Rognes M, Wells G. The Fenics project version 1.5. *Archive of Numerical Software* 2015; **3**(100).
13. Balay S, Brown J, Buschelman K, Eijkhout V, Gropp WD, Kaushik D, Knepley MG, McInnes LC, Smith BF, Zhang H. PETSc users manual. *Technical Report ANL-95/11 - Revision 3.4*, Argonne National Laboratory 2013.
14. Brezzi F. On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique* 1974; **8**(2):129–151.
15. Braess D. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 2001.
16. Fortin M. An analysis of the convergence of mixed finite element methods. *ESAIM: Mathematical Modelling and Numerical Analysis - Modlisation Mathmatique et Analyse Numrique* 1977; **11**(4):341–354.
17. Arnold D, Brezzi F, Fortin M. A stable finite element for the Stokes equations. *Calcolo* 1984; **21**(4):337–344.
18. Málek J, Strakoš Z. *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs*. Society for Industrial and Applied Mathematics: Philadelphia, PA, 2014.
19. Kuchta M, Mardal KA, Mortensen M. Characterisation of the space of rigid motions in arbitrary domains. *Proc. of 8th National Conference on Computational Mechanics*, CIMNE: Barcelona, Spain, 2015.
20. Gurtin M. *An Introduction to Continuum Mechanics*. Mathematics in Science and Engineering, Elsevier Science, 1982.
21. Bergh J, Löfström J. *Interpolation spaces: an introduction*. Grundlehren der mathematischen Wissenschaften, Springer, 1976.
22. Benzi M, Golub GH, Liesen J. Numerical solution of saddle point problems. *ACTA NUMERICA* 2005; **14**:1–137.
23. Paige CC, Saunders MA. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis* 1975; **12**(4):617–629.
24. Mardal KA, Haga JB. Block preconditioning of systems of PDEs. *Automated Solution of Differential Equations by the Finite Element Method*, Logg A, Mardal KA, Wells GNea (eds.). Springer, 2012.
25. Liesen J, Tichỳ P. Convergence analysis of Krylov subspace methods. *GAMM-Mitteilungen* 2004; **27**(2):153–173.
26. Trefethen LN, Bau D. *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, 1997.

27. Brenner S, Scott R. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics, Springer New York, 2007.
28. Penrose R. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* Oct 2008; **51**(3):406413.
29. Lanczos C. *Linear Differential Operators*. Dover books on mathematics, Dover Publications, 1997.
30. Shewchuk JR. An introduction to the conjugate gradient method without the agonizing pain 1994.
31. Mardal KA, Tai XC, Winther R. A robust finite element method for Darcy–Stokes flow. *SIAM Journal on Numerical Analysis* 2002; **40**(5):1605–1631.
32. Mardal KA, Winther R. An observation on Korn's inequality for nonconforming finite element methods. *Mathematics of computation* 2006; **75**(253):1–6.
33. Schöberl J. Multigrid methods for a parameter dependent problem in primal variables. *Numerische Mathematik* 1999; **84**(1):97–119.
34. Klawonn A. Block-triangular preconditioners for saddle point problems with a penalty term. *SIAM Journal on Scientific Computing* 1998; **19**(1):172–184.
35. Klawonn A. An optimal preconditioner for a class of saddle point problems with a penalty term. *SIAM Journal on Scientific Computing* 1998; **19**(2):540–552.