SYMMETRIC INTERIOR PENALTY DISCONTINUOUS GALERKIN DISCRETIZATIONS AND BLOCK PRECONDITIONING FOR HETEROGENEOUS STOKES FLOW[*]

D. E. CHARRIER[†], D. A. MAY[‡], AND S. M. SCHNEPP[§]

Abstract. Provable stable arbitrary order symmetric interior penalty (SIP) discontinuous Galerkin discretizations of heterogeneous, incompressible Stokes flow utilizing Q_k^2–Q_{k-1} elements and hierarchical Legendre basis polynomials are developed and investigated. For solving the resulting linear system, a block preconditioned iterative method is proposed. The nested viscous problem is solved by a hp-multilevel preconditioned Krylov subspace method. For the p-coarsening, a two-level method utilizing element-block Jacobi preconditioned iterations as a smoother is employed. Piecewise bilinear (Q_1^2) and piecewise constant (Q_0^2) p-coarse spaces are considered. Finally, Galerkin h-coarsening is proposed and investigated for the two p-coarse spaces considered. Through a number of numerical experiments, we demonstrate that utilizing the Q_1^2 coarse space results in the most robust hp-multigrid method for heterogeneous Stokes flow. Using this Q_1^2 coarse space we observe that the convergence of the overall Stokes solver appears to be robust with respect to the jump in the viscosity and only mildly depending on the polynomial order k. It is demonstrated and supported by theoretical results that the convergence of the SIP discretizations and the iterative methods rely on a sharp choice of the penalty parameter based on local values of the viscosity.

1. Introduction.

1.1. Background and motivations. Earth exhibits a diverse range of unique geological processes: mountain building, subduction and continental rifting, earthquakes, and volcanism. These phenomena are the result of multiphase, history-dependent, large-deformation processes spanning million-year time scales.

Computational models provide a viable technique to study the evolution in both space and time of geological processes. A prototypical continuum description of the behavior of rocks is stationary, incompressible Stokes flow with Boussinesq approximation [44, 40]:

$$(1.1) \quad \begin{aligned} -\mathbf{div}\big[\bar{\eta}(\boldsymbol{u}, p, T, \theta(x))\,\boldsymbol{\varepsilon}(\boldsymbol{u})\big] + \nabla p &= \rho_0(\theta)\big(1 - \alpha(\theta)(T - T_0)\big)\hat{\boldsymbol{g}}, \\ -\mathrm{div}(\boldsymbol{u}) &= 0, \end{aligned}$$

[†]Fachbereich Mathematik, Technische Universität Darmstadt, AG Numerik und Wissenschaftliches Rechnen, 64293 Darmstadt, Germany, Institute of Geophysics, ETH Zürich, 8092 Zürich, Switzerland, and School of Engineering and Computing Sciences, Durham University, Lower Mountjoy South Road, Durham DH1 3LE, UK (dominic.e.charrier@durham.ac.uk).

[‡]Institute of Geophysics, ETH Zürich, 8092 Zürich, Switzerland, and Department of Earth Sciences, University of Oxford, OX1 3AN Oxford, UK (david.may@earth.ox.ac.uk).

[§]Institute of Geophysics, ETH Zürich, 8092 Zürich, Switzerland (mail@saschaschnepp.net).

where $\boldsymbol{u}$, p, $\boldsymbol{\varepsilon}(\cdot)$ are the velocity, pressure, and strain rate, respectively, T is the temperature, θ is the material composition, $\bar{\eta}$ is the effective viscosity, ρ_0 is the reference density at reference temperature T_0, and $\hat{\boldsymbol{g}}$ is the gravity vector. The conservation of momentum and mass for the creeping fluid is coupled with the conservation of energy equation,

$$(1.2) \qquad \rho_0(\theta)C_p \frac{DT}{Dt} = \mathrm{div}\big(k(\theta)\nabla T\big) + Q(\theta),$$

where C_p is the specific heat at constant pressure, k the conductivity, and Q the external heat source, and the evolution of the composition,

$$(1.3) \qquad \frac{\partial \theta}{\partial t} + \boldsymbol{u} \cdot \nabla \theta = 0.$$

From high-pressure and temperature laboratory experiments of minerals, it is known that rocks exhibit thermally activated creep and follow an Arrhenius type law [37, 26]:

$$(1.4) \qquad \bar{\eta}(u, p, T, \theta) = A(\theta)\varepsilon_{II}^{n(\theta)} \exp\left[\frac{E(\theta) + pV(\theta)}{n(\theta)RT}\right],$$

where A is a compositional dependent experimentally determined constant, ε_{II} is the second invariant of the strain rate tensor, E, V are the activation energy and activation volume, R is the universal gas constant, and n is the power-law exponent. To facilitate brittle behavior at low temperature, the ductile creep laws are augmented with a plasticity model (e.g., Drucker–Prager [37]).

When such a composite flow law is applied to geodynamics scenarios, the effective viscosity ($\bar{\eta}$) is highly heterogeneous. At depths > 200 km, ductile behavior dominates and the viscosity profile can to first order be characterized by a smooth, exponential function. Above, due to material failure or compositional variations associated with crustal layers, the viscosity profile will be discontinuous and can possess jumps on the order of 10^4–10^8 Pa s. Realistic forward models of both mantle- and crust-scale simulations are adversely affected by the degree of heterogeneity within the viscosity due to both accuracy concerns associated with the particular spatial discretization used and a lack of solver (linear and nonlinear) robustness.

1.2. Related work. Geodynamics forward models of incompressible Stokes which permit highly heterogeneous viscosity structures have traditionally utilized finite difference (FD) (e.g., [48, 49]), finite volume (e.g., [23, 41, 42]), or finite element (FE) (e.g., [35, 21, 33, 36, 9, 28, 31]) spatial discretizations. The relative merits of the FD and FE methods for geodynamics applications can be broadly summarized as follows.

Staggered grid FD methods are "cheap" (few nonzeros in the stencil), and the general implementation is rather straightforward. However, geometric flexibility is limited, and boundary condition imposition is nontrivial. Introducing new physics may further require the development of a modified stencil; see, e.g., [22, 18]. In the context of nonlinear problems, Newton linearization causes stencil growth and thus increases the overall cost of the discretization.

While more expensive than FD methods (on the same grid), inf-sup stable FE methods permit geometric versatility and natural boundary conditions are trivial to impose. Spatial adaptivity (h) can be readily introduced without requiring redeveloping the underlying numerical method (e.g., [30, 13, 27]). Newton linearization does not cause the equivalent of stencil growth. For both FD and FE discretizations, robust multilevel preconditioners suitable for highly heterogeneous viscosity structures exist [42, 13, 27, 31].

Inf-sup stable discontinuous Galerkin (DG) methods of the interior penalty type for the Stokes equations can be constructed by using the tensor product element pairs Q_k^2–Q_{k-1} and Q_k^2–Q_{k-2} [43, 38], as well as the $H(\mathrm{div};\Omega)$-conforming Raviart–Thomas, Brezzi–Douglas–Marini, and Brezzi–Douglas–Fortin–Marini kind element pairs; see [47, 14, 3, 25].

The fact that adaptivity in space and approximation order (k or "p" in the preconditioner context) is realized with comparably less effort than for other discretizations makes DG very appealing for geodynamics applications given the very nature of mantle-lithosphere-crust systems. The inf-sup constants of stable DG discretizations are further not sensitive to the element aspect ratio which is highly desirable, for example, within a crustal scale model with a domain spanning $1000 \times 1000 \times 20$ km where the domain itself possesses a high aspect ratio, or if anisotropic refinement was employed [39, Theorem 9].

Regarding the solution of the equation system arising from symmetric interior penalty DG (SIP) [1, 2] based discretizations of incompressible Stokes flow, we note that for $H(\mathrm{div};\Omega)$-conforming discretizations, efficient preconditioners have been introduced very recently [3, 25]. Recent advances in developing efficient and robust solvers for interior penalty DG discretizations of second order elliptic problems with heterogeneous coefficients involve the algebraic multigrid preconditioner proposed in [8, 7], as well as the two-level methods proposed in [15] and [45, 46].

To the best of our knowledge, employing DG methods for heterogeneous Stokes flow problems in geodynamics was so far considered only in [29]. Preconditioning was not discussed there.

1.3. Contributions. We examine the applicability of using mixed SIP Stokes discretizations for studying heterogeneous, incompressible Stokes flow problems associated with prototype problems arising in geodynamics. Through comparison with an analytic solution employing a discontinuous viscosity structure, we numerically demonstrate that the discretization yields optimal order of accuracy for Q_k^2–Q_{k-1} elements for $k = 1, \ldots, 6$.

Our main contribution is the development of a preconditioned iterative method for the discrete saddle point system resulting from the Stokes discretization. To this end, we follow a block preconditioning approach, e.g., [20]. The nested viscous problem is solved by an hp-multilevel preconditioned Krylov subspace method. For the p-coarsening, a two-level method utilizing element-block Jacobi preconditioned iterations as a smoother is employed. Two p-coarse spaces are considered: the space of elementwise constants and the space of continuous, elementwise bilinear functions. Through numerical experiments with heterogeneous viscosity, we demonstrate that the variant utilizing the elementwise bilinear coarse space has a convergence rate which is independent of the number of elements, largely insensitive to the jump in viscosity, and only weakly dependent on the polynomial order.

The heterogeneous nature of the viscosity in geodynamics applications requires a careful choice of the SIP penalty parameters. We provide a brief analysis of the influence of the penalty parameters on the discretization error as well as on the quality of the element-block Jacobi smoothers.

1.4. Limitations. We restrict ourselves to linear problems with elementwise constant viscosity distributions in this study.

2. Governing equations. Neglecting nonlinearities and the effect of temperature on the viscosity, we restrict ourselves to a model that is solely dependent on the

material composition of the rocks in Earth's mantle:

$$-\mathbf{div}(2\,\eta(\theta)\,\boldsymbol{\varepsilon}(\boldsymbol{u})) + \nabla p = \boldsymbol{f}(\theta) \qquad \text{in } \Omega, \tag{2.1a}$$

$$\mathrm{div}(\boldsymbol{u}) = 0 \qquad \text{in } \Omega, \tag{2.1b}$$

where $\boldsymbol{u}$ is a velocity field, p is a pressure, η denotes the viscosity, $\boldsymbol{f}$ denotes a volumetric force, θ denotes the material composition with $0 \le \theta \le 1$, and $\boldsymbol{\varepsilon}(\boldsymbol{u})$ denotes the (linearized) strain rate tensor with $\varepsilon_{ij}(\boldsymbol{u}) = \frac{1}{2}(\partial_{x_j} u_i + \partial_{x_i} u_j)$, $i, j = 1, \ldots, d$. Further, $\Omega \subset \mathbb{R}^2$ denotes a rectangular domain with boundary $\partial\Omega = \overline{\partial\Omega_{\mathrm{M}} \cup \partial\Omega_{\mathrm{N}}}$ consisting of a Neumann part $\partial\Omega_{\mathrm{N}}$ and a Navier part $\partial\Omega_{\mathrm{M}}$. We require the velocity and pressure to satisfy homogeneous Neumann boundary conditions,

$$(2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\boldsymbol{I})\boldsymbol{n} = 0 \qquad \text{on } \partial\Omega_{\mathrm{N}}, \tag{2.1c}$$

as well as homogeneous Navier boundary conditions,

$$\boldsymbol{u} \cdot \boldsymbol{n} = 0, \qquad \boldsymbol{t} \cdot 2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{u})\boldsymbol{n} = 0 \qquad \text{on } \partial\Omega_{\mathrm{M}}, \tag{2.1d}$$

where $\boldsymbol{n}$ denotes the outward normal to the boundary, and $\boldsymbol{t}$ denotes a vector belonging to the tangential space of $\partial\Omega$. In case the Neumann boundary is empty, we ensure uniqueness of the pressure solution by enforcing the following constraint:

$$\int_\Omega p\,\mathrm{d}\boldsymbol{x} = 0. \tag{2.1e}$$

We further require that the domain Ω is restrained against rigid motions (assumption (A1)).

Let us denote by $L^2(\Omega)$ and $H^1(\Omega)$ the usual Sobolev spaces and by $\|\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_{H^1(\Omega)}$ their norms.

Let us assume that the viscosity $\eta \in L^2(\Omega)$ is bounded according to

$$0 < \eta_{\min} \le \eta \le \eta_{\max}, \tag{2.2}$$

(assumption (AS1)) and that $\boldsymbol{f} \in L^2(\Omega)^2$ (assumption (AS2)); then it is well-known that problem (2.1a)–(2.1d) subject to the other named constraints admits a unique weak solution; see, e.g., [11].

3. Computational grid and trace operators. Let $\mathcal{T}_h$ be a regular Cartesian grid on Ω (assumption (Ah1)). We refer to the disjoint open sets $K \in \mathcal{T}_h$ as elements and denote their diameter by h. The number of elements is denoted by N_K. Finally, $\boldsymbol{n}$ denotes the outward normal unit vector to the element boundary ∂K.

An interior face of $\mathcal{T}_h$ is the $d-1$ dimensional intersection $\partial K^+ \cap \partial K^-$, where K^+ and K^- are two adjacent elements of $\mathcal{T}_h$. Similarly, a boundary face of $\mathcal{T}_h$ is the $d-1$ dimensional intersection $\partial K \cap \partial\Omega$ which consists of entire faces of ∂K. We denote by Γ_h the union of all interior faces of $\mathcal{T}_h$, denote by Γ_{N} and Γ_{M} the union of all boundary faces belonging to the Neumann part and the Navier part of the boundary, respectively, and set $\Gamma = \Gamma_h \cup \Gamma_{\mathrm{N}} \cup \Gamma_{\mathrm{M}}$. Here and in the following, we refer generically to a "face" although we consider only two-dimensional problems in this paper.

Let $K \in \mathcal{T}_h$; we denote by $H^s(K)$ the space of real-valued functions $v \in L^2(K)$ such that the function v and its weak derivatives up to order s are measurable and square integrable in K. We will denote the norms on all three spaces $H^s(K)$, $H^s(K)^2$, and $H^s(K)^{2\times2}$ by the symbol $\|\cdot\|_{s,K}$. We further introduce the broken Sobolev space

$$(3.1) \qquad H^s(\mathcal{T}_h) = \{v \in L^2(\Omega) \colon v|_K \in H^s(K),\, K \in \mathcal{T}_h\}.$$

Let us denote the norms of $H^s(\mathcal{T}_h)$, $H^s(\mathcal{T}_h)^2$, and $H^s(\mathcal{T}_h)^{2\times 2}$ by the symbol $\|\cdot\|_{H^s(\mathcal{T}_h)}$.

Let $q \in H^1(\mathcal{T}_h)$, and let φ belong to either $H^1(\mathcal{T}_h)$, $H^1(\mathcal{T}_h)^2$, or $H^1(\mathcal{T}_h)^{2\times 2}$. Let $e \in \Gamma_h$ be an interior face shared by the elements K^+ and K^-. Let $\varphi^\pm$ and $q^\pm$ denote the traces of φ and q on e from the interior of $K^\pm$, respectively. Further, let $\boldsymbol{n}^\pm$ denote the outward normal unit vector to the boundary $\partial K^\pm$. We define the mean value $\{\!\{\varphi\}\!\}$ and the jump $[\![q\boldsymbol{n}]\!]$ at $\boldsymbol{x} \in e$ by

$$(3.2) \qquad \{\!\{\varphi\}\!\} = \frac{1}{2}(\varphi^+ + \varphi^-), \quad [\![q\boldsymbol{n}]\!] = q^+\boldsymbol{n}^+ + q^-\boldsymbol{n}^-.$$

Let $\boldsymbol{w}$ denote a vector-valued function in $H^1(\mathcal{T}_h)^2$, and let $\boldsymbol{w}^\pm$ denote its traces on e from the interior of $K^\pm$. We define the jumps $[\![\boldsymbol{w} \otimes \boldsymbol{n}]\!]$ and $[\![\boldsymbol{w} \cdot \boldsymbol{n}]\!]$ at $\boldsymbol{x} \in e$ by

$$(3.3) \qquad [\![\boldsymbol{w} \otimes \boldsymbol{n}]\!] = \boldsymbol{w}^+ \otimes \boldsymbol{n}^+ + \boldsymbol{w}^- \otimes \boldsymbol{n}^-, \qquad [\![\boldsymbol{w} \cdot \boldsymbol{n}]\!] = \boldsymbol{w}^+ \cdot \boldsymbol{n}^+ + \boldsymbol{w}^- \cdot \boldsymbol{n}^-,$$

where "$\otimes$" denotes the dyadic product.

4. Discretization of the Stokes problem. Let us introduce $\eta^K = \eta|_K$, $K \in \mathcal{T}_h$, and define

$$(4.1) \qquad \eta_{\max}^e = \begin{cases} \max\left\{\eta^{K^+}, \eta^{K^-}\right\}, & e \in \Gamma_h, \\ \eta^K, & e \in \Gamma_{\mathrm{M}}. \end{cases}$$

For simplicity, we assume here that the viscosity is elementwise constant (assumption (ASh1)). Let us additionally define the constants

$$(4.2) \qquad C_{\mathrm{tr}}(k,e) = \begin{cases} \max\left\{C_{\mathrm{tr}}(k,e,K^+), C_{\mathrm{tr}}(k,e,K^-)\right\}, & e \in \Gamma_h, \\ C_{\mathrm{tr}}(k,e,K), & e \in \Gamma_{\mathrm{M}}, \end{cases}$$

where the constants $C_{\mathrm{tr}}(k,e,K)$, $K \in \mathcal{T}_h$, stem from the following discrete trace inequality [24].

LEMMA 4.1 (discrete trace inequality). *Let K be an affine quadrilateral, and let e be an edge belonging to the boundary of K. Then, it holds that*

$$(4.3) \qquad \|\varphi_h\|_{L^2(e)}^2 \le C_{\mathrm{tr}}(k,e,K)\,\|\varphi_h\|_{L^2(K)}^2 \qquad \forall \varphi_h \in Q_k(K)$$

with the trace inequality constant

$$(4.4) \qquad C_{\mathrm{tr}}(k,e,K) = (k+1)^2\,\frac{|e|}{|K|}$$

We approximate velocity and pressure in the discontinuous FE spaces

$$(4.5) \qquad \boldsymbol{V}_k = \left\{\boldsymbol{v} \in L^2(\Omega)^2 \colon \boldsymbol{v}|_K \in Q_k(K)^2,\, K \in \mathcal{T}_h\right\},$$

$$(4.6) \qquad M_{k-1} = \begin{cases} \left\{q \in L^2(\Omega) \colon q|_K \in Q_{k-1}(K),\, K \in \mathcal{T}_h\right\}, & |\partial\Omega_{\mathrm{N}}| > 0, \\ \left\{q \in L^2(\Omega) \setminus \mathbb{R} \colon q|_K \in Q_{k-1}(K),\, K \in \mathcal{T}_h\right\} & \text{else,} \end{cases}$$

where $Q_k(K)$ is the space of polynomials of maximum degree k in each variable on the mesh cell $K \in \mathcal{T}_h$.

As approximation to (2.1a)–(2.1e), we then consider the problem of finding $\boldsymbol{u_h} \in \boldsymbol{V}_k$ and $p_h \in M_{k-1}$ such that

$$(4.7) \qquad \mathcal{A}_h(\boldsymbol{u}_h, \boldsymbol{v}_h) + \mathcal{B}_h(\boldsymbol{v}_h, p_h) = \mathcal{F}_h(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_k,$$

$$(4.8) \qquad \mathcal{B}_h(\boldsymbol{u}_h, q_h) \qquad\qquad = \mathcal{G}_h(q_h) \qquad \forall q_h \in M_{k-1},$$

where we use a SIP form $\mathcal{A}_h$ and a form $\mathcal{B}_h$ similar to the one used in [43]:

$$
\begin{aligned}
\mathcal{A}_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = &\sum_{K \in \mathcal{T}_h} \int_K 2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{u}_h) : \boldsymbol{\varepsilon}(\boldsymbol{v}_h)\,\mathrm{d}\boldsymbol{x} \\
(4.9) \qquad & - \sum_{e \in \Gamma_h} \int_e \{2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{u}_h)\} : [\![\boldsymbol{v}_h \otimes \boldsymbol{n}]\!]\,\mathrm{d}s - \sum_{e \in \Gamma_h} \int_e \{2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{v}_h)\} : [\![\boldsymbol{u}_h \otimes \boldsymbol{n}]\!]\,\mathrm{d}s \\
& - \sum_{e \in \Gamma_M} \int_e (\boldsymbol{n} \cdot 2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{u}_h)\boldsymbol{n})(\boldsymbol{v}_h \cdot \boldsymbol{n})\mathrm{d}s - \sum_{e \in \Gamma_M} \int_e (\boldsymbol{n} \cdot 2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{v}_h)\boldsymbol{n})(\boldsymbol{u}_h \cdot \boldsymbol{n})\mathrm{d}s \\
& + \sum_{e \in \Gamma_h} \delta_e \int_e [\![\boldsymbol{u}_h \otimes \boldsymbol{n}]\!] : [\![\boldsymbol{v}_h \otimes \boldsymbol{n}]\!]\,\mathrm{d}s + \sum_{e \in \Gamma_M} \delta_e \int_e (\boldsymbol{u}_h \cdot \boldsymbol{n})\,(\boldsymbol{v}_h \cdot \boldsymbol{n})\,\mathrm{d}s
\end{aligned}
$$

and

$$
\begin{aligned}
(4.10) \qquad \mathcal{B}_h(\boldsymbol{u}_h, q_h) = &\sum_{K \in \mathcal{T}_h} \int_K -\mathrm{div}(\boldsymbol{u}_h)\,q_h\,\mathrm{d}\boldsymbol{x} + \sum_{e \in \Gamma_h} \int_e \{q_h\}\,[\![\boldsymbol{u}_h \cdot \boldsymbol{n}]\!]\,\mathrm{d}s \\
& + \sum_{e \in \Gamma_M} \int_e q_h\,(\boldsymbol{u}_h \cdot \boldsymbol{n})\,\mathrm{d}s,
\end{aligned}
$$

$$(4.11) \qquad \mathcal{F}_h(\boldsymbol{v}_h) = \sum_{K \in \mathcal{T}_h} \int_K \boldsymbol{f} \cdot \boldsymbol{v}_h\,\mathrm{d}\boldsymbol{x}, \qquad \mathcal{G}_h(q_h) = 0,$$

with $\boldsymbol{u}_h, \boldsymbol{v}_h \in \boldsymbol{V}_k$, $q_h \in M_{k-1}$, and the facewise penalties $\delta_e = \sigma_e\,C_{\mathrm{tr}}(k, e)$, $e \in \Gamma$. The parameters σ_e are the so-called penalty or stability parameters that must be chosen sufficiently large (to be specified below) to guarantee that the bilinear form $\mathcal{A}_h$ is coercive on the discrete space $\boldsymbol{V}_k$.

Consistency of the discrete variational problem can be shown by following the proof of [43, Lemma 7.5].

4.1. Stability of the Stokes discretization. For the analysis of the Stokes discretization, it is necessary to introduce the functionals

$$(4.12) \qquad \|q\|_{0,h}^2 = \sum_{K \in \mathcal{T}_h} \|q\|_{L^2(K)}^2 + \sum_{e \in \Gamma_h} k^{-2}\,h\,\|\{q\}\|_{L^2(e)}^2 + \sum_{e \in \Gamma_M} k^{-2}\,h\,\|q\|_{L^2(e)}^2,$$

$$(4.13) \qquad \|\boldsymbol{v}\|_{1,h}^2 = \sum_{K \in \mathcal{T}_h} \|\nabla \boldsymbol{v}\|_{L^2(K)}^2 + \sum_{e \in \Gamma_h} k^2\,h^{-1}\,\|[\![\boldsymbol{v} \otimes \boldsymbol{n}]\!]\|_{L^2(e)}^2$$
$$+ \sum_{e \in \Gamma_M} k^2\,h^{-1}\,\|\boldsymbol{v} \cdot \boldsymbol{n}\|_{L^2(e)}^2,$$

$$(4.14) \qquad \|\!|\boldsymbol{v}|\!\|_{1,h}^2 = \|\boldsymbol{v}\|_{1,h}^2 + \sum_{e \in \Gamma_h} k^{-2}\,h\,\|\{\nabla \boldsymbol{v}\}\|_{L^2(e)}^2 \sum_{e \in \Gamma_M} k^{-2}\,h\,\|\nabla \boldsymbol{v}\|_{L^2(e)}^2$$

with $\boldsymbol{v} \in H^2(\mathcal{T}_h)^2$ and $q \in L^2(\Omega)$. By definition, $\|\cdot\|_{0,h}$ is a norm on $L^2(\Omega)$. Under assumption (A1), we assume that $\|\cdot\|_{1,h}$ is a norm on $H^2(\mathcal{T}_h)^2$ for the considered boundary conditions (assumption (ASh2)).

Stability of the Stokes discretization follows from the discrete inf-sup stability of $\mathcal{B}_h$, the discrete coercivity of $\mathcal{A}_h$, and Brezzi's lemma.

LEMMA 4.2 (discrete inf-sup stability). *Let the assumptions* (Ah1) *be satisfied and let* $k \geq 1$, $K \in \mathcal{T}_h$. *Then, it holds that*

$$(4.15) \qquad \sup_{\mathbf{0} \neq \boldsymbol{u}_h \in \boldsymbol{V}_k} \frac{\mathcal{B}_h(\boldsymbol{u}_h, q_h)}{\|\boldsymbol{u}_h\|_{1,h}} \geq \beta_h \|q_h\|_{L^2(\Omega)} = c\, k^{-1} \|q_h\|_{L^2(\Omega)} \qquad \forall q_h \in M_{k-1}$$

with $c > 0$ *independent of* h *and* k.

Proof. This follows from [38, Theorem 6.2] for $k \geq 2$. The case $k = 1$ is covered by [38, Theorem 6.12.]. □

We rely on a discrete Korn inequality to show coercivity of the bilinear form $\mathcal{A}_h$.

LEMMA 4.3 (discrete Korn inequality [10]). *Let the assumptions* (A1) *and* (Ah1) *be satisfied. Then, it holds* $\forall \boldsymbol{v} \in H^1(\mathcal{T}_h)^2$ *that*

$$(4.16) \qquad \left(\|\boldsymbol{\varepsilon}(\boldsymbol{v})\|_{L^2(\Omega)}^2 + \sum_{e \in \Gamma_h} h^{-1} \|[\![\boldsymbol{v} \otimes \boldsymbol{n}]\!]\|_{L^2(e)}^2 \right) \geq C_{\mathrm{Kn},h} \|\nabla \boldsymbol{v}\|_{L^2(\Omega)}^2$$

with a constant $C_{\mathrm{Kn},h} > 0$ *independent of* h.

Proof. This is a slight modification to the result in [10] relying on assumption (A1) as well as on the affinity and shape regularity of the grid elements (Ah1). □

LEMMA 4.4 (discrete coercivity). *Let the assumptions* (A1), (Ah1), *and* (ASh1)–(ASh2) *hold. Assume that the penalty parameters are chosen according to*

$$(4.17) \qquad \sigma_e > N_e\, \eta_{\max}^e\, (1 + \tau) \begin{cases} 1, & e \in \Gamma_h, \\ 2, & e \in \Gamma_{\mathrm{M}}, \end{cases}$$

with $\eta_{\max}^e$ *from* (4.1) *and with* $N_e = 4$ *denoting the number of faces of an element. Then, it holds that*

$$(4.18) \qquad \mathcal{A}_h(\boldsymbol{v}_h, \boldsymbol{v}_h) \geq \alpha_h \|\boldsymbol{v}_h\|_{1,h}^2 \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_k$$

with a constant $\alpha_h = C\, 2\, \eta_{\min} > 0$ *independent of* h *and* k. *Here,* $\eta_{\min}$ *is from assumption* (AS1). *The parameter* $\tau > 0$ *is a small value that is necessary to use the discrete Korn inequality.*

Proof. The proof is based for the most part for the stability analysis for the *elementwise* penalty approach presented in [16, section 3.3.3.]. Let $\boldsymbol{u}_h = \boldsymbol{v}_h$, $\boldsymbol{u}_h, \boldsymbol{v}_h \in \boldsymbol{V}_k$; we apply a Young's inequality to the second and third terms of (4.9). Thus, we obtain

$$\mathcal{A}_h(\boldsymbol{v}_h, \boldsymbol{v}_h) \geq \sum_{K \in \mathcal{T}_h} \int_K 2\, \eta\, \boldsymbol{\varepsilon}(\boldsymbol{v}_h) : \boldsymbol{\varepsilon}(\boldsymbol{v}_h)\, \mathrm{d}s - \sum_{e \in \Gamma_h} \epsilon_e^{-1} \int_e \{2\, \eta\, \boldsymbol{\varepsilon}(\boldsymbol{v}_h)\}^2\, \mathrm{d}s$$

$$- \sum_{e \in \Gamma_{\mathrm{M}}} \epsilon_e^{-1} \int_e (2\, \eta\, \boldsymbol{\varepsilon}(\boldsymbol{v}_h))^2\, \mathrm{d}s + \sum_{e \in \Gamma_h} (\delta_e - \epsilon_e) \|[\![\boldsymbol{v}_h \otimes \boldsymbol{n}]\!]\|_{L^2(e)}^2$$

$$(4.19) \qquad\qquad\qquad + \sum_{e \in \Gamma_{\mathrm{M}}} (\delta_e - \epsilon_e) \|\boldsymbol{v}_h \cdot \boldsymbol{n}\|_{L^2(e)}^2,$$

where $\epsilon_e > 0\ \forall e \in \Gamma_h \cup \Gamma_{\mathrm{M}}$. As a next step, we will bound the second and third terms of (4.19) below. We have for interior faces $e \in \Gamma_h$

$$(4.20) \qquad \int_e \{2\,\eta\,\varepsilon(\boldsymbol{v}_h)\}^2 \,\mathrm{d}s \leq \int_e \frac{1}{2}\left((2\,\eta\,\varepsilon(\boldsymbol{v}_h))|_{K^+}^2 + (2\,\eta\,\varepsilon(\boldsymbol{v}_h))|_{K^-}^2\right)\,\mathrm{d}s$$

$$\overset{\mathrm{(I)}}{\leq} 2\,\eta_{\max}^e\,\frac{1}{2}C_{\mathrm{tr}}(k,e)\left(\int_{K^+} 2\,\eta\,\varepsilon(\boldsymbol{v}_h) : \varepsilon(\boldsymbol{v}_h)\,\mathrm{d}\boldsymbol{x}\right.$$

$$\left.+ \int_{K^-} 2\,\eta\,\varepsilon(\boldsymbol{v}_h) : \varepsilon(\boldsymbol{v}_h)\,\mathrm{d}\boldsymbol{x}\right).$$

In step (I), we have applied the trace inequality (4.3) from lemma 4.1 and further have introduced constant $C_{\mathrm{tr}}(k,e)$ defined as in (4.2), as well as constant $\eta_{\max}^e$ defined as in (4.1). Analogously, we obtain for boundary faces $e \in \Gamma_\mathrm{M}$

$$(4.21) \qquad \int_e (2\,\eta\,\varepsilon(\boldsymbol{v}_h))^2\,\mathrm{d}\boldsymbol{x} \leq 2\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e)\int_K 2\,\eta\,\varepsilon(\boldsymbol{v}_h) : \varepsilon(\boldsymbol{v}_h)\,\mathrm{d}s.$$

Inserting (4.20)–(4.21) into (4.19) leads to

(4.22)

$$\mathcal{A}_h(\boldsymbol{v}_h, \boldsymbol{v}_h) \geq \sum_{K\in\mathcal{T}_h}\int_K 2\,\eta\,\varepsilon(\boldsymbol{v}_h) : \varepsilon(\boldsymbol{v}_h)$$

$$\times\left(1 - \sum_{e\subset\partial K, e\in\Gamma_h}\epsilon_e^{-1}\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e) - \sum_{e\subset\partial K, e\in\Gamma_\mathrm{M}}\epsilon_e^{-1}\,2\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e)\right)\mathrm{d}\boldsymbol{x}$$

$$+ \sum_{e\in\Gamma_h}\left(\sigma_e\,C_{\mathrm{tr}}(k,e) - \epsilon_e\right)\|[\![\boldsymbol{v}_h\otimes\boldsymbol{n}]\!]\|_{L^2(e)}^2$$

$$+ \sum_{e\in\Gamma_\mathrm{M}}\left(\sigma_e\,C_{\mathrm{tr}}(k,e) - \epsilon_e\right)\|\boldsymbol{v}_h\cdot\boldsymbol{n}\|_{L^2(e)}^2,\ \boldsymbol{v}_h\in\boldsymbol{V}_k.$$

We see that this expression is positive for any $\boldsymbol{v}_h \in \boldsymbol{V}_k$ if

$$(4.23) \qquad \sigma_e\,C_{\mathrm{tr}}(k,e) > \epsilon_e > \sigma_e^*\,C_{\mathrm{tr}}(k,e) = N_e\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e), e\in\Gamma_h,$$

$$(4.24) \qquad \sigma_e\,C_{\mathrm{tr}}(k,e) > \epsilon_e > \sigma_e^{**}\,C_{\mathrm{tr}}(k,e) = N_e\,2\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e),\ e\in\Gamma_\mathrm{M},$$

where N_e denotes the number of faces of a quadrilateral element. In the following, we choose the penalty parameters according to (4.17). We add a small value $\tau\,\eta_{\max}^e/N_e \geq \tau\,\eta_{\min}/N_e$ once/twice to each penalty parameter σ_e in order to use the discrete Korn inequality (4.16) from Lemma 4.3 in step (I) of the next derivations. Thus, if we choose the penalty value according to (4.17), we obtain from (4.22) that

$$(4.25) \qquad \mathcal{A}_h(\boldsymbol{v}_h, \boldsymbol{v}_h) \overset{\mathrm{(I)}}{\geq} C\left(C_{\mathrm{Kn},h}\,2\,\eta_{\min}\sum_{K\in\mathcal{T}_h}\|\nabla\boldsymbol{v}_h\|_{L^2(K)}^2\right.$$

$$+ \sum_{e\in\Gamma_h} 2\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e)\,\|[\![\boldsymbol{v}_h\otimes\boldsymbol{n}]\!]\|_{L^2(e)}^2$$

$$\left.+ \sum_{e\in\Gamma_\mathrm{M}} 2\,\eta_{\max}^e\,C_{\mathrm{tr}}(k,e)\,\|\boldsymbol{v}_h\cdot\boldsymbol{n}\|_{L^2(e)}^2\right)$$

$$\overset{\mathrm{(II)}}{\geq} C\,2\,\eta_{\min}\,\|\boldsymbol{v}_h\|_{1,h}^2,\ \boldsymbol{v}_h\in\boldsymbol{V}_k,$$

where $\eta_{\min}$ is from assumption (AS1), and $C_{\mathrm{Kn},h}$ is from the discrete Korn inequality (4.16) from Lemma 4.3. In step (II), we have used that due to the affinity and shape

regularity of the grid elements (Ah1), it holds that $h^{-1} \leq |e|/|K| \leq \gamma^2 \, h^{-1}$ for $K \in \mathcal{T}_h$ and $e \subset \partial K$, with γ denoting the shape regularity constant. We have further bounded σ_e below by $\eta_{\min} N_e$. The constant C is assigned a different value in every step and independent of h and k. $\square$

We typically choose the penalty parameters as the lower bound since estimate (4.17) is not totally sharp due to the utilized inequalities. Even smaller values can be chosen in practice [24]. We note that the parameter τ is set to zero in our computations. We further remark that estimate (4.17) assumes piecewise constant viscosity distributions. For piecewise polynomial viscosity distributions, it is necessary to replace k by $k \leftarrow k + m_\eta$ in the discrete trace inequality constant.

4.2. A priori error estimates for the Stokes discretization. Let us state the continuity properties of the forms $\mathcal{A}_h$ and $\mathcal{B}_h$.

LEMMA 4.5 (continuity of $\mathcal{A}_h$ and $\mathcal{B}_h$). *Under assumptions* (A1)–(AS2), (Ah1), *and* (ASh1), *bilinear forms* $\mathcal{A}_h$ *and* $\mathcal{B}_h$ *are continuous in the sense that*

$$(4.26) \qquad \mathcal{A}_h(\boldsymbol{u}, \boldsymbol{v}) \leq C_{a,h} \, |\!|\!|\boldsymbol{u}|\!|\!|_{1,h} \, |\!|\!|\boldsymbol{v}|\!|\!|_{1,h} \qquad \forall (\boldsymbol{u}, \boldsymbol{v}) \in H^2(\mathcal{T}_h)^2 \times H^2(\mathcal{T}_h)^2,$$

$$(4.27) \qquad \mathcal{B}_h(\boldsymbol{u}, q) \leq C_{b,h} \, \|\boldsymbol{u}\|_{1,h} \, \|q\|_{0,h} \qquad \forall (\boldsymbol{u}, q) \in H^2(\mathcal{T}_h)^2 \times H^1(\mathcal{T}_h)$$

with $C_{a,h} = (2\,\eta_{\max} + \max_{e \in \Gamma_h \cup \Gamma_{\mathrm{M}}}\{\sigma_e\}) > 0$ *and* $C_{b,h} = \sqrt{d} > 0$. *Here,* $\eta_{\max}$ *is from assumption* (AS1), *and* σ_e *denotes the penalty value on face* e. *Both constants* $C_{a,h}$ *and* $C_{b,h}$ *are independent of* h *and* k.

Proof. The proof of both inequalities follows from standard inequalities. See, e.g., the proofs of [43, Lemma 7.1] and [43, Lemma 7.2] for more details. $\square$

Further note that the norms $\|\cdot\|_{1,h}$ and $|\!|\!| \cdot |\!|\!|_{1,h}$ are equivalent on the discrete spaces $\boldsymbol{V}_k$ and M_{k-1}. This is also the case for the norms $\|\cdot\|_{0,h}$ and $\|\cdot\|_{L^2(\Omega)}$.

Using the consistency of the discretization, the discrete inf-sup stability of $\mathcal{B}_h$, the discrete coercivity of $\mathcal{A}_h$, the continuity of both bilinear forms together with the discrete equivalency of norms, as well as suitable elementwise hp-interpolants, we can derive the following a priori error estimate.

LEMMA 4.6. *Let the assumptions* (A1), (AS1)–(AS2), (Ah1), *and* (ASh1) *hold. Assume that the weak solution* $(\boldsymbol{u}, p)$ *to* (2.1a)–(2.1e) *belongs to* $H^m(K)^2 \times H^n(K)$, $K \in \mathcal{T}_h$, *with* $m \geq 2$ *and* $n \geq 1$. *Further, let* $\boldsymbol{u}_h \in \boldsymbol{V}_k$ *and* $p_h \in M_{k-1}$ *denote the discrete solution to problem* (4.7)–(4.8). *Then, it holds that*

$$(4.28) \qquad \begin{aligned} &\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,h} + \|p - p_h\|_{0,h} \\ &\leq C \sum_{K \in \mathcal{T}_h} \left(\frac{C_{a,h}}{\alpha_h} \frac{1}{\beta_h} \frac{h^{s-1}}{k^{m-3/2}} \|\boldsymbol{u}\|_{H^m(K)} + \frac{h^r}{k^n} \|p\|_{H^n(K)} \right) \end{aligned}$$

with $1 \leq s \leq \min\{k+1, m\}$ *and* $1 \leq r \leq \min\{k, n\}$. *The constant* $C > 0$ *is independent of* h *and* k *but depends on the shape regularity of the grid elements. The constant* $C_{a,h}$ *is from Lemma* 4.5, *and* α_h *is from Lemma* 4.4. *Note that this estimate holds up to the discrete inf-sup constant* β_h *from Lemma* 4.2 *that depends on* k.

Proof. The proof requires arguments analogous to the ones used for the proofs of [43, Lemma 8.1] and [43, Lemma 8.2]. $\square$

Utilizing a duality argument, additional requirements on the regularity of the solution to the continuous variational problem (and on the solution to the adjoint

problem), as well as hp-interpolants for the whole domain Ω, we can derive the following L^2 error estimate for the discrete velocity solution.

LEMMA 4.7 (L^2 *error*).   *Let the assumptions* (A1)–(AS2), (Ah1), *and* (ASh1) *hold. Assume further that the weak solution* $(\boldsymbol{u}, p)$ *to* (2.1a)–(2.1e) *belongs to* $H^{n+1}(\Omega)^2$ $\times H^n(\Omega)$ *with* $m \geq 1$. *Choose the penalty values according to* (4.17) *in Lemma* 4.4. *Let* $\boldsymbol{u}_h \in \boldsymbol{V}_k$ *and* $p_h \in M_{k-1}$ *denote the discrete solution to problem* (4.7)–(4.8). *Then, it holds that*

(4.29)

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^2(\Omega)} \leq C\, h^{\min(k+1,n+1)} \left(\frac{C_{a,h}^2}{\alpha_h} \frac{1}{\beta_h^2} \|\boldsymbol{u}\|_{H^{n+1}(\Omega)} + C_{a,h} \frac{1}{\beta_h} \|p\|_{H^n(\Omega)} \right),$$

where the constant $C > 0$ *is independent of* h *and* k *but depends on the shape regularity of the grid. See Lemma* 4.6 *for a definition of the remaining constants.*

Remark 4.8. We emphasize that the discretization errors depend on the size of the largest penalty parameter as well as on the viscosity contrast by means of the constants $C_{a,h}$ and α_h.

5. Preconditioning the Stokes system. The discrete variational problem (4.7)–(4.8) is equivalent to the saddle point problem

(5.1)
$$\begin{bmatrix} \underline{A} & \underline{B} \\ \underline{B}^{\mathrm{T}} & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{p} \end{bmatrix} = \begin{bmatrix} \underline{f} \\ \underline{g} \end{bmatrix}.$$

We solve (5.1) using a right preconditioned Krylov method, with an upper block triangular preconditioner $\underline{\mathcal{P}}$ of the form

(5.2)
$$\underline{\mathcal{P}} = \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{0} & \underline{S} \end{bmatrix},$$

where $\underline{S} = \underline{B}^T \underline{A}^{-1} \underline{B}$ is the pressure Schur complement.

Our implementation employs tensor products of pairwise orthogonal Legendre polynomials as basis functions for the pressure and the velocity components. The zero pressure average is not build into the pressure basis functions. In case the Neumann boundary is empty, we thus solve a singular system. As we typically use Krylov methods which are mathematically equivalent to GMRES, we then require that the right-hand side is consistent (e.g., we remove the constant pressure null space) [12, Theorem 2.4].

Noting that both $\underline{A}$ is symmetric positive definite (stemming from the SIP formulation) and $\underline{S}$ is symmetric positive definite (stemming from the inf-sup stability), this choice for $\underline{\mathcal{P}}$ will result in convergence in at most two iterations in exact arithmetic [19]. While optimal (in the sense of iteration counts), the definition of $\underline{\mathcal{P}}$ is not practical as it involves an exact inverse for $\underline{A}$ and $\underline{S}$. A practical Stokes preconditioner replaces $\underline{A}^{-1}$ and $\underline{S}^{-1}$ by spectrally equivalent operators such that their application of the inverse on a vector is significantly cheaper.

In our computations, we replace the Schur complement $\underline{S}$ by the pressure mass matrix scaled by the inverse of the element viscosity ($\underline{S}^*$). The proof of spectral equivalence between $\underline{S}$ and $\underline{S}^*$ for our DG spaces stems immediately from [19, Theorem 5.22]. Due to the use of an orthonormal basis for the pressure space, $\underline{S}^*$ is diagonal. The definition of $\underline{y} = \underline{A}^{-1} \underline{x}$ is further replaced with a preconditioned Krylov method with a fixed relative stopping condition which is described in detail in section 6.

6. Preconditioning the viscous block. When used in conjunction with a Krylov method, we are required to apply the action of $\underline{\mathcal{P}}^{-1}$ on an arbitrary vector $\underline{x} = (\underline{x}_u, \underline{x}_p)$. We consider replacing the definition $\underline{y}_u = \underline{A}^{-1}\underline{x}_u$ with a spectrally equivalent operation: solve $\underline{A}\underline{y}_u = \underline{x}_u$ for $\underline{y}_u$ using a preconditioned Krylov method such that at the ith iteration, $\|\underline{A}\underline{y}_u^i - \underline{x}_u\|/\|\underline{x}_u\| < \epsilon$. In order to develop an optimal and scalable preconditioner for $\underline{A}$, in this work we utilize an hp-multilevel preconditioner. The hp-multilevel preconditioner employs coarsening with respect to both the polynomial order k of the velocity function space ("p-coarsening") and the spatial resolution h. The rational for coarsening in both p and h will be elaborated below.

To introduce the hp-multilevel preconditioner, we first recall the basic two-level multigrid method (see Algorithm 1). The essential components of the multigrid algorithm are the fine level operator $\underline{A}$, the coarse level operator $\underline{\bar{A}}$, the restriction and prolongation operators $\underline{R}$, $\underline{P}$, which map vectors from the fine level to the coarse level (and vice versa), and the smoothing operator $\underline{\mathcal{M}}$.

Algorithm 1 Two-level multigrid.

1: **procedure** MGVCYCLE($\underline{A}$, $\underline{f}$, $\underline{\mathcal{M}}$, $\underline{R}$, $\underline{P}$, $\underline{\bar{A}}$, $\underline{y}$)
2:     Set $i = 0, \underline{u}^0 = \underline{y}$
3:     **repeat**
4:         $\underline{u}^i = \underline{u}^i + \underline{\mathcal{M}}^{-1}(\underline{f} - \underline{A}\,\underline{u}^i)$ ▷ presmooth m times
5:         $\underline{\bar{r}} = \underline{R}\,(\underline{f} - \underline{A}\,\underline{u}^i)$ ▷ restrict residual
6:         Solve $\underline{\bar{A}}\,\underline{\bar{e}} = \underline{\bar{r}}$ ▷ solve for the coarse grid correction
7:         $\underline{u}^i = \underline{u}^i + \underline{P}\,\underline{\bar{e}}$ ▷ prolongate error
8:         $\underline{u}^i = \underline{u}^i + \underline{\mathcal{M}}^{-\mathrm{T}}(\underline{f} - \underline{A}\,\underline{u}^i)$ ▷ postsmooth m times
9:         $\underline{u}^{i+1} = \underline{u}^i$ ▷ update for next iteration
10:        $i = i + 1$
11:    **until** converged
12:    $\underline{u} = \underline{u}^i$
13: **end procedure**

In the context of p-multigrid, restriction refers to mapping a discrete vector defined using a function space of order k to a function space of order s, where $0 \le s < k$. Contrary to traditional p-coarsening strategies with hierarchies like $\boldsymbol{V}_k \to \boldsymbol{V}_{k/2} \to \boldsymbol{V}_{k/4}\ldots$, here we follow [45, 46] and consider "aggressive" coarsening from $\boldsymbol{V}_k$ to an a priori defined p-coarse space in a single step. This implicitly defines a two-level hierarchy in p-space. In this work, we study two different p-coarse spaces, namely, the space of elementwise constants $\boldsymbol{V}_0$ and the space of d-linear functions $\boldsymbol{V}_1$. We will use the symbols $\underline{A}_0$ and $\underline{A}_1$ for the associated coarse grid operators.

The construction of the restriction and prolongation operators between different order basis functions is natural to implement as we have employed a hierarchical basis. Furthermore, the prolongation operators are identical to the transposed restriction operators. We define coarse operators via Galerkin projection. Denoting the prolongation from polynomial degree k to s via P_k^s, the coarse level operators we consider are thus given by

$$\underline{A}_0 = \left(\underline{P}_k^0\right)^{\mathrm{T}}\underline{A}\,\underline{P}_k^0, \quad \underline{A}_1 = \left(\underline{P}_k^1\right)^{\mathrm{T}}\underline{A}\,\underline{P}_k^1.$$

The smoother is defined as a Chebyshev iteration preconditioned with an element-block Jacobi operator that consists of the diagonal blocks of $\underline{A}$. The minimum (λ_0)

and maximum (λ_1) eigenvalue bounds required by Chebyshev are defined in the following manner. First, we estimate the maximum eigenvalue (λ^*) of $\underline{A}$ by performing 10 iterations of GMRES with a random right-hand-side vector. We then choose $\lambda_0 = 0.1\lambda^*$ and $\lambda_1 = 1.1\lambda^*$, respectively. The choice of factors 0.1 and 1.1 has been determined empirically; however, they are robust for variable coefficient scalar/vector elliptic problems and in fact are the default values used in PETSc's Chebyshev implementation.

The bilinear form associated with the element-block Jacobi operator is

(6.1)

$$
\begin{aligned}
\hat{\mathcal{A}}_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = \sum_{K \in \mathcal{T}_h} \Bigg(& \int_K 2\,\eta\,\boldsymbol{\varepsilon}(\boldsymbol{u}_h) : \boldsymbol{\varepsilon}(\boldsymbol{v}_h)\,\mathrm{d}\boldsymbol{x} \\
& - \sum_{\substack{e \subset \partial K \\ e \in \Gamma_h}} \int_e \eta^K\,\boldsymbol{\varepsilon}(\boldsymbol{u}_h^K) : (\boldsymbol{v}_h^K \otimes \boldsymbol{n}^K)\,\mathrm{d}s - \sum_{\substack{e \subset \partial K \\ e \in \Gamma_h}} \int_e \eta^K\,\boldsymbol{\varepsilon}(\boldsymbol{v}_h^K) : (\boldsymbol{u}_h^K \otimes \boldsymbol{n}^K)\,\mathrm{d}s \\
& + \sum_{\substack{e \subset \partial K \\ e \in \Gamma_h}} \delta_e \int_e (\boldsymbol{u}_h^K \otimes \boldsymbol{n}^K) : (\boldsymbol{v}_h^K \otimes \boldsymbol{n}^K)\,\mathrm{d}s + \text{ boundary terms} \Bigg)
\end{aligned}
$$

with $\boldsymbol{u}_h, \boldsymbol{v}_h \in \boldsymbol{V}_k$, $\boldsymbol{v}_h^K = \boldsymbol{v}_h|_K$, and $\boldsymbol{n}^K$ denoting the outward normal to the boundary of element $K \in \mathcal{T}_h$. Notice the division by two in the second and third terms stemming from the averaging on interior faces.

From similar arguments as in the proof of Lemma 4.4 follows that the form (6.1) is elliptic on $\boldsymbol{V}_k$. Consequently, the element-block Jacobi operator $\underline{\mathcal{M}}_k$ is symmetric and positive definite. Following the proofs of [46, Corollary 1 and equation (54)], one can then show that

(6.2)
$$
\kappa_2\left(\underline{\mathcal{M}}_k^{-1}\,\underline{A}\right) \le 1 + C_1 \max_{K \in \mathcal{T}_h} \left\{ \frac{\max_{e \subset \partial K}\sigma_e}{\min_{\mathbf{x} \in K}\eta} \right\} + C_2\,\frac{1}{\eta_{\min}}
$$

with constants $C_1 > 0$ and $C_2 > 0$ independent of h. This result emphasizes the importance of choosing the penalty parameters based on local values of the viscosity.

6.1. Preconditioning the coarse problem. In the context of high-resolution simulations, a pure p-multilevel preconditioner will never yield optimal $O(n)$ solve times due to the increasing cost of performing the solve on the coarsest level. This motivates us to employ a h-multigrid preconditioner for the coarse operators $\underline{A}_0$ and $\underline{A}_1$, respectively. To realize this, we have adopted standard multigrid techniques which have been developed for FD discretizations and low order finite discretizations. Below we elaborate on how these techniques are fused with our SIP spatial discretization.

6.1.1. Elementwise constant coarse problem. Let us denote by $\mathcal{A}_{h.0}$ the bilinear form inherited from $\mathcal{A}_h$ that corresponds to the elementwise constants. Discretizing $\mathcal{A}_{h.0}$ for an isoviscous fluid yields a stencil which mimics a standard five-point FD stencil [45]. Hence, heuristically it seems plausible to assume that any geometric multigrid preconditioner suitable for a five-point FD stencil should be appropriate to use as a preconditioner for the coarse grid solver associated with $\underline{A}_0$.

We first generate a hierarchy of meshes (with differing h) by isotropically coarsening the mesh defining $\underline{A}_0$. The maximum number of times coarsening can be applied, and thus the number of levels in the h-multilevel preconditioner, is determined by the spatial resolution of the grid. Note, however, that our implementation does not

support semicoarsening, and thus the finest grid must always employ an odd number of elements in the i and j directions. Between each mesh in the hierarchy, we have a restriction operator $\underline{R}_h$ defined by bilinear interpolation (Q_1^2) and again we will use $\underline{P}_h = \underline{R}_h^T$.

As in the p-multigrid implementation, we define coarse operators via Galerkin projection, e.g., $\underline{\bar{A}}_h = \underline{P}_h^T \underline{A}_0 \underline{P}_h$. The construction of Galerkin coarse operators is applied recursively for all levels in the mesh hierarchy. The smoother used within the h-multilevel hierarchy is Chebyshev preconditioned with Jacobi. The Chebyshev bounds are estimated similarly as for the p-coarsening smoother. On the coarsest level of the h-hierarchy, we will apply an exact LU factorization.

6.1.2. Elementwise bilinear coarse problem. In the case when the coarse space is Q_1^2, rather than leverage an FD analogue to build an h-multigrid preconditioner, we will exploit methods designed for low order finite elements. Specifically, we consider projecting the Q_1^2 discontinuous space into the space of continuous bilinear functions.

We first define the discontinuous to continuous projector as the transpose of the continuous to discontinuous projector. The latter is a simple elementwise nodal to modal projection. Let us denote the continuous-to-discontinuous and discontinuous-to-continuous projectors by $\underline{P}_{cd}$ and $\underline{R}_{dc} = \underline{P}_{cd}^T$, respectively. We then project the discontinuous coarse level problem $\underline{A}_1$ into Q_1^2 via $\underline{A}_{1,c} = \underline{P}_{cd}^T \underline{A}_1 \underline{P}_{cd}$. As in the elementwise constant case, a mesh hierarchy is created via isotropic coarsening and again we utilize linear interpolation and transposed restriction between each level. All coarse operators are then constructed from $\underline{A}_{1,c}$ and recursive application of Galerkin projection. The same smoother and coarse level solver are used as in the elementwise constant case.

It is important to note that the size of the continuous p-coarse grid problem $\underline{A}_{1,c}$ equals the size of the elementwise constant coarse grid problem $\underline{A}_c$ for comparable problem sizes. For a 33×33 element mesh, e.g., the elementwise constant p-coarse space is spanned by $33 \times 33 \times 2$ constants. While for a 32×32 element mesh, the continuous p-coarse space is spanned by $33 \times 33 \times 2$ bilinear (hat) functions.

7. The heterogeneous viscosity Stokes benchmark SolCx. In order to verify the theoretical approximation properties of our SIP Stokes discretization for heterogeneous problems, we consider the *SolCx* benchmark which has been considered extensively for both solver and discretization developments [32, 17, 27]. The analytic solution to the above problem is described in [50] and is available as part of the `Underworld` package [34].

Let Ω be the unit square and let the viscosity η contain a jump in the lateral direction located along the line $0 < x_c < 1$. We consider the problem of finding a solution to (2.1a)–(2.1b) subject to homogeneous Navier boundary conditions. The free parameters of the model are chosen according to $x_c = 0.5$, $\eta_2 = 1$, and $\eta_1 = 10^6$.

Let us denote by $e_{\boldsymbol{u}} = \boldsymbol{u} - \boldsymbol{u}_h$ and $e_p = p - p_h$ the discretization error of the velocity and the pressure fields, respectively. The errors measured in the $L^2(\Omega)$ norm for grids employing an even number of elements in each direction and for different polynomial orders are reported in Table 1. The discrete problem is solved to machine (double) precision.

For grids with an even number of elements in each coordinate direction, the jump in viscosity is aligned with the edges of the elements. Consequently the discontinuous basis functions accurately resolve the pressure field and thus optimal convergence in h is obtained. This was predicted in Lemma 4.6. From the same lemma, we could

TABLE 1

SolCx Stokes benchmark. Results for grids with an even number of elements in the x and y directions ($\Delta x = \Delta y$). All values below the horizontal lines in the second part of the table are considered as affected by machine precision.

Δx	Q_1^2–Q_0		Q_2^2–Q_1		Q_3^2–Q_2	
	$\|e_{\boldsymbol{u}}\|_{L^2}$	$\|e_p\|_{L^2}$	$\|e_{\boldsymbol{u}}\|_{L^2}$	$\|e_p\|_{L^2}$	$\|e_{\boldsymbol{u}}\|_{L^2}$	$\|e_p\|_{L^2}$
1/2	1.3×10^{-3}	6.7×10^{-2}	6.5×10^{-4}	1.4×10^{-2}	9.9×10^{-5}	2.0×10^{-3}
1/4	7.6×10^{-4}	3.5×10^{-2}	9.7×10^{-5}	3.7×10^{-3}	7.0×10^{-6}	2.6×10^{-4}
1/8	2.2×10^{-4}	1.7×10^{-2}	1.2×10^{-5}	9.4×10^{-4}	4.5×10^{-7}	3.2×10^{-5}
1/16	5.7×10^{-5}	8.7×10^{-3}	1.5×10^{-6}	2.3×10^{-4}	2.9×10^{-8}	4.0×10^{-6}
1/32	1.4×10^{-5}	4.4×10^{-3}	1.9×10^{-7}	5.9×10^{-5}	1.8×10^{-9}	5.1×10^{-7}
1/64	3.6×10^{-6}	2.2×10^{-3}	2.4×10^{-8}	1.5×10^{-5}	1.1×10^{-10}	6.3×10^{-8}
1/128	9.1×10^{-7}	1.1×10^{-3}	3.0×10^{-9}	3.7×10^{-6}	7.0×10^{-12}	7.9×10^{-9}
	$\mathcal{O}(h^{1.98})$	$\mathcal{O}(h^{1.00})$	$\mathcal{O}(h^{3.00})$	$\mathcal{O}(h^{2.02})$	$\mathcal{O}(h^{3.97})$	$\mathcal{O}(h^{3.00})$

Δx	Q_4^2–Q_3		Q_5^2–Q_4		Q_6^2–Q_5	
	$\|e_{\boldsymbol{u}}\|_{L^2}$	$\|e_p\|_{L^2}$	$\|e_{\boldsymbol{u}}\|_{L^2}$	$\|e_p\|_{L^2}$	$\|e_{\boldsymbol{u}}\|_{L^2}$	$\|e_p\|_{L^2}$
1/2	7.1×10^{-6}	2.1×10^{-4}	4.8×10^{-7}	1.4×10^{-5}	3.7×10^{-8}	9.7×10^{-7}
1/4	2.5×10^{-7}	1.3×10^{-5}	9.4×10^{-9}	4.5×10^{-7}	3.4×10^{-10}	1.6×10^{-8}
1/8	8.2×10^{-9}	8.4×10^{-7}	1.6×10^{-10}	1.4×10^{-8}	5.4×10^{-12}	2.8×10^{-10}
1/16	2.6×10^{-10}	5.3×10^{-8}	4.8×10^{-12}	4.8×10^{-10}		
1/32	8.3×10^{-12}	3.3×10^{-9}				
1/64	1.8×10^{-12}	2.2×10^{-10}				
	$\mathcal{O}(h^{4.96})$	$\mathcal{O}(h^{4.01})$	$\mathcal{O}(h^{5.88})$	$\mathcal{O}(h^{5.01})$	$\mathcal{O}(h^{6.77})$	$\mathcal{O}(h^{5.92})$

only expect $L^2(\Omega)$ convergence in velocity that is suboptimal by one order however we observe optimal convergence rates. Note that the $L^2(\Omega)$ error estimate from Lemma 4.7 is not applicable here since the pressure solution is discontinuous. The measured order of accuracy is shown in the final row within Table 1.

For grids employing N^2 elements, where N is an odd number, convergence degrades to first order convergence in velocity and convergence by half an order in pressure (results not shown). Degraded convergence was expected due to standard interpolation error results.

8. Solver performance. In this section, we evaluate the robustness and scalability of the Stokes solver discussed in section 6. Specifically we consider four variants of the preconditioner associated with the $\underline{A}$ operator described in section 6. The first two configurations $\mathcal{A}_p(Q_0^2)$ and $\mathcal{A}_{hp}(Q_0^2)$ employ a two-level p-coarsening strategy in which the polynomial order is aggressively coarsened until we obtain a Q_0^2 basis. The latter additionally applies geometric coarsening as detailed in section 6.1.1. Similarly we introduce $\mathcal{A}_p(Q_1^2)$ and $\mathcal{A}_{hp}(Q_1^2)$, where the latter uses a geometric coarsening strategy as outlined in section 6.1.2. The number of Chebyshev-accelerated element-block Jacobi smoothing steps for the two-level p-coarsening is set to 2 for all preconditioners (up and down smoothing each). The smoothers employed in the h-coarsening part of $\mathcal{A}_{hp}(Q_0^2)$ and $\mathcal{A}_{hp}(Q_1^2)$ run three iterations (up and down smoothing each). In all experiments, the coarsest level in both the p-multigrid and the hp-multigrid variants employed LU factorization.

All numerical experiments were performed on a single node of "Hamilton," located at Durham University (UK), equipped with two Intel Xeon E5-2650 v2 (Ivy Bridge)

8-core 2.6-GHz processors with 64 GB of RAM. Experiments that state solve times have been performed using only a single processor. If not otherwise indicated, we will perform experiments in double precision.

8.1. SolCx. We consider the SolCx benchmark (section 7) with parameters $x_c = 0.5$, $\eta_2 = 1$. As free parameters we use the viscosity contrast $\Delta\eta = \eta_1 : \eta_2$ and the grid resolution in our tests. We discretize the problem using a second order velocity space and a first order pressure space (Q_2^2–Q_1 elements). For preconditioners $\mathcal{A}_p(Q_0^2)$, $\mathcal{A}_p(Q_1^2)$, and $\mathcal{A}_{hp}(Q_1^2)$, we consider meshes with the sizes 64^2, 128^2, 256^2, and 512^2. We remark that we have to use odd numbers of elements in each coordinate direction if we want to use the preconditioner $\mathcal{A}_{hp}(Q_0)$; see section 6.1.1. We then consider the mesh with element resolutions of 65^2, 129^2, 257^2, and 513^2. We further perform an L^2 projection of the original viscosity on the elementwise constants since the mesh does not align with the viscosity structure for these meshes. For both preconditioners $\mathcal{A}_{hp}(Q_0^2)$ and $\mathcal{A}_{hp}(Q_1^2)$, we use 3, 4, 5, and 6 h-multigrid levels for the considered mesh sizes, respectively.

We used right preconditioned FGMRES to solve the Stokes problem. Convergence of the saddle point problem is deemed to have occurred when the 2-norm of the residual is 10^6 times smaller than the initial residual (which we denote via rtol(FGMRES) $= 10^{-6}$). The inner solver applied to $\underline{A}$ is preconditioned CG and is terminated according to a relative tolerance criterion of rtol(CG) $= 10^{-3}$. A flexible Krylov method is not required for the viscous block solve since the p- and hp-multigrid preconditioners are linear operators. The wall-clock time and iterations required to solve the Stokes problem, as well as the iterations required by the viscous block solve, are reported in Table 2.

TABLE 2

SolCx: Performance of the Stokes solver using different viscous block preconditioners (see text for details) as a function viscosity jump ($\Delta\eta$) and the number of elements N_K. The polynomial order of the velocity is fixed to $k = 2$. Here, #it indicates the number of outer iterations applied to the Stokes operator, while numbers in brackets indicate the average and maximum iterations required by the viscous block solver. t denotes the CPU time required for the solve. See the text for details on the stopping criteria.

Preconditioner for $\underline{A}$	N_K	$\Delta\eta = 10^0$		$\Delta\eta = 10^6$	
		#it	t (s)	#it	t (s)
$\mathcal{A}_p(Q_0^2)$	64^2	3 (21.3, 23)	2.76	5 (34.2, 43)	12.23
	128^2	3 (20.7, 22)	12.24	5 (36.6, 45)	54.19
	256^2	3 (20.7, 23)	54.71	5 (35.4, 46)	212.55
	512^2	3 (20.0, 23)	302.21	6 (33.5, 46)	965.76
$\mathcal{A}_{hp}(Q_0^2)$	65^2	3 (21.3, 23)	4.90	5 (36.6, 45)	13.61
	129^2	3 (21.0, 22)	19.26	5 (39.0, 46)	57.74
	257^2	3 (20.7, 22)	75.43	5 (40.4, 46)	237.43
	513^2	3 (20.7, 23)	301.30	5 (43.8, 54)	1028.10
$\mathcal{A}_p(Q_1^2)$	64^2	3 (3.3, 4)	1.47	5 (4.8, 6)	3.55
	128^2	3 (3.3, 4)	21.74	5 (4.8, 6)	30.87
	256^2	3 (3.3, 4)	323.77	5 (4.6, 6)	355.21
	512^2	3 (3.3, 4)	2681.80	5 (4.6, 6)	2805.20
$\mathcal{A}_{hp}(Q_1^2)$	64^2	3 (4.0, 4)	1.14	5 (5.2, 7)	2.22
	128^2	3 (3.7, 4)	4.29	5 (5.4, 7)	10.75
	256^2	3 (3.7, 4)	17.66	5 (5.2, 6)	37.18
	512^2	4 (3.7, 4)	70.59	5 (5.4, 7)	155.52

The overall Stokes solver is observed to be scalable for all four preconditioners as the number of outer and inner iterations are virtually independent of the grid resolution for a given viscosity contrast. The viscous block preconditioners $\mathcal{A}_p(Q_1^2)$ and $\mathcal{A}_{hp}(Q_1^2)$ yield significantly fewer inner iterations than the other two preconditioners.

We note that due to the usage of an LU factorization on the coarsest level problem, we do not observe optimal (e.g., $O(n)$) solve times for the variants which do not employ hp-multigrid. Clearly, $\mathcal{A}_p(Q_1^2)$ has a significantly larger coarse grid problem than the other solvers (no discontinuous-to-continuous projection is performed); thus CPU time is far from optimal. We note that the timings for $\mathcal{A}_p(Q_0^2)$ are close to optimal. However, for increasing problem sizes, we can expect further departure from $O(n)$ as observed when using $\mathcal{A}_p(Q_1^2)$. Variant $\mathcal{A}_{hp}(Q_1^2)$ clearly outperforms the three other preconditioners in terms of solve time, and the inner solver needs only one more iteration (at maximum) than $\mathcal{A}_p(Q_1^2)$.

8.2. SolCx checkerboard. In the last section we observed that the preconditioners based on an elementwise bilinear p-coarse space are significantly more efficient in terms of iterations than their counterparts using an elementwise constant coarse space. The preconditioner $\mathcal{A}_{hp}(Q_1^2)$ applying h-coarsening was further found to yield solve times that scale optimally and which are significantly smaller than those of the other three preconditioners. In the following tests, we will thus only consider $\mathcal{A}_{hp}(Q_1^2)$.

In order to demonstrate the robustness of this preconditioner for harder problems, we again solve a SolCx setting but this time with an additional viscosity jump in the y-direction at $y = 0.5$. The resulting viscosity structure is a 2×2 checkerboard. All solver components are configured as detailed in the previous section.

This time we further investigate the influence of the polynomial order on the convergence, and we consider the viscosity contrasts $\eta_2 : \eta_1 = \Delta\eta \in \{10^3, 10^6, 10^8\}$.

The wall-clock time and iterations required to solve the Stokes problem, as well as the iterations required by the viscous block solve, are reported in Table 3.

TABLE 3

SolCx checkerboard: Performance of the Stokes solver using configuration $\mathcal{A}_{hp}(Q_1^2)$ (see text for details) as a function of element order, viscosity jump ($\Delta\eta$), and the number of elements N_K. Refer to Table 2 for the definition of the data reported. Columns reporting iterations which are marked with a ($*$) indicate jobs which required > 64 GB of RAM and thus could not be executed.

$\Delta\eta$	N_K	Q_1^2–Q_0		Q_2^2–Q_1		Q_3^2–Q_2	
		#it	t (s)	#it	t (s)	#it	t (s)
10^3	64^2	14 (4.0, 5)	0.97	17 (7.9, 10)	10.47	18 (11.7, 15)	47.50
	128^2	15 (4.2, 5)	5.38	17 (8.1, 11)	43.35	18 (11.4, 16)	187.51
	256^2	16 (4.9, 6)	27.20	18 (8.2, 11)	188.94	22 (11.0, 17)	892.24
	512^2	15 (5.1, 7)	108.45	17 (8.4, 12)	731.49	23 (11.0, 18)	3697.00
	1024^2	15 (5.1, 7)	458.02	$*$		$*$	
10^6	64^2	15 (4.1, 5)	1.08	17 (8.4, 11)	11.15	17 (12.4, 16)	47.85
	128^2	19 (4.6, 6)	7.27	17 (8.3, 11)	44.60	17 (12.4, 17)	193.43
	256^2	15 (4.7, 6)	24.79	17 (8.6, 12)	187.26	20 (11.3, 17)	833.87
	512^2	17 (5.2, 7)	122.68	19 (8.7, 12)	847.84	19 (12.4, 18)	3846.70
	1024^2	17 (5.4, 8)	508.84	$*$		$*$	
10^8	64^2	15 (4.2, 6)	1.10	16 (9.0, 11)	11.20	18 (13.1, 17)	52.84
	128^2	15 (4.9, 6)	8.03	16 (9.2, 12)	46.58	18 (13.3, 18)	218.42
	256^2	14 (5.1, 6)	24.69	17 (9.3, 12)	201.54	19 (13.3, 18)	918.59
	512^2	19 (5.2, 7)	136.65	17 (9.6, 13)	829.67	18 (13.5, 19)	3795.30
	1024^2	21 (5.6, 8)	814.33	$*$		$*$	

Incrementing the polynomial order by one increments the iterations necessary to converge the viscous solve by around 5 independent of the viscosity jump. A slight dependence of the outer iterations on the polynomial order k is observed for this problem. The most noticeable dependence of the outer iterations on k can be observed between the elements Q_1^2–Q_0 and Q_2^2–Q_1.

For a given Q_k^2–Q_{k-1} element pair, the following additional observations can be made: (1) close to optimal solve times are observed under mesh refinement, and (2) the solve times are almost independent of the jump in viscosity.

8.3. Sedimenting viscous circular inclusions. In our final test, we place six circular inclusions with viscosity $\eta_2 \in \{10^3, 10^6, 10^8\}$ and density $\rho_2 = 1.2$ in a medium with viscosity $\eta_1 = 1$ and density $\rho_1 = 1$. We then consider a forcing term $\boldsymbol{f} = (0, -g\,\rho)$, where the gravity constant is chosen as $g = 10$. The inclusions are placed at $(0.84, 0.39)$, $(0.79, 0.91)$, $(0.33, 0.76)$, $(0.55, 0.47)$, $(0.14, 0.60)$, $(0.24, 0.13)$ and have radii 0.089, 0.059, 0.063, 0.081, 0.05, and 0.09, respectively. The model configuration is such that the inclusions sediment under gravity. At the top boundary, homogeneous Neumann boundary conditions are imposed, while on the remaining parts of the boundary, homogeneous Navier boundary conditions are imposed.

An approximation of the model and a corresponding numerical velocity solution computed using Q_2^2–Q_1 discontinuous finite elements are depicted in Figure 1. The model is approximated using a uniform mesh consisting of 256^2 elements.

A Cartesian grid can never resolve these viscosity and density distributions exactly; we thus always take the elementwise maximum viscosity and the elementwise minimum density. Thus, the discrete problem has by construction a mesh dependence.

We consider meshes with element resolutions of 64^2, 128^2, 256^2, 512, and 1024^2 and employ 2, 3, 4, 5, and 6 h-coarsening levels, one level less than in the previous experiments. We use GCR for both the Stokes (outer) problem and the viscous block (inner) problem.
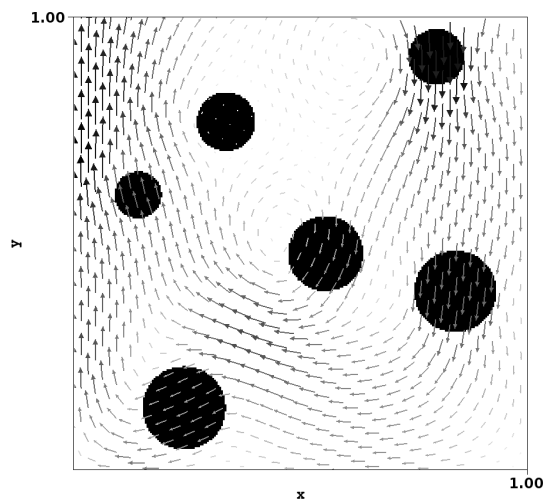

FIG. 1. *Sedimenting circular inclusions: Approximate material composition (pseudocolor plot) and a corresponding numerical velocity solution (vector plot) computed using discontinuous Q_2^2–Q_1 elements. The model is discretized using a uniform mesh consisting of 256^2 elements. Viscosity and density are chosen $\eta_1 = 1$ and $\rho_1 = 1$, respectively, for the background material, while they are chosen $\eta_2 = 10^6$ and $\rho_2 = 1.2$ for the circular inclusions.*

The wall-clock time and iterations required to solve the Stokes problem, as well as the iterations required by the viscous block solve, are reported in Table 4. We vary again the polynomial order $k \in 1, 2, 3$ and the viscosity contrast $\Delta\eta \in \{10^3, 10^6, 10^8\}$. As in the previous two experiments which used a simpler viscosity structure, using $\mathcal{A}_{hp}(Q_1^2)$ with the sinker model configuration we observe that the Stokes preconditioner is both scalable and near optimal. Outer iterations are observed to be only very weakly dependent on the jump in viscosity and mildly dependent on the polynomial degree. The inner iterations required to converge the viscous block are approximately independent of the viscosity jump for each polynomial degree considered. As per other experiments, the average number of iterates required to converge the viscous block are mildly dependent on the polynomial degree.

8.3.1. Quadruple-precision floating point arithmetic. GCR was adopted in the previous experiment as we observed that the orthogonalization procedures of CG, GMRES, and FGMRES would break down for viscosity jumps $\Delta\eta > 10^3$ at a given mesh size. We further note that this breakdown behavior appears to be independent of the viscous block preconditioner as it also occurred when using $\mathcal{A}_p(Q_1^2)$.

TABLE 4

Six circular inclusions: Performance of the Stokes solver using configuration $\mathcal{A}_{hp}(Q_1^2)$ (see text for details) as a function of element order, viscosity jump ($\Delta\eta$), and the number of elements N_K. Refer to Table 2 for the definition of the data reported.

$\Delta\eta$	N_K	Q_1^2–Q_0 #it	t (s)	Q_2^2–Q_1 #it	t (s)	Q_3^2–Q_2 #it	t (s)
10^3	64^2	17 (3.3, 6)	0.97	20 (6.7, 11)	9.54	21 (8.9, 16)	38.12
	128^2	16 (3.5, 7)	4.47	19 (6.6, 12)	36.36	22 (9.0, 16)	163.31
	256^2	17 (3.8, 8)	21.16	19 (6.8, 12)	151.00	22 (8.9, 17)	650.42
	512^2	19 (3.8, 9)	96.98	20 (6.7, 12)	668.22	26 (9.4, 35)	3502.30
	1024^2	25 (4.0, 9)	589.39	*		*	
10^6	64^2	17 (3.4, 6)	0.98	20 (6.7, 11)	9.50	20 (8.9, 16)	36.28
	128^2	17 (3.6, 8)	4.92	18 (6.8, 12)	35.47	21 (8.7, 16)	151.37
	256^2	19 (4.1, 10)	25.43	18 (6.9, 13)	145.18	24 (9.0, 17)	715.33
	512^2	19 (4.4, 11)	108.61	21 (6.9, 12)	710.49	25 (10.2, 35)	4606.20
	1024^2	22 (4.6, 11)	603.12	*		*	

TABLE 5

Six circular inclusions (quad precision): Performance of the Stokes solver using configuration $\mathcal{A}_{hp}(Q_1^2)$ (see text for details) as a function of element order, extreme viscosity jumps ($\Delta\eta$), and the number of elements N_K. Refer to Table 2 for the definition of the data reported. CPU time columns marked with (–) indicate the job required longer than 24 hours to complete.

$\Delta\eta$	N_K	Q_1^2–Q_0 #it	t (s)	Q_2^2–Q_1 #it	t (s)	Q_3^2–Q_2 #it	t (s)
10^{10}	64^2	17 (4.1, 7)	57.32	18 (10.1, 15)	571.04	18 (13.3, 20)	2154.60
	128^2	16 (4.6, 9)	236.45	17 (9.9, 17)	2087.30	18 (13.9, 21)	9153.80
	256^2	18 (5.6, 11)	1236.20	17 (8.8, 15)	7325.30	18 (14.7, 23)	37673.00
	512^2	16 (6.1, 12)	4693.20	17 (8.4, 16)	27888.00	-	-
10^{20}	64^2	18 (8.3, 10)	118.30	21 (14.9, 19)	984.84	26 (17.8, 26)	4257.80
	128^2	18 (10.7, 13)	596.19	18 (17.6, 21)	4088.10	25 (19.6, 27)	18110.00
	256^2	18 (12.3, 16)	2697.60	19 (15.0, 19)	14058.00	28 (17.7, 27)	78188.00
	512^2	18 (13.4, 17)	11480.00	16 (14.8, 20)	60247.00	-	-

We could trace the breakdown back to being related to a loss of floating point precision. Using quadruple-precision floating point arithmetic, breakdown of the orthogonalization does not occur and our preconditioner is able to solve problems with extreme viscosity contrasts ($\Delta\eta \sim 10^{20}$). A selected number of results using $\Delta\eta = 10^{10}, 10^{20}$ are reported in Table 5. As per the results obtained with double precision, the Stokes and viscous block preconditioners are observed to be scalable, and solve times are close to optimal. For these experiments, the outer solver was chosen as FGMRES and the inner solver used was CG.

9. Conclusions. We have investigated high order SIP discretizations of variable viscosity Stokes flow. We have demonstrated that the discretizations are optimally convergent in h for prototypical geodynamics problems where the viscosity discontinuity can be resolved by the grid.

For the solution of the saddle point system arising from the discretization of the Stokes equations, we proposed an iterative method based on block preconditioned FGMRES for the overall linear system and hp-multilevel preconditioned CG for the viscous block. We considered coarsening the polynomial degree of the viscous block to the space of either piecewise constants (Q_0^2) or bilinear functions (Q_1^2), and for each coarse space, an h-multigrid preconditioner was proposed.

Through a series of numerical experiments with heterogeneous viscosity, we have demonstrated that the second strategy yields a more robust coarse level preconditioner. This was attributed to fact that the Q_0^2 coarse space by construction excludes cross derivatives which appear in the definition of the stress tensor when the viscosity is a function of space. Neglecting these terms in the coarse space does not result in error corrections which drive the fine level residual to zero. In contrast, the h-multigrid variant considered for the Q_1^2 coarse space problem results in a solver with a convergence rate that was observed to be independent of the number of elements, largely insensitive to both the viscosity structure and the jump in viscosity, and only weakly dependent on the approximation order.

Last, we have outlined the importance of choosing the facewise SIP penalty parameters depending on the local viscosity and close to the lower bound of the stable regime in order to minimize discretization errors and the number of iterations of the nested inner solver.

9.1. Outlook. A subject of future research could be an extension of the methodology to h- and p-adaptive methods, to three dimensions, as well as to distributed and shared memory parallelism. Furthermore, it might be interesting to perform an analysis of the p-coarse level operator defined on the space of continuous, elementwise bilinear functions. The operator might be related to a Nitsche type discretization. In the context of extreme viscosity contrast problems ($\Delta\eta > 10^6$), one might want to analyze for which substeps and constituents of the considered preconditioned Krylov methods high precision is required.

REFERENCES

[1] D. ARNOLD, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.

[2] D. ARNOLD, F. BREZZI, B. COCKBURN, AND L. MARINI, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., 39 (2002), pp. 1749–1779.

[3] B. AYUSO DE DIOS, F. BREZZI, L. D. MARINI, J. XU, AND L. ZIKATANOV, *A simple preconditioner for a discontinuous Galerkin method for the Stokes problem*, J. Sci. Comput., 58 (2014), pp. 517–547.

[4] S. BALAY, S. ABHYANKAR, M. F. ADAMS, J. BROWN, P. BRUNE, K. BUSCHELMAN, L. DALCIN, V. EIJKHOUT, W. D. GROPP, D. KAUSHIK, M. G. KNEPLEY, L. C. MCINNES, K. RUPP, B. F. SMITH, S. ZAMPINI, H. ZHANG, AND H. ZHANG, *PETSc Users Manual*, Tech. Report ANL-95/11, Revision 3.7, Argonne National Laboratory, 2016.

[5] S. BALAY, S. ABHYANKAR, M. F. ADAMS, J. BROWN, P. BRUNE, K. BUSCHELMAN, L. DALCIN, V. EIJKHOUT, W. D. GROPP, D. KAUSHIK, M. G. KNEPLEY, L. C. MCINNES, K. RUPP, B. F. SMITH, S. ZAMPINI, H. ZHANG, AND H. ZHANG, *PETSc*, http://www.mcs.anl.gov/petsc (2016).

[6] S. BALAY, W. D. GROPP, L. C. MCINNES, AND B. F. SMITH, *Modern Software Tools in Scientific Computing*, in Efficient Management of Parallelism in Object Oriented Numerical Software Libraries, E. Arge, A. M. Bruaset, and H. P. Langtangen, eds., Birkhäuser, Basel, 1997, pp. 163–202.

[7] P. BASTIAN, *A fully-coupled discontinuous Galerkin method for two-phase flow in porous media with discontinuous capillary pressure*, Comput. Geosci., 18 (2014), pp. 779–796.

[8] P. BASTIAN, M. BLATT, AND R. SCHEICHL, *Algebraic multigrid for discontinuous Galerkin discretizations of heterogeneous elliptic problems: AMG4dg*, Numer. Linear Algebra Appl., 19 (2012), pp. 367–388.

[9] J. BRAUN, C. THIEULOT, P. FULLSACK, M. DEKOOL, C. BEAUMONT, AND R. HUISMANS, *DOUAR: A new three-dimensional creeping flow numerical model for the solution of geological problems*, Phys. Earth Planetary Interiors, 171 (2008), pp. 76–91.

[10] S. C. BRENNER, *Korn's inequalities for piecewise H^1 vector fields*, Math. Comp., 73 (2004), pp. pp. 1067–1087.

[11] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Element Methods*, Springer, New York, 1991.

[12] P. N. BROWN AND H. F. WALKER, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.

[13] C. BURSTEDDE, O. GHATTAS, G. STADLER, T. TU, AND L. C. WILCOX, *Parallel scalable adjoint-based adaptive solution of variable-viscosity Stokes flow problems*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1691–1700.

[14] B. COCKBURN, G. KANSCHAT, AND D. SCHÖTZAU, *A note on discontinuous Galerkin divergence-free solutions of the Navier-Stokes equations*, J. Sci. Comput., 31 (2007), pp. 61–73.

[15] V. A. DOBREV, R. D. LAZAROV, P. S. VASSILEVSKI, AND L. T. ZIKATANOV, *Two-level preconditioning of discontinuous Galerkin approximations of second-order elliptic equations*, Numer. Linear Algebra Appl., 13 (2006), pp. 753–770.

[16] M. DROSSON AND K. HILLEWAERT, *On the stability of the symmetric interior penalty method for the Spalart–Allmaras turbulence model*, J. Comput. Appl. Math., 246 (2013), pp. 122–135.

[17] T. DURETZ, D. A. MAY, T. V. GERYA, AND P. J. TACKLEY, *Discretization errors and free surface stabilization in the finite difference and marker-in-cell method for applied geodynamics: A numerical study: FD-MIC scheme discretization errors*, Geochemistry Geophysics Geosystems, 12 (2011).

[18] T. DURETZ, D. A. MAY, AND P. YAMATO, *A free surface capturing discretization for the staggered grid finite difference scheme*, Geophys. J. Internat., 204 (2016), pp. 1518–1530.

[19] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers*, Oxford University Press, New York, 2005.

[20] H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, New York, 2014.

[21] P. FULLSACK, *An arbitrary Lagrangian-Eulerian formulation for creeping flows and its application in tectonic models*, Geophys. J. Interat., 120 (1995), pp. 1–23.

[22] T. V. GERYA, D. A. MAY, AND T. DURETZ, *An adaptive staggered grid finite difference method for modeling geodynamic stokes flows with strongly variable viscosity*, Geochemistry Geophysics Geosystems, 14 (2013), pp. 1200–1225.

[23] T. V. GERYA AND D. A. YUEN, *Charaterictics-based marker method with conservative finite-difference schemes for modeling geological flows with strongly variable transport properties*, Phys. Earth Planetary Interiors, 140 (2003), pp. 293–318.

[24] K. HILLEWAERT, *Development of the Discontinuous Galerkin Method for High-Resolution, Large Scale CFD and Acoustics in Industrial Geometries*, Ph.D. thesis, Université Catholique de Louvain, 2013.

[25] G. KANSCHAT AND Y. MAO, *Multigrid methods for H^{div}-conforming discontinuous Galerkin methods for the Stokes equations*, J. Numer. Math., 23 (2015), pp. 51–66.

[26] S.-I. KARATO, *Phase transformations and rheological properties of mantle minerals*, Earths Deep Interior, 7 (1997), pp. 223–272.

[27] M. KRONBICHLER, T. HEISTER, AND W. BANGERTH, *High accuracy mantle convection simulation through modern numerical methods: High accuracy mantle convection simulation*, Geophys. J. Internat., 191 (2012), pp. 12–29.

[28] S. M. LECHMANN, D. A. MAY, B. J. P. KAUS, AND S. M. SCHMALHOLZ, *Comparing thin-sheet models with 3-D multilayer models for continental collision*, Geophys. J. Internat., 187 (2011), pp. 10–33.

[29] R. S. LEHMANN, M. LUKÁČOVÁ-MEDVID'OVÁ, B. J. P. KAUS, AND A. A. POPOV, *Comparison of continuous and discontinuous Galerkin approaches for variable-viscosity Stokes flow*, ZAMM Z. Angew. Math. Mech., 96 (2016), pp. 733–746.

[30] W. LENG AND S. J. ZHONG, *Implementation and application of adaptive mesh refinement for thermochemical mantle convection studies*, Geochemistry Geophysics Geosystems, 12 (2011), Q04006.

[31] D. A. MAY, J. BROWN, AND L. LE POURHIET, *A scalable, matrix-free multigrid preconditioner for finite element discretizations of heterogeneous stokes flow*, Comput. Methods Appl. Mech. Engrg., 290 (2015), pp. 496–523.

[32] D. A. MAY AND L. MORESI, *Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics*, Phys. Earth Planetary Interiors, 171 (2008), pp. 33–47.

[33] L. MORESI, F. DUFOUR, AND H.-B. MÜHLHAUS, *A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials*, J. Comput. Phys., 184 (2003), pp. 476–497.

[34] L. MORESI, S. QUENETTE, V. LEMIALE, C. MÉRIAUX, B. APPELBE, AND H.-B. MÜHLHAUS, *Computational approaches to studying non-linear dynamics of the crust and mantle*, Phys. Earth Planetary Interiors, 163 (2007), pp. 69–82.

[35] A. POLIAKOV AND Y. PODLADCHIKOV, *Diapirism and topography*, Geophys. J. Internat., 109 (1992), pp. 553–564.

[36] A. A. POPOV AND S. V. SOBOLEV, *SLIM3D: A tool for three-dimensional thermomechanical modeling of lithospheric deformation with elasto-visco-plastic rheology*, Phys. Earth Planetary Interiors, 171 (2008), pp. 55–75.

[37] G. RANALLI, *Rheology of the Earth*, Springer Science & Business Media, New York, 1995.

[38] D. SCHÖTZAU, C. SCHWAB, AND A. TOSELLI, *Mixed hp-DGFEM for incompressible flows*, SIAM J. Numer. Anal., 40 (2002), pp. 2171–2194.

[39] D. SCHÖTZAU, C. SCHWAB, AND A. TOSELLI, *Mixed hp-DGFEM for incompressible flows* II: *Geometric edge meshes*, IMA J. Numer. Anal., 24 (2004), pp. 273–308.

[40] G. SCHUBERT, D. L. TURCOTTE, AND P. OLSON, *Mantle Convection in the Earth and Planets*, Cambridge University Press, Cambridge, UK, 2001.

[41] K. STEMMER, H. HARDER, AND U. HANSEN, *A new method to simulate convection with strongly temperature- and pressure-dependent viscosity in a spherical shell: Applications to the Earth's mantle*, Phys. Earth Planetary Interiors, 157 (2006), pp. 223–249.

[42] P. J. TACKLEY, *Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid*, Phys. Earth Planetary Interiors, 171 (2008), pp. 7–18.

[43] A. TOSELLI, *hp discontinuous Galerkin approximations for the Stokes problem*, Math. Models Methods Appl. Sci., 12 (2002), pp. 1565–1597.

[44] D. L. TURCOTTE, K. E. TORRANCE, AND A. T. HSUI, *Convection in the Earth's Mantle*, in Methods in Computational Physics: Geophysics, Academic Press, New York, 1973, pp. 431–454.

[45] P. VAN SLINGERLAND AND C. VUIK, *Fast linear solver for diffusion problems with applications to pressure computation in layered domains*, Comput. Geosci., 18 (2014), pp. 343–356.

[46] P. VAN SLINGERLAND AND C. VUIK, *Scalable two-level preconditioning and deflation based on a piecewise constant subspace for (SIP)DG systems for diffusion problems*, J. Comput. Appl. Math., 275 (2015), pp. 61–78.

[47] J. Wang and X. Ye, *New finite element methods in computational fluid dynamics by H(div) elements*, SIAM J. Numer. Anal., 45 (2007), pp. 1269–1286.

[48] R. F. Weinberg and H. Schmeling, *Polydiapirs: Multiwave length gravity structures*, J. Structural Geol., 14 (1992), pp. 425–436.

[49] S. Zaleski and P. Julien, *Numerical simulation of Rayleigh-Taylor instability for single and multiple salt diapirs*, Tectonophysics, 206 (1992), pp. 55–69.

[50] S. Zhong, *Analytic solutions for Stokes' flow with lateral variations in viscosity*, Geophys. J. Internat., 124 (1996), pp. 18–28.