

Weighted BFBT Preconditioner for Stokes Flow Problems with Highly Heterogeneous Viscosity*

Johann Rudi[†]Georg Stadler[‡]Omar Ghattas^{†§}

January 31, 2017

Abstract

We present a weighted BFBT approximation (w-BFBT) to the inverse Schur complement of a Stokes system with highly heterogeneous viscosity. When used as part of a Schur complement-based Stokes preconditioner, we observe robust fast convergence for Stokes problems with smooth but highly varying (up to 10 orders of magnitude) viscosities, optimal algorithmic scalability with respect to mesh refinement, and only a mild dependence on the polynomial order of high-order finite element discretizations ($\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$, order $k \geq 2$). For certain difficult problems, we demonstrate numerically that w-BFBT significantly improves Stokes solver convergence over the widely used inverse viscosity-weighted pressure mass matrix approximation of the Schur complement. In addition, we derive theoretical eigenvalue bounds to prove spectral equivalence of w-BFBT. Using detailed numerical experiments, we discuss modifications to w-BFBT at Dirichlet boundaries that decrease the number of iterations. The overall algorithmic performance of the Stokes solver is governed by the efficacy of w-BFBT as a Schur complement approximation and, in addition, by our parallel hybrid spectral-geometric-algebraic multigrid (HMG) method, which we use to approximate the inverses of the viscous block and variable-coefficient pressure Poisson operators within w-BFBT. Building on the scalability of HMG, our Stokes solver achieves a parallel efficiency of 90% while weak scaling over a more than 600-fold increase from 48 to all 30,000 cores of TACC's Lonestar 5 supercomputer.

1 Introduction

1.1 Motivation and governing equations

Many problems in science and engineering involve creeping flows of non-Newtonian fluids [19, 35]. Important examples can be found in geophysical fluid flows, where the incompressible Stokes equations with power-law rheology have become a prototypical continuum mechanical description for creeping flows occurring in mantle convection [37], magma dynamics [32], and ice flow [22]. The linearization of the nonlinear momentum equations, for instance within a Newton method, leads to incompressible Stokes-like equations with highly heterogeneous viscosity fields.

In particular, simulations of Earth's mantle convection at global scale [39] exhibit extreme computational challenges due to a highly heterogeneous viscosity stemming from its dependence on temperature and strain rate as well as sharp viscosity gradients in narrow regions modeling tectonic plate boundaries (six orders of magnitude drop in ~ 5 km) [6, 36]. This leads to a wide range of spatial scales since small localized features at plate boundaries of size $\mathcal{O}(1$ km) influence plate motion at continental scales of $\mathcal{O}(1000$ km). The complex character of the flow presents severe computational challenges for iterative solvers due to poor conditioning of the linear systems that arise.

*This research was partially supported by NSF grants CMMI-1028889 and ARC-0941678 and U.S. Department of Energy grants DE-FC02-13ER26128 and the Scientific Discovery through Advanced Computing (SciDAC) program.

[†]Institute for Computational Engineering and Sciences, The University of Texas at Austin, USA (johann@ices.utexas.edu, omar@ices.utexas.edu).

[‡]Courant Institute of Mathematical Sciences, New York University, USA (stadler@cims.nyu.edu).

[§]Jackson School of Geosciences and Department of Mechanical Engineering, The University of Texas at Austin, USA.

Since we focus here on preconditioning linearized Stokes-like problems that arise at each step of a nonlinear solver, we can simplify our problem setup by taking the viscosity to be independent of the strain rate, but otherwise exhibiting severe spatial heterogeneity. Given a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, right-hand side forcing $\mathbf{f}(\mathbf{x})$, and spatially-varying viscosity $\mu(\mathbf{x}) \geq \mu_{\min} > 0$ for all $\mathbf{x} \in \Omega$, we consider the incompressible Stokes equations with homogeneous Dirichlet boundary conditions

$$-\nabla \cdot [\mu(\mathbf{x})(\nabla \mathbf{u} + \nabla \mathbf{u}^T)] + \nabla p = \mathbf{f} \quad \text{in } \Omega, \quad (1a)$$

$$-\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega, \quad (1b)$$

$$\mathbf{u} = 0 \quad \text{on } \partial\Omega, \quad (1c)$$

where \mathbf{u} and p are the unknown velocity and pressure fields, respectively.¹

1.2 Discretization and computational challenges

Discretizing (1) leads to a linear algebraic system of equations of the form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}, \quad (2)$$

where \mathbf{A} , \mathbf{B} , and \mathbf{B}^T are matrices corresponding to discretizations of the viscous stress, divergence, and gradient operators, respectively. The discretization is carried out by high-order finite elements on (possibly aggressively adaptively refined) hexahedral meshes with velocity–pressure pairings $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$ of polynomial order $k \geq 2$ with a continuous, nodal velocity approximation \mathbb{Q}_k and a discontinuous, modal pressure approximation $\mathbb{P}_{k-1}^{\text{disc}}$. These pairings yield optimal asymptotic convergence rates of the finite element approximation to the infinite-dimensional solution with decreasing mesh element size, are inf-sup stable on general, non-conforming hexahedral meshes with “hanging nodes,” and have the advantage of preserving mass locally at the element level due to the discontinuous pressure [16, 21, 40]. While these properties have been recognized to be important for geophysics applications (e.g., see [29, 30]), the high-order discretization, adaptivity, and discontinuous pressure approximation present significant additional difficulties for iterative solvers (relative to low order, uniform grid, continuous discretizations). Finally, a number of frontier geophysical problems, such as global mantle convection with plate boundary-resolving meshes and continental ice sheet models with grounding line-resolving meshes, result in billions of degrees of freedom, demanding efficient execution and scalability on leading edge supercomputers [6, 36, 39].

The applications we target (such as global mantle convection) exhibit all of the difficulties described above (severe heterogeneity, very large scale, need for aggressively-adapted meshes, need for high order, mass-conserving discretization) and thus demand robust and effective preconditioners for (2), resulting in iterative solvers with optimal (or nearly optimal) algorithmic and parallel scalability. This paper describes the design of such a preconditioner and its analysis and performance evaluation for problems with highly heterogeneous viscosity. The preconditioner—which we call *weighted BFBT* (w-BFBT)—is of Schur complement type, and we study its robustness as well as its algorithmic and parallel scalability.

1.3 Iterative methods and Schur complement approximations

An effective approximation of the Schur² complement $\mathbf{S} := \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ is an essential ingredient for attaining fast convergence of Schur complement-based iterative solvers for (2). More precisely, a sufficiently good approximation of the inverse Schur complement $\tilde{\mathbf{S}}^{-1} \approx \mathbf{S}^{-1}$ is sought, which, together with an approximation

¹Note that Newton linearization (as opposed to, e.g., Picard) results in a fourth-order anisotropic viscosity tensor, with the anisotropic component proportional to the tensor product of the second-order strain rate tensor with itself. This anisotropic viscosity tensor also appears in adjoint Stokes equations, which are important for inverse problems [23, 34, 43]. For power-law rheologies that are typical of many geophysical flows, the anisotropic component of the viscosity tensor is dominated by the isotropic component $\mu(\mathbf{x})\mathbf{I}$ [24, 36]. Thus for many forward and inverse modeling applications, the performance of our preconditioner for isotropic viscosities is indicative of its performance for operators with anisotropic viscosity tensors. Hence we focus on the former in this paper.

²Strictly speaking, our definition is the negative Schur complement. However, as in [16], we prefer to work with positive-definite operators and thus define the Schur complement to be positive rather than negative definite.

of the inverse viscous block, $\tilde{\mathbf{A}}^{-1} \approx \mathbf{A}^{-1}$, is used in an iterative scheme with right preconditioning based on an upper triangular block matrix:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\top \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{u}} \\ \tilde{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}. \quad (3)$$

Note that the original solution to (2) is recovered by applying the preconditioner once to the solution of (3). For the preconditioned Stokes system (3), we use GMRES as the Krylov subspace solver. This particular combination of Krylov method and preconditioner is known to converge in just two iterations for exact choices of $\tilde{\mathbf{A}}^{-1}$ and $\tilde{\mathbf{S}}^{-1}$ [4].

The most widely used approximation of the Schur complement (for variable viscosity Stokes systems) is the inverse viscosity-weighted mass matrix of the pressure space [8, 24, 27, 30], denoted by $\mathbf{M}_p(1/\mu)$, with entries $[\mathbf{M}_p(1/\mu)]_{i,j} = \int_{\Omega} q_i(\mathbf{x}) q_j(\mathbf{x}) / \mu(\mathbf{x}) d\mathbf{x}$, where $q_i, q_j \in \mathbb{P}_{k-1}^{\text{disc}}$ are global basis functions of the finite dimensional space $\mathbb{P}_{k-1}^{\text{disc}}$. Since the basis functions of $\mathbb{P}_{k-1}^{\text{disc}}$ are modal and not orthogonal to each other, the mass matrix is not diagonal, and thus $\mathbf{M}_p(1/\mu)$ is typically diagonalized to further simplify its inversion. One common way to obtain a diagonalized version is mass lumping. For nodal discretizations, the corresponding diagonal elements are computed by summation of the entries of each matrix row, i.e., $\mathbf{M}_p(1/\mu)\mathbf{1}$, where $\mathbf{1}$ is the vector with ones in all entries. For modal discretizations, we generalize the lumping procedure by using the coefficient vector, $\mathbf{1}_{\{q_i\}_i}$, representing the constant function having value 1 in the associated basis $\{q_i\}_i$, i.e.,

$$\tilde{\mathbf{M}}_p(1/\mu) := \text{diag}(\mathbf{M}_p(1/\mu)\mathbf{1}_{\{q_i\}_i}). \quad (4)$$

Provided that μ is sufficiently smooth, $\mathbf{M}_p(1/\mu)$ can be an effective approximation of \mathbf{S} in numerical experiments [7] and spectral equivalence can be shown [20]. However, it has been observed in applications with highly heterogeneous viscosities μ (e.g., mantle convection [31, 36]) that convergence slows down significantly due to a poor Schur complement approximation by $\mathbf{M}_p(1/\mu)$. Therefore, we propose a new approximation, w-BFBT, that remains robust when $\mathbf{M}_p(1/\mu)$ fails.

Preconditioners based on BFBT approximations for the Schur complement were initially proposed in [13] for the Navier–Stokes equations. Over the years, these ideas were refined and extended [14, 15, 17, 26, 38] to arrive at a class of closely related Schur complement approximations: Pressure Convection–Diffusion, BFBT, and Least Squares Commutator. The underlying principle, now in a Stokes setting, is that one seeks a commutator matrix \mathbf{X} such that the following commutator nearly vanishes,

$$\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X} \approx \mathbf{0}, \quad (5)$$

for a given diagonal matrix \mathbf{D}^{-1} . The Navier–Stokes case differs from Stokes in that the viscous stress matrix \mathbf{A} contains an additional convection term. The motivation for seeking a near-commutator \mathbf{X} is that (5) can be rearranged by multiplying (5) with $\mathbf{B}\mathbf{A}^{-1}$ from the left and, provided the inverse exists, with \mathbf{X}^{-1} from the right to obtain $\mathbf{S} \approx \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top\mathbf{X}^{-1}$, where the closer the commutator is to zero, the more accurate the approximation [16]. The goal of finding a vanishing commutator can be recast as solving the following least-squares minimization problem:

$$\text{Find matrix } \mathbf{X} \text{ minimizing } \|\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top\mathbf{e}_j - \mathbf{B}^\top\mathbf{X}\mathbf{e}_j\|_{\mathbf{C}^{-1}}^2 \quad \text{for all } j, \quad (6)$$

where \mathbf{e}_j is the j -th Cartesian unit vector and the norm arises from a symmetric and positive definite matrix \mathbf{C} . The solution is given by $\mathbf{X} = (\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top)$. Then the BFBT approximation of the inverse Schur complement is derived by algebraic rearrangement of the commutator (5):

$$\tilde{\mathbf{S}}_{\text{BFBT}}^{-1} := (\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top)^{-1} (\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top) (\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}. \quad (7)$$

In the literature cited above (which addresses preconditioning of Navier–Stokes equations with constant viscosity), the diagonal weighting matrices are chosen as $\mathbf{C} = \mathbf{D} = \tilde{\mathbf{M}}_u$, i.e., a diagonalized version of the velocity space mass matrix; hence we call this the \mathbf{M}_u -BFBT approximation of the Schur complement. \mathbf{M}_u -BFBT can be used for Stokes problems with constant viscosities providing convergence similar to $\mathbf{M}_p(1/\mu)$. However, the computational cost of applying (7) is significantly higher than the (cheap) application of a possibly diagonalized inverse of $\mathbf{M}_p(1/\mu)$. Moreover, \mathbf{M}_u -BFBT is not an option for heterogeneous

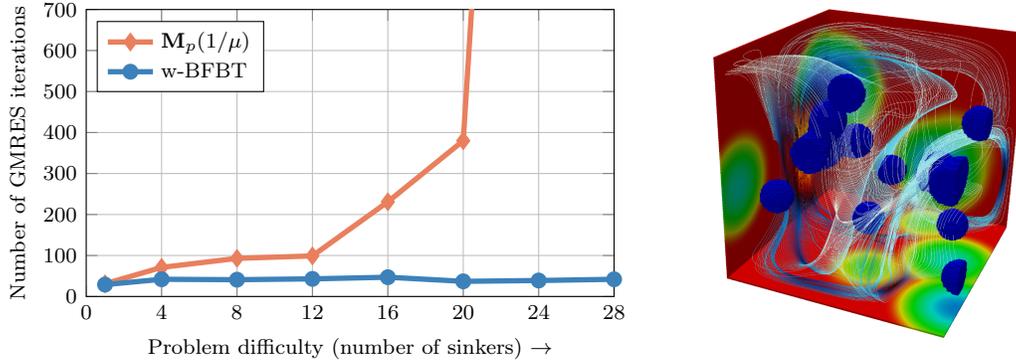


Figure 1: *Left image* shows the improvement in convergence obtained with the proposed w-BFBT preconditioner over a preconditioner using the inverse viscosity-weighted pressure mass matrix as Schur complement approximation. The number of randomly placed sinks (high viscosity inclusions in low viscosity medium) increases along the horizontal axis. The vertical axis depicts the number of GMRES iterations required for 10^6 residual reduction for the most popular $\mathbf{M}_p(1/\mu)$ and the proposed w-BFBT preconditioner. Fixed problem parameters are the dynamic ratio $\text{DR}(\mu) = \max(\mu)/\min(\mu) = 10^8$, discretization order $k = 2$, and the mesh refinement level $\ell = 7$, resulting in 128^3 finite elements. *Right image* shows an example viscosity field with 16 sinks (*blue spheres* depict highly viscous centers of Gaussian-like sinks, low viscosity medium in *red color*) and the *streamlines* of the computed velocity field.

viscosities because convergence becomes extremely slow or stagnates, as observed in [31]. Instead, in [31] for finite element and in [18] for staggered grid finite difference discretizations, a re-scaling of the discrete Stokes system (1) was performed, which essentially alters the diagonal weighting matrices \mathbf{C} , \mathbf{D} . By choosing entries from \mathbf{A} for these weighting matrices, it was possible to demonstrate improved convergence with BFBT compared to $\mathbf{M}_p(1/\mu)$ for certain benchmark problems with strong viscosity variations. Building on ideas from [31], [36] chose the weighting matrices such that $\mathbf{C} = \mathbf{D} = \text{diag}(\mathbf{A})$, which led to superior performance compared to $\mathbf{M}_p(1/\mu)$ for highly heterogeneous mantle convection problems. Hence we refer to this approach as $\text{diag}(\mathbf{A})$ -BFBT.

However, even $\text{diag}(\mathbf{A})$ -BFBT can fail to achieve fast convergence for some problems/discretizations, as shown below (Section 2.2). Moreover, choosing the weighting matrices as $\text{diag}(\mathbf{A})$ is problematic for high-order discretizations, where $\text{diag}(\mathbf{A})$ becomes a poor approximation of \mathbf{A} . These drawbacks lead us to propose the following w-BFBT approximation for the inverse Schur complement:

$$\tilde{\mathbf{S}}_{\text{w-BFBT}}^{-1} := (\mathbf{B}\mathbf{C}_{w_l}^{-1}\mathbf{B}^\top)^{-1} (\mathbf{B}\mathbf{C}_{w_l}^{-1}\mathbf{A}\mathbf{D}_{w_r}^{-1}\mathbf{B}^\top) (\mathbf{B}\mathbf{D}_{w_r}^{-1}\mathbf{B}^\top)^{-1}, \quad (8)$$

where $\mathbf{C}_{w_l} = \tilde{\mathbf{M}}_{\mathbf{u}}(w_l)$ and $\mathbf{D}_{w_r} = \tilde{\mathbf{M}}_{\mathbf{u}}(w_r)$ are lumped velocity space mass matrices (lumping analogously to (4)) that are weighted by the square root of the viscosity, $w_l(\mathbf{x}) = \sqrt{\mu(\mathbf{x})} = w_r(\mathbf{x})$, $\mathbf{x} \in \Omega$.

1.4 Outline and summary of key results

After defining a class of benchmark problems (Section 2.1), we compare the convergence obtained with different Schur complement approximations to motivate preconditioning with w-BFBT (Section 2.2). Theoretical estimates for spectral equivalence of w-BFBT are derived in Section 3. This is followed by a detailed numerical study showing when w-BFBT is advantageous over $\mathbf{M}_p(1/\mu)$ (Section 4), and a discussion of boundary modifications for w-BFBT that accelerate convergence (Section 5). In Section 6 we describe an algorithm for w-BFBT-based Stokes preconditioning, which uses hybrid spectral-geometric-algebraic multigrid (HMG). Finally, in Section 7, we provide numerical evidence for near-optimal algorithmic and parallel scalability. In particular, we demonstrate that the preconditioner's parallel efficiency remains high when weak scaling out to tens of thousands of threads and even millions of threads.³

³See [36], which demonstrated parallel scalability for a BFBT-type method. w-BFBT deviates from [36] because of a different choice of diagonal weighting matrices in (7). However, the work per application of both preconditioners is the same, resulting in comparable parallel scalability.

To motivate our study of w-BFBT, we give an example for a possible improvement in convergence in Figure 1. There, a comparison is drawn between the $\mathbf{M}_p(1/\mu)$ and w-BFBT approximations for the Schur complement. The Stokes problem that is being solved is the multi-sinker benchmark problem from Section 2.1. The difficulty of the problem can be increased by adding more and more high-viscosity inclusions, called sinkers, into a low-viscosity background medium, which, as a result, introduces more variation in the viscosity. As can be seen in the figure, the number of GMRES iterations remain flat when preconditioning with w-BFBT, whereas the number of GMRES iterations increases significantly for higher sinker counts when using $\mathbf{M}_p(1/\mu)$, rendering $\mathbf{M}_p(1/\mu)$ inefficient for these types of difficult problems. Therefore we propose w-BFBT as an alternative Schur complement approximation for Stokes flow problems with a highly varying viscosity.

2 Benchmark problem and comparison of Schur complement approximations

This section further motivates the need for more effective Schur complement preconditioners. We first present a class of benchmark problems that range from relatively mild viscosity variations to highly heterogeneous. Then a challenging problem is used to compare Stokes solver convergence with different Schur complement approximations to demonstrate the limitations of established methods and motivate the development of w-BFBT.

2.1 Multi-sinker benchmark problem

The design of suitable benchmark problems is critical to conduct studies that can give useful convergence estimates for challenging applications. We seek complex geometrical structures in the viscosity that generate irregular, nonlocal, multiscale flow fields. Additionally, the viscosity should exhibit sharp gradients and its dynamic ratio $\text{DR}(\mu) := \max(\mu)/\min(\mu)$ (also commonly referred to as viscosity contrast) can be six orders of magnitude or higher in demanding applications. As in [29], we use a multi-sinker test problem with randomly positioned inclusions (e.g., as in Figure 1, *right image*) to study solver performance. We find that the arising viscosity structure is a suitable test for challenging, highly heterogeneous coefficient Stokes problems, and that the solver performance observed for such models can be indicative of the performance for other challenging applications.

In the (open) unit cube domain $\Omega = (0, 1)^3$, we define the viscosity coefficient $\mu(\mathbf{x}) \in [\mu_{\min}, \mu_{\max}]$, $\mathbf{x} \in \Omega$, $0 < \mu_{\min} < \mu_{\max} < \infty$, with dynamic ratio $\text{DR}(\mu) = \mu_{\max}/\mu_{\min}$ by means of rescaling a C^∞ indicator function $\chi_n(\mathbf{x}) \in [0, 1]$ that accumulates n sinkers via a product of modified Gaussian functions:

$$\begin{aligned} \mu(\mathbf{x}) &:= (\mu_{\max} - \mu_{\min})(1 - \chi_n(\mathbf{x})) + \mu_{\min}, \quad \mathbf{x} \in \Omega, \\ \chi_n(\mathbf{x}) &:= \prod_{i=1}^n 1 - \exp\left(-\delta \max\left(0, |\mathbf{c}_i - \mathbf{x}| - \frac{\omega}{2}\right)^2\right), \quad \mathbf{x} \in \Omega, \end{aligned}$$

where $\mathbf{c}_i \in \Omega$, $i = 1, \dots, n$, are the centers of the sinkers, $\delta > 0$ controls the exponential decay of the Gaussian smoothing, and $\omega \geq 0$ is the diameter of a sinker where μ_{\max} is attained. Since all sinkers are equal in size, inserting more of them inside the domain will eventually result in overlapping with each other and possible intersections with the domain's boundary. Throughout the paper, we fix $\delta = 200$, $\omega = 0.1$, and use the same set of precomputed random points \mathbf{c}_i in all numerical experiments. Two parameters are varied: (i) the number of sinkers n at random positions (the label S_n -rand indicates a multi-sinker problem with n randomly positioned sinkers) and (ii) the dynamic ratio $\text{DR}(\mu)$ which in turn determines $\mu_{\min} := \text{DR}(\mu)^{-1/2}$ and $\mu_{\max} := \text{DR}(\mu)^{1/2}$. The right-hand side of (2), $\mathbf{f}(\mathbf{x}) := (0, 0, \beta(\chi_n(\mathbf{x}) - 1))$, $\beta = 10$ constant, is such that it forces the high-viscosity sinkers downward, similarly to a gravity that pulls on high-density inclusions within a medium of lower density.

2.2 Comparison of Schur complement approximations

We compare convergence of the Stokes solver using the Schur complement approximation $\mathbf{M}_p(1/\mu)$ with $\text{diag}(\mathbf{A})$ -BFBT and with the proposed w-BFBT. The problem parameters are held fixed to S16-rand and

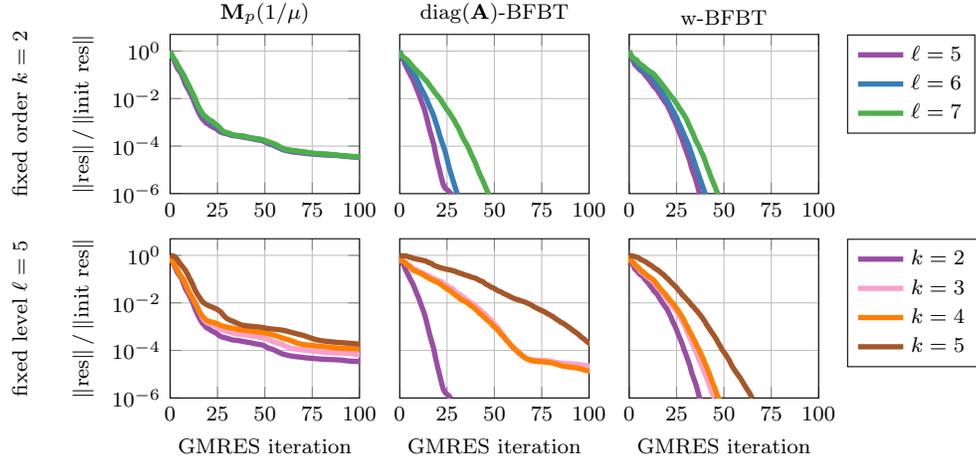


Figure 2: Comparison of Stokes solver convergence with $\mathbf{M}_p(1/\mu)$ (left column), $\text{diag}(\mathbf{A})$ -BFBT (middle column), and w-BFBT (right column) preconditioning. We fix the problem S16-rand, $\text{DR}(\mu) = 10^8$ while varying mesh refinement level ℓ (top row) and discretization order k (bottom row). This comparison shows that w-BFBT combines robust convergence of $\text{diag}(\mathbf{A})$ -BFBT with improved algorithmic scalability when k is increased.

$\text{DR}(\mu) = 10^8$. The numerical experiments are carried out using different levels of mesh refinement $\ell = 5, \dots, 7$ (for fixed order $k = 2$) and different discretization orders $k = 2, \dots, 5$ (for fixed level $\ell = 5$). A level ℓ corresponds to a mesh of $2^{3\ell}$ elements due to uniform refinement. Note that for these tests, the applications of \mathbf{A}^{-1} , $(\mathbf{B}\mathbf{C}_{w_l}^{-1}\mathbf{B}^T)^{-1}$, and $(\mathbf{B}\mathbf{D}_{w_r}^{-1}\mathbf{B}^T)^{-1}$ are approximated using a multigrid method (introduced in Section 6). These approximations are sufficiently accurate, such that the comparison is indicative of the effectiveness of the different Schur complement approximations. In particular, improving the approximation does not change the results, which are presented in Figure 2. In the left two plots, the poor Schur complement approximation by $\mathbf{M}_p(1/\mu)$ for this problem setup can be observed clearly. Convergence stagnates (similar results are found in [29, 36]).

Preconditioner $\text{diag}(\mathbf{A})$ -BFBT (Figure 2, middle) is able to achieve fast convergence for discretization order $k = 2$. A limitation of $\text{diag}(\mathbf{A})$ -BFBT is a strong dependence on the order k . This can be explained by the decreasing diagonal dominance in the viscous block \mathbf{A} with increasing order k : for higher k the approximation of \mathbf{A} by $\text{diag}(\mathbf{A})$ deteriorates. Note that numerical experiments with \mathbf{M}_u -BFBT are not presented, because it performs poorly in the presence of spatially-varying viscosities. This leads to the conclusion that the choice of the weighting matrices \mathbf{C} , \mathbf{D} in $\tilde{\mathbf{S}}_{\text{BFBT}}^{-1}$ crucially affects the quality of the Schur complement approximation.

The w-BFBT approximation delivers convergence that is nearly as fast as in the $\text{diag}(\mathbf{A})$ -BFBT, $k = 2$ case, but without the severe deterioration when k is increased (see Figure 2, right). Thus, w-BFBT exhibits the robustness of $\text{diag}(\mathbf{A})$ -BFBT and additionally shows superior algorithmic scalability with respect to k . Having illustrated the efficacy of w-BFBT for certain problem parameters, we next establish spectral equivalence of w-BFBT (Section 3) and then analyze in more detail how crucial parameters influence convergence in Section 4.

3 Spectral equivalence of w-BFBT

Before we show spectral equivalence, we introduce notation and basic definitions.

3.1 Basic definitions

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$. We denote as $L^2(\Omega)$ the class of real-valued square-integrable functions, equipped with the usual L^2 -inner product $(u, v)_{L^2(\Omega)}$ and induced norm

$\|u\|_{L^2(\Omega)}$, $u, v \in L^2(\Omega)$. We also consider corresponding spaces of d -dimensional vector-valued functions and $(d \times d)$ -dimensional tensor-valued functions with component-wise multiplication, denoted by $L^2(\Omega)^d$ and $L^2(\Omega)^{d \times d}$. The subspace of $L^2(\Omega)$ that does not contain constant functions is denoted by $L^2(\Omega)/\mathbb{R}$. A bounded function, say $\alpha = \alpha(\mathbf{x})$, belongs to the space $L^\infty(\Omega)$ by satisfying the following finite norm: $\|\alpha\|_{L^\infty(\Omega)} := \text{ess sup}_{\mathbf{x} \in \Omega} |\alpha(\mathbf{x})| < \infty$. We generalize the L^2 -norms to classes of weighted L^2_α -norms for functions $f \in L^2(\Omega)^n$, $n \in \{1, d, d \times d\}$, defined by

$$\|f\|_{L^2_\alpha(\Omega)^n} := \left\| \alpha^{\frac{1}{2}} f \right\|_{L^2(\Omega)^n} \quad \text{for } \alpha \in L^\infty(\Omega), 0 < \alpha(\mathbf{x}) \text{ a.e. in } \Omega.$$

Next, we introduce $H^m(\Omega)$, with $m \geq 0$, which is the Sobolev space of m derivatives in $L^2(\Omega)$, and for $m = 1$ we use the inner product $(u, v)_{H^1(\Omega)} := (u, v)_{L^2(\Omega)} + (\nabla u, \nabla v)_{L^2(\Omega)}$, inducing the norm $\|u\|_{H^1(\Omega)}$. Functions in $H^m(\Omega)$ with vanishing trace on the boundary $\partial\Omega$ belong to the space $H_0^m(\Omega)$. Finally, we say that a function belongs to the class of $C^\infty(\Omega)$ if it has partial derivatives of any order in Ω , and these derivatives are continuous.

We transition from abstract definitions to fluid mechanics. The differential operators acting on velocity $\mathbf{u} \in (H^1(\Omega))^d$ and pressure $p \in L^2(\Omega)$ within the Stokes equations are defined in the sense of distributions:

$$\nabla_s \mathbf{u} := \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^\top), \quad B\mathbf{u} := -\nabla \cdot \mathbf{u}, \quad B^*p := \nabla p.$$

Moreover, assume a sufficiently regular, bounded viscosity $\mu \in H^1(\Omega) \cap L^\infty(\Omega)$ such that $0 < \mu_{\min} \leq \mu(\mathbf{x})$ a.e. in Ω and then define the viscous stress tensor $\boldsymbol{\tau} := 2\mu \nabla_s \mathbf{u}$. We denote the function space for velocity by

$$V := \{ \mathbf{u} \in (H^1(\Omega))^d \mid \mathbf{n} \cdot \mathbf{u} = 0 \text{ on } \partial\Omega \}, \quad (9)$$

where $\mathbf{n} \in \mathbb{R}^d$ is the outward unit normal vector at the boundary $\partial\Omega$, and the function space for pressure by $Q := L^2(\Omega)/\mathbb{R}$, and we introduce the viscous stress operator with a heterogeneous viscosity

$$A_\mu : V \rightarrow V', \quad A_\mu \mathbf{u} := -\nabla \cdot (2\mu \nabla_s \mathbf{u}) = -\nabla \cdot \boldsymbol{\tau}.$$

Given exterior forces acting on the fluid $\mathbf{f} \in V'$, we consider the incompressible Stokes problem with free-slip and no-normal flow boundary conditions:

$$A_\mu \mathbf{u} + B^*p = \mathbf{f} \quad \text{in } \Omega, \quad (10a)$$

$$B\mathbf{u} = 0 \quad \text{in } \Omega, \quad (10b)$$

$$\mathbf{T}[\boldsymbol{\tau} - p\mathbf{I}]\mathbf{n} = 0 \quad \text{on } \partial\Omega, \quad (10c)$$

$$\mathbf{u} \cdot \mathbf{n} = 0 \quad \text{on } \partial\Omega, \quad (10d)$$

in which we seek the velocity $\mathbf{u} \in V$ and pressure $p \in Q$. On the boundary, we have outward unit normal vectors $\mathbf{n} \in \mathbb{R}^d$ and tangential projectors $\mathbf{T} := \mathbf{I} - \mathbf{n}\mathbf{n}^\top$. For the theoretical analysis sections, we altered the boundary conditions compared to (1c) to achieve a cleaner presentation.

For the definition of the w-BFBT approximation of the Schur complement, we introduce a Poisson operator for higher regularity pressure functions

$$K_w^* : H^2(\Omega) \rightarrow L^2(\Omega), \quad K_w^*p := BwB^*p, \quad (11)$$

with an appropriate coefficient w (see below) and augmented with homogeneous Neumann boundary conditions, $\mathbf{n} \cdot B^*p = 0$. The L^2 -adjoint of K_w^* is denoted by K_w . Finally, we define the w-BFBT approximation of the Schur complement $S = BA_\mu^{-1}B^*$ by:

$$\tilde{S}_{\text{w-BFBT}} := K_{w_r}^* (Bw_l A_\mu w_r B^*)^{-1} K_{w_l}, \quad (12)$$

with sufficiently regular, bounded weight functions $w_l, w_r \in H^1(\Omega) \cap L^\infty(\Omega)$ such that $0 < w_{\min} \leq w_l(\mathbf{x}), w_r(\mathbf{x})$ a.e. in Ω . Note that the definitions of the w-BFBT weight functions in the discrete case (8) are reciprocal to definitions in (11) and (12), because in the discrete case the weight functions were embedded into inverses of mass matrices.

3.2 Main theorem on spectral equivalence of w-BFBT

One measure for the efficacy of a preconditioner consists of the ratio of the maximal to minimal eigenvalues of the preconditioned system $\tilde{\mathbf{S}}_{\text{w-BFBT}}^{-1}\mathbf{S}$. This section establishes inequalities for spectral equivalence of w-BFBT by providing bounds on that ratio. The derivations are carried out in an infinite-dimensional setting. We begin by stating the main result of this section in Theorem 3.1 and continue with proving this result using a sequence of lemmas.

Theorem 3.1 (Main result). *Let $\hat{Q} = L^2(\Omega)/\mathbb{R} \cap H^1(\Omega)$. If the left and right w-BFBT weight functions are equal to*

$$w_l(\mathbf{x}) = \mu(\mathbf{x})^{-\frac{1}{2}} = w_r(\mathbf{x}) \quad \text{for a.a. } \mathbf{x} \in \Omega,$$

then the exact Schur complement is equivalent to the w-BFBT approximation such that

$$\left(\tilde{\mathcal{S}}_{\text{w-BFBT}} q, q\right) \leq (Sq, q) \leq C_{\text{w-BFBT}} \left(\tilde{\mathcal{S}}_{\text{w-BFBT}} q, q\right) \quad \text{for all } q \in \hat{Q},$$

where

$$C_{\text{w-BFBT}} := \left(1 + \frac{1}{4} \|\nabla \mu\|_{L^\infty(\Omega)^d}^2\right) \left(C_{P,\mu}^2 + 1\right) C_{K,\mu}^2$$

and the constants $C_{P,\mu}, C_{K,\mu} > 0$ stem from weighted Poincaré–Friedrichs’ and Korn’s inequalities, respectively (see Remark 3.7 for more information); the viscosity μ assumes the role of the weight function in the weighted inequalities.

If the viscosity and the w-BFBT weight functions are constant,

$$\mu \equiv 1, \quad w_l \equiv 1 \equiv w_r,$$

then the exact Schur complement is equivalent to the w-BFBT approximation such that

$$\left(\tilde{\mathcal{S}}_{\text{w-BFBT}} q, q\right) \leq (Sq, q) \leq (C_P^2 + 1) C_K^2 \left(\tilde{\mathcal{S}}_{\text{w-BFBT}} q, q\right) \quad \text{for all } q \in \hat{Q}$$

with constants $C_P, C_K > 0$ stemming from (the classical) Poincaré–Friedrichs’ and Korn’s inequalities, respectively.

3.3 Proofs

The proof of Theorem 3.1 is established in the remainder of this section. In what follows, suprema are understood over spaces excluding operator kernels that would cause a supremum to blow up. The following basic, but hereafter frequently used, result is shown for completeness of the discussion.

Lemma 3.2 (sup-form of inverse operator). *Let V be a complete Hilbert space and $W \subseteq V$ be a dense subspace. Assume the linear operator $T : V \rightarrow V'$ to be bounded, invertible, symmetric, and positive definite. Then for any $f \in V'$ follows*

$$(T^{-1}f, f) = \sup_{w \in W} \frac{(w, f)^2}{(Tw, w)}.$$

Proof. Let $w \in W$, then with Hölder’s inequality follows

$$(w, f)^2 = \left(T^{1/2}w, T^{-1/2}f\right)^2 \leq \left\|T^{1/2}w\right\|^2 \left\|T^{-1/2}f\right\|^2 = (Tw, w) (T^{-1}f, f).$$

Additionally, let $v = T^{-1}f$ and since $W \subseteq V$ dense, there exists a sequence $\{w_k\}_k \subset W$ such that $w_k \rightarrow v = T^{-1}f \in V$, hence

$$\frac{(w_k, f)^2}{(Tw_k, w_k)} \rightarrow \frac{(v, f)^2}{(Tv, v)} = \frac{(T^{-1}f, f)^2}{(f, T^{-1}f)} = (T^{-1}f, f),$$

which shows that equality is achieved in the limit. □

The next lemma establishes Schur complement properties that are essential for deriving lower and upper bounds in the spectral equivalence estimates.

Lemma 3.3 (sup-form of Schur complement). *With the definitions from Section 3.1, the following two equalities hold:*

$$\left(\tilde{S}_{\text{w-BFBT}} q, q\right) = \sup_{p \in \hat{P}} \frac{(B^*p, w_r B^*q)^2}{(w_l A_\mu w_r B^*p, B^*p)} \quad \text{for all } q \in \hat{Q}, \quad (13)$$

where $\hat{P} := C^\infty(\Omega)$ and $\hat{Q} := L^2(\Omega)/\mathbb{R} \cap H^1(\Omega)$, and

$$(Sq, q) = \sup_{\mathbf{v} \in V} \frac{(\mathbf{v}, w_r B^*q)^2}{(w_l A_\mu w_r \mathbf{v}, \mathbf{v})} \quad \text{for all } q \in Q. \quad (14)$$

Proof. For $q \in \hat{Q}$, we use integration by parts on the left hand side of (13) to obtain

$$\begin{aligned} \left(\tilde{S}_{\text{w-BFBT}} q, q\right) &= \int_{\Omega} (w_r B^*(B w_l A_\mu w_r B^*)^{-1} K_{w_l} q) (B^*q) \, d\mathbf{x} + b_1(q) \\ &= \int_{\Omega} ((B w_l A_\mu w_r B^*)^{-1} K_{w_l} q) (K_{w_r} q) \, d\mathbf{x} + b_1(q) + b_2(q) \end{aligned}$$

with boundary terms

$$\begin{aligned} b_1(q) &:= - \int_{\partial\Omega} (\mathbf{n} \cdot w_r B^*(B w_l A_\mu w_r B^*)^{-1} K_{w_l} q) q \, d\mathbf{x}, \\ b_2(q) &:= \int_{\partial\Omega} ((B w_l A_\mu w_r B^*)^{-1} K_{w_l} q) (\mathbf{n} \cdot w_r B^*q) \, d\mathbf{x}. \end{aligned}$$

Using that $\hat{P} \subset H^2(\Omega)$ is dense, application of Lemma 3.2 and further integration by parts yields

$$\begin{aligned} \left(\tilde{S}_{\text{w-BFBT}} q, q\right) &= \sup_{p \in \hat{P}} \frac{(p, K_{w_r} q)^2}{(B w_l A_\mu w_r B^*p, p)} + b_1(q) + b_2(q) \\ &= \sup_{p \in \hat{P}} \frac{((B^*p, w_r B^*q) + b_3(p, q))^2}{(w_l A_\mu w_r B^*p, B^*p) + b_4(p)} + b_1(q) + b_2(q) \end{aligned}$$

with boundary terms

$$\begin{aligned} b_3(p, q) &:= - \int_{\partial\Omega} p (\mathbf{n} \cdot w_r B^*q) \, d\mathbf{x}, \\ b_4(p) &:= - \int_{\partial\Omega} (\mathbf{n} \cdot w_l A_\mu w_r B^*p) p \, d\mathbf{x}. \end{aligned}$$

Because the operator $K_{w_r}^*$ from (11) is augmented with homogeneous Neumann boundary conditions, $\mathbf{n} \cdot B^*p = 0$, the boundary terms $b_1(q)$, $b_2(q)$, and $b_3(p, q)$ vanish. In addition, $p \in \hat{P}$ is sufficiently regular for the term $b_4(p)$ to be well-defined and it equals to zero because the velocity \mathbf{u} satisfies $\mathbf{n} \cdot \mathbf{u} = 0$ on $\partial\Omega$. Hence, (13) follows.

To show (14), let $q \in Q$, then for the exact Schur complement we apply integration by parts with a vanishing boundary term

$$(Sq, q) = (BA_\mu^{-1}B^*q, q) = (A_\mu^{-1}B^*q, B^*q) = (w_r^{-1}A_\mu^{-1}w_l^{-1}w_l B^*q, w_r B^*q)$$

and (14) follows from Lemma 3.2. \square

A direct consequence of Lemma 3.3 is the following lower bound.

Corollary 3.4 (Lower bound, $\tilde{S}_{w\text{-BFBT}} \lesssim S$). *The exact Schur complement is bounded by the w-BFBT approximation from below, i.e.,*

$$\left(\tilde{S}_{w\text{-BFBT}} q, q\right) \leq (Sq, q) \quad \text{for all } q \in \hat{Q}, \quad (15)$$

where $\hat{Q} := L^2(\Omega)/\mathbb{R} \cap H^1(\Omega)$.

Proof. Let $\hat{P} = C^\infty(\Omega)$ and $q \in \hat{Q}$. Since B^* maps \hat{P} into V , we combine (13) and (14) to get

$$\left(\tilde{S}_{w\text{-BFBT}} q, q\right) = \sup_{p \in \hat{P}} \frac{(B^*p, w_r B^*q)^2}{(w_l A_\mu w_r B^*p, B^*p)} \leq \sup_{\mathbf{v} \in V} \frac{(\mathbf{v}, w_r B^*q)^2}{(w_l A_\mu w_r \mathbf{v}, \mathbf{v})} = (Sq, q),$$

and obtain the result (15). \square

We begin the derivation of an upper bound for the case of constant viscosity $\mu \equiv 1$. Note that $\tilde{S}_{w\text{-BFBT}}$ is scaling invariant with respect to constants multiplied to the w-BFBT weight functions w_l, w_r . Hence, it always assumes the correct scaling of S independent of the viscosity constant. The result for constant viscosity presented below in Lemma 3.5 is generalized to variable viscosity in Lemma 3.6. While Lemma 3.5 is a special case of Lemma 3.6, we first prove the result for constant viscosity as the arguments are less technical and easier to follow. In the proof of the result for variable viscosity, we build on some of the arguments from the constant viscosity case and thus avoid unnecessary duplication.

Lemma 3.5 (Upper bound, $S \lesssim \tilde{S}_{w\text{-BFBT}}$, for constant μ). *Assume a constant viscosity $\mu \equiv 1$ and constant w-BFBT weight functions $w_l \equiv 1 \equiv w_r$, and, as before in Lemma 3.3, let $\hat{Q} = L^2(\Omega)/\mathbb{R} \cap H^1(\Omega)$. Then the exact Schur complement is bounded by the w-BFBT approximation from above by*

$$(Sq, q) \leq (C_P^2 + 1) C_K^2 \left(\tilde{S}_{w\text{-BFBT}} q, q\right) \quad \text{for all } q \in \hat{Q} \quad (16)$$

with constants $C_P, C_K > 0$ stemming from Poincaré–Friedrichs’ and Korn’s inequalities, respectively.

Proof. Let $\hat{P} = C^\infty(\Omega)$ and $q \in \hat{Q}$, then due to (13) we can write

$$\left(\tilde{S}_{w\text{-BFBT}} q, q\right) = \sup_{p \in \hat{P}} \frac{(B^*p, B^*q)^2}{\|B^*p\|_{(H^1(\Omega))^d}^2} \frac{\|B^*p\|_{(H^1(\Omega))^d}^2}{(A_1 B^*p, B^*p)}. \quad (17)$$

To estimate the second factor on the right-hand side of (17), note that

$$(A_1 B^*p, B^*p) = 2(\nabla_s B^*p, \nabla_s B^*p) = 2(\nabla B^*p, \nabla B^*p) = 2\|\nabla B^*p\|_{(L^2(\Omega))^{d \times d}}^2,$$

where we used that $\mathbf{n} \cdot B^*p = 0$ on the boundary and that ∇B^*p is symmetric. Thus,

$$(A_1 B^*p, B^*p) \leq 2\|B^*p\|_{(H^1(\Omega))^d}^2. \quad (18)$$

For the first factor on the right-hand side of (17), observe that for any $\mathbf{v} \in V$ there exists a sequence $\{p_i\}_i \subset \hat{P} \subset H^2(\Omega)$ such that $K_1^* p_i = B B^* p_i \rightarrow B\mathbf{v}$, since K_1^* is invertible, where convergence is with respect to the L^2 -norm. Thus,

$$\sup_{p \in \hat{P}} \frac{(B^*p, B^*q)}{\|B^*p\|_{(H^1(\Omega))^d}} = \sup_{\mathbf{v} \in V} \frac{(\mathbf{v}, B^*q)}{\|\mathbf{v}\|_{(H^1(\Omega))^d}} = \|B^*q\|_{(H^{-1}(\Omega))^d}. \quad (19)$$

Combining (18) and (19) provides the following estimate for the w-BFBT Schur complement approximation (17):

$$\left(\tilde{S}_{w\text{-BFBT}} q, q\right) \geq \frac{1}{2} \|B^*q\|_{(H^{-1}(\Omega))^d}^2. \quad (20)$$

The exact Schur complement, on the other hand, in the form (14) from Lemma 3.3, can be bounded by

$$(Sq, q) = \sup_{\mathbf{v} \in V} \frac{(\mathbf{v}, B^*q)^2}{(A_1 \mathbf{v}, \mathbf{v})} \leq \sup_{\mathbf{v} \in V} \frac{\|\mathbf{v}\|_{(H^1(\Omega))^d}^2 \|B^*q\|_{(H^{-1}(\Omega))^d}^2}{2\|\nabla_s \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2}. \quad (21)$$

With Poincaré–Friedrichs’ inequality,

$$\|\mathbf{v}\|_{L^2(\Omega)^d} \leq C_P \|\nabla \mathbf{v}\|_{(L^2(\Omega))^{d \times d}},$$

where the constant $C_P > 0$ depends on the domain Ω , and Korn’s inequality,

$$\|\nabla \mathbf{v}\|_{(L^2(\Omega))^{d \times d}} \leq C_K \|\nabla_s \mathbf{v}\|_{(L^2(\Omega))^{d \times d}},$$

with constant $C_K > 0$, we obtain

$$\frac{1}{(C_P^2 + 1)} \|\mathbf{v}\|_{(H^1(\Omega))^d}^2 \leq \|\nabla \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2 \leq C_K^2 \|\nabla_s \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2,$$

and substituting this into (21) gives

$$(Sq, q) \leq \frac{(C_P^2 + 1) C_K^2}{2} \|B^* q\|_{(H^{-1}(\Omega))^d}^2. \quad (22)$$

Together with (20), this yields the desired result (16). \square

We complete the presentation of spectral equivalence by deriving an upper bound for problems with variable viscosities.

Lemma 3.6 (Upper bound, $S \lesssim \tilde{S}_{w\text{-BFBT}}$, for variable μ). *As before in Lemma 3.3, let $\hat{Q} = L^2(\Omega)/\mathbb{R} \cap H^1(\Omega)$. If the left and right w-BFBT weight functions are equal to*

$$w_l(\mathbf{x}) = \mu(\mathbf{x})^{-\frac{1}{2}} = w_r(\mathbf{x}) \quad \text{for a.a. } \mathbf{x} \in \Omega, \quad (23)$$

then the exact Schur complement is bounded by the w-BFBT approximation from above by

$$(Sq, q) \leq \left(1 + \frac{1}{4} \|\nabla \mu\|_{L^\infty(\Omega)^d}^2\right) (C_{P,\mu}^2 + 1) C_{K,\mu}^2 \left(\tilde{S}_{w\text{-BFBT}} q, q\right) \quad \text{for all } q \in \hat{Q} \quad (24)$$

with constants $C_{P,\mu}, C_{K,\mu} > 0$ stemming from weighted Poincaré–Friedrichs’ and Korn’s inequalities, respectively, where μ assumes the role of the weight function.

Proof. Let the weight functions be equal, $w_l \equiv w \equiv w_r$, but (for now) otherwise arbitrary subject to the condition $0 < w_{\min} \leq w(\mathbf{x})$ for a.a. $\mathbf{x} \in \Omega$. At the end of this proof we will argue the special role of the choice (23) for the weight functions. In addition, let $\hat{P} = C^\infty(\Omega)$ and $q \in \hat{Q}$, then due to (13) we can write

$$\left(\tilde{S}_{w\text{-BFBT}} q, q\right) = \sup_{p \in \hat{P}} \frac{(B^* p, w B^* q)^2}{\|B^* p\|_{(H^1(\Omega))^d}^2} \frac{\|B^* p\|_{(H^1(\Omega))^d}^2}{(w A_\mu w B^* p, B^* p)}. \quad (25)$$

We begin by estimating the second factor on the right-hand side of (25). For an arbitrary $\mathbf{v} \in (H^1(\Omega))^d$, observe that $\nabla_s(w\mathbf{v}) = w \nabla_s \mathbf{v} + \nabla w \otimes \mathbf{v}$, where “ \otimes ” denotes the outer product of two vectors in \mathbb{R}^d , and thus

$$(w A_\mu w \mathbf{v}, \mathbf{v}) = 2 (\mu \nabla_s(w\mathbf{v}), \nabla_s(w\mathbf{v})) = 2 \|\sqrt{\mu} w \nabla_s \mathbf{v} + \sqrt{\mu} \nabla w \otimes \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2.$$

Applying the triangle inequality and then Hölder’s inequality to the resulting terms,

$$\begin{aligned} \|\sqrt{\mu} w \nabla_s \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2 &\leq \|\sqrt{\mu} w\|_{L^\infty(\Omega)}^2 \|\nabla_s \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2, \\ \|\sqrt{\mu} \nabla w \otimes \mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2 &\leq \|\sqrt{\mu} \nabla w\|_{L^\infty(\Omega)^d}^2 \|\mathbf{v}\|_{L^2(\Omega)^d}^2, \end{aligned}$$

and thus we obtain the estimate

$$(w A_\mu w \mathbf{v}, \mathbf{v}) \leq 2 C_{A,\mu,w} \|\mathbf{v}\|_{(H^1(\Omega))^d}^2,$$

where

$$C_{A,\mu,w} := \|\sqrt{\mu} w\|_{L^\infty(\Omega)}^2 + \|\sqrt{\mu} \nabla w\|_{L^\infty(\Omega)^d}^2. \quad (26)$$

Similarly to (19) and (20), we obtain the following estimate for the w-BFBT Schur complement approximation (25):

$$\left(\tilde{S}_{w\text{-BFBT}} q, q\right) \geq \frac{1}{2C_{A,\mu,w}} \|wB^*q\|_{(H^{-1}(\Omega))^d}^2. \quad (27)$$

Proceeding with the exact Schur complement, we obtain from (14) in Lemma 3.3 that

$$(Sq, q) = \sup_{\mathbf{v} \in V} \frac{(w^{-1}\mathbf{v}, wB^*q)^2}{(A_\mu\mathbf{v}, \mathbf{v})} \leq \sup_{\mathbf{v} \in V} \frac{\|w^{-1}\mathbf{v}\|_{(H^1(\Omega))^d}^2 \|wB^*q\|_{(H^{-1}(\Omega))^d}^2}{2 \|\sqrt{\mu}\nabla_s\mathbf{v}\|_{(L^2(\Omega))^{d \times d}}^2}. \quad (28)$$

We require a weighted Poincaré–Friedrichs’ inequality (see Remark 3.7 for details),

$$\|\mathbf{v}\|_{L_{w^{-2}}^2(\Omega)^d} \leq C_{P,w^{-2}} \|\nabla\mathbf{v}\|_{L_{w^{-2}}^2(\Omega)^{d \times d}}, \quad (29)$$

and also a weighted Korn’s inequality (see Remark 3.7 for more information),

$$\|\nabla\mathbf{v}\|_{L_\mu^2(\Omega)^{d \times d}} \leq C_{K,\mu} \|\nabla_s\mathbf{v}\|_{L_\mu^2(\Omega)^{d \times d}}. \quad (30)$$

With (29) and (30), we are able to bound (28) from above:

$$(Sq, q) \leq \frac{(C_{P,w^{-2}}^2 + 1)C_{K,\mu}^2}{2} \left(\sup_{\mathbf{v} \in V} \frac{\|\nabla\mathbf{v}\|_{L_{w^{-2}}^2(\Omega)^{d \times d}}}{\|\nabla\mathbf{v}\|_{L_\mu^2(\Omega)^{d \times d}}} \right) \|wB^*q\|_{(H^{-1}(\Omega))^d}^2. \quad (31)$$

The supremum term in (31) and the constant $C_{A,\mu,w}$ in (26) motivate the choice for the weight w to be

$$w := \mu^{-\frac{1}{2}}.$$

Then the supremum in (31) vanishes and (26) simplifies to

$$C_{A,\mu,w} = 1 + \frac{1}{4} \|\nabla\mu\|_{L^\infty(\Omega)^d}.$$

Substituting this into (27) together with inequality (31) yields the desired result (24). \square

Remark 3.7. In the proof of Lemma 3.6 we utilized a weighted Poincaré–Friedrichs’ inequality, for which the optimal constant is

$$C_{P,\mu} = \sup_{\mathbf{v} \in V} \frac{\|\mathbf{v}\|_{L_\mu^2(\Omega)^d}}{\|\nabla\mathbf{v}\|_{L_\mu^2(\Omega)^{d \times d}}},$$

where the viscosity takes the role of the weight function. While weighted Poincaré and Friedrichs’ inequalities have been investigated in the literature numerous times, usually they are proven by contradiction and scaling arguments, which does not provide information about the constants. If explicit constants are found, they depend, in general, on the weight such that the resulting estimates are too pessimistic, e.g., $C_{P,\mu} = \mathcal{O}(\text{DR}(\mu))$. Knowledge of constants that are robust with respect to weight functions is limited. In the context of a posteriori error estimates for finite elements, weight-independent constants could be found for convex domains and weights that are a positive power of a non-negative concave function [10]. These results were refined for star-shaped domains under certain assumptions for the weights [42]. For another class of weights, namely quasi-monotone piecewise constant weight functions, robust constants were derived in [33].

In addition to weighted Poincaré–Friedrichs’, we utilized a weighted Korn’s inequality in the proof of Lemma 3.6. The optimal constant for this inequality is

$$C_{K,\mu} = \sup_{\mathbf{v} \in V} \frac{\|\nabla\mathbf{v}\|_{L_\mu^2(\Omega)^{d \times d}}}{\|\nabla_s\mathbf{v}\|_{L_\mu^2(\Omega)^{d \times d}}}.$$

As for $C_{P,\mu}$, straightforward estimation results in an overly pessimistic weight-dependent constant, namely $C_{K,\mu} = \mathcal{O}(\text{DR}(\mu))$, [25]. Other work utilizing weighted Korn’s inequalities usually aims to derive inequalities for special domain shapes, e.g., [1].

In summary, robust constants for weighted Poincaré–Friedrichs’ and Korn’s inequalities for general weight functions are difficult to obtain and limitations exist in the form of assumptions on the weights. Further work on this topic could improve the constants for the spectral equivalence of w-BFBT but is beyond the scope of this paper.

Table 1: Robustness classification for Schur complement approximations a $\mathbf{M}_p(1/\mu)$ and b w-BFBT in terms of number of GMRES iterations (10^{-6} residual reduction, GMRES restart every 100 iterations). Number of randomly placed sinkers ($\#\text{sinkers}$) is increased across rows, while dynamic ratio ($\text{DR}(\mu)$) is increased across columns. Discretization is fixed at $k = 2, \ell = 7$.

(a) $\mathbf{M}_p(1/\mu)$					(b) w-BFBT				
$\#\text{sinkers} \setminus \text{DR}(\mu)$	10^4	10^6	10^8	10^{10}	$\#\text{sinkers} \setminus \text{DR}(\mu)$	10^4	10^6	10^8	10^{10}
S1-rand	29	31	31	29	S1-rand	29	29	29	30
S4-rand	53	63	71	80	S4-rand	39	41	42	44
S8-rand	64	79	93	165	S8-rand	38	40	41	44
S12-rand	70	86	99	180	S12-rand	38	40	43	45
S16-rand	85	167	231	891	S16-rand	40	45	47	48
S20-rand	84	167	380	724	S20-rand	34	36	37	38
S24-rand	117	286	3279	5983	S24-rand	31	32	39	55
S28-rand	108	499	2472	>10000	S28-rand	29	31	42	60

4 Robustness of w-BFBT

In this section, we analyze the robustness properties of the widely used Schur complement approximation $\mathbf{M}_p(1/\mu)$ and the new w-BFBT via numerical experiments. Furthermore, we calculate the spectra for both approaches and thus support the discussion in Section 3 about theoretical eigenvalue bounds with numerical results. The comparison of $\mathbf{M}_p(1/\mu)$ and w-BFBT is of particular importance, because of the widespread use of the inverse viscosity-weighted mass matrix. It is therefore of interest to determine when convergence with $\mathbf{M}_p(1/\mu)$ deteriorates and using w-BFBT becomes beneficial. A comparison with $\text{diag}(\mathbf{A})$ -BFBT was not performed because Section 2.2 already showed that $\text{diag}(\mathbf{A})$ -BFBT performs similarly or worse than w-BFBT, hence there are no advantages in using $\text{diag}(\mathbf{A})$ -BFBT over w-BFBT.

For the numerical experiments in this section, we return to the definitions and setup from Section 2.1. To apply the inverse of $\mathbf{M}_p(1/\mu)$, we diagonalize the mass matrix of the discontinuous, modal pressure space by forming its lumped version (4). Moreover, to apply the approximate inverse of the viscous block in (3) we use the same multigrid method for each of the two Schur approximations; this multigrid method is also used for the inverse operators of w-BFBT in (8). The details of the multigrid method are provided in Section 6. To compare the robustness, we vary two problem parameters: (i) the number of randomly placed sinkers n and (ii) the dynamic ratio $\text{DR}(\mu)$. The parameter n influences the geometric complexity of the viscosity μ while $\text{DR}(\mu)$ controls the magnitude of viscosity gradients.

Tables 1a and 1b present the number of GMRES iterations for a 10^{-6} residual reduction in the Euclidean norm. Observe that for the S1-rand problem, the iteration count is essentially the same for both $\mathbf{M}_p(1/\mu)$ and w-BFBT, and that it stays stable across all dynamic ratios $\text{DR}(\mu) = 10^4, \dots, 10^{10}$. Hence for this simple problem, w-BFBT has no advantages and its additional computational cost makes it less efficient than $\mathbf{M}_p(1/\mu)$. However, the limitations of the $\mathbf{M}_p(1/\mu)$ approach become apparent by increasing the number of randomly positioned sinkers. Two observations for $\mathbf{M}_p(1/\mu)$ can be made from Table 1a. First, the number of GMRES iterations rises with increasing number of sinkers (factor ~ 80 increase for $n = 1, \dots, 28$, $\text{DR}(\mu) = 10^8$). Second, in a multi-sinker setup the dependence on $\text{DR}(\mu)$ becomes more severe (factor ~ 50 increase for $n = 24$, $\text{DR}(\mu) = 10^4, \dots, 10^{10}$). This demonstrates that $\mathbf{M}_p(1/\mu)$ is a poor approximation of the Schur complement for certain classes of problems with highly heterogeneous viscosities.

The advantages in robustness of the w-BFBT preconditioner are demonstrated in Table 1b. Compared to $\mathbf{M}_p(1/\mu)$, the number of GMRES iterations is stable and the increase over the whole range of problem parameters is just a factor of 2. Only 60 iterations are needed for the most extreme problem, namely S28-rand, $\text{DR}(\mu) = 10^{10}$, for which convergence with $\mathbf{M}_p(1/\mu)$ essentially stagnated.

More insight concerning the different convergence behaviors can be gained from the eigenvalues in Figure 3. The plots in that figure are for two-dimensional multi-sinker problems, which are analogous to the three-dimensional benchmark problems from Section 2.1. We discretize the problems on triangular meshes utilizing the FEniCS library [28]. We choose $\mathbb{P}_2^{\text{bubble}} \times \mathbb{P}_1^{\text{disc}}$ finite elements [11, 12] because they represent a close analog to the $\mathbb{Q}_2 \times \mathbb{P}_1^{\text{disc}}$ elements, which are employed on (three-dimensional) hexahedral meshes. Each plot shows the eigenvalues of the exact Schur complement and the eigenvalues of the preconditioned Schur complement for the

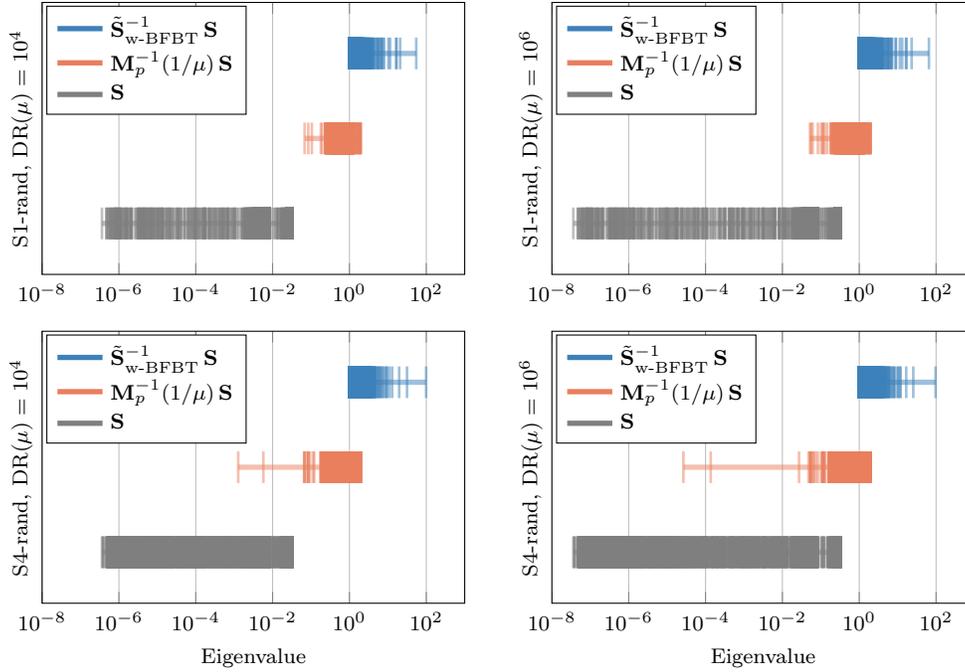


Figure 3: Spectra of the Schur complement (*gray*), $\mathbf{M}_p(1/\mu)$ -preconditioned Schur complement (*red*), and w-BFBT-preconditioned Schur complement (*blue*); zero eigenvalues corresponding to the null space of the Schur complement matrix are omitted. Results for viscosities with one sinker (S1-rand) are shown in *top row*, and with four sinkers (S4-rand) in the *bottom row* of plots; $\text{DR}(\mu) = 10^4$ in the *left column* and $\text{DR}(\mu) = 10^6$ in the *right column*. The two-dimensional Stokes equations are discretized with $\mathbb{P}_2^{\text{bubble}} \times \mathbb{P}_1^{\text{disc}}$ finite elements on a uniform triangular mesh consisting of 512 triangles using the FEniCS library. As the problem difficulty increases from one to four sinkers, the spreading of small eigenvalues for $\mathbf{M}_p(1/\mu)$ becomes more severe, which is disadvantageous for solver convergence. For w-BFBT, the spectrum remains largely unaffected by increased sinker counts, which contributes to convergence that is robust with respect to viscosity variations.

$\mathbf{M}_p(1/\mu)$ and w-BFBT approximations, where all inverse matrices, e.g., the viscous block matrix \mathbf{A} and the pressure Poisson matrices within w-BFBT, are inverted with a direct solver. Effective preconditioners exhibit a strong clustering of eigenvalues, whereas the convergence of Krylov methods deteriorates if the eigenvalues are spread out. We recognize different characteristics in the spectra associated with $\mathbf{M}_p(1/\mu)$ and w-BFBT preconditioning. For $\mathbf{M}_p(1/\mu)$, the dominant eigenvalues are clustered around one while smaller eigenvalues, i.e., eigenvalues $\ll 1$, are spread out. The behavior for w-BFBT is the opposite: the dominant eigenvalues are spread out and the smaller eigenvalues are tightly clustered around one. Now, as the problem difficulty is increased by introducing more viscosity anomalies in the domain, the spreading of smaller eigenvalues associated with $\mathbf{M}_p(1/\mu)$ becomes more severe (compare in Figure 3 the *top row* of plots and the *bottom row* of plots). We postulate that this is the property that is responsible for the deteriorating convergence with $\mathbf{M}_p(1/\mu)$ that was observed in Table 1a. With w-BFBT on the other hand, the spectrum remains largely unaffected by increased sinker counts. The clustering of smaller eigenvalues around one remains stable, which is likely the reason for the robustness of w-BFBT. The lower bound on the eigenvalues that we observe here numerically supports the theoretical estimates on spectral equivalence in Section 3. Therefore we find the lower bound to be sharp and, moreover, to be essential for the robustness of the w-BFBT preconditioner.

Remark 4.1. In addition to the $\mathbb{P}_2^{\text{bubble}} \times \mathbb{P}_1^{\text{disc}}$ discretization used for the results in Figure 3, we also calculated the spectra using $\mathbb{P}_2 \times \mathbb{P}_1$ Taylor-Hood finite elements. We obtained very similar results for this discretization, which uses continuous elements to approximate the pressure. Therefore, both the efficacy of w-BFBT as a preconditioner and the issues with $\mathbf{M}_p(1/\mu)$ seem to be largely unaffected by the specific type of discretization, at least for the two cases that we tested.

Remark 4.2. The convergence of the Stokes solver with the $\mathbf{M}_p(1/\mu)$ preconditioner can be improved by

approximating the heterogeneous viscosity $\mu(\mathbf{x})$ with elementwise constants, computed by averaging μ over each element [3]. The benefit of faster convergence comes at the cost of slower asymptotic convergence of the discrete finite element solution and an altered constitutive relationship, which might be a less accurate representation of the physics; this is, however, problem-dependent. Moreover, for a nonlinear (e.g., power-law) rheology, elementwise averaging of μ can introduce non-physical, artificial disturbances in the effective viscosity during Newton or Picard-type nonlinear solves. We observed such a behavior in mantle convection simulations, which are governed by a nonlinear power-law rheology. Here, viscosity averaging led to non-physical checkerboard-like patterns upon convergence of the nonlinear Newton solver.

Remark 4.3. In practice, the convergence of w-BFBT can be improved for coarse meshes, where the viscosity variations over elements are large. This is achieved by alternative choices for the diagonal weighting matrices \mathbf{C}_{w_l} and \mathbf{D}_{w_r} from (8) with the weights

$$w_l(\mathbf{x}) = \left(\mu^2(\mathbf{x}) + |\nabla\mu(\mathbf{x})|^2 \right)^{\frac{1}{4}} = w_r(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega,$$

where $|\cdot|$ denotes the Euclidean norm in \mathbb{R}^d . These viscosity gradient-based w-BFBT weights have the advantage of performing at least as well as the pure viscosity-based weights proposed in Section 1.3, but they exhibit superior robustness on coarser meshes. They are, however, challenging to analyze theoretically.

5 Modifications for Dirichlet boundary conditions

In Section 2.2, deteriorating approximation properties of w-BFBT for increasing discretization order and mesh refinement level could be observed. The numerical experiments in Figure 2, *right* did show slightly slower convergence when k and ℓ were increased. This can stem from w-BFBT representing a poor approximation to the exact Schur complement at the boundary $\partial\Omega$ in the presence of Dirichlet boundary conditions for the velocity. This section investigates modifications to w-BFBT near a Dirichlet boundary and aims at obtaining mesh independence and only a mild dependence on discretization order in terms of Stokes solver convergence.

Consider the commutator that leads to the w-BFBT formulation in an infinite-dimensional form: $AB^* - B^*X \approx 0$, where A represents the viscous stress operator, B^* the gradient operator and X the sought commuting operator. In case of an unbounded domain $\Omega = \mathbb{R}^d$ and constant viscosity $\mu \equiv 1$, this commutator is exactly satisfied since $(\nabla \cdot \nabla)\nabla - \nabla(\nabla \cdot \nabla) = 0$. For Dirichlet boundary conditions, the commutator does not, in general, vanish at the boundary. Therefore a possible source for deteriorating Schur complement approximation properties of w-BFBT is a commutator mismatch for mesh elements that are touching the boundary $\partial\Omega$. A similar observation was also made in [16, 17]. A possible remedy is to modify the norm in the least-squares minimization problem (6), which is represented by the matrix \mathbf{C}^{-1} , such that a damping factor is applied to the matrix entries near the boundary. By damping the influence of the boundary in the minimization objective, more emphasis is given to the domain interior, and the w-BFBT approximation is improved.

Damping near Dirichlet boundaries can be incorporated by modifying the matrices $\mathbf{C}_{w_l}^{-1}$ or $\mathbf{D}_{w_r}^{-1}$ of the w-BFBT inverse Schur complement approximation (8). A similar idea for BFBT in a Navier–Stokes setting is presented in [17], where a damping to the weighting matrix \mathbf{D}^{-1} in (7) is introduced to achieve mesh independence (\mathbf{C}^{-1} is not changed). There, damping affects the normal components of the velocity space inside mesh elements touching $\partial\Omega$ and simply a constant damping factor of 1/10 is set regardless of mesh refinement ℓ . Also, only the discretization order $k = 2$ was considered (in addition to $\mathbb{Q}_2 \times \mathbb{Q}_1$ and MAC discretizations).

Now, we attempt to enhance our understanding of how modifications at a Dirichlet boundary $\partial\Omega$ influence convergence and therefore the efficacy of w-BFBT as a Schur complement approximation. Let $\Omega_D := \bigcup_{e \in D} \Omega_e$, $D := \{e \mid \bar{\Omega}_e \cap \partial\Omega \neq \emptyset\}$ be the set of all mesh elements Ω_e touching the Dirichlet boundary. Given values $a_l, a_r \geq 1$, extend the previous definition of the weights $w_l = w_r = \sqrt{\mu}$ (see Section 1.3) to a version with boundary modification:

$$w_l(\mathbf{x}) := \begin{cases} a_l \sqrt{\mu(\mathbf{x})} & \mathbf{x} \in \Omega_D, \\ \sqrt{\mu(\mathbf{x})} & \mathbf{x} \notin \Omega_D, \end{cases} \quad \text{and} \quad w_r(\mathbf{x}) := \begin{cases} a_r \sqrt{\mu(\mathbf{x})} & \mathbf{x} \in \Omega_D, \\ \sqrt{\mu(\mathbf{x})} & \mathbf{x} \notin \Omega_D. \end{cases} \quad (32)$$

Table 2: Influence of boundary modification factors a_l , a_r on the Stokes solver convergence with w-BFBT for discretizations: $k = 2, \ell = 5, \dots, 7$ (see a, c, e) and $k = 2, \dots, 5, \ell = 5$ (see a, b, d, f). Reported are the number of GMRES iterations for 10^{-6} residual reduction for the problem S16-rand, $\text{DR}(\mu) = 10^6$. Colors highlight solves within $\sim 5\%$ of iterations above the lowest iteration count. Increase of mesh refinement level ℓ or discretization order k demands larger boundary amplification a_r or a_l to maintain fast convergence.

(a) $k = 2, \ell = 5$							(b) $k = 3, \ell = 5$						
$a_l \setminus a_r$	1	2	4	8	16	32	$a_l \setminus a_r$	1	2	4	8	16	32
1	33	33	34	34	34	35	1	41	38	37	37	37	37
2	33	33	34	34	34	34	2	38	37	38	38	39	39
4	33	34	34	36	38	39	4	37	38	40	42	44	46
8	34	34	36	39	43	44	8	36	38	42	47	50	51
16	34	34	38	43	46	49	16	37	39	44	50	53	56
32	34	34	39	44	49	53	32	37	39	45	51	56	59
(c) $k = 2, \ell = 6$							(d) $k = 4, \ell = 5$						
$a_l \setminus a_r$	1	2	4	8	16	32	$a_l \setminus a_r$	1	2	4	8	16	32
1	37	34	33	34	34	34	1	44	39	36	36	36	36
2	34	34	34	34	34	34	2	39	39	39	40	41	41
4	33	33	34	35	36	37	4	36	39	43	47	49	51
8	34	34	35	38	39	39	8	36	40	47	52	56	58
16	34	34	36	39	40	41	16	36	41	49	56	60	63
32	34	34	37	39	41	42	32	36	41	50	58	63	66
(e) $k = 2, \ell = 7$							(f) $k = 5, \ell = 5$						
$a_l \setminus a_r$	1	2	4	8	16	32	$a_l \setminus a_r$	1	2	4	8	16	32
1	45	37	34	34	34	34	1	63	53	46	43	43	44
2	37	36	35	36	36	36	2	53	51	51	51	52	53
4	34	36	38	39	40	41	4	47	51	55	59	62	64
8	34	36	39	42	44	44	8	44	51	59	65	69	72
16	34	36	40	44	45	46	16	43	52	62	69	75	78
32	34	36	41	44	46	47	32	44	53	64	72	78	82

We obtain matrices $\mathbf{C}_{w_l} = \tilde{\mathbf{M}}_{\mathbf{u}}(w_l)$ and $\mathbf{D}_{w_r} = \tilde{\mathbf{M}}_{\mathbf{u}}(w_r)$ in (8) that may differ at boundary elements in Ω_D due to possibly different values for a_l and a_r . Note that amplifying the weight functions w_l, w_r at the boundary is similar to damping at the boundary after taking the inverses $\mathbf{C}_{w_l}^{-1}, \mathbf{D}_{w_r}^{-1}$.

The Stokes solver convergence under the influence of boundary amplifications a_l, a_r is summarized in Table 2. The table shows that the boundary amplification is most effective when performed non-symmetrically, i.e., either $a_l > 1$ or $a_r > 1$ but not both. Further, we deduce that with higher mesh refinement level ℓ , the boundary amplification should increase roughly proportional to 2^ℓ (or proportional to the reciprocal element size, here $h^{-1} = 2^\ell$). Similar observations can be made for the discretization order k , i.e., amplification needs to increase for larger k to avoid higher iteration counts. These implications were made based on extensive numerical experiments for which Table 2 serves as a representative summary.

Remark 5.1. The theoretical derivations of spectral equivalence from Section 3 and the necessity for damping at Dirichlet boundaries appear inconsistent. However, spectral equivalence was shown in infinite dimensions whereas boundary damping is applied to the discretized problem. Therefore, we believe that the necessity for damping is introduced through the discretization. It still remains an open question what might be causing the slowdown in convergence that is avoided by damping.

6 Parallel hybrid spectral-geometric-algebraic multigrid (HMG) for w-BFBT

Two aspects of the Stokes preconditioner with w-BFBT have not been discussed yet. One is the approximation of the inverse viscous block $\tilde{\mathbf{A}}^{-1}$ required in (3) and the other is the approximation of inverses $\tilde{\mathbf{K}}_{w_l}^{-1} \approx \mathbf{K}_{w_l}^{-1} :=$

$(\mathbf{B}\mathbf{C}_{w_l}^{-1}\mathbf{B}^\top)^{-1}$ and $\tilde{\mathbf{K}}_{w_r}^{-1} \approx \mathbf{K}_{w_r}^{-1} := (\mathbf{B}\mathbf{D}_{w_r}^{-1}\mathbf{B}^\top)^{-1}$ in (8). These approximations are crucial for overall Stokes solver performance and scalability and are addressed in this section. For brevity, we limit our discussion to $\tilde{\mathbf{K}}_{w_r}^{-1}$; the results also hold for $\tilde{\mathbf{K}}_{w_l}^{-1}$.

The approximation of the inverse viscous block $\tilde{\mathbf{A}}^{-1}$ is well suited for multigrid V-cycles. To this end, in [36], we developed a hybrid spectral-geometric-algebraic multigrid (HMG) method, which exhibits extreme parallel scalability and retains nearly optimal algorithmic scalability (see Section 7 for scalability results). While traversing the HMG hierarchy shown in Figure 4, HMG initially reduces the discretization order (spectral multigrid); after arriving at order one, it continues by coarsening mesh elements (geometric multigrid); once the degrees of freedom fall below a threshold, algebraic multigrid (AMG) carries out further coarsening until a direct solve can be computed efficiently ([41]). Parallel forest-of-octrees algorithms, implemented in the p4est parallel adaptive mesh refinement library [5, 9], are used for efficient, scalable mesh refinement/coarsening, mesh balancing, and repartitioning in the geometric HMG phase. During parallel geometric coarsening, the number of compute cores and the size of the MPI communicator is reduced successively to minimize communication. Re-discretization of the differential equations is performed on each coarser spectral and geometric level. The viscosity values in each element are stored at the quadrature points of the velocity discretization, and are thus local to each element. The viscosity coarsening is done level-by-level during the setup phase. The coarsening operator is the adjoint of the refinement operator, which performs element-wise interpolation. This adjoint is computed with respect to the L^2 -inner products, and since viscosity values are not shared amongst elements, this does not require (an approximation of) a global mass matrix solve. The transition from geometric to algebraic multigrid is done at a sufficiently small core count and small MPI communicator. AMG continues to further reduce problem size (via Galerkin coarse grid projection) and the number of cores down to a single core for the direct solver.

The operator \mathbf{K}_{w_r} can be regarded as a discrete, variable-coefficient Poisson operator on the discontinuous pressure space $\mathbb{P}_{k-1}^{\text{disc}}$ with Neumann boundary conditions. Therefore, multigrid V-cycles can also be employed to approximate the inverse $\tilde{\mathbf{K}}_{w_r}^{-1}$. However, it turned out to be problematic to apply multigrid coarsening directly due to the discontinuous, modal discretization of the pressure. We took a novel approach in [36] by considering the underlying infinite-dimensional, variable-coefficient Poisson operator, where the coefficient is derived from the diagonal weighting matrix (here, $\mathbf{C}_{w_l}^{-1}$ or $\mathbf{D}_{w_r}^{-1}$). Then we re-discretize with continuous, nodal high-order finite elements in \mathbb{Q}_k . An alternative would be to use \mathbb{Q}_{k-1} , but we prefer to use \mathbb{Q}_k since the corresponding data structures are readily available from the discretization of the velocity. Hence, this choice avoids the setup cost related to discretization-specific parameters and their storage. Additionally, the HMG hierarchy of the preconditioner acting on the velocity can be partially reused, again saving setup time and memory. This continuous, nodal discretization of the Poisson operator is then approximately inverted with an HMG V-cycle that is similar to the one described above for the inverse viscous block approximation $\tilde{\mathbf{A}}^{-1}$. Additional smoothing is applied in the discontinuous pressure space (Figure 4, *green level*) to account for high frequency modes in residuals that are introduced through projections between \mathbb{Q}_k and $\mathbb{P}_{k-1}^{\text{disc}}$.

Our hybrid multigrid method combines high-order L^2 -restrictions/interpolations and employs Chebyshev-accelerated point-Jacobi smoothers. This results in optimal or nearly optimal algorithmic multigrid performance (see Section 7), i.e., iteration numbers are independent of mesh size and only mildly dependent on discretization order, while maintaining robustness with respect to highly heterogeneous coefficients. In addition, the efficacy of the HMG preconditioner does not deteriorate with increasing core counts, because the spectral and geometric multigrid is by construction independent of the number of cores and AMG is invoked for prescribed small problem sizes on essentially fixed small core counts.

For all numerical experiments presented here, three pre- and post-smoothing iterations with a Chebyshev accelerated point-Jacobi smoother are performed. AMG is always invoked on just one MPI rank after geometrically coarsening the uniform mesh to $\ell = 2$ with 64 elements (using a direct solver for such small problems is also reasonable). GMRES is restarted after every 100 iterations throughout all experiments. PETSc's [2] implementations of Chebyshev acceleration, direct solver, AMG (called GAMG), and GMRES are used.

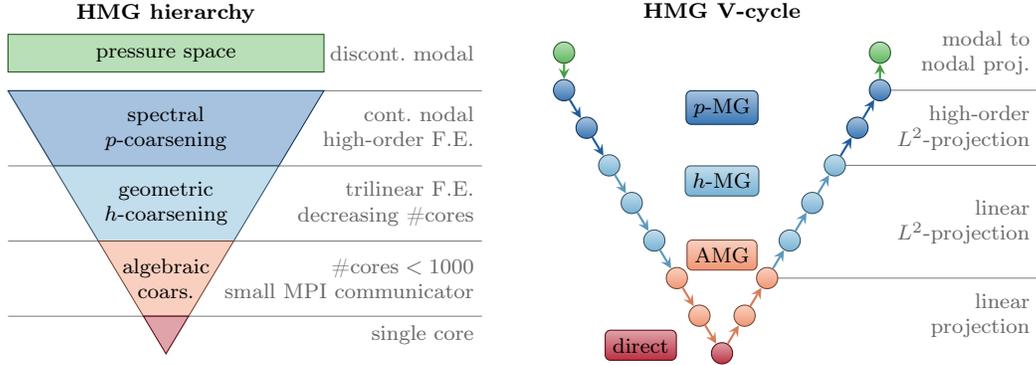


Figure 4: Hybrid spectral-geometric-algebraic multigrid (HMG). *Left*: Illustration of multigrid hierarchy. From top to bottom, first, the multigrid levels are obtained by spectral coarsening (*dark blue*). Next, the mesh is geometrically coarsened and repartitioned on successively fewer cores to minimize communication (*light blue*). Finally, AMG further reduces problem size and core count (*light red*). The multigrid hierarchy for the pressure Poisson operator \mathbf{K}_{w_r} additionally involves smoothing in the discontinuous, modal pressure space (*green*). The projection from the discontinuous, modal to a continuous finite element nodal basis uses a lumped mass matrix in the nodal space to avoid the global mass matrix system solve. *Right*: The multigrid V-cycle consists of smoothing at each level of the hierarchy (*circles*) and intergrid transfer operators (*arrows downward* for restriction and *arrows upward* for interpolation). To enhance efficacy of the the V-cycle as a preconditioner, different types of projection operators are employed for these operators depending on the phase within the V-cycle.

7 Algorithmic and parallel scalability for HMG + w-BFBT Stokes preconditioner

After establishing the robustness of the Stokes solver with w-BFBT preconditioning in theory (Section 3) and numerically (Section 4), addressing issues associated with Dirichlet boundary conditions (Section 5), and introducing an effective and scalable multigrid method (Section 6), in this section we finally study the scalability of the Stokes solver building on HMG+w-BFBT. One aspect of scalability is algorithmic scalability, i.e., the dependence of Krylov iterations on the mesh resolution and the discretization order. The second aspect is parallel scalability of the implementation, i.e., runtime measured on increasing numbers of compute cores. Studying both aspects is required to fully assess the performance of a solver at scale.

The algorithmic scalability in Table 3 shows results for the Stokes solver as well as its individual components by reporting iteration numbers for solving the systems $\mathbf{A}\mathbf{u} = \mathbf{f}$ and $\mathbf{K}_{w_r}\mathbf{p} = \mathbf{g}$. Studying the individual components allows us to observe HMG performance in isolation and to compare it to the algorithmic scalability of the full Stokes system, which is indicative of the quality of the w-BFBT Schur complement approximation. All systems are solved with preconditioned GMRES down to a relative tolerance of 10^{-6} . The preconditioners for \mathbf{A} and \mathbf{K}_{w_r} are HMG-V-cycles as described in Section 6. For the w-BFBT preconditioner, we set a constant left boundary amplification $a_l = 1$ and vary the right boundary amplification a_r according to results from Section 5. The iteration counts in Table 3a show textbook mesh independence when increasing the level of refinement ℓ . This holds for each component, \mathbf{A} and \mathbf{K}_{w_r} , and also the whole Stokes solver, and hence we conclude that the Schur complement approximation by w-BFBT is mesh-independent. When the discretization order k is increased, the iteration counts presented in Table 3b increase mildly. The convergence of both components \mathbf{A} and \mathbf{K}_{w_r} exhibits a moderate dependence on k . Since the increase in number of iterations is slightly larger for the full Stokes solve than for \mathbf{A} and \mathbf{K}_{w_r} , we suspect a mild deterioration of w-BFBT as a Schur complement approximation.

The following parallel scalability results complement the results presented in [36], where scalability to millions of threads was demonstrated on IBM's BlueGene/Q architecture, which is specifically designed for large-scale supercomputers and thus differs from conventional clusters. The parallel scalability results here were obtained on the full Lonestar 5 peta-scale system, which represents a rather conventional cluster, housed

Table 3: Algorithmic scalability for Stokes solver with HMG+w-BFBT preconditioning while a varying mesh refinement level ℓ and b varying discretization order k (problem S16-rand, $\text{DR}(\mu) = 10^6$). Computational cost is expressed in number of GMRES iterations (abbreviated by It.) for full Stokes solve (10^{-6} residual reduction). Left boundary amplification for \mathbf{C}_{w_l} is fixed to $a_l = 1$; right boundary amplification a_r for \mathbf{D}_{w_r} varies. Additionally, the numbers of GMRES iterations for solving only the sub-systems $\mathbf{A}\mathbf{u} = \mathbf{f}$ and $\mathbf{K}_{w_r}\mathbf{p} = \mathbf{g}$ are given for demonstration of HMG efficacy (here, \mathbf{f} is the right-hand side of the momentum equation and \mathbf{g} is the discrete representation of $\nabla \cdot \mathbf{f}$; however, random right-hand sides would give similar convergence results).

(a) Algorithmic scalability (fixed order $k = 2$)								(b) Algorithmic scalability (fixed level $\ell = 5$)							
ℓ	a_r	\mathbf{u} -DOF [$\times 10^6$]	It. \mathbf{A}	\mathbf{p} -DOF [$\times 10^6$]	It. \mathbf{K}_{w_r}	DOF [$\times 10^6$]	It. Stokes	k	a_r	\mathbf{u} -DOF [$\times 10^6$]	It. \mathbf{A}	\mathbf{p} -DOF [$\times 10^6$]	It. \mathbf{K}_{w_r}	DOF [$\times 10^6$]	It. Stokes
4	1	0.11	18	0.02	8	0.12	40	2	2	0.82	18	0.13	7	0.95	33
5	2	0.82	18	0.13	7	0.95	33	3	4	2.74	20	0.32	8	3.07	37
6	4	6.44	18	1.05	6	7.49	33	4	8	6.44	20	0.66	7	7.10	36
7	8	50.92	18	8.39	6	59.31	34	5	16	12.52	23	1.15	12	13.67	43
8	16	405.02	18	67.11	6	472.12	34	6	32	21.56	23	1.84	12	23.40	50
9	32	3230.67	18	536.87	6	3767.54	34	7	64	34.17	22	2.75	10	36.92	54
10	64	25807.57	18	4294.97	6	30102.53	34	8	128	50.92	22	3.93	10	54.86	67

at the Texas Advanced Computing Center (TACC). The Lonestar 5 supercomputer entered production in January 2016 and is a Cray XC40 system consisting of 1252 compute nodes. Each node is equipped with two Intel Haswell 12-core processors (Xeon E5-2680v3) and 64 GBytes of memory. Inter-node communication is based on an Aries Dragonfly topology network that provides dynamic routing and thus enables optimal use of the system bandwidth.

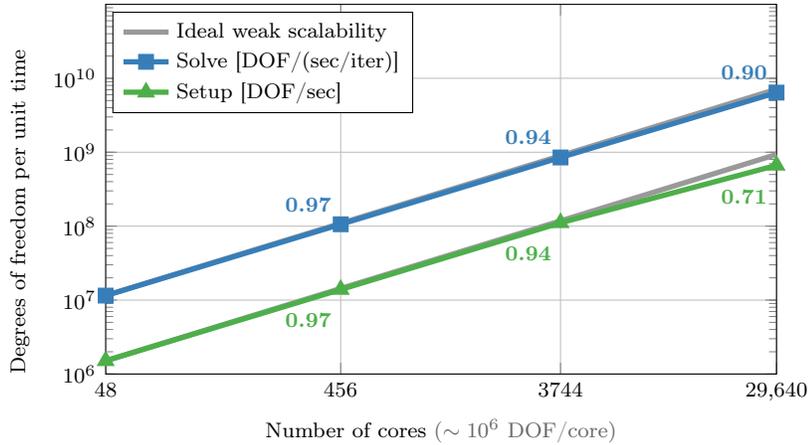
Results for weak scalability (DOF/core fixed to ~ 1 million) in Figure 5a show that the Stokes solver with w-BFBT (*blue curve*) maintains 90% parallel efficiency over a 618-fold increase in degrees of freedom along with cores. Even for the setup of the Stokes solver (*green curve*), which mainly involves generation of the HMG hierarchy, we observe 71% parallel efficiency. These are excellent results for such a complex implicit multilevel solver with optimal algorithmic performance (when the mesh is refined, or nearly algorithmically optimal when the order is increased) and with convergence that is independent of the number of cores.

Finally, Figure 5b reports strong scalability results (overall DOF fixed to 59 million) and how the number of OpenMP threads substituting MPI ranks influences speedup.⁴ Over the 78-fold increase from 48 to 3744 cores, efficiency reduces moderately, to a worst-case 68% for $24 \times \text{OMP1}$. However, note that in the largest run with 29,640 cores, the granularity is only ~ 2000 DOF/core, which is extremely challenging for strong scalability. In this case, due to the increased communication volume, overlapping with decreased amounts of computation becomes impossible and communication dominates the runtime. This behavior is expected for an implicit solver, especially for a multilevel method that does not sacrifice algorithmic optimality for parallel scalability.

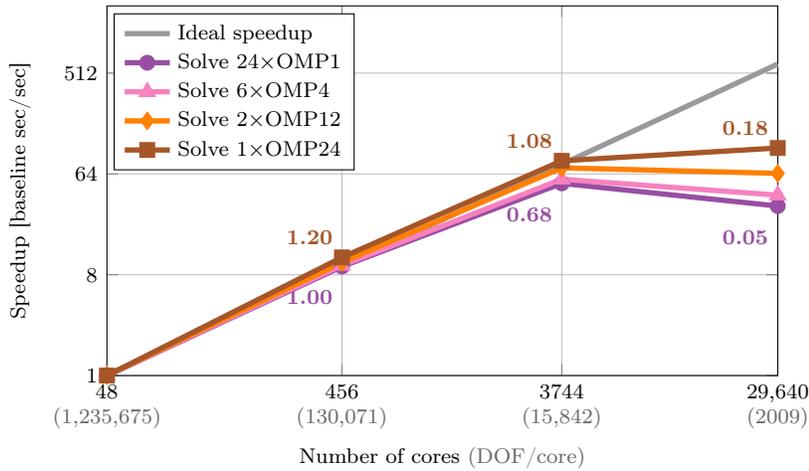
8 Conclusions

For Stokes flow problems with highly heterogeneous viscosity, the commonly used inverse viscosity-weighted mass matrix approximation of the Schur complement can be insufficient. As a consequence, convergence of Schur complement-based Krylov solvers can be extremely slow. While each iteration with the weighted BFBT (w-BFBT) Schur complement approximation proposed in this paper is computationally more costly, we observe that it results in robust and fast convergence even for complex viscosity structures and for up to ten orders of magnitude viscosity contrasts—properties than occur, for instance, in problems involving non-Newtonian geophysical fluids. Our numerical findings are supported by theoretical spectral equivalence results. To the best of our knowledge, a similar analysis has not been shown for any BFBT-type method before.

⁴Even though the processors of Lonestar 5 support two threads per physical core (Intel Hyper-Threading Technology), assigning more than one OpenMP thread per core did not improve performance.



(a) Weak scalability ($k = 2, \ell = 7, \dots, 10$)



(b) Strong scalability ($k = 2, \ell = 7$)

Figure 5: Parallel scalability on Lonestar 5 for Stokes solver with HMG+w-BFBT preconditioning (problem S16-rand, $DR(\mu) = 10^6$ as in Table 3a). a Weak scalability of setup and solve phases (normalized w.r.t. deviations from const. DOF/core). *Numbers along the graph lines* indicate weak parallel efficiency w.r.t. ideal weak scalability (baseline is 48 cores result). DOF/core is ~ 1 million; the largest problem size on 29,640 cores has 30 billion DOF. b Strong scalability of solve phase for different configurations of OpenMP threads (OMP) substituting MPI ranks on each node consisting of 24 cores. *Numbers along the graph lines* indicate strong efficiency w.r.t. ideal speedup (baseline is 48 cores result).

At Dirichlet boundaries, we use a modification of w-BFBT that is necessary for mesh-independent convergence. In this modification, we dampen the influence of elements at these boundaries, which would otherwise dominate in w-BFBT and degrade its efficacy as a preconditioner. Finding an alternative remedy is a subject of current research.

Indefinite problems with highly heterogeneous coefficients and high-order discretizations present significant challenges for efficient solvers, especially in parallel. Nevertheless, we have demonstrated that with careful attention paid to algorithm design and scalable implementation, viscosity-robust and mesh-independent solvers can be designed that exhibit nearly optimal algorithmic and parallel scalability over a wide range of problems sizes and core counts.

Acknowledgments

We wish to offer our deep thanks to the Texas Advanced Computing Center (TACC) for granting us early user access to Lonestar 5 and enabling runs on the full system. We appreciate many helpful discussions with Tobin Isaac, who also allowed us to build on his previous work on Stokes solvers. We also thank Carsten Burstedde for his dedicated work on the p4est library.

References

- [1] GABRIEL ACOSTA, RICARDO G. DURÁN, AND ARIEL L. LOMBARDI, *Weighted Poincaré and Korn inequalities for Hölder α domains*, *Mathematical Methods in the Applied Sciences*, 29 (2006), pp. 387–400.
- [2] SATISH BALAY, SHRIRANG ABHYANKAR, MARK F. ADAMS, JED BROWN, PETER BRUNE, KRIS BUSCHELMAN, LISANDRO DALCIN, VICTOR ELJKHOUT, WILLIAM D. GROPP, DINESH KAUSHIK, MATTHEW G. KNEPLEY, LOIS CURFMAN MCINNES, KARL RUPP, BARRY F. SMITH, STEFANO ZAMPINI, AND HONG ZHANG, *PETSc users manual*, Tech. Report ANL-95/11 - Revision 3.6, Argonne National Laboratory, 2015.
- [3] WOLFGANG BANGERTH AND TIMO HEISTER, *ASPECT: Advanced Solver for Problems in Earth’s ConvecTion*, *Computational Infrastructure in Geodynamics*, 2015.
- [4] MICHELE BENZI, GENE H. GOLUB, AND JÖRG LIESEN, *Numerical solution of saddle point problems*, *Acta Numerica*, 14 (2005), pp. 1–137.
- [5] CARSTEN BURSTEDDE, OMAR GHATTAS, MICHAEL GURNIS, TOBIN ISAAC, GEORG STADLER, TIM WARBURTON, AND LUCAS C. WILCOX, *Extreme-scale AMR*, in SC10: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, ACM/IEEE, 2010.
- [6] CARSTEN BURSTEDDE, OMAR GHATTAS, MICHAEL GURNIS, EH TAN, TIANKAI TU, GEORG STADLER, LUCAS C. WILCOX, AND SHIJIE ZHONG, *Scalable adaptive mantle convection simulation on petascale supercomputers*, in SC08: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, ACM/IEEE, 2008.
- [7] CARSTEN BURSTEDDE, OMAR GHATTAS, GEORG STADLER, TIANKAI TU, AND LUCAS C. WILCOX, *Parallel scalable adjoint-based adaptive solution for variable-viscosity Stokes flows*, *Computer Methods in Applied Mechanics and Engineering*, 198 (2009), pp. 1691–1700.
- [8] CARSTEN BURSTEDDE, GEORG STADLER, LAURA ALISIC, LUCAS C. WILCOX, EH TAN, MICHAEL GURNIS, AND OMAR GHATTAS, *Large-scale adaptive mantle convection simulation*, *Geophysical Journal International*, 192 (2013), pp. 889–906.
- [9] CARSTEN BURSTEDDE, LUCAS C. WILCOX, AND OMAR GHATTAS, *p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees*, *SIAM Journal on Scientific Computing*, 33 (2011), pp. 1103–1133.
- [10] SENG-KEE CHUA AND RICHARD L. WHEEDEN, *Estimates of best constants for weighted Poincaré inequalities on convex domains*, *Proceedings of the London Mathematical Society*, 93 (2006), pp. 197–226.

- [11] MICHEL CROUZEIX AND PIERRE-ARNAUD RAVIART, *Conforming and nonconforming finite element methods for solving the stationary Stokes equations. I*, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge, 7 (1973), pp. 33–75.
- [12] JEAN DONEA AND ANTONIO HUERTA, *Finite Element Methods for Flow Problems*, John Wiley & Sons, 2003.
- [13] HOWARD C. ELMAN, *Preconditioning for the steady-state Navier–Stokes equations with low viscosity*, SIAM Journal on Scientific Computing, 20 (1999), pp. 1299–1316.
- [14] HOWARD C. ELMAN, VICTORIA E. HOWLE, JOHN SHADID, ROBERT SHUTTLEWORTH, AND RAYMOND S. TUMINARO, *Block preconditioners based on approximate commutators*, SIAM Journal on Scientific Computing, 27 (2006), pp. 1651–1668.
- [15] ———, *A taxonomy and comparison of parallel block multi-level preconditioners for the incompressible Navier–Stokes equations*, Journal of Computational Physics, 227 (2008), pp. 1790–1808.
- [16] HOWARD C. ELMAN, DAVID J. SILVESTER, AND ANDREW J. WATHEN, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Oxford University Press, 2014.
- [17] HOWARD C. ELMAN AND RAYMOND S. TUMINARO, *Boundary conditions in approximate commutator preconditioners for the Navier–Stokes equations*, Electronic Transactions on Numerical Analysis, 35 (2009), pp. 257–280.
- [18] MIKITO FURUICHI, DAVE A. MAY, AND PAUL J. TACKLEY, *Development of a Stokes flow solver robust to large viscosity jumps using a Schur complement approach with mixed precision arithmetic*, Journal of Computational Physics, 230 (2011), pp. 8835–8851.
- [19] ROLAND GLOWINSKI AND JINCHAO XU, *Numerical Methods for Non-Newtonian Fluids: Special Volume*, vol. 16 of Handbook of Numerical Analysis, North-Holland, 2011.
- [20] PIOTR P. GRINEVICH AND MAXIM A. OLSHANSKII, *An iterative method for the Stokes-type problem with variable viscosity*, SIAM Journal on Scientific Computing, 31 (2009), pp. 3959–3978.
- [21] VINCENT HEUVELINE AND FRIEDHELM SCHIEWECK, *On the inf-sup condition for higher order mixed FEM on meshes with hanging nodes*, ESAIM: Mathematical Modelling and Numerical Analysis, 41 (2007), pp. 1–20.
- [22] KOLUMBAN HUTTER, *Theoretical Glaciology*, Mathematical Approaches to Geophysics, D. Reidel Publishing Company, 1983.
- [23] TOBIN ISAAC, NOEMI PETRA, GEORG STADLER, AND OMAR GHATTAS, *Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet*, Journal of Computational Physics, 296 (2015), pp. 348–368.
- [24] TOBIN ISAAC, GEORG STADLER, AND OMAR GHATTAS, *Solution of nonlinear Stokes equations discretized by high-order finite elements on nonconforming and anisotropic meshes, with application to ice sheet dynamics*, SIAM Journal on Scientific Computing, 37 (2015), pp. B804–B833.
- [25] VOLKER JOHN, KRISTINE KAISER, AND JULIA NOVO, *Finite element methods for the incompressible Stokes equations with variable viscosity*, ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik, 96 (2016), pp. 205–216.
- [26] DAVID KAY, DANIEL LOGHIN, AND ANDREW WATHEN, *A preconditioner for the steady-state Navier–Stokes equations*, SIAM Journal on Scientific Computing, 24 (2002), pp. 237–256.
- [27] MARTIN KRONBICHLER, TIMO HEISTER, AND WOLFGANG BANGERTH, *High accuracy mantle convection simulation through modern numerical methods*, Geophysical Journal International, 191 (2012), pp. 12–29.

- [28] ANDERS LOGG, KENT-ANDRE MARDAL, AND GARTH WELLS, *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS book*, vol. 84, Springer Science & Business Media, 2012.
- [29] DAVE A. MAY, JED BROWN, AND LAETITIA LE POURHIET, *pTatin3D: High-performance methods for long-term lithospheric dynamics*, in SC14: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, IEEE Press, 2014, pp. 274–284.
- [30] DAVE A. MAY, JED BROWN, AND LAETITIA LE POURHIET, *A scalable, matrix-free multigrid preconditioner for finite element discretizations of heterogeneous Stokes flow*, *Computer Methods in Applied Mechanics and Engineering*, 290 (2015), pp. 496–523.
- [31] DAVE A. MAY AND LOUIS MORESI, *Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics*, *Physics of the Earth and Planetary Interiors*, 171 (2008), pp. 33–47.
- [32] DAN MCKENZIE, *The generation and compaction of partially molten rock*, *Journal of Petrology*, 25 (1984), pp. 713–765.
- [33] CLEMENS PECHSTEIN AND ROBERT SCHEICHL, *Weighted Poincaré inequalities*, *IMA Journal of Numerical Analysis*, 33 (2013), pp. 652–686.
- [34] NOEMI PETRA, HONGYU ZHU, GEORG STADLER, THOMAS J. R. HUGHES, AND OMAR GHATTAS, *An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model*, *Journal of Glaciology*, 58 (2012), pp. 889–903.
- [35] KUMBAKONAM R. RAJAGOPAL, *Mechanics of non-Newtonian fluids*, in *Recent Developments in Theoretical Fluid Mechanics*, G.P. Galdi and J. Necas, eds., vol. 291, Longman’s Scientific and Technical, 1993, pp. 129–162.
- [36] JOHANN RUDI, A. CRISIANO I. MALOSSI, TOBIN ISAAC, GEORG STADLER, MICHAEL GURNIS, YVES INEICHEN, COSTAS BEKAS, ALESSANDRO CURIONI, AND OMAR GHATTAS, *An extreme-scale implicit solver for complex PDEs: Highly heterogeneous flow in earth’s mantle*, in SC15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, ACM, 2015, pp. 5:1–5:12.
- [37] GERALD SCHUBERT, DONALD L. TURCOTTE, AND PETER OLSON, *Mantle Convection in the Earth and Planets*, Cambridge University Press, 2001.
- [38] DAVID J. SILVESTER, HOWARD C. ELMAN, DAVID KAY, AND ANDREW J. WATHEN, *Efficient preconditioning of the linearized Navier–Stokes equations for incompressible flow*, *Journal of Computational and Applied Mathematics*, 128 (2001), pp. 261–279. *Numerical analysis 2000*, Vol. VII, Partial differential equations.
- [39] GEORG STADLER, MICHAEL GURNIS, CARSTEN BURSTEDDE, LUCAS C. WILCOX, LAURA ALISIC, AND OMAR GHATTAS, *The dynamics of plate tectonics and mantle flow: From local to global scales*, *Science*, 329 (2010), pp. 1033–1038.
- [40] ROLF STENBERG AND MANIL SURI, *Mixed hp finite element methods for problems in elasticity and Stokes flow*, *Numerische Mathematik*, 72 (1996), pp. 367–389.
- [41] HARI SUNDAR, GEORGE BIROS, CARSTEN BURSTEDDE, JOHANN RUDI, OMAR GHATTAS, AND GEORG STADLER, *Parallel geometric-algebraic multigrid on unstructured forests of octrees*, in SC12: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Salt Lake City, UT, 2012, ACM/IEEE.
- [42] RÜDIGER VERFÜRTH, *A Posteriori Error Estimation Techniques for Finite Element Methods*, Oxford University Press, 2013.
- [43] JENNIFER WORTHEN, GEORG STADLER, NOEMI PETRA, MICHAEL GURNIS, AND OMAR GHATTAS, *Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow*, *Physics of the Earth and Planetary Interiors*, 234 (2014), pp. 23–34.