MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

FINITE ELEMENT APPROXIMATION AND AUGMENTED LAGRANGIAN PRECONDITIONING FOR ANISOTHERMAL IMPLICITLY-CONSTITUTED NON-NEWTONIAN FLOW

PATRICK FARRELL, PABLO ALEXEI GAZCA OROZCO, AND ENDRE SÜLI

ABSTRACT. We devise 3-field and 4-field finite element approximations of a system describing the steady state of an incompressible heat-conducting fluid with implicit non-Newtonian rheology. We prove that the sequence of numerical approximations converges to a weak solution of the problem. We develop a block preconditioner based on augmented Lagrangian stabilisation for a discretisation based on the Scott–Vogelius finite element pair for the velocity and pressure. The preconditioner involves a specialised multigrid algorithm that makes use of a space decomposition that captures the kernel of the divergence and non-standard intergrid transfer operators. The preconditioner exhibits robust convergence behaviour when applied to the Navier–Stokes system, including temperature-dependent viscosity, heat conductivity and viscous dissipation.

1. Introduction

For $d \in \{2,3\}$, let $\Omega \subset \mathbb{R}^d$ be a bounded polytopal domain with a Lipschitz boundary. The steady form of the Oberbeck–Boussinesq [48, 8] approximation used in the modelling of natural convection reads:

$$-\operatorname{div} \mathbf{S} + \rho_0 \operatorname{div}(\boldsymbol{u} \otimes \boldsymbol{u}) + \nabla p = -\rho_0 \beta g (\theta - \theta_C)\boldsymbol{e}_d \quad \text{in } \Omega, \tag{1.1a}$$

$$\operatorname{div} \boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{1.1b}$$

$$-\operatorname{div}(\hat{\kappa}(\theta)\nabla\theta) + \rho_0 c_p \operatorname{div}(\boldsymbol{u}\theta) + \beta\rho_0 g\theta \boldsymbol{u}\cdot\boldsymbol{e}_d = \mathbf{S} : \mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega, \tag{1.1c}$$

where $\boldsymbol{e}_d$ is the unit vector pointing against gravity, $\mathbf{D}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^\top)$ denotes the symmetric gradient, and the quantities appearing in the equations are as follows:

$\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$	velocity field	β	thermal expansion coefficient
$\mathbf{S}\colon \Omega \to \mathbb{R}^{d\times d}_{\mathrm{sym,tr}}$	shear stress	c_p	specific heat capacity
$p\colon \Omega \to \mathbb{R}$	pressure	g	acceleration due to gravity
$\theta\colon \Omega \to \mathbb{R}$	temperature	ρ_0	reference density
$\hat{\kappa}\colon \mathbb{R} \to \mathbb{R}$	heat conductivity	θ_C	reference temperature

Received by the editor July 16, 2020.
1991 *Mathematics Subject Classification.* Primary 65N30, 65F08; Secondary 65N55, 76A05.
This research was supported by the Engineering and Physical Sciences Research Council, grants EP/K030930/1 and EP/R029423/1, and by the EPSRC Centre for Doctoral Training in Partial Differential Equations: Analysis and Applications, grant EP/L015811/1. The second author was supported by CONACyT (Scholarship 438269).

The symbol $\mathbb{R}^{d\times d}_{\mathrm{sym,tr}}$ above denotes the set of $d \times d$ symmetric and traceless matrices. The system is supplemented with the boundary conditions

$$\boldsymbol{u}|_{\partial\Omega} = \boldsymbol{0}, \qquad \theta|_{\Gamma_D} = \theta_b, \qquad \hat{\kappa}(\theta)\nabla\theta\cdot\boldsymbol{n}|_{\partial\Omega\setminus\Gamma_D} = 0, \tag{1.2}$$

where Γ_D is a relatively open subset of $\partial\Omega$ with $|\Gamma_D| \neq 0$, $\boldsymbol{n}$ is the unit outward-pointing normal vector to the boundary, and θ_b is a given temperature distribution on Γ_D. In many applications the effects of viscous dissipation are ignored, i.e. only the first two terms in the temperature equation (1.1c) are kept. However, it has been observed that in some cases the effects of the viscous dissipation term $\mathbf{S}:\mathbf{D}(\boldsymbol{u})$ are non-negligible and should be taken into account [31, 62, 63, 49]. Furthermore, as noted in [3, 62], the viscous dissipation must be balanced with the adiabatic heating term $\beta\rho_0 g\theta\boldsymbol{u}\cdot\boldsymbol{e}_d$; for a mathematically rigorous derivation of the system (1.1) see [34]. The existence of distributional solutions of (1.1) with non-Newtonian rheology of power-law type was shown in [55, 47].

The system must be closed with a constitutive relation that relates the stress $\mathbf{S}$ and the symmetric velocity gradient $\mathbf{D}(\boldsymbol{u})$. The most commonly considered closure is the Newtonian constitutive relation $\mathbf{S} = 2\hat{\mu}(\theta)\mathbf{D}(\boldsymbol{u})$, where $\hat{\mu}\colon \mathbb{R}\to\mathbb{R}$ is the viscosity. In this work we consider much more general implicit constitutive relations of the form $\mathbf{G}(\mathbf{S},\mathbf{D}(\boldsymbol{u}),\theta) = \boldsymbol{0}$ and $\mathbf{H}(\mathbf{S},\mathbf{D}(\boldsymbol{u})) = \boldsymbol{0}$. The framework of implicitly constituted fluids is a generalisation of classical continuum mechanics that allows the study of a much wider class of materials in a thermodynamically consistent manner (see [51, 52, 53]).

The first rigorous existence results within the implicitly constituted framework for the isothermal system can be found in [10, 11] (see also [6]), while an extension to a temperature-dependent system was carried out in [46]. Regarding the numerical approximation of these systems, only the isothermal case has been considered so far. The finite element approximation of the steady isothermal system was analysed in [16, 38], and extensions to the unsteady problem can be found in [60, 22]. An augmented Lagrangian preconditioner was proposed for a 3-field formulation of the isothermal system in [21].

Our analysis focuses on the constitutive relation defined by

$$\mathbf{G}(\mathbf{S},\mathbf{D}(\boldsymbol{u}),\theta) := 2\hat{\mu}(\theta)\frac{(|\mathbf{D}(\boldsymbol{u})| - \hat{\sigma}(\theta))^+}{|\mathbf{D}(\boldsymbol{u})|}\mathbf{D}(\boldsymbol{u}) - \frac{(|\mathbf{S}| - \hat{\tau}(\theta))^+}{|\mathbf{S}|}\mathbf{S}, \tag{1.3}$$

where $\hat{\tau}\hat{\sigma} = 0$ (the precise assumptions on $\hat{\mu},\hat{\tau},\hat{\sigma}$ will be introduced later). The relation (1.3) describes a fluid with either Bingham or activated-Euler rheology in which the viscosity and activation parameters may depend on the temperature. Naturally, this family of constitutive relations also includes the Navier–Stokes model with a temperature-dependent viscosity (when $\hat{\tau} \equiv 0 \equiv \hat{\sigma}$). The relation (1.3) was introduced to [46], where existence of weak solutions to the unsteady version of a similar system was shown. Our results will also cover relations with more general power-law behaviour if one restricts the system to have either an explicit constitutive relation or constant rheological parameters.

We make two main contributions in this work. First, we introduce a finite element approximation of the system (1.1) and prove convergence of the sequence of finite element approximations to a weak solution. This represents the first finite element convergence result for heat-conducting implicitly constituted fluids. For the sake of simplicity, we will neglect the viscous dissipation in the analysis. However, this can be included in the numerical algorithm without any difficulties. The main challenge associated with this term in the analysis stems from the fact that $\mathbf{S}:\mathbf{D}(\boldsymbol{u})$ belongs *a priori* to $L^1(\Omega)$ only, and hence a suitable notion of renormalised solution must be employed for the temperature equation. We note that this difficulty has been circumvented in the PDE analysis of similar systems in the transient case [9, 12]. We would expect that by imposing certain restrictions on the mesh and for $\mathbb{P}_1$ elements, a similar convergence result would hold for an appropriately defined renormalised solution (c.f. [14]). When restricted to constant rheological parameters and the isothermal problem, the convergence result here improves on the result for r-graphs from [16] by extending it to cover the whole admissible range $r > \frac{2d}{d+2}$, even without pointwise divergence-free elements. This is possible by making use of reconstruction operators, which in recent years were introduced to restore the pressure-robustness in the finite element formulations (see e.g. [33]).

The second main contribution is the development of a preconditioner based on augmented Lagrangian stabilisation for linearisations of the discretisation of (1.1), including the viscous dissipation term. After Newton linearisation the system takes the following form

$$\begin{bmatrix} A & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ p \end{bmatrix} = \begin{bmatrix} \boldsymbol{f} \\ g \end{bmatrix}, \tag{1.4}$$

where $\boldsymbol{z} = (\theta, \boldsymbol{u})^\top$ or $\boldsymbol{z} = (\mathbf{S}, \theta, \boldsymbol{u})^\top$, depending on whether a 3-field or a 4-field formulation is employed, and B represents the divergence operator acting on the velocity space. After performing Gaussian elimination on the blocks, the problem of solving (1.4) reduces to solving smaller systems involving A and the Schur complement $S := -BA^{-1}B^\top$. In many cases, such as in a velocity-pressure formulation of the Stokes system, A represents a symmetric and coercive operator which can be inverted efficiently, and so the challenge is to develop an effective and efficient approximation for the Schur complement inverse $\tilde{S}^{-1}$. For the Stokes system with constant viscosity ν it is known that the choice $\tilde{S}^{-1} = -\nu M_p^{-1}$, where M_p is the pressure mass matrix, results in a spectrally equivalent preconditioner [59, 45]. When the convective term is introduced to the formulation, the performance of this strategy degrades as the Reynolds number Re gets larger (meaning that the number of Krylov subspace iterations per nonlinear iteration grows with Re) [18]. This loss of robustness occurs also with other well-known preconditioners, such as the PCD [35] and LSC [18] preconditioners (see e.g. [19]). Block preconditioners based on PCD for the system (1.1) without viscous dissipation were proposed in [32, 36], where it was observed that the number of linear iterations increased strongly with the Rayleigh number Ra.

Alternatively, one can consider the system with an augmented Lagrangian term, with $\gamma > 0$:

$$\begin{bmatrix} A + \gamma B^\top M_p^{-1} B & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ p \end{bmatrix}, = \begin{bmatrix} \boldsymbol{f} + \gamma B^\top M_p^{-1} g \\ g \end{bmatrix}, \tag{1.5}$$

which has the same solution as (1.4), since $B\boldsymbol{z} = g$. The advantage of this is that using the Sherman–Morrison–Woodbury formula (see e.g. [1]), the Schur complement can be approximated in a straightforward way:

$$\begin{aligned} S^{-1} &= (-B(A + \gamma B^\top M_p^{-1} B)^{-1} B^\top)^{-1} = -(BA^{-1}B^\top)^{-1} - \gamma M_p^{-1} \\ &\approx -(\nu + \gamma) M_p^{-1}, \end{aligned}$$

and the approximation gets better as $\gamma \to \infty$. The difficulty now becomes solving the linear system associated with top block $A + \gamma B^\top M_p^{-1} B$ efficiently, since the augmented Lagrangian term possesses a large kernel (the set of all discretely divergence-free velocities). This approach was used for the 2D Navier–Stokes system by Benzi and Olshanskii [5] and later extended to three dimensions by Farrell, Mitchell and Wechsung [26]. The strategy for efficiently solving the top block in these works was based on ideas developed by Schöberl in the context of nearly incompressible elasticity [57, 56], where it became clear that constructing robust relaxation and transfer operators is essential for obtaining a γ-robust multigrid algorithm.

These ideas will be applied here to develop a preconditioner for the anisothermal system based on a discretisation using the Scott–Vogelius pair for the velocity and pressure, which has the advantage of preserving the divergence constraint exactly (to machine precision and solver tolerances). This builds on previous work for the Navier–Stokes system [24] and a stress-velocity-pressure formulation for isothermal non-Newtonian fluids with implicit rheology [21]. An augmented Lagrangian-based preconditioner (AL) for buoyancy-driven flow was already presented in [37] for a stabilised $\mathbb{P}_1$–$\mathbb{P}_1$ velocity-pressure pair, in which the augmented velocity block was substituted by $A+\gamma B^\top \operatorname{diag}(M_p)^{-1}B$ and handled by GMRES preconditioned with algebraic multigrid; in that work it was shown that the AL preconditioner performed better than non-augmented variants, at least for Prandtl and Rayleigh numbers in the ranges $0.04 \leq \mathrm{Pr} \leq 1$, $500 \leq \mathrm{Ra} \leq 10000$. Numerical experiments with the preconditioner will show good performance with the Navier–Stokes and power-law models for a wider range of non-dimensional numbers, even with temperature-dependent viscosity, heat conductivity, and viscous dissipation. It is remarkable that the robustness properties of the preconditioner hold in this case, given that the available parameter-robust multigrid theory pioneered by Schöberl does not apply, since the block A is non-symmetric and non-coercive.

2. Preliminaries

2.1. **Function spaces.** Throughout this work we will employ standard notation for Sobolev and Lebesgue spaces (e.g. $(W^{k,s}(\Omega), \|\cdot\|_{W^{k,s}(\Omega)})$ and $(L^q(\Omega), \|\cdot\|_{L^q(\Omega)})$). The space $W^{k,r}_0(\Omega)$, for $r \in [1,\infty)$, is defined as the closure of the space of smooth functions with compact support $C^\infty_0(\Omega)$ with respect to $\|\cdot\|_{k,r}$; its dual space will be denoted by $W^{-1,r'}(\Omega)$, where r' is the Hölder-conjugate of the number r, i.e. $1/r' + 1/r = 1$. When $r = 2$ we will write $W^{k,2}(\Omega) = H^k(\Omega)$ and $W^{-1,2}(\Omega) = H^{-1}(\Omega)$. Let us also define the following useful subspaces for $r > 1$ and $\Gamma \subset \partial\Omega$:

$$
\begin{aligned}
L^r_0(\Omega) &:= \left\{ q \in L^r(\Omega) \colon \int_\Omega q = 0 \right\}, \\
W^{1,r}_{0,\mathrm{div}}(\Omega)^d &:= \overline{\{\boldsymbol{v} \in C^\infty_0(\Omega)^d \,:\, \operatorname{div} \boldsymbol{v} = 0\}}^{\|\cdot\|_{W^{1,r}(\Omega)}}, \\
W^{1,r}_{\Gamma}(\Omega) &:= \overline{\{w \in C^\infty(\Omega) \,:\, w|_\Gamma = 0\}}^{\|\cdot\|_{W^{1,r}(\Omega)}}, \\
L^r_{\mathrm{sym}}(\Omega)^{d\times d} &:= \{\boldsymbol{\tau} \in L^r(\Omega)^{d\times d} \colon \boldsymbol{\tau}^\top = \boldsymbol{\tau}\}, \\
L^r_{\mathrm{sym,tr}}(\Omega)^{d\times d} &:= \{\boldsymbol{\tau} \in L^r_{\mathrm{sym}}(\Omega)^{d\times d} \colon \operatorname{tr} \boldsymbol{\tau} = 0\}, \\
W^{1/r',r}_{00}(\Gamma) &:= \{w|_\Gamma \colon w \in W^{1,r}(\Omega),\, w = 0 \text{ on } \partial\Omega \setminus \overline{\Gamma}\}.
\end{aligned}
$$

The operator tr in the definition of $L^r_{\mathrm{sym,tr}}(\Omega)^{d\times d}$ denotes the trace of a $d\times d$ matrix. The letter c will be used in various estimates to denote a generic positive constant whose value might change from line to line (the dependence on the parameters will be made explicit whenever necessary).

2.2. **Implicit constitutive relations.** Let us assume now that the material parameters $\hat{\mu}, \hat{\tau}, \hat{\sigma}, \hat{\kappa}$ are continuous functions of one variable such that

$$\begin{gathered} 0 \leq \hat{\tau}(s), \hat{\sigma}(s) \leq c_0, \\ c_1 \leq \hat{\mu}(s), \hat{\kappa}(s) \leq c_2, \\ \hat{\tau}(s)\hat{\sigma}(s) = 0, \end{gathered} \tag{2.1}$$

for all $s \in \mathbb{R}$, and some positive constants c_0, c_1, c_2. It is not difficult to show that under these assumptions, the relation (1.3) defines a monotone and coercive 2-graph [46, Lemma 3].

Lemma 2.1. *Let $\mathbf{G}\colon \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R} \to \mathbb{R}$ be the function defined by* (1.3) *and suppose that $\hat{\mu}, \hat{\tau}, \hat{\sigma} \in C(\mathbb{R})$ satisfy* (2.1)*. Then there exist two constants $\alpha, \beta > 0$ such that*

$$\mathbf{S} : \mathbf{D} \geq \alpha(|\mathbf{S}|^2 + |\mathbf{D}|^2) - \beta \tag{2.2}$$

$$(\mathbf{S} - \overline{\mathbf{S}}) : (\mathbf{D} - \overline{\mathbf{D}}) \geq 0, \tag{2.3}$$

for any $(\mathbf{S}, \mathbf{D}, \theta), (\overline{\mathbf{S}}, \overline{\mathbf{D}}, \theta) \in \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R}$ such that $\mathbf{G}(\mathbf{S}, \mathbf{D}, \theta) = \mathbf{0} = \mathbf{G}(\overline{\mathbf{S}}, \overline{\mathbf{D}}, \theta)$.

In the same spirit as [16, 22], in the numerical scheme we will employ a sequence of continuous explicit approximations of the implicit constitutive relation (1.3). Let us define for $n \in \mathbb{N}$ the approximations as follows:

$$\begin{aligned} \boldsymbol{\mathcal{D}}^n(\mathbf{S}, \theta) &:= \min\left\{ n + \frac{1}{2\hat{\mu}(\theta)}, \frac{\frac{1}{2\hat{\mu}(\theta)}(|\mathbf{S}| - \hat{\tau}(\theta))^+ + \hat{\sigma}(\theta)}{|\mathbf{S}|} \right\}, \\ \boldsymbol{\mathcal{S}}^n(\mathbf{D}, \theta) &:= \min\left\{ n + 2\hat{\mu}(\theta), \frac{2\hat{\mu}(\theta)(|\mathbf{D}| - \hat{\sigma}(\theta))^+ + \hat{\tau}(\theta)}{|\mathbf{D}|} \right\}. \end{aligned} \tag{2.4}$$

Either of the two can be chosen, depending on whether one wishes to consider explicit approximations of the stress in terms of the symmetric velocity gradient or vice-versa. The functions $\boldsymbol{\mathcal{D}}^n$ and $\boldsymbol{\mathcal{S}}^n$ satisfy the same monotonicity and coercivity conditions as those stated in Lemma 2.1, uniformly in n. More importantly, the following localised Minty's lemma is available for these approximations, which will be useful when proving that the limit of the numerical approximations satisfies the constitutive relation.

Lemma 2.2 ([46], Lemma 6)**.** *Let $M \subset \Omega$ be measurable and let $\mathbf{G}$ be defined by* (1.3)*. Now suppose that $\{\mathbf{D}^n\}_{\mathbb{N}}$ and $\{\theta^n\}_{\mathbb{N}}$ are sequences of measurable functions on Ω and let $\mathbf{S}^n := \boldsymbol{\mathcal{S}}^n(\mathbf{D}^n, \theta^n)$. Assume that the following conditions hold:*

$$\begin{aligned} \boldsymbol{\mathcal{S}}^n(\mathbf{D}^n, \theta^n) &= \mathbf{0} && \text{a.e. in } M, \\ \mathbf{S}^n &\rightharpoonup \mathbf{S} && \text{weakly in } L^2(M)^{d\times d}, \\ \mathbf{D}^n &\rightharpoonup \mathbf{D} && \text{weakly in } L^2(M)^{d\times d}, \\ \theta^n &\rightharpoonup \theta && \text{a.e. in } M, \end{aligned}$$

$$\limsup_{n\to\infty} \int_M \mathbf{S}^n : \mathbf{D}^n \leq \int_M \mathbf{S} : \mathbf{D}.$$

Then $\mathbf{G}(\mathbf{S},\mathbf{D},\theta)=\mathbf{0}$ *and* $\mathbf{S}^n:\mathbf{D}^n \rightharpoonup \mathbf{S}:\mathbf{D}$ *weakly in* $L^1(M)$. *An analogous statement holds for* $\boldsymbol{\mathcal{D}}^n$.

If the rheological parameters are constant (i.e. do not depend on the temperature), it is possible to generalise the convergence result to cover implicit relations of the form $\mathbf{H}(\cdot,\mathbf{S},\mathbf{D}(\boldsymbol{u}))=\mathbf{0}$, where $\mathbf{H}\colon \Omega\times\mathbb{R}^{d\times d}_{\mathrm{sym}}\times\mathbb{R}^{d\times d}_{\mathrm{sym}}\to\mathbb{R}^{d\times d}_{\mathrm{sym}}$, that satisfy the coercivity condition (2.2) with an exponent other than 2; this would for instance capture the Herschel–Bulkley constitutive relation. For convenience, the assumptions will be written in terms of the graph induced by $\mathbf{H}$, which is defined in the standard way:

$$(\mathbf{D},\mathbf{S})\in\mathcal{A}(\cdot)\Longleftrightarrow \mathbf{H}(\cdot,\mathbf{S},\mathbf{D})=\mathbf{0}.$$

Assumption 2.3. The graph $\mathcal{A}$ is a maximal monotone r-graph for some $r>\frac{2d}{d+2}$. More precisely, the following properties hold for almost every $x\in\Omega$:

- ($\mathcal{A}$ contains the origin). $(\mathbf{0},\mathbf{0})\in\mathcal{A}(x)$;
- ($\mathcal{A}$ is a monotone graph). For every $(\mathbf{D}_1,\mathbf{S}_1),(\mathbf{D}_2,\mathbf{S}_2)\in\mathcal{A}(x)$,
$$(\mathbf{S}_1-\mathbf{S}_2):(\mathbf{D}_1-\mathbf{D}_2)\geq 0;$$
- ($\mathcal{A}$ is maximal monotone). If $(\mathbf{D},\mathbf{S})\in\mathbb{R}^{d\times d}_{\mathrm{sym}}\times\mathbb{R}^{d\times d}_{\mathrm{sym}}$ is such that
$$(\hat{\mathbf{S}}-\mathbf{S}):(\hat{\mathbf{D}}-\mathbf{D})\geq 0 \quad \text{for all } (\hat{\mathbf{D}},\hat{\mathbf{S}})\in\mathcal{A}(x),$$
then $(\mathbf{D},\mathbf{S})\in\mathcal{A}(x)$;
- ($\mathcal{A}$ is an r-graph). There is a non-negative function $m\in L^1(\Omega)$ and a constant $c>0$ such that
$$\mathbf{S}\colon\mathbf{D}\geq -m+c(|\mathbf{D}|^r+|\mathbf{S}|^{r'}) \quad \text{for all } (\mathbf{D},\mathbf{S})\in\mathcal{A}(x);$$
- (Measurability). The set-valued map $x\mapsto\mathcal{A}$ is $\mathcal{L}(\Omega)$–$(\mathcal{B}(\mathbb{R}^{d\times d}_{\mathrm{sym}})\otimes\mathcal{B}(\mathbb{R}^{d\times d}_{\mathrm{sym}}))$ measurable; here $\mathcal{L}(\Omega)$ denotes the family of Lebesgue measurable subsets of Ω and $\mathcal{B}$ is the family of Borel subsets of $\mathbb{R}^{d\times d}_{\mathrm{sym}}$;
- (Compatibility). For any $(\mathbf{D},\mathbf{S})\in\mathcal{A}(x)$ we have that
$$\mathrm{tr}(\mathbf{D})=0\Longleftrightarrow\mathrm{tr}(\mathbf{S})=0.$$

2.3. **Finite element spaces.** Let $\{\mathcal{T}_n\}_{n\in\mathbb{N}}$ be a family of shape-regular triangulations such that the mesh size $h_n:=\max_{K\in\mathcal{T}_n}h_K$ tends to zero as $n\to\infty$, where h_K denotes the diameter of an element $K\in\mathcal{T}_n$. We define the following conforming families of finite element spaces:

$$\begin{aligned}
\Sigma^n &:= \left\{\boldsymbol{\sigma}\in L^\infty_{\mathrm{sym}}(\Omega)^{d\times d}\colon \boldsymbol{\sigma}|_K\in\mathbb{P}_{\mathbb{S}}(K)^{d\times d},\, K\in\mathcal{T}_n\right\},\\
V^n &:= \left\{\boldsymbol{v}\in W^{1,\infty}_0(\Omega)^d\colon \boldsymbol{v}|_K\in\mathbb{P}_{\mathbb{V}}(K)^d,\, K\in\mathcal{T}_n\right\},\\
M^n &:= \left\{q\in L^\infty(\Omega)\colon q|_K\in\mathbb{P}_{\mathbb{M}}(K),\, K\in\mathcal{T}_n\right\},\\
U^n &:= \left\{w\in W^{1,\infty}_{\Gamma_D}(\Omega)\colon w|_K\in\mathbb{P}_{\mathbb{U}}(K),\, K\in\mathcal{T}_n\right\},
\end{aligned}$$

where $\mathbb{P}_{\mathbb{S}}(K),\mathbb{P}_{\mathbb{V}}(K),\mathbb{P}_{\mathbb{M}}(K),\mathbb{P}_{\mathbb{U}}(K)$ are spaces of polynomials on the element $K\in\mathcal{T}_n$. It will be convenient to define the following subspaces:

$$M^n_0:=M^n\cap L^2_0(\Omega),\quad \Sigma^n_{\mathrm{tr}}:=\Sigma^n\cap L^2_{\mathrm{sym,tr}}(\Omega)^{d\times d},$$
$$V^n_{\mathrm{div}}:=\left\{\boldsymbol{v}\in V^n\colon \int_\Omega q\,\mathrm{div}\,\boldsymbol{v}=0\quad\forall q\in M^n\right\}.$$

Assumption 2.4 (Approximability)**.** For every $s \in [1,\infty)$ we have that

$$\begin{aligned}
\inf_{\overline{\boldsymbol{v}} \in V^n} \|\boldsymbol{v} - \overline{\boldsymbol{v}}\|_{W^{1,s}(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, \boldsymbol{v} \in W^{1,s}_0(\Omega)^d,\\
\inf_{\overline{q} \in M^n} \|q - \overline{q}\|_{L^s(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, q \in L^s(\Omega),\\
\inf_{\overline{\boldsymbol{\sigma}} \in \Sigma^n} \|\boldsymbol{\sigma} - \overline{\boldsymbol{\sigma}}\|_{L^s(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, \boldsymbol{\sigma} \in L^s(\Omega)^{d\times d},\\
\inf_{\overline{w} \in U^n} \|w - \overline{w}\|_{W^{1,s}(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, w \in W^{1,s}_{\Gamma_D}(\Omega).
\end{aligned}$$

Assumption 2.5 (Fortin Projector Π^n_Σ)**.** For each $n \in \mathbb{N}$ there is a linear projector $\Pi^n_\Sigma \colon L^1_{\text{sym}}(\Omega)^{d\times d} \to \Sigma^n$ such that:

- (Preservation of divergence). For any $\boldsymbol{\sigma} \in L^1_{\text{sym}}(\Omega)^{d\times d}$ we have that

$$\int_\Omega \boldsymbol{\sigma} : \mathsf{D}(\boldsymbol{v}) = \int_\Omega \Pi^n_\Sigma(\boldsymbol{\sigma}) : \mathsf{D}(\boldsymbol{v}) \quad \forall\, \boldsymbol{v} \in V^n_{\text{div}}.$$

- (L^s–stability). For every $s \in (1,\infty)$ there is a constant $c > 0$, independent of n, such that:

$$\|\Pi^n_\Sigma \boldsymbol{\sigma}\|_{L^s(\Omega)} \le c\|\boldsymbol{\sigma}\|_{L^s(\Omega)} \qquad \forall\, \boldsymbol{\sigma} \in L^s_{sym}(\Omega)^{d\times d}.$$

Assumption 2.6 (Fortin Projector Π^n_V)**.** For each $n \in \mathbb{N}$ there is a linear projector $\Pi^n_V : W^{1,1}_0(\Omega)^d \to V^n$ such that the following properties hold:

- (Preservation of divergence). For any $\boldsymbol{v} \in W^{1,1}_0(\Omega)^d$ we have that

$$\int_\Omega q \,\operatorname{div} \boldsymbol{v} = \int_\Omega q \,\operatorname{div}(\Pi^n_V \boldsymbol{v}) \quad \forall\, q \in M^n.$$

- ($W^{1,s}$–stability). For every $s \in (1,\infty)$ there is a constant $c > 0$, independent of n, such that:

$$\|\Pi^n_V \boldsymbol{v}\|_{W^{1,s}(\Omega)} \le c\|\boldsymbol{v}\|_{W^{1,s}(\Omega)} \qquad \forall\, \boldsymbol{v} \in W^{1,s}_0(\Omega)^d.$$

Assumption 2.7 (Projectors Π^n_M, Π^n_U)**.** For each $n \in \mathbb{N}$ there is a linear projector $\Pi^n_M : L^1(\Omega) \to M^n$ and a linear projector $\Pi^n_U \colon W^{1,1}_{\Gamma_D}(\Omega) \to U^n$ such that for all $s \in (1,\infty)$ there is a constant $c > 0$, independent of n, such that:

$$\begin{aligned}
\|\Pi^n_M q\|_{L^s(\Omega)} &\le c\|q\|_{L^s(\Omega)} &&\forall\, q \in L^s(\Omega),\\
\|\Pi^n_U w\|_{L^s(\Omega)} &\le c\|w\|_{W^{1,s}(\Omega)} &&\forall\, w \in W^{1,s}_{\Gamma_D}(\Omega).
\end{aligned}$$

The stability and approximability assumptions above imply immediately that for any $s \in [1,\infty)$ we have:

$$\begin{aligned}
\|\boldsymbol{\sigma} - \Pi^n_\Sigma \boldsymbol{\sigma}\|_{L^s(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, \boldsymbol{\sigma} \in L^s_{\text{sym}}(\Omega)^{d\times d},\\
\|\boldsymbol{v} - \Pi^n_V \boldsymbol{v}\|_{W^{1,s}(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, \boldsymbol{v} \in W^{1,s}_0(\Omega)^d,\\
\|q - \Pi^n_M q\|_{L^s(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, q \in L^s(\Omega),\\
\|w - \Pi^n_U w\|_{W^{1,s}(\Omega)} &\to 0 \quad \text{as } n \to \infty \quad \forall\, w \in W^{1,s}_{\Gamma_D}(\Omega).
\end{aligned} \tag{2.5}$$

In addition, the assumptions guarantee that the velocity-pressure and stress-velocity pairs are inf-sup stable: for any $s \in (1, \infty)$ there are two constants $\beta_s, \gamma_s > 0$, independent of n, such that the following inf-sup conditions are satisfied:

$$\inf_{q \in M^n \setminus \{0\}} \sup_{\boldsymbol{v} \in V^n \setminus \{0\}} \frac{\int_\Omega q \operatorname{div} \boldsymbol{v}}{\|\boldsymbol{v}\|_{W^{1,s}(\Omega)} \|q\|_{L^{s'}(\Omega)}} \geq \beta_s, \tag{2.6}$$

$$\inf_{\boldsymbol{v} \in V^n_{\mathrm{div}} \setminus \{0\}} \sup_{\boldsymbol{\tau} \in \Sigma^n_{\mathrm{sym,tr}} \setminus \{0\}} \frac{\int_\Omega \boldsymbol{\tau} : \mathbf{D}(\boldsymbol{v})}{\|\boldsymbol{\tau}\|_{L^{s'}(\Omega)} \|\boldsymbol{v}\|_{W^{1,s}(\Omega)}} \geq \gamma_s. \tag{2.7}$$

In the literature there are several well-known examples of velocity-pressure pairs V^n–M^n that satisfy the approximability and stability assumptions above. They include, among others, the MINI element, the Taylor–Hood element $\mathbb{P}_k$–$\mathbb{P}_{k-1}$, and the conforming Crouzeix–Raviart element (see e.g. [7, 29, 15]). The Scott–Vogelius pair $\mathbb{P}_k$–$\mathbb{P}^{\mathrm{disc}}_{k-1}$ is another example that in addition has the remarkable property that discretely divergence-free functions are also pointwise divergence-free [58]. This element can be shown to be inf-sup stable for instance on barycentrically refined meshes [50, 67], and the preconditioner to be introduced in Section 4 will be based on a discretisation using this pair. As for the stress variable, if the velocity space consists of continuous piecewise polynomials of degree k (as is the case of the Scott–Vogelius element), then a space satisfying Assumption 2.5 is [22]:

$$\Sigma^n = \{\boldsymbol{\sigma} \in L^\infty_{\mathrm{sym}}(\Omega)^{d\times d} : \boldsymbol{\sigma}|_K \in \mathbb{P}_{k-1}(K)^{d\times d}, \text{ for all } K \in \mathcal{T}_n\}. \tag{2.8}$$

The space of discrete temperatures U^n is not required to satisfy any inf-sup stability conditions, and so it suffices to choose any H^1-conforming space for which the expected order of accuracy is consistent with that of the other variables.

2.4. **Convective term.** A useful property in the analysis of systems describing incompressible fluids is that the convective term vanishes when testing with the divergence-free velocity itself. This is a consequence of the identity

$$-\int_\Omega (\boldsymbol{v} \otimes \boldsymbol{v}) : \mathbf{D}(\boldsymbol{v}) = 0 \quad \text{for all } \boldsymbol{v} \in C^\infty_0(\Omega)^d \text{ with } \operatorname{div} \boldsymbol{v} = 0. \tag{2.9}$$

Such an identity will not be satisfied in general with only discretely divergence-free elements. In order to recover this cancellation property at the discrete level let us define a skew-symmetric form of the convective term as follows:

$$\mathcal{B}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) := \begin{cases} -\displaystyle\int_\Omega \boldsymbol{u} \otimes \boldsymbol{v} : \nabla \boldsymbol{w}, & \text{if } V^n_{\mathrm{div}} \subset W^{1,1}_{0,\mathrm{div}}(\Omega)^d, \\ \dfrac{1}{2} \displaystyle\int_\Omega \boldsymbol{u} \otimes \boldsymbol{w} : \nabla \boldsymbol{v} - \boldsymbol{u} \otimes \boldsymbol{v} : \nabla \boldsymbol{w}, & \text{otherwise.} \end{cases}$$

This new trilinear form now satisfies $\mathcal{B}(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{v}) = 0$ for any $\boldsymbol{v} \in W^{1,\infty}_0(\Omega)^d$, regardless of whether $\boldsymbol{v}$ is divergence-free or not, and it reduces to the original trilinear form $-\int_\Omega (\boldsymbol{u} \otimes \boldsymbol{w}) : \nabla \boldsymbol{w}$ if $\operatorname{div} \boldsymbol{v} = 0$.

Let us now define

$$\tilde{r} := \min\{r', r^*/2\}, \text{ where } r^* := \begin{cases} \frac{dr}{d-r} & \text{if } r < d, \\ \infty, & \text{otherwise.} \end{cases}$$

Observe that the condition $\tilde{r} > 1$ is equivalent to $r > \frac{2d}{d+2}$, which is the natural condition required to have a well-defined weak form of the convective term, because it ensures that $W^{1,r}(\Omega)^d \hookrightarrow L^2(\Omega)^d$. In this case, for exactly divergence-free functions $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V^n_{\mathrm{div}}$ one has that

$$|\mathcal{B}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})| \leq \int_\Omega |\boldsymbol{u} \otimes \boldsymbol{v} : \nabla \boldsymbol{w}| \leq c \|\boldsymbol{u}\|_{W^{1,r}(\Omega)} \|\boldsymbol{v}\|_{W^{1,r}(\Omega)} \|\boldsymbol{w}\|_{W^{1,\tilde{r}'}(\Omega)}. \tag{2.10}$$

Otherwise one needs the stronger assumption $r > \frac{2d}{d+1}$; this ensures that there is an $s \in (1, \infty)$ such that $\frac{1}{r} + \frac{1}{2\tilde{r}} + \frac{1}{s} = 1$ and so (c.f. [16])

$$\begin{aligned} \int_\Omega |\boldsymbol{u} \otimes \boldsymbol{w} : \nabla \boldsymbol{v}| &\leq \|\boldsymbol{u}\|_{L^{2\tilde{r}}(\Omega)} \|\boldsymbol{v}\|_{W^{1,r}(\Omega)} \|\boldsymbol{w}\|_{L^s(\Omega)} \\ &\leq c \|\boldsymbol{u}\|_{W^{1,r}(\Omega)} \|\boldsymbol{v}\|_{W^{1,r}(\Omega)} \|\boldsymbol{w}\|_{W^{1,\tilde{r}'}(\Omega)}, \end{aligned} \tag{2.11}$$

for any $\boldsymbol{u}, \boldsymbol{v} \in W^{1,r}(\Omega)^d, \boldsymbol{w} \in W^{1,\tilde{r}'}(\Omega)^d$. Thus we deduce that the trilinear form $\mathcal{B}(\cdot,\cdot,\cdot)$ is bounded on $W^{1,r}(\Omega)^d \times W^{1,r}(\Omega)^d \times W^{1,\tilde{r}'}(\Omega)^d$ if $r > \frac{2d}{d+2}$ when using exactly divergence-free elements and if $r > \frac{2d}{d+1}$ otherwise. This does not pose a problem when working with the constitutive relation (1.3) (for which $r = 2$), but for relations with more general r-growth the more demanding requirement that $r > \frac{2d}{d+1}$ would impose a restriction on the convergence result that can be obtained (see [16, Thm. 18]). In order to circumvent this issue we shall make use of a reconstruction operator.

Assumption 2.8 (Reconstruction operator π^n)**.** Let X^n be an auxiliary $H(\mathrm{div};\Omega)$-conforming finite element space. There exists a map $\pi^n \colon W^{1,1}(\Omega)^d \to V^n + X^n$ (usually called a reconstruction operator) that satisfies:

- (Preservation of Divergence). If $\boldsymbol{v} \in V^n_{\mathrm{div}}$ then $\mathrm{div}(\pi^n \boldsymbol{v}) = 0$ pointwise.
- (Consistency). For every $\boldsymbol{v} \in V^n$ and $K \in \mathcal{T}_n$ it holds that

$$\|\boldsymbol{v} - \pi^n \boldsymbol{v}\|_{L^s(K)} \leq c h_K^m |\boldsymbol{v}|_{W^{m,s}(K)}, \qquad \text{for } s \in [1,\infty),\ m \in \{0,1,2\}.$$

Operators with the properties described above have been constructed in [41, 43, 42, 44, 33] for elements with discontinuous pressures; the construction is based on the interpolation operators associated with the Raviart–Thomas and Brezzi–Douglas–Marini elements. A slightly more complicated construction for elements with continuous pressures was introduced in [39]. These reconstruction operators have been employed to obtain pressure-robust discretisations by “repairing” the L^2-orthogonality between discretely divergence-free functions and gradient fields; see [33] for more details. In order to exploit the advantages of this framework one has to replace the L^2 inner products in the discrete formulation in the following way:

$$\int_\Omega \boldsymbol{w} \cdot \boldsymbol{v} \mapsto \int_\Omega \boldsymbol{w} \cdot \pi^n \boldsymbol{v}, \tag{2.12}$$

where $\boldsymbol{v} \in V^n$ is a test function. As for the convective term, let us define

$$\tilde{\mathcal{B}}_n(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) := \begin{cases} -\displaystyle\int_\Omega \boldsymbol{u} \otimes \boldsymbol{v} : \nabla \boldsymbol{w}, & \text{if } V^n_{\mathrm{div}} \subset W^{1,1}_{0,\mathrm{div}}(\Omega)^d, \\ -\int_\Omega \boldsymbol{u} \otimes \pi^n \boldsymbol{v} : \nabla \boldsymbol{w}, & \text{otherwise.} \end{cases} \tag{2.13}$$

From the properties of π^n stated in Assumption 2.8 one readily sees that the trilinear form $\tilde{\mathcal{B}}_n$ is bounded on $W^{1,r}(\Omega)^d \times W^{1,r}(\Omega)^d \times W^{1,\tilde{r}'}(\Omega)^d$, and that $\tilde{\mathcal{B}}_n(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{v}) = 0$ for any $\boldsymbol{v} \in V^n_{\mathrm{div}}$.

For the advective term for the temperature one can analogously define the trilinear form

$$\mathcal{C}(\boldsymbol{u},\theta,\eta) := \begin{cases} -\displaystyle\int_\Omega \boldsymbol{u}\theta\cdot\nabla\eta, & \text{if } V^n_{\mathrm{div}} \subset W^{1,1}_{0,\mathrm{div}}(\Omega)^d, \\ \dfrac{1}{2}\displaystyle\int_\Omega \boldsymbol{u}\eta\cdot\nabla\theta - \boldsymbol{u}\theta\cdot\nabla\eta, & \text{otherwise,} \end{cases}$$

which is well defined and bounded on $W^{1,r}(\Omega)^d \times H^1(\Omega) \times W^{1,\infty}(\Omega)$ assuming that $r > \frac{2d}{d+2}$. In addition, this form satisfies $\mathcal{C}(\boldsymbol{u},\eta,\eta) = 0$ for any $\eta \in W^{1,\infty}(\Omega)$, regardless of whether $\boldsymbol{u}$ is divergence-free or not. The form $\mathcal{C}$ does not impose additional restrictions like $\mathcal{B}$ does for small r, but a trilinear form using a reconstruction operator $\tilde{\mathcal{C}}_n$ could be used instead (and defined analogously).

3. Finite Element Approximation

Let us now set the physical constants to unity for ease of readability (appropriate non-dimensional forms of the system will be employed in Section 5). Suppose that $\theta_b \in H^{1/2}_{00}(\Gamma_D) := W^{1/2,2}_{00}(\Gamma_D)$, and let $\hat{\theta}_b \in H^1(\Omega)$ be such that $\hat{\theta}_b|_{\Gamma_D} = \theta_b$. We can now define the weak formulation of the system (without viscous heating).

Formulation $\mathbf{A_0}$. Find $(\mathbf{S},\theta,\boldsymbol{u},p) \in L^2_{\mathrm{sym,tr}}(\Omega)^{d\times d} \times (\hat{\theta}_b + H^1_{\Gamma_D}(\Omega)) \times H^1_0(\Omega)^d \times L^2_0(\Omega)$ such that:

$$\int_\Omega \mathbf{S} : \mathbf{D}(\boldsymbol{v}) - \int_\Omega \boldsymbol{u}\otimes\boldsymbol{u} : \mathbf{D}(\boldsymbol{v}) - \int_\Omega p \,\mathrm{div}\, \boldsymbol{v} = \int_\Omega \theta\boldsymbol{v}\cdot\boldsymbol{e}_d \quad \forall\, \boldsymbol{v} \in C^\infty_0(\Omega)^d, \tag{3.1a}$$

$$-\int_\Omega q \,\mathrm{div}\, \boldsymbol{u} = 0 \quad \forall\, q \in C^\infty_0(\Omega), \tag{3.1b}$$

$$\int_\Omega \hat{\kappa}(\theta)\nabla\theta\cdot\nabla\eta - \boldsymbol{u}\theta\cdot\nabla\eta = 0 \quad \forall\, \eta \in C^\infty_{\Gamma_D}(\Omega), \tag{3.1c}$$

$$\mathbf{G}(\mathbf{S},\mathbf{D}(\boldsymbol{u}),\theta) = \mathbf{0} \quad \text{a.e. in } \Omega. \tag{3.1d}$$

Let $\hat{\theta}^n_b$ be the standard Scott–Zhang interpolant of $\hat{\theta}_b$ into $\hat{U}^n$, where $\hat{U}^n$ is the same finite element space as U^n, but without strongly imposed boundary conditions. We have everything in place to state the finite element approximation of the problem.

Formulation $\mathbf{A^n_0}$. Find $(\theta^n,\boldsymbol{u}^n,p^n) \in (\hat{\theta}^n_b + U^n) \times V^n \times M^n_0$ such that:

(3.2a)
$$\int_\Omega \boldsymbol{\mathcal{S}}^n(\mathbf{D}(\boldsymbol{u}^n),\theta^n) : \mathbf{D}(\boldsymbol{v}) + \mathcal{B}(\boldsymbol{u}^n,\boldsymbol{u}^n,\boldsymbol{v}) - \int_\Omega p^n \,\mathrm{div}\, \boldsymbol{v} = \int_\Omega \theta^n\boldsymbol{v}\cdot\boldsymbol{e}_d \quad \forall\, \boldsymbol{v} \in V^n,$$

$$-\int_\Omega q \,\mathrm{div}\, \boldsymbol{u}^n = 0 \quad \forall\, q \in M^n, \tag{3.2b}$$

$$\int_\Omega \hat{\kappa}(\theta^n)\nabla(\theta^n)\cdot\nabla\eta + \mathcal{C}(\boldsymbol{u}^n,\theta^n,\eta) = 0 \quad \forall\, \eta \in U^n. \tag{3.2c}$$

In case one wishes to compute the shear stress directly, a 4-field formulation may be employed instead. We refer to this formulation as Formulation B^n_0. We will prove that the solutions to the discrete formulations A^n_0 and B^n_0 converge to a weak solution of Formulation A_0.

Formulation $\mathbf{B_0^n}$. Find $(\mathsf{S}^n, \theta^n, \boldsymbol{u}^n, p^n) \in \Sigma^n \times (\hat{\theta}_b^n + U^n) \times V^n \times M_0^n$ such that:

$$\int_\Omega (\boldsymbol{\mathcal{D}}^n(\mathsf{S}^n, \theta^n) - \mathsf{D}(\boldsymbol{u}^n)) : \boldsymbol{\tau} = 0 \qquad \forall \boldsymbol{\tau} \in \Sigma^n, \tag{3.3a}$$

$$\int_\Omega \mathsf{S}^n : \mathsf{D}(\boldsymbol{v}) + \mathcal{B}(\boldsymbol{u}^n, \boldsymbol{u}^n, \boldsymbol{v}) - \int_\Omega p^n \operatorname{div} \boldsymbol{v} = \int_\Omega \theta^n \boldsymbol{v} \cdot \boldsymbol{e}_d \quad \forall \boldsymbol{v} \in V^n, \tag{3.3b}$$

$$-\int_\Omega q \operatorname{div} \boldsymbol{u}^n = 0 \qquad \forall q \in M^n, \tag{3.3c}$$

$$\int_\Omega \hat{\kappa}(\theta^n) \nabla \theta^n \cdot \nabla \eta + \mathcal{C}(\boldsymbol{u}^n, \theta^n) = 0 \qquad \forall \eta \in U^n. \tag{3.3d}$$

We define Formulations $\tilde{\mathrm{A}}_0^n$ and $\tilde{\mathrm{B}}_0^n$ as the analogues of the formulations A_0^n and B_0^n, respectively, in which we replace $\mathcal{B}$ and $\mathcal{C}$ by $\tilde{\mathcal{B}}_n$ and $\tilde{\mathcal{C}}_n$. The following lemma asserts that all of these formulations have a solution.

Lemma 3.1. *Suppose the material parameters satisfy condition* (2.1) *and suppose that* $\{U^n, V^n, M^n\}_{n \in \mathbb{N}}$ *(respectively* $\{\Sigma^n, U^n, V^n, M^n\}_{n \in \mathbb{N}}$*) is a family of finite element spaces satisfying Assumptions 2.4 and 2.6–2.7 (resp. 2.4–2.7). In the case of formulations* $\tilde{\mathrm{A}}_0^n$ *and* $\tilde{\mathrm{B}}_0^n$ *suppose further that Assumption 2.8 holds. Then, for every* $n \in \mathbb{N}$*, Formulations* A_0^n *and* $\tilde{\mathrm{A}}_0^n$ *(resp.* B_0^n *and* $\tilde{\mathrm{B}}_0^n$*) admit a solution* $(\theta^n, \boldsymbol{u}^n, p^n) \in (\hat{\theta}_b^n + U^n) \times V^n \times M_0^n$ *(resp.* $(\mathsf{S}^n, \theta^n, \boldsymbol{u}^n, p^n) \in \Sigma^n \times (\hat{\theta}_b^n + U^n) \times V^n \times M_0^n$*). Moreover, the following a priori estimate holds:*

$$\|\boldsymbol{u}^n\|_{H^1(\Omega)} + \|\theta^n\|_{H^1(\Omega)} + \|p^n\|_{L^2(\Omega)} + \|\mathsf{S}^n\|_{L^2(\Omega)} \leq c, \tag{3.4a}$$

where the constant c *is independent of* n*; we denote* $\mathsf{S}^n := \boldsymbol{\mathcal{S}}^n(\mathsf{D}(\boldsymbol{u}^n), \theta^n)$ *in the case of Formulations* A_0^n *and* $\tilde{\mathrm{A}}_0^n$*. In addition, for Formulations* B_0^n *and* $\tilde{\mathrm{B}}_0^n$ *we have*

$$\|\boldsymbol{\mathcal{D}}^n(\mathsf{S}^n, \theta^n)\|_{L^2(\Omega)} \leq c. \tag{3.4b}$$

Proof. We will carry out the proof for Formulation B_0^n; the proof for the other formulations is analogous with some simplifications. The existence proof will make use of a fixed point argument. Let θ_0^n be an arbitrary nonzero element of $\hat{\theta}_b^n + U^n$ and define, for $j \in \mathbb{N}$, the function $\theta_j^n \in \hat{\theta}_b^n + U^n$ as follows: given θ_{j-1}^n we first find $(\mathsf{S}_j^n, \boldsymbol{u}_j^n, p_j^n) \in \Sigma^n \times V^n \times M_0^n$ by solving

$$\int_\Omega (\boldsymbol{\mathcal{D}}^n(\mathsf{S}_j^n, \theta_{j-1}^n) - \mathsf{D}(\boldsymbol{u}_j^n)) : \boldsymbol{\tau} = 0 \qquad \forall \boldsymbol{\tau} \in \Sigma^n, \tag{3.5a}$$

(3.5b)
$$\int_\Omega \left(\frac{1}{j} \mathsf{D}(\boldsymbol{u}_j^n) + \mathsf{S}_j^n \right) : \mathsf{D}(\boldsymbol{v}) + \mathcal{B}(\boldsymbol{u}_j^n, \boldsymbol{u}_j^n, \boldsymbol{v}) - \int_\Omega p_j^n \operatorname{div} \boldsymbol{v} = \int_\Omega \theta_{j-1}^n \boldsymbol{v} \cdot \boldsymbol{e}_d \ \forall \boldsymbol{v} \in V^n,$$

$$-\int_\Omega q \operatorname{div} \boldsymbol{u}_j^n = 0 \qquad \forall q \in M^n, \tag{3.5c}$$

and then θ_j^n is defined as $\hat{\theta}_b^n + \tilde{\theta}_j^n$, where $\tilde{\theta}_j^n \in U^n$ is the solution of the nonlinear problem

$$\int_\Omega \hat{\kappa}(\tilde{\theta}_j^n + \hat{\theta}_b^n) \nabla(\tilde{\theta}_j^n + \hat{\theta}_b^n) \cdot \nabla \eta + \mathcal{C}(\boldsymbol{u}_j^n, \tilde{\theta}_j^n + \hat{\theta}_b^n, \eta) = 0 \qquad \forall \eta \in U^n. \tag{3.6}$$

In order to show that the problem (3.5) is well-posed, let us define a mapping $F^n_j\colon \Sigma^n \times V^n_{\mathrm{div}} \to (\Sigma^n \times V^n_{\mathrm{div}})^*$ by

$$\begin{aligned}\langle F^n_j(\boldsymbol{\sigma},\boldsymbol{v});(\boldsymbol{\tau},\boldsymbol{w})\rangle := \int_\Omega &(\boldsymbol{\mathcal{D}}^n(\boldsymbol{\sigma},\theta^n_{j-1}):\boldsymbol{\tau} - \mathbf{D}(\boldsymbol{v}):\boldsymbol{\tau} + \frac{1}{j}\mathbf{D}(\boldsymbol{v}):\mathbf{D}(\boldsymbol{w}) \\ &+ \boldsymbol{\sigma}:\mathbf{D}(\boldsymbol{w}) + \mathcal{B}(\boldsymbol{v},\boldsymbol{v},\boldsymbol{w}) - \theta^n_j \boldsymbol{v}\cdot\boldsymbol{e}_d).\end{aligned}$$

By using the coercivity of $\boldsymbol{\mathcal{D}}^n$ and the fact that $\mathcal{B}(\boldsymbol{v},\boldsymbol{v},\boldsymbol{v}) = 0$, one obtains using the inequalities of Young, Korn and Poincaré that there exists a $\delta(j) > 0$ such that

$$\langle F^n_j(\boldsymbol{\sigma},\boldsymbol{v}),(\boldsymbol{\sigma},\boldsymbol{v})\rangle > 0 \quad \text{if} \quad \|(\boldsymbol{\sigma},\boldsymbol{v})\| = \delta(j).$$

A corollary of Brouwer's fixed point theorem [29, Ch. 4, Cor. 1.1] guarantees the existence of functions $(\mathbf{S}^n_j,\boldsymbol{u}^n_j) \in \Sigma^n \times V^n_{\mathrm{div}}$ satisfying $F^n_j(\mathbf{S}^n_j,\boldsymbol{u}^n_j) = 0$ (which is equivalent to (3.5) with divergence-free test functions) and such that $\|(\mathbf{S}^n_j,\boldsymbol{u}^n_j)\| \leq \delta(j)$. The existence of $p^n_j \in M^n_0$ then follows from the inf-sup condition (2.6). A similar argument can be used to prove the well-posedness of the problem (3.6).

Now, the inf-sup condition (2.7) and the discrete form of the constitutive relation (3.3a) allow us to control, uniformly in j and n, the norm of the velocity in terms of the stress:

$$\gamma_2\|\boldsymbol{u}^n_j\|_{H^1(\Omega)} \leq \|\mathbf{S}^n_j\|_{L^2(\Omega)}. \tag{3.7}$$

Therefore, testing (3.5) with $(\mathbf{S}^n_j,\boldsymbol{u}^n_k,p^n_j)$ yields the estimate

$$\|\boldsymbol{\mathcal{D}}^n(\mathbf{S}^n_j,\theta^n_{j-1})\|^2_{L^2(\Omega)} + \|\mathbf{S}^n_j\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^n_j\|^2_{H^1(\Omega)} \leq c\|\theta^n_{j-1}\|^2_{L^2(\Omega)}, \tag{3.8}$$

where $c > 0$ is independent of j and n. The inf-sup condition (2.6) and the discrete momentum equation in turn imply an estimate for the pressure:

$$\|p^n_j\|^2_{L^2(\Omega)} \leq c\|\theta^n_{j-1}\|^2_{L^2(\Omega)}. \tag{3.9}$$

Furthermore, testing (3.6) with $\theta^n_j - \hat{\theta}^n_b$ results in

$$\|\theta^n_j\|^2_{H^1(\Omega)} \leq c\|\boldsymbol{u}^n_j\|^2_{H^1(\Omega)}. \tag{3.10}$$

Hence, up to a subsequence, we have as $j \to \infty$ that

$$\begin{aligned} \boldsymbol{\mathcal{D}}^n(\mathbf{S}^n_j,\theta^n_{j-1}) &\rightharpoonup \overline{\mathbf{D}}^n && \text{weakly in } L^2_{\mathrm{sym}}(\Omega)^{d\times d},\\ \mathbf{S}^n_j &\to \mathbf{S}^n && \text{strongly in } L^2_{\mathrm{sym}}(\Omega)^{d\times d},\\ \boldsymbol{u}^n_j &\to \boldsymbol{u}^n && \text{strongly in } H^1(\Omega)^d,\\ p^n_j &\to p^n && \text{strongly in } L^2(\Omega),\\ \theta^n_j &\to \theta^n && \text{strongly in } H^1(\Omega), \end{aligned} \tag{3.11}$$

where we used the fact that weak and strong convergence are equivalent in finite-dimensional spaces. Since $\boldsymbol{\mathcal{D}}^n$ is continuous and the convergences are strong, one can straightforwardly identify $\overline{\mathbf{D}}^n = \boldsymbol{\mathcal{D}}^n(\mathbf{S}^n,\theta^n)$ and pass to the limit to show that $(\mathbf{S}^n,\theta^n,\boldsymbol{u}^n,p^n)$ solve Formulation B^n_0. Now, testing Formulation B^n_0 with $(\mathbf{S}^n,\theta^n - \hat{\theta}^n_b,\boldsymbol{u}^n,p^n)$ allows one to obtain the estimate (3.4). Note that the inf-sup conditions were essential to obtain estimates that are uniform in n. □

Having shown that the discrete problems are well-posed, we now consider the question of convergence.

Theorem 3.2. *Suppose the same assumptions as in Lemma 3.1 hold and suppose that* $\{(\theta^n, \boldsymbol{u}^n, p^n)\}_{n\in\mathbb{N}}$ *(respectively* $(\{\mathbf{S}^n, \theta^n, \boldsymbol{u}^n, p^n\}_{\mathbb{N}})$*) is a sequence of solutions of Formulation* A_0^n *or* $\tilde{\mathrm{A}}_0^n$ *(resp. Formulation* B_0^n *or* $\tilde{\mathrm{B}}_0^n$*). Then there exists a solution* $(\mathbf{S}, \theta, \boldsymbol{u}, p) \in L^2_{\mathrm{sym,tr}}(\Omega)^{d\times d} \times (\hat{\theta}_b + H^1_{\Gamma_D}(\Omega)) \times H^1_0(\Omega)^d \times L^2_0(\Omega)$ *of Formulation* A_0 *such that, up to a subsequence, as* $n \to \infty$*:*

$$
\begin{aligned}
\mathbf{S}^n &\rightharpoonup \mathbf{S} && \textit{weakly in } L^2_{\mathrm{sym}}(\Omega)^{d\times d},\\
\boldsymbol{u}^n &\rightharpoonup \boldsymbol{u} && \textit{weakly in } H^1(\Omega)^d,\\
p^n &\rightharpoonup p && \textit{weakly in } L^2(\Omega),\\
\theta^n &\rightharpoonup \theta && \textit{weakly in } H^1(\Omega),
\end{aligned}
\tag{3.12}
$$

where in the case of Formulations A_0^n *and* $\tilde{\mathrm{A}}_0^n$ *we denote* $\mathbf{S}^n := \boldsymbol{\mathcal{S}}^n(\mathbf{D}(\boldsymbol{u}^n), \theta^n)$*.*

Proof. We will once again focus on Formulation B_0^n, since the other cases are completely analogous. From the *a priori* estimate (3.4) and the fact that $\hat{\theta}_b^n \to \hat{\theta}_b$ in $H^1(\Omega)$, we immediately obtain the convergences (3.12) (for a not relabelled subsequence) for some $(\mathbf{S}, \theta, \boldsymbol{u}, p) \in L^2_{\mathrm{sym}}(\Omega)^{d\times d} \times (\hat{\theta}_b + H^1_{\Gamma_D}(\Omega)) \times H^1_0(\Omega)^d \times L^2_0(\Omega)$, and that

$$
\boldsymbol{\mathcal{D}}^n(\mathbf{S}^n, \theta^n) \rightharpoonup \overline{\mathbf{D}} \quad \text{weakly in } L^2_{\mathrm{sym}}(\Omega)^{d\times d}. \tag{3.13}
$$

All that is left to prove is that the limiting functions are a solution of Formulation A_0.

Let $\boldsymbol{\tau} \in L^2_{\mathrm{sym}}(\Omega)^{d\times d}$ be arbitrary. Then (3.12) and (2.5) result in

$$
0 = \int_\Omega (\boldsymbol{\mathcal{D}}^n(\mathbf{S}^n, \theta^n) - \mathbf{D}(\boldsymbol{u}^n)) : \Pi^n_\Sigma \boldsymbol{\tau} \xrightarrow[n\to\infty]{} \int_\Omega (\overline{\mathbf{D}} - \mathbf{D}(\boldsymbol{u})) : \boldsymbol{\tau}, \tag{3.14}
$$

and therefore $\overline{\mathbf{D}} = \mathbf{D}(\boldsymbol{u})$ almost everywhere. Similarly, for an arbitrary $q \in L^2_0(\Omega)$ one obtains that

$$
0 = \int_\Omega \operatorname{div} \boldsymbol{u}^n\, \Pi^n_M q \xrightarrow[n\to\infty]{} \int_\Omega \operatorname{div} \boldsymbol{u}\, q, \tag{3.15}
$$

and so $\boldsymbol{u}$ is pointwise divergence-free. One can pass to the limit in (3.3b) and (3.3d) in a similar manner, but perhaps the convective terms are worth looking at in more detail. To that end, first note that the Sobolev embedding theorem ensures that (up to a subsequence) we have, for any $p \in [1, 2^*)$,

$$
\begin{aligned}
\boldsymbol{u}^n &\to \boldsymbol{u} && \text{strongly in } L^p(\Omega)^d,\\
\theta^n &\to \theta && \text{strongly in } L^p(\Omega),\\
\theta^n &\to \theta && \text{a.e. in } \Omega.
\end{aligned}
\tag{3.16}
$$

The strong convergence of $\boldsymbol{u}^n$ suffices to prove that, for an arbitrary $\boldsymbol{v} \in H^1_0(\Omega)^d$:

$$
\mathcal{B}(\boldsymbol{u}^n, \boldsymbol{u}^n, \Pi^n_V \boldsymbol{v}) \xrightarrow[n\to\infty]{} \frac{1}{2}\int_\Omega \boldsymbol{u}\otimes\boldsymbol{v} : \nabla\boldsymbol{u} - \boldsymbol{u}\otimes\boldsymbol{u} : \nabla\boldsymbol{v} = -\int_\Omega \boldsymbol{u}\otimes\boldsymbol{u} : \mathbf{D}\boldsymbol{v}, \tag{3.17}
$$

where the last equality is a consequence of the fact that $\operatorname{div}\boldsymbol{u} = 0$. Now, from testing the discrete momentum equation with $\boldsymbol{u}^n$ and taking (3.16) into account we observe that

$$
\limsup_{n\to\infty} \int_\Omega \mathbf{S}^n : \mathbf{D}(\boldsymbol{u}^n) = \lim_{n\to\infty} \int_\Omega \theta^n \boldsymbol{u}^n \cdot \boldsymbol{e}_d = \int_\Omega \theta \boldsymbol{u} \cdot \boldsymbol{e}_d = \int_\Omega \mathbf{S} : \mathbf{D}(\boldsymbol{u}), \tag{3.18}
$$

and hence by Lemma 2.2 we conclude that $\mathbf{G}(\mathbf{S}, \mathbf{D}(\boldsymbol{u}), \theta) = \mathbf{0}$. Finally, by taking traces on both sides of the constitutive relation we also obtain that $\operatorname{tr} \mathbf{S} = 0$ and so $\mathbf{S} \in L^2_{\mathrm{sym,tr}}(\Omega)^{d\times d}$, which concludes the proof. □

In the proof of Theorem 3.2 it becomes clear that the only bottleneck that prevents one from considering constitutive laws with more general r-coercivity (e.g. a power-law with temperature dependent consistency), is the fact that Lemma 2.2 is tied to the particular function $\mathbf{G}$ defined in (1.3). Using Minty's trick it is possible to show that if an explicit constitutive relation is available, an analogous convergence result will hold.

Assumption 3.3. Let $\boldsymbol{\mathcal{S}} : \Omega \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ be a continuous function satisfying, for some $r > \frac{2d}{d+2}$:

- (Monotonicity). For every $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$:
$$(\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}_1, s) - \boldsymbol{\mathcal{S}}(\boldsymbol{\tau}_2, s)) : (\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2) \geq 0 \text{ for fixed } s \in \mathbb{R};$$
- (Coercivity). There is a non-negative function $m \in L^1(\Omega)$ and a constant $c > 0$ such that
$$\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, s) : \boldsymbol{\tau} \geq -m + c(|\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, s)|^{r'} + |\boldsymbol{\tau}|^r) \quad \text{ for all } \boldsymbol{\tau} \in \mathbb{R}^{d\times d}_{\mathrm{sym}}, s \in \mathbb{R};$$
- (Growth). There is a function $n \in L^{r'}(\Omega)$ and a constant $c > 0$ such that
$$|\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, s)| \leq c(|\boldsymbol{\tau}|^{r'-1} + n);$$
- (Compatibility). For a fixed $s \in \mathbb{R}$ we have that $\operatorname{tr}(\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, s)) = 0$ if and only if $\operatorname{tr}(\boldsymbol{\tau}) = 0$, for any $\boldsymbol{\tau} \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$.

When $r < \frac{3d}{d+2}$ the velocity $\boldsymbol{u}$ is not an admissible test function anymore and so obtaining an identity such as (3.18) is not straightforward. This difficulty can be overcome by testing instead with a discrete Lipschitz truncation of the error $e^n := \boldsymbol{u} - \boldsymbol{u}^n$. The discrete Lipschitz truncation was introduced in [16], and the idea is that it turns $\boldsymbol{e}^n$ into a Lipschitz function belonging to V^n in such a way that the size of the set where the truncation does not equal the original function can be controlled. We note that the construction of this discrete Lipschitz truncation requires a refined version of Assumption 2.6.

Assumption 3.4 (Fortin Projector Π^n_V)**.** For each $n \in \mathbb{N}$ there is a linear projector $\Pi^n_V : W^{1,1}_0(\Omega)^d \to V^n$ such that it preserves the divergence in the same sense as in assumption 2.6, but the stability condition is replaced by:

- (Local $W^{1,1}$-stability). For every $s \in (1, \infty)$ there is a constant $c > 0$, independent of n, such that
$$\frac{1}{|K|}\int_K |\nabla \Pi^n_V \boldsymbol{v}| \leq c \frac{1}{|\Omega^n_K|}\int_{\Omega^n_K} |\nabla \boldsymbol{v}| \qquad \forall\, \boldsymbol{v} \in W^{1,s}_0(\Omega)^d, K \in \mathcal{T}_n,$$
where Ω^n_K denotes the patch of elements in $\mathcal{T}_n$ whose intersection with K is nonempty.

It can be shown that the local $W^{1,1}$-stability from Assumption 3.4 implies the global $W^{1,s}$-stability of Assumption 2.6 [4, 16]. Some examples of finite elements satisfying Assumption 3.4 include the conforming Crouzeix–Raviart element, the MINI element, the Bernardi–Raugel element, the $\mathbb{P}_2$–$\mathbb{P}_0$ and the Taylor–Hood pair $\mathbb{P}_k$–$\mathbb{P}_{k-1}$ for $k \geq d$ [4]; the lowest order Taylor–Hood pair in 3D also satisfies the assumption if the mesh has a certain macroelement structure [28]. As for exactly divergence-free elements, this assumption can also be verified for low order Guzmán–Neilan elements and the Scott–Vogelius pair [16, 61].

Corollary 3.5. *Let $r > \frac{2d}{d+2}$ and let $\boldsymbol{\mathcal{S}} : \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ be a function satisfying Assumption 3.3 and suppose that $\{U^n, V^n, M^n\}_{n\in\mathbb{N}}$ is a family of finite element subspaces satisfying Assumptions 2.4, 2.7, 2.8, and 3.4. Then, for any $n \in \mathbb{N}$, the finite element formulation obtained by replacing $\boldsymbol{\mathcal{S}}^n$ by $\boldsymbol{\mathcal{S}}$ in Formulation $\tilde{\mathrm{A}}^n_0$ admits a solution $(\theta^n, \boldsymbol{u}^n, p^n) \in (\hat{\theta}^n_b + U^n) \times V^n \times M^n_0$ and we have, up to subsequences, that*

$$\begin{aligned}
\boldsymbol{u}^n &\rightharpoonup \boldsymbol{u} && \text{weakly in } W^{1,r}(\Omega)^d,\\
p^n &\rightharpoonup p && \text{weakly in } L^{\tilde{r}}(\Omega),\\
\theta^n &\rightharpoonup \theta && \text{weakly in } H^1(\Omega),\\
\boldsymbol{\mathcal{S}}(\mathsf{D}(\boldsymbol{u}^n), \theta^n) &\rightharpoonup \mathsf{S} && \text{weakly in } L^{r'}_{\mathrm{sym}}(\Omega)^{d\times d},
\end{aligned}$$

where $(\mathsf{S}, \theta, \boldsymbol{u}, p) \in L^{r'}_{\mathrm{sym,tr}}(\Omega)^{d\times d} \times (\hat{\theta}_b + H^1_{\Gamma_D}(\Omega)) \times W^{1,r}_0(\Omega)^d \times L^{\tilde{r}}_0(\Omega)$ is a solution of Formulation A_0.

Proof. The proof is entirely analogous to the proofs of Lemma 3.1 and Theorem 3.2, with a couple of small differences. Firstly, the *a priori* estimate (3.4) changes to

$$\|\boldsymbol{u}^n\|_{W^{1,r}(\Omega)^d} + \|\theta^n\|_{H^1(\Omega)} + \|p^n\|_{L^{\tilde{r}}(\Omega)} + \|\mathsf{S}^n\|_{L^{r'}(\Omega)} \leq c, \tag{3.19}$$

which implies the desired weak convergences. On the other hand, since $r > \frac{2d}{d+2}$, for a small enough $\varepsilon > 0$ we have that $r > \frac{(2+\varepsilon)d}{d+(2+\varepsilon)}$, which implies that $\boldsymbol{u}^n \to \boldsymbol{u}$ strongly in $L^{2+\varepsilon}(\Omega)^d$ as $n \to \infty$. Furthermore, from the consistency condition in Assumption 2.8 we see that

$$\|\pi^n \boldsymbol{u}^n - \boldsymbol{u}\|_{L^{2+\varepsilon}(K)} \leq \|\boldsymbol{u}^n - \boldsymbol{u}\|_{L^{2+\varepsilon}(K)} + c h_K^{1+d(\frac{1}{2+\varepsilon} - \frac{1}{r})} \|\boldsymbol{u}^n\|_{W^{1,r}(K)},$$

where we have used a standard local inverse inequality; the exponent of h_K is positive by the choice of ε, which implies that $\pi^n \boldsymbol{u}^n \to \boldsymbol{u}$ strongly in $L^{2+\varepsilon}(\Omega)^d$ as $n \to \infty$. This is enough to pass to the limit in the convective term:

$$\tilde{\mathcal{B}}_n(\boldsymbol{u}^n, \boldsymbol{u}^n, \Pi^n \boldsymbol{v}) \xrightarrow[n\to\infty]{} -\int_\Omega \boldsymbol{u} \otimes \boldsymbol{u} : \mathsf{D}(\boldsymbol{v}), \tag{3.20}$$

for any $\boldsymbol{v} \in W^{1,(\frac{2+\varepsilon}{2})'}_0(\Omega)^d$. As for the identification of the constitutive relation, by testing the discrete momentum equation with the discrete Lipschitz truncation of the error $\boldsymbol{e}^n := \boldsymbol{u} - \boldsymbol{u}^n$ it is possible to prove that (see [61] for a similar argument)

$$\limsup_{n\to\infty} \int_\Omega \boldsymbol{\mathcal{S}}(\mathsf{D}(\boldsymbol{u}^n), \theta^n) : \mathsf{D}(\boldsymbol{u}^n) \leq \int_\Omega \mathsf{S} : \mathsf{D}(\boldsymbol{u}). \tag{3.21}$$

Furthermore, from the growth condition of $\boldsymbol{\mathcal{S}}$ and the dominated convergence theorem (note that, up to a subsequence, we have that $\theta^n \to \theta$ almost everywhere, c.f. (3.16)) we see, that for any $\boldsymbol{\tau} \in L^r_{\mathrm{sym}}(\Omega)^{d\times d}$,

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, \theta^n) \to \boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, \theta) \quad \text{strongly in } L^{r'}(\Omega)^{d\times d}, \tag{3.22}$$

as $n \to \infty$. Combining the monotonicity of $\boldsymbol{\mathcal{S}}$ with (3.21) and (3.22) yields for an arbitrary $\boldsymbol{\tau} \in L^r_{\mathrm{sym}}(\Omega)^{d\times d}$:

$$\begin{aligned}0 &\leq \limsup_{n\to\infty} \int_\Omega (\boldsymbol{\mathcal{S}}(\mathbf{D}(\boldsymbol{u}^n), \theta^n) - \boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, \theta^n)) : (\mathbf{D}(\boldsymbol{u}^n) - \boldsymbol{\tau}) \\ &\leq \int_\Omega (\mathbf{S} - \boldsymbol{\mathcal{S}}(\boldsymbol{\tau}, \theta)) : (\mathbf{D}(\boldsymbol{u}^n) - \boldsymbol{\tau}).\end{aligned}$$

Choosing $\boldsymbol{\tau} = \mathbf{D}(\boldsymbol{u}) \pm \varepsilon\boldsymbol{\sigma}$ with an arbitrary $\boldsymbol{\sigma} \in C^\infty_0(\Omega)^{d\times d}$ and letting $\varepsilon \to 0$ concludes the proof. □

Remark 3.6. The use of the discrete Lipschitz truncation is only necessary when the velocity $\boldsymbol{u}$ is not an admissible test function in the momentum equation, which occurs when $r < \frac{3d}{d+2}$. If $r \geq \frac{3d}{d+2}$ then one can substitute Assumption 3.4 with Assumption 2.6. It is also important to note that if the trilinear form $\mathcal{B}$ is used instead, the stronger assumption $r > \frac{2d}{d+1}$ is required (see (2.11)).

Remark 3.7. If the constitutive relation can be written in the form $\mathbf{D}(\boldsymbol{u}) = \boldsymbol{\mathcal{D}}(\mathbf{S}, \theta)$, where $\boldsymbol{\mathcal{D}}$ satisfies analogous conditions to the ones stated in Assumption 3.3, then the corresponding 4-field formulation will also satisfy an analogous convergence result. An example of a constitutive relation captured by these assumptions is the Ostwald–de Waele power-law model with $r > \frac{2d}{d+2}$:

$$\begin{aligned}\boldsymbol{\mathcal{S}}(\mathbf{D}, \theta) &:= K(\theta)|\mathbf{D}|^{r-2}\mathbf{D}, \\ \boldsymbol{\mathcal{D}}(\mathbf{S}, \theta) &:= \frac{1}{K(\theta)}\left|\frac{\mathbf{S}}{K(\theta)}\right|^{r'-2}\mathbf{S},\end{aligned}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a continuous function satisfying $c_1 \leq K(s) \leq c_2$ for any $s \in \mathbb{R}$, where c_1, c_2 are two positive constants.

As mentioned in Section 2.2, if the rheological parameters are not temperature-dependent, the convergence result can cover very general constitutive relations defined by maximal monotone r-graphs (which include, for instance, Herschel–Bulkley fluids). For this problem let us define Formulation C_0 in exactly the same way as Formulation A_0, but replacing (3.1d) with

$$\mathbf{H}(\cdot, \mathbf{S}, \mathbf{D}(\boldsymbol{u})) = \mathbf{0} \qquad \text{a.e. in } \Omega. \tag{3.23}$$

In order to introduce the finite element formulation, the only necessary ingredient is an approximation to the graph $\mathcal{A}$, for which a result analogous to Lemma 2.2 holds. This is the case e.g. for the generalised Yosida approximation described in [61]:

$$\boldsymbol{\mathcal{D}}^n(x, \mathbf{S}) := \{\mathbf{D} \in \mathbb{R}^{d\times d}_{\mathrm{sym}} \,:\, (\mathbf{D}, \mathbf{S}) \in \mathcal{A}^n(x)\}, \tag{3.24}$$

where the approximate graph $\mathcal{A}^n$ is defined as follows

$$\mathcal{A}^n(x) := \{\big(\mathbf{D}, \mathbf{S} + \tfrac{1}{n}|\mathbf{D}|^{r-2}\mathbf{D}\big) \in \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \,:\, (\mathbf{D}, \mathbf{S}) \in \mathcal{A}(x)\}, \tag{3.25}$$

where $x \in \Omega$. The relation (3.24) defines in fact a single-valued function that can be employed in the definition of the finite element formulation. Formulation $\tilde{\mathrm{C}}^n_0$ is then defined in the same way as Formulation $\tilde{\mathrm{B}}^n_0$, but with $\boldsymbol{\mathcal{D}}^n(\mathbf{S}^n, \theta^n)$ replaced by

(3.24). However, it is worth pointing out that in the numerical computations one can simply work with the implicit function directly by writing

$$\int_\Omega \mathbf{H}(\cdot, \mathbf{S}^n, \mathbf{D}(\boldsymbol{u}^n)) : \boldsymbol{\tau} = 0 \quad \forall\, \boldsymbol{\tau} \in \Sigma^n, \tag{3.26}$$

instead of (3.3a).

Corollary 3.8. *Let $r > \frac{2d}{d+2}$ and let $\mathbf{H}\colon \Omega \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ be a function satisfying Assumption 2.3. Suppose that $\{\Sigma^n, U^n, V^n, M^n\}_{n\in\mathbb{N}}$ is a family of finite element subspaces satisfying Assumptions 2.4, 2.5, 2.7, 2.8, and 3.4. Then, for any $n \in \mathbb{N}$, Formulation $\tilde{\mathrm{C}}^n_0$ admits a solution $(\mathbf{S}^n, \theta^n, \boldsymbol{u}^n, p^n) \in \Sigma^n \times (\hat{\theta}^n_b + U^n) \times V^n \times M^n$, and we have, up to subsequences, that*

$$\begin{aligned}
\boldsymbol{u}^n &\rightharpoonup \boldsymbol{u} && \text{weakly in } W^{1,r}(\Omega)^d,\\
p^n &\rightharpoonup p && \text{weakly in } L^{\tilde{r}}(\Omega),\\
\theta^n &\rightharpoonup \theta && \text{weakly in } H^1(\Omega),\\
\mathbf{S}^n &\rightharpoonup \mathbf{S} && \text{weakly in } L^{r'}_{\mathrm{sym}}(\Omega)^{d\times d},
\end{aligned}$$

where $(\mathbf{S}, \theta, \boldsymbol{u}, p) \in L^{r'}_{\mathrm{sym,tr}}(\Omega)^{d\times d} \times (\hat{\theta}_b + H^1_{\Gamma_D}(\Omega)) \times W^{1,r}_0(\Omega)^d \times L^{\tilde{r}}_0(\Omega)$ is a solution of Formulation C_0.

Remark 3.9. When restricted to the isothermal case, the convergence result from Corollary 3.8 improves the one presented in [16] in two respects: the graph is not required to be strictly monotone here, which allows models with a yield stress, for instance, and the result holds for the whole admissible range $r > \frac{2d}{d+2}$ even without the use of pointwise divergence-free elements, thanks to the modified convective term $\tilde{\mathcal{B}}_n$. In addition, the argument used here in the identification of the constitutive relation avoids the use of Young measures, simplifying the proof.

4. Augmented Lagrangian Preconditioner

Henceforth we employ the Scott–Vogelius pair for the velocity and pressure, and discontinuous and continuous elements for the stress and temperature, respectively, with $k \geq d$:

$$\begin{aligned}
\Sigma^h &= \{\boldsymbol{\sigma} \in L^\infty_{\mathrm{sym,tr}}(\Omega)^{d\times d} \,:\, \boldsymbol{\sigma}|_K \in \mathbb{P}_{k-1}(K)^{d\times d} \text{ for all } K \in \mathcal{T}_n\},\\
U^h &= \{\eta \in W^{1,\infty}_{\Gamma_D}(\Omega) \,:\, \eta|_K \in \mathbb{P}_{k-1}(K) \text{ for all } K \in \mathcal{T}_n\},\\
V^h &= \{\boldsymbol{w} \in W^{1,\infty}_0(\Omega)^d \,:\, \boldsymbol{w}|_K \in \mathbb{P}_k(K)^d \text{ for all } K \in \mathcal{T}_n\},\\
M^h &= \{q \in L^\infty_0(\Omega) \,:\, q|_K \in \mathbb{P}_{k-1}(K) \text{ for all } K \in \mathcal{T}_n\}.
\end{aligned} \tag{4.1}$$

In order to ensure the inf-sup stability of the velocity-pressure pair, each level $\mathcal{T}_n$ of the mesh hierarchy is barycentrically refined, with the hierarchy itself constructed by uniform refinement, to prevent the appearance of degenerate elements (Figure 1). A drawback of this approach is that the resulting mesh hierarchy is non-nested, which introduces some difficulties when dealing with the transfer operators in the multigrid algorithm.

As mentioned in Section 2.3, this choice of finite element space for the stress satisfies the inf-sup condition (2.7). In fact, since the Scott–Vogelius element has the property that on barycentrically refined meshes discretely divergence-free velocities are exactly divergence-free, one can work with traceless stresses and hence

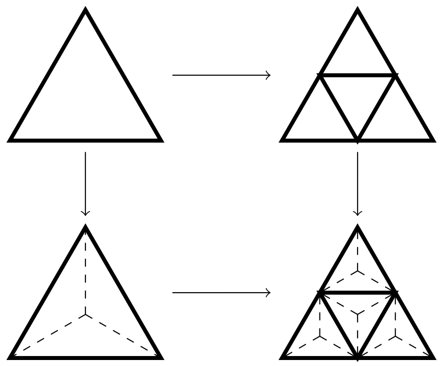


FIGURE 1. Non-nested two-level barycentrically refined mesh hierarchy.

fewer degrees of freedom will be required (c.f. [21]). This exact enforcement of the divergence constraint was one of the motivations behind our choice of elements; it is known that a failure to enforce the divergence-free constraint appropriately can lead to unphysical behaviour in the solution of buoyancy-driven flow [33].

At this point the viscous dissipation and the adiabatic heating terms can be incorporated into the formulation. For instance, when working with the setting described by Corollary 3.5, in the finite element formulation we seek $(\theta^n, \boldsymbol{u}^n, p^n) \in (\hat{\theta}_b + U^n) \times V^n \times M_0^n$ such that

$$\begin{aligned} \int_\Omega \boldsymbol{\mathcal{S}}(\mathsf{D}(\boldsymbol{u}^n), \theta^n) : \mathsf{D}(\boldsymbol{v}) - \int_\Omega (\boldsymbol{u}^n \otimes \boldsymbol{u}^n) : \mathsf{D}(\boldsymbol{v}) - \int_\Omega p^n \operatorname{div} \boldsymbol{v} &= \int_\Omega \theta^n \boldsymbol{v} \cdot \boldsymbol{e}_d \;\; \forall\, \boldsymbol{v} \in V^n, \\ -\int_\Omega q \operatorname{div} \boldsymbol{u}^n &= 0 \qquad \forall\, q \in M^n, \\ \int_\Omega (\hat{\kappa}(\theta^n)\nabla\theta^n - \boldsymbol{u}^n\theta^n) \cdot \nabla\eta + \int_\Omega \theta^n \boldsymbol{u}^n \cdot \boldsymbol{e}_d \eta &= \int_\Omega \boldsymbol{\mathcal{S}}(\mathsf{D}(\boldsymbol{u}^n), \theta^n) : \mathsf{D}(\boldsymbol{u}^n)\eta \;\forall\, \eta \in U^n, \end{aligned} \tag{4.2a}$$

with analogous modifications for the other formulations. Note that the form of the convective term could be simplified since the elements are exactly divergence-free. The nonlinear finite element formulations are linearised using Newton's method; for instance, if the current guess for the solution of (4.2) is $(\tilde{\theta}, \tilde{\boldsymbol{u}}, \tilde{p})$, then the method is defined by the correction step $(\tilde{\theta}, \tilde{\boldsymbol{u}}, \tilde{p}) \mapsto (\tilde{\theta}, \tilde{\boldsymbol{u}}, \tilde{p}) + (\theta, \boldsymbol{u}, p)$ where $(\theta, \boldsymbol{u}, p)$ is the solution of a linear system whose matrix has the block structure

$$\begin{bmatrix} A_1 & C & 0 \\ E & A_2 & \tilde{B}^\top \\ 0 & \tilde{B} & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \boldsymbol{u} \\ p \end{bmatrix}. \tag{4.3}$$

The blocks in (4.3) are defined through the linear operators:

$$\begin{aligned}
\langle A_1\theta,\eta\rangle &:= \int_\Omega \theta\hat\kappa'(\tilde\theta)\nabla\tilde\theta\cdot\nabla\eta + \int_\Omega \hat\kappa(\tilde\theta)\nabla\theta\cdot\nabla\eta - \int_\Omega \tilde{\boldsymbol{u}}\theta\cdot\nabla\eta \\
&\quad + \int_\Omega \tilde{\boldsymbol{u}}\theta\cdot\boldsymbol{e}_d\eta - \int_\Omega \boldsymbol{\mathcal{S}}_\theta(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta) : \mathsf{D}(\tilde{\boldsymbol{u}})\theta\eta && \forall\,\theta,\eta\in U^n, \\
\langle C\boldsymbol{u},\eta\rangle &:= \int_\Omega \tilde\theta\boldsymbol{u}\cdot(\boldsymbol{e}_d\eta-\nabla\eta) - \int_\Omega \boldsymbol{\mathcal{S}}_{\mathsf{D}}(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta)\mathsf{D}(\boldsymbol{u}):\mathsf{D}(\boldsymbol{v})\eta \\
&\quad - \int_\Omega \boldsymbol{\mathcal{S}}(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta):\mathsf{D}(\boldsymbol{u})\eta && \forall\,\boldsymbol{u}\in V^n,\eta\in U^n, \\
\langle E\theta,\boldsymbol{v}\rangle &:= \int_\Omega \boldsymbol{\mathcal{S}}_\theta(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta)\theta:\mathsf{D}(\boldsymbol{v}) - \int_\Omega \theta\boldsymbol{v}\cdot\boldsymbol{e}_d && \forall\,\theta\in U^n,\boldsymbol{v}\in V^n, \\
\langle A_2\boldsymbol{u},\boldsymbol{v}\rangle &:= \int_\Omega \Big(\boldsymbol{\mathcal{S}}_{\mathsf{D}}(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta)\ \mathsf{D}(\boldsymbol{u}) - \tilde{\boldsymbol{u}}\otimes\boldsymbol{u} - \boldsymbol{u}\otimes\tilde{\boldsymbol{u}}\Big):\mathsf{D}(\boldsymbol{v}) && \forall\,\boldsymbol{u},\boldsymbol{v}\in V^n, \\
\langle \tilde B\boldsymbol{v},q\rangle &:= -\int_\Omega q\,\mathrm{div}\,\boldsymbol{v} && \forall\,\boldsymbol{v}\in V^n,q\in M^n.
\end{aligned}$$

We use the notation $\boldsymbol{\mathcal{S}}_{\mathsf{D}}, \boldsymbol{\mathcal{S}}_\theta$ to denote the partial derivatives of $\boldsymbol{\mathcal{S}}$; for instance, for the Navier–Stokes model one would have $\boldsymbol{\mathcal{S}}_{\mathsf{D}}(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta) = 2\hat\mu(\tilde\theta)\boldsymbol{I}$ and $\boldsymbol{\mathcal{S}}_\theta(\mathsf{D}(\tilde{\boldsymbol{u}}),\tilde\theta) = 2\hat\mu'(\tilde\theta)\mathsf{D}(\tilde{\boldsymbol{u}})$, where $\boldsymbol{I}$ is the fourth-order identity tensor.

4.1. **Robust relaxation and prolongation.** Keeping (4.3) as an illustrative example, we see that after augmentation the top block can be written in the form

$$A + \gamma B^\top M_p^{-1} B = \begin{bmatrix} A_1 & C \\ E & A_2 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ \tilde B^\top \end{bmatrix} M_p^{-1} \begin{bmatrix} 0 & \tilde B \end{bmatrix}, \tag{4.5}$$

where A is invertible and $\gamma B^\top M_p^{-1}B$ is symmetric and semi-definite. Let us define $Z^n := U^n \times V^n$ whenever the 3-field formulation is employed and $Z^n := \Sigma^n \times U^n \times V^n$ otherwise. Relaxation methods in multigrid algorithms are often studied as subspace correction methods [65, 66]. Consider the space decomposition

$$Z^n = \sum_i Z_i^n, \tag{4.6}$$

where the sum is not necessarily direct. The key insight from [57, 40] is that, assuming A is symmetric and coercive, the subspace correction method induced by the decomposition (4.6) will be robust in γ if the decomposition stably captures the kernel $\mathcal{N}^n$ of the semi-definite term:

$$\mathcal{N}^n = \sum_i Z_i^n \cap \mathcal{N}^n. \tag{4.7}$$

Here $\mathcal{N}^n$ consists of the elements of the form $(\theta,\boldsymbol{v})^\top$ and $(\boldsymbol{\sigma},\theta,\boldsymbol{v})^\top$ for the 3-field and 4-field formulations, respectively, where $\boldsymbol{v}\in V^n_{\mathrm{div}}$, and $\boldsymbol{\sigma}\in\Sigma^n$, $\theta\in U^n$ are arbitrary. This means that the decomposition must allow for sufficiently rich subspaces such that divergence-free velocities can be written as combinations of divergence-free elements of the subspaces. A local characterisation of the kernel of the divergence for Scott–Vogelius elements on meshes with the macro element structure shown in Figure 1 was presented in [25] and used to construct a preconditioner for a system of nearly incompressible elasticity; this construction was then employed in [24] and [21] to precondition the isothermal Navier–Stokes system and

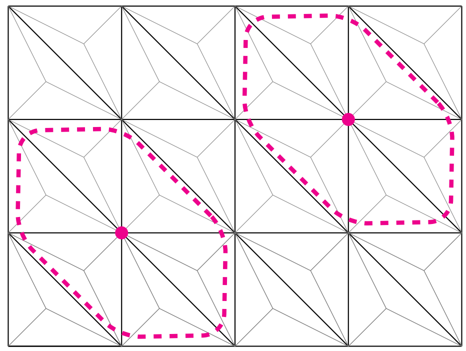


FIGURE 2. Macrostar patches on a barycentrically refined mesh.

a 3-field non-Newtonian formulation, respectively. In [25] it was shown that the kernel is captured by using a decomposition based on the subspaces

$$Z_i^n := \{\boldsymbol{z} \in Z^n : \operatorname{supp}(\boldsymbol{z}) \subset \operatorname{macrostar}(q_i)\}, \tag{4.8}$$

where for a vertex q_i, the macrostar patch $\operatorname{macrostar}(q_i)$ is defined as the union of all macro cells touching the vertex (Figure 2). In the algorithm presented here the relaxation solves based on the decomposition (4.8) are performed additively.

The work of Schöberl [57] also revealed the necessity of controlling the continuity constant of the prolongation operator in order to obtain a robust solver. In our setting, this entails ensuring that the prolongation operator $P_N : V^N \to V^n$ mapping coarse grid functions in V^N into fine grid functions in V^n has the property that divergence-free velocities get mapped to (nearly) divergence-free velocities; note that when using a standard prolongation based on interpolation, the condition $\operatorname{div} \boldsymbol{v}^N = 0$ does not necessarily imply that $\operatorname{div}(P_N \boldsymbol{v}^n) = 0$. For the setting described here a modified prolongation operator can be defined by computing a correction using local Stokes solves on the macro cells (see [25] for details).

For the formulations including the stress there is an additional difficulty: it is not obvious how to transfer piecewise discontinuous fields between non-nested meshes. Here we employ the supermesh projection described in [21]. For the temperature variables we employ a standard interpolation-based prolongation operator.

While the macrostar iteration mentioned above results in a robust relaxation scheme for the linear elasticity problem considered in [25], on its own it ceases to be effective when applied to the substantially more complex problem (4.5). However, we find that a handful of GMRES iterations preconditioned by the macrostar iteration are very effective for the problem under consideration.

5. Numerical experiments

Let us suppose that the parameters in the constitutive relation (1.3) can be written as

$$\frac{\hat{\mu}(\theta)}{\mu_0} = \mu(\theta), \quad \frac{\hat{\kappa}(\theta)}{\kappa_0} = \kappa(\theta), \quad \frac{\hat{\tau}(\theta)}{\tau_0} = \tau(\theta), \quad \frac{\hat{\sigma}(\theta)}{\sigma_0} = \sigma(\theta), \tag{5.1}$$

where $\mu_0, \kappa_0 > 0$ are reference values for the viscosity and heat conductivity, $\tau, \sigma \geq 0$ are reference values for the activation parameters, and $\mu, \kappa, \tau, \sigma$ are then non-dimensional functions. In practice the system can be non-dimensionalised in

distinct ways to give more importance to different physical regimes. For example, suppose that the time scale is chosen based on the diffusion of heat, and that the non-dimensional variables are introduced in the following way:
(5.2)
$$\tilde{t} := \frac{\alpha}{L^2}t, \quad \tilde{x} := \frac{x}{L}, \quad \tilde{\boldsymbol{u}} := \frac{L}{\alpha}\boldsymbol{u}, \quad \tilde{p} := \frac{L^2}{\rho_0\alpha^2}p, \quad \tilde{\theta} := \frac{\theta-\theta_C}{\theta_H-\theta_C}, \quad \tilde{\mathbf{S}} := \frac{L^2}{\mu_0\alpha}\mathbf{S},$$
where L is a characteristic length scale, θ_H is a reference temperature (e.g. the temperature of the hot plate in a Bénard problem), and $\alpha = \frac{\kappa_0}{\rho_0 c_p}$ is the thermal diffusion rate. Then, the non-dimensional form of the system reads (dropping the tildes):

$$-\mathrm{Pr}\,\mathrm{div}\,\mathbf{S} + \mathrm{div}(\boldsymbol{u}\otimes\boldsymbol{u}) + \nabla p = \mathrm{Ra}\,\mathrm{Pr}\,\theta\boldsymbol{e}_d \quad \text{in } \Omega, \tag{5.3a}$$
$$\mathrm{div}\,\boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{5.3b}$$
$$-\mathrm{div}(\kappa(\theta)\nabla\theta) + \mathrm{div}(\boldsymbol{u}\theta) + \mathrm{Di}(\theta+\Theta)\boldsymbol{u}\cdot\boldsymbol{e}_d = \frac{\mathrm{Di}}{\mathrm{Ra}}\mathbf{S}:\mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega, \tag{5.3c}$$
where the Rayleigh, Prandtl, Dissipation and Theta numbers are defined respectively as
$$\mathrm{Ra} = \frac{\beta g(\theta_H-\theta_C)L^3}{\nu\alpha}, \quad \mathrm{Pr} = \frac{\nu_0}{\alpha}, \quad \mathrm{Di} = \frac{\beta g L}{c_p}, \quad \Theta = \frac{\theta_C}{\theta_H-\theta_C}, \tag{5.4}$$
where $\nu_0 := \frac{\mu_0}{\rho_0}$ is the reference kinematic viscosity (more non-dimensional numbers could arise with a non-Newtonian constitutive relation). Alternatively, if one assumes that the gravitational potential energy is completely transformed into kinetic energy [31, 49], the characteristic velocity is chosen as $U = (gL\beta(\theta_H-\theta_C))^{1/2}$ and the resulting non-dimensional system becomes
$$-\frac{1}{\sqrt{\mathrm{Gr}}}\,\mathrm{div}\,\mathbf{S} + \mathrm{div}(\boldsymbol{u}\otimes\boldsymbol{u}) + \nabla p = \theta\boldsymbol{e}_d \quad \text{in } \Omega, \tag{5.5a}$$
$$\mathrm{div}\,\boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{5.5b}$$
(5.5c)
$$-\frac{1}{\mathrm{Pr}\sqrt{\mathrm{Gr}}}\,\mathrm{div}(\kappa(\theta)\nabla\theta) + \mathrm{div}(\boldsymbol{u}\theta) + \mathrm{Di}(\theta+\Theta)\boldsymbol{u}\cdot\boldsymbol{e}_d = \frac{\mathrm{Di}}{\sqrt{\mathrm{Gr}}}\mathbf{S}:\mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega,$$
where the Grashof number is defined as
$$\mathrm{Gr} = \frac{gL^3\beta(\theta_H-\theta_C)}{\nu_0^2}. \tag{5.6}$$
In the following section we will test the solver using the different forms (5.3), (5.5) with a heated cavity problem. The computational examples were implemented in Firedrake [54], and PCPATCH [23] (a recently developed tool for subspace decomposition in multigrid in PETSc [2]) was employed for the macrostar patch solves in the multigrid algorithm. The augmented Lagrangian parameter was set to $\gamma = 10^4$, and unless specified otherwise, the Newton solver was deemed to have converged when the Euclidean norm of the residual fell below 1×10^{-8} and the corresponding tolerance for the linear solver in 2D was set to 1×10^{-10} (1×10^{-8} in 3D). In the implementation the uniqueness of the pressure was enforced by orthogonalizing against the nullspace of constants in the Krylov solver, instead of enforcing a zero mean condition.

5.1. **Heated cavity.** The problem is solved on the unit square/cube $\Omega = (0,1)^d$ with boundary data

$$\boldsymbol{u} = \mathbf{0} \quad \text{on } \partial\Omega, \quad \nabla\theta \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \setminus (\Gamma_H \cup \Gamma_C), \quad \theta = \begin{cases} 1, & \text{on } \Gamma_H, \\ 0, & \text{on } \Gamma_C, \end{cases}$$

where $\Gamma_H := \{x_1 = 0\}$ and $\Gamma_C := \{x_1 = 1\}$. For the problems with temperature-dependent viscosity and conductivity we choose the following functional dependences:

$$\mu(\theta) := \mathrm{e}^{-\frac{\theta}{10}}, \tag{5.7a}$$

$$\kappa(\theta) := \frac{1}{2} + \frac{\theta}{2} + \theta^2. \tag{5.7b}$$

The viscosity defined by (5.7a) decreases with temperature, as is the case with most liquids [27]; heat conductivities of the form (5.7b) are a good fit for most liquid metals and gases [20]. Let us denote the problem solved with $\mu(\theta) \equiv 1 \equiv \kappa(\theta)$ by (P1), the one using (5.7a) and $\kappa(\theta) \equiv 1$ by (P2), and by (P3) the one using both forms in (5.7).

A simple continuation algorithm was used to reach the different values of the parameters; for instance, the solution corresponding to a Rayleigh number Ra was used as an initial guess in Newton's method for the problem with $\mathrm{Ra}+\mathrm{Ra}_{\mathrm{step}}$, where $\mathrm{Ra}_{\mathrm{step}}$ is some predetermined step. In some cases (most notably shear-thinning fluids) the use of advective stabilisation was essential; here we have added to the formulation an advective stabilisation term based on penalising the jumps between facets [13, 17]:

$$S_h(\boldsymbol{v}, \boldsymbol{w}) := \sum_{K \in \mathcal{M}_h} \frac{1}{2} \int_{\partial K} \delta\, h_{\partial K}^2 \, [\![\nabla \boldsymbol{v}]\!] : [\![\nabla \boldsymbol{w}]\!], \tag{5.8}$$

where $[\![\boldsymbol{z}]\!]$ denotes the jump of $\boldsymbol{z}$ across ∂K, $h_{\partial K}$ is the diameter of each face in ∂K, and δ is an arbitrary stabilization parameter. In the numerical experiments the stabilization parameter was chosen to be cell-dependent and set to $5 \times 10^{-3} \|\tilde{u}\|_{L^\infty(K)}$; an analogous term was added to the temperature equation. The choice of stabilisation (5.8) was preferred over the more common SUPG stabilisation because the latter introduces additional couplings between the velocity and the pressure in the momentum equation, and between the velocity, stress and temperature in the energy equation, which can spoil the convergence of the nonlinear solver (this was already observed in the isothermal case in [24]). The disadvantage is that (5.8) introduces an additional kernel consisting of C^1 functions, that might not be captured by the relaxation. This means that unless $k \geq 3$ in 2D or $k \geq 5$ in 3D, a slight loss of robustness might be expected [24].

Tables 1–3 show the average number of Krylov iterations for the problem with non-dimensional form (5.5) and increasingly large Grashof number, comparing with different values of the Dissipation number; Tables 4–6 show the same for the three-dimensional problem. It can be observed that the iteration count remains under control, and the previously mentioned loss of robustness occurs when $k = 2$. Figure 3 shows the streamlines and temperature contours for the problem (P2); it can be observed that the presence of the viscous dissipation term has a stabilising effect on the flow. Table 7 shows the number of iterations for the problem using the

Di	k	# refs	# dofs	Gr 1×10^6	Gr 5×10^6	Gr 1×10^7	Gr 1.5×10^7
0	2	1	1.8×10^4	5	7.66	10	22
		2	7.2×10^4	4.25	7	8	8.5
	3	1	4.1×10^4	2	3.5	4	4.5
		2	1.6×10^5	1.66	2	2.5	3
0.6	2	1	1.8×10^4	4.75	8	13.3	18.7
		2	7.2×10^4	4	7	7.5	7
	3	1	4.1×10^4	2	3.5	4.5	5.5
		2	1.6×10^5	1.67	2	3.5	4
1.3	2	1	1.8×10^4	5.67	8	12.67	18.67
		2	7.2×10^4	4	6.5	6.5	7
	3	1	4.1×10^4	2	2.5	4	4
		2	1.6×10^5	1.67	2	2.5	2.5
2.0	2	1	1.8×10^4	5.67	9.33	12.67	18.67
		2	7.2×10^4	4	6.5	6.5	8
	3	1	4.1×10^4	2	2.5	3	3
		2	1.6×10^5	1.67	2	2	2

TABLE 1. Average number of Krylov iterations per Newton step as Gr increases for the 2D problem (P1) with Pr = 1, obtained using 2 multigrid cycles with 4 relaxation sweeps.

temperature-dependent power-law relation

$$\mathsf{S} = \boldsymbol{\mathcal{S}}(\mathsf{D}(\boldsymbol{u}), \theta) := \mathrm{e}^{-\frac{\theta}{10}} |\mathsf{D}(\boldsymbol{u})|^{r-2} \mathsf{D}(\boldsymbol{u}), \tag{5.9}$$

using $r = 1.6$ and the streamlines are shown in Figure 5 alongside the ones of the Newtonian problem ($r = 2$). In this case the tolerances for the linear and nonlinear iterations were set to 1×10^{-10}.

5.2. **Bingham flow in a cooling channel.** Let $\Omega := (0, 40)\times(-1, 1)$, and consider the following boundary conditions for the temperature:

$$\nabla\theta \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \setminus (\Gamma_H \cup \Gamma_C), \quad \theta = \begin{cases} \theta_H, & \text{on } \Gamma_H, \\ 0, & \text{on } \Gamma_C, \end{cases}$$

where $\theta_H > 0$, and $\Gamma_H := \{(x_1, x_2)^\top \in \partial\Omega : x_1 \le 10\}$ and $\Gamma_C := \{(x_1, x_2)^\top \in \partial\Omega : x_2 \in \{-1, 1\}, 10 < x_1\}$. The Bingham constitutive relation for viscoplastic fluids is obtained by setting $\hat{\sigma} \equiv 0$ in (1.3). This relation is not described by a Fréchet differentiable function and therefore Newton's method cannot be directly applied; for this reason we will introduce a regularised version of the constitutive relation. In this example we will consider a forced convection regime, in which the buoyancy effects are not taken into account. The non-dimensional form of the system then

Di	k	# refs	# dofs	Gr 1×10^6	Gr 5×10^6	Gr 1×10^7	Gr 1.25×10^7
0	2	1	1.8×10^4	5.25	8.33	18	23.25
		2	7.2×10^4	4.25	7.5	9	9.5
	3	1	4.1×10^4	2	3.5	4.5	5
		2	1.6×10^5	1.67	2	2.5	2.5
0.6	2	1	1.8×10^4	4.75	8.67	15	15.5
		2	7.2×10^4	4.33	7	7.5	7.5
	3	1	4.1×10^4	2.33	3.5	5.5	5.5
		2	1.6×10^5	1.67	2	3.5	4.5
1.3	2	1	1.8×10^4	4.75	9.33	15.67	20.67
		2	7.2×10^4	4	7	6.5	6.5
	3	1	4.1×10^4	2	3.5	4	4.5
		2	1.6×10^5	1.67	2	3	3.5
2.0	2	1	1.8×10^4	5.67	10.67	16.33	19
		2	7.2×10^4	4	7	6.5	7.5
	3	1	4.1×10^4	2	2.5	3	3
		2	1.6×10^5	1.67	2	2.5	2.5

TABLE 2. Average number of Krylov iterations per Newton step as Gr increases for the 2D problem (P2) with Pr = 1, obtained using 2 multigrid cycles with 4 relaxation sweeps.

reads:

$$-\operatorname{div}\mathbf{S} + \mathrm{Re}\operatorname{div}(\boldsymbol{u}\otimes\boldsymbol{u}) + \nabla p = 0 \quad \text{in } \Omega, \tag{5.10a}$$

$$\operatorname{div}\boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{5.10b}$$

$$-\frac{1}{\mathrm{Pe}}\operatorname{div}(\nabla\theta) + \operatorname{div}(\boldsymbol{u}\theta) = \frac{\mathrm{Br}}{\mathrm{Pe}}\mathbf{S}:\mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega, \tag{5.10c}$$

$$\sqrt{\varepsilon^2 + |\mathbf{D}(\boldsymbol{u})|^2}\mathbf{S} = (\mathrm{Bn}\,\tau(\theta) + 2\mu(\theta)|\mathbf{D}(\boldsymbol{u})|)\mathbf{D}(\boldsymbol{u}), \text{ in } \Omega, \tag{5.10d}$$

where $\varepsilon > 0$ is the regularisation parameter, and the Reynolds, Péclet, Bingham and Brinkman numbers are defined as

$$\mathrm{Re} = \frac{\rho_0 U R}{\nu_0}, \quad \mathrm{Pe} = \frac{\rho_0 c_p U R}{\kappa_0}, \quad \mathrm{Bn} = \frac{\tau_0 R}{\nu_0 U}, \quad \mathrm{Br} = \frac{\nu_0 U^2}{\kappa_0 \theta_H}, \tag{5.11}$$

where R is the radius of the pipe, U is the average velocity at the inlet and τ_0 is the value of the yield stress at the inlet. Two choices for the (non-dimensional) viscosity and yield stress are considered here:

$$\text{Problem (Q1): } \mu(\theta) := a_1\theta + a_2 \quad \tau(\theta) := 1.$$
$$\text{Problem (Q2): } \mu(\theta) := 1 \quad \tau(\theta) := b_1\theta + b_2.$$

The values of a_1 and a_2 are chosen so that the viscosity is unity at the inlet and increases by a factor of 20 at the outlet (which means that the effective Bingham number decreases by the same factor). The constants b_1 and b_2 are such that the Bingham number is 1.5 at the inlet, and 9 at the outlet when a temperature drop

Di	k	# refs	# dofs	Gr 1×10^6	5×10^6	1×10^7	1.25×10^7
0	2	1	1.8×10^4	5.75	9	17.75	23
		2	7.2×10^4	4.25	6.33	10	11
	3	1	4.1×10^4	2	3	5	6
		2	1.6×10^5	1.67	2	1.5	1.5
0.6	2	1	1.8×10^4	5.5	9	17.33	24.4
		2	7.2×10^4	4.67	8	9	9.5
	3	1	4.1×10^4	2.33	3.5	5	6
		2	1.6×10^5	1.67	2.5	3.5	4
1.3	2	1	1.8×10^4	4.75	9.67	18	23.67
		2	7.2×10^4	4	8	9.5	9
	3	1	4.1×10^4	2.33	2.5	4	4
		2	1.6×10^5	1.67	2	2.5	2.5
2.0	2	1	1.8×10^4	5.66	10.33	18.33	24.33
		2	7.2×10^4	4.33	10	10	8.5
	3	1	4.1×10^4	2.33	2.5	2.5	3
		2	1.6×10^5	1.67	2	1.5	1.5

TABLE 3. Average number of Krylov iterations per Newton step as Gr increases for the 2D problem (P3) with $\mathrm{Pr} = 1$, obtained using 2 multigrid cycles with 4 relaxation sweeps.

Di	# refs	# dofs	Gr 2.52×10^5	6.30×10^5	9.45×10^5	1.26×10^6
0	1	3.2×10^5	3.33	4	4.5	9
	2	2.6×10^6	6	4.5	3.5	3.5
0.6	1	3.2×10^5	3.33	4	4	10.5
	2	2.6×10^6	4.33	5	4.5	4.5
1.3	1	3.2×10^5	3.33	4	4	10.5
	2	2.6×10^6	6	4.5	4.5	4
2	1	3.2×10^5	3	4	4.5	12
	2	2.6×10^6	6	4.5	3.5	3.5

TABLE 4. Average number of Krylov iterations per Newton step as Gr increases for the 3D problem (P1) with $\mathrm{Pr} = 1$ and $k = 3$, obtained using 2 multigrid cycles with 4 relaxation sweeps.

of 15 is applied. As for the velocity, we impose the following boundary conditions:

$$(\mathsf{S} - p\boldsymbol{I})(1,0)^\top \cdot (1,0)^\top = 0,\ \boldsymbol{u}\cdot(0,1)^\top = 0 \text{ on } \Gamma_{\text{out}}, \quad \boldsymbol{u} = \boldsymbol{u}_B \text{ on } \Gamma_{\text{in}},$$
$$\boldsymbol{u} = \boldsymbol{0} \text{ on } \partial\Omega \setminus (\Gamma_{\text{in}} \cup \Gamma_{\text{out}}),$$

Di	# refs	# dofs	Gr 2.52×10^5	6.30×10^5	9.45×10^5	1.26×10^6
0	1	3.2×10^5	3.67	4	5	13.5
	2	2.6×10^6	5	5.5	5.5	5.5
0.6	1	3.2×10^5	3.33	4	4.5	14
	2	2.6×10^6	4.33	5.5	4.5	5
1.3	1	3.2×10^5	3.33	4	5	16.5
	2	2.6×10^6	6	4.5	4.5	4.5
2	1	3.2×10^5	3.67	4	5	13.5
	2	2.6×10^6	6	4.5	3.5	4

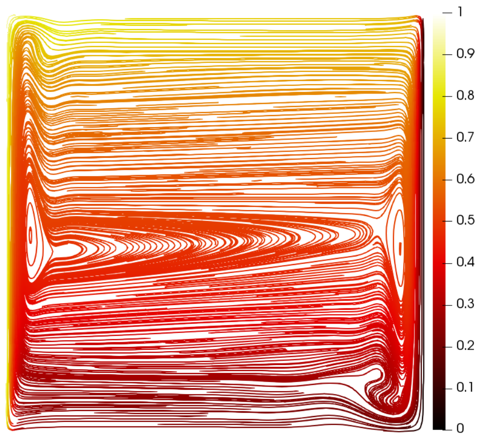
TABLE 5. Average number of Krylov iterations per Newton step as Gr increases for the 3D problem (P2) with $\mathrm{Pr} = 1$ and $k = 3$, obtained using 2 multigrid cycles with 4 relaxation sweeps.

Di	# refs	# dofs	Gr 2.52×10^5	6.30×10^5	9.45×10^5	1.26×10^6
0	1	3.2×10^5	3.67	5	7.5	19
	2	2.6×10^6	5	6	6	9.5
0.6	1	3.2×10^5	3.33	4	4	10.5
	2	2.6×10^6	4.33	5.5	4.5	7
1.3	1	3.2×10^5	3.33	4	10	28.5
	2	2.6×10^6	6	4.5	4.5	4
2	1	3.2×10^5	3	4	11.5	41.5
	2	2.6×10^6	6	4.5	3.5	3.5

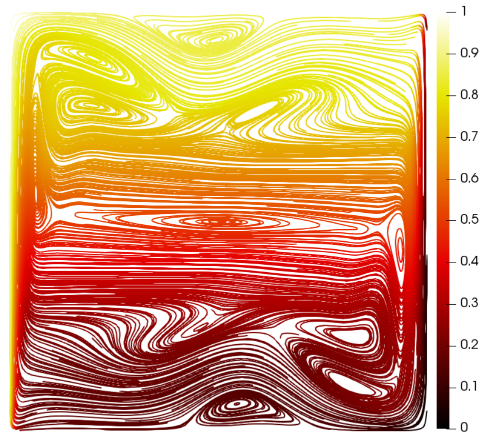
TABLE 6. Average number of Krylov iterations per Newton step as Gr increases for the 3D problem (P3) with $\mathrm{Pr} = 1$ and $k = 3$, obtained using 2 multigrid cycles with 4 relaxation sweeps.

k	# refs	# dofs	Ra 5000	10000	15000	20000
2	1	1.8×10^4	3.64	5.25	6.42	6.38
	2	7.2×10^4	3.78	5.78	7	9.75
	3	2.9×10^5	3.22	4.8	6.3	8.3
3	1	7.3×10^4	2.57	3.11	3.5	4.25
	2	1.6×10^5	2.5	2.8	3.33	4.75
	3	6.5×10^5	1.9	2.22	2.44	4

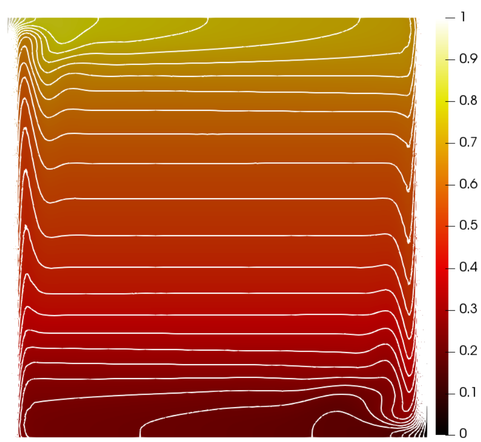
TABLE 7. Average number of Krylov iterations per Newton step as Ra increases for the constitutive relation - with $r = 1.6$ and $\mathrm{Di} = 0$, obtained using 7 multigrid cycles with 7 relaxation sweeps.



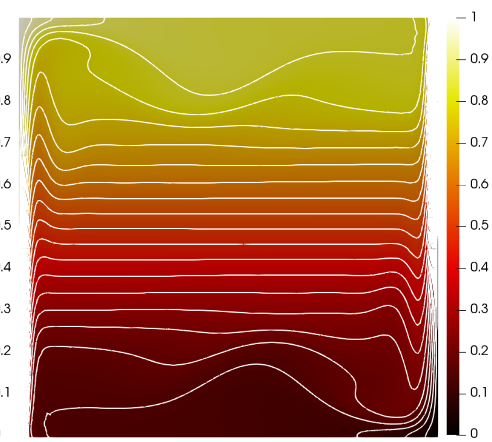
(A) Streamlines for $\mathrm{Di} = 2$.



(B) Streamlines for $\mathrm{Di} = 0$.



(C) Temperature contour for $\mathrm{Di} = 2$.



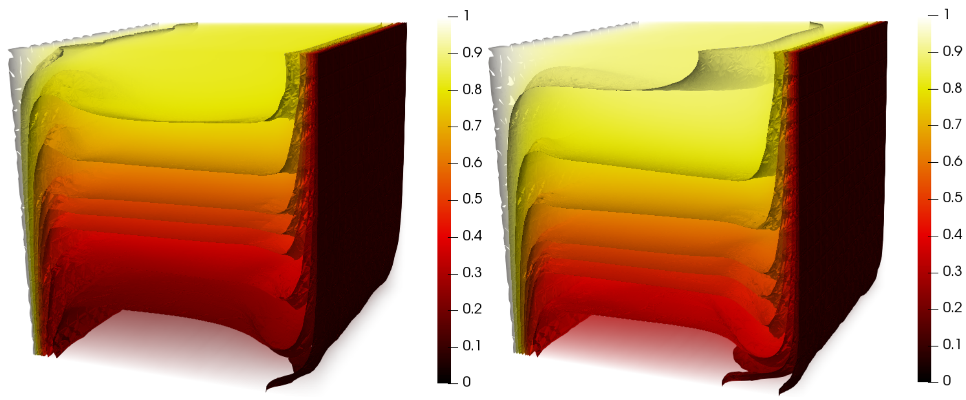
(D) Temperature contour for $\mathrm{Di} = 0$.

FIGURE 3. Streamlines and temperature contours for the heated cavity with temperature dependent viscosity and $\mathrm{Gr} = 1.25 \times 10^7$.

where $\Gamma_{\text{in}} := \{x_1 = 0\}$, $\Gamma_{\text{out}} := \{x_1 = 40\}$, and $\boldsymbol{u}_B$ is the fully developed Poiseuille flow for the isothermal problem with $\mathrm{Bn} = 1.5$, for which the exact solution is available (see e.g. [30]). In order to obtain better initial guesses for Newton's method, secant continuation was employed: given two previously computed solutions $\boldsymbol{z}_1, \boldsymbol{z}_2$ corresponding to the parameters $\varepsilon_1, \varepsilon_2$, respectively, the initial guess for Newton's method at ε is chosen as

$$\frac{\varepsilon - \varepsilon_2}{\varepsilon_2 - \varepsilon_1}(\boldsymbol{z}_2 - \boldsymbol{z}_1) + \boldsymbol{z}_2.$$

For this (arguably more complex) problem, the multigrid algorithm for the top block ceased to be effective. Tables 8–9 show the average number of Krylov iterations per Newton step obtained when using a sparse direct solver for the top block. It can be observed that for large values of the augmented Lagrangian parameter γ it is still possible to have an excellent control of the Schur complement. This



(A) Problem (P1). (B) Problem (P3).

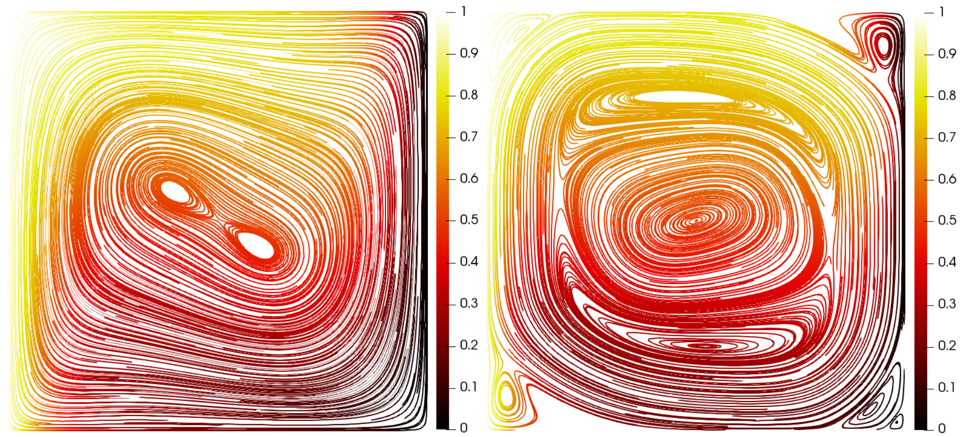
FIGURE 4. Temperature contours for the 3D heated cavity with $\mathrm{Gr} = 1.26 \times 10^6$.

γ	# refs	# dofs	ε: 1×10^{-3}	1×10^{-4}	2×10^{-5}	1×10^{-5}
10^3	1	2.7×10^4	12.8	22	51	48
	2	1.0×10^5	14.8	33.5	55	49
	3	4.3×10^5	13.5	17	25	17
	4	1.7×10^6	11.71	8.8	13	12
10^5	1	2.7×10^4	2.6	2	2	1.33
	2	1.0×10^5	2.6	2.25	1.4	1.15
	3	4.3×10^5	2	1.33	1.14	1
	4	1.7×10^6	1.75	1.33	1.15	1.07

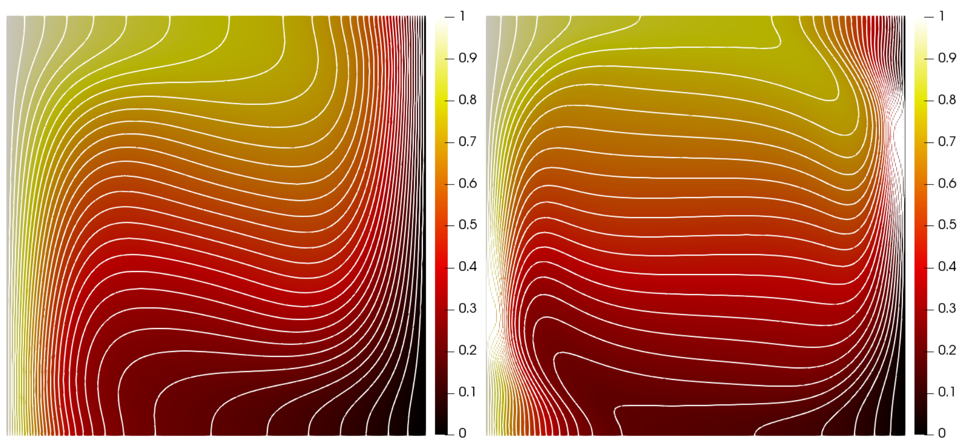
TABLE 8. Average number of Krylov iterations per Newton step as ε decreases for Problem (Q1) with $k = 2$, $\mathrm{Pe} = 10$, $\theta_H = 10$, $\mathrm{Br} = 0.1$.

suggests that it might be worthwile to follow the same strategy of using a block preconditioner that singles out the pressure, while attempting a different strategy for constructing a scalable solver for the top block.

Figures 6–7 show the temperature field and the yielded/unyielded regions of the fluid. The results are qualitatively similar to those found in [64], where an algorithm based on the augmented Lagrangian method was applied to a similar problem (neglecting the convective term and viscous dissipation). While it is known that a method based on regularisation, such as the one applied here, is not the most appropriate if one wishes to locate the exact position of the yield surfaces, it can still be useful to obtain the general features of the flow. For example, the solutions found here show no unyielded regions in the transition zone where the temperature field varies with the mean flow direction, which is the expected behaviour [64].



(A) Streamlines for $r = 2$. (B) Streamlines for $r = 1.6$.



(C) Temperature contour for $r = 2$. (D) Temperature contour for $r = 1.6$.

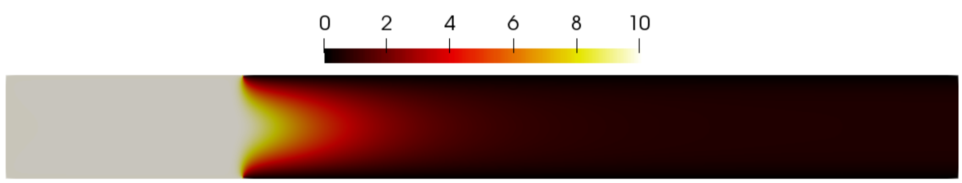
FIGURE 5. Streamlines and temperature contours for the heated cavity with the power-law constitutive relation (5.9) and $\mathrm{Ra} = 2 \times 10^4$.

REFERENCES

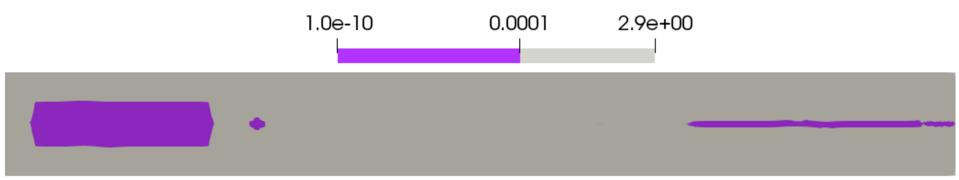
1. C. Bacuta. A unified approach for Uzawa algorithms. *SIAM J. Numer. Anal.*, 44(6):2633–2649, 2006.
2. S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, S. Smith, B. F. Zampini, H. Zhang, and H. Zhang. PETSc users manual. *Tech. Report ANL–95/11–Revision 3.8, Argonne National Laboratory*, 2017. http://www.mcs.anl.gov/petsc.
3. B. J. Bayly, C. D. Levermore, and T. Passot. Density variations in weakly compressible flows. *Phys. Fluids A*, 4(5):945–954, 1992.
4. L. Belenki, L. Berselli, L. Diening, and M. Růžička. On the finite element approximation of p-Stokes systems. *SIAM J. Numer. Anal.*, 50:373–397, 2012.
5. M. Benzi and M. A. Olshanskii. An augmented Lagrangian-based approach to the Oseen problem. *SIAM J. Sci. Comput.*, 28(6):2005–2113, 2006.

γ	# refs	# dofs	ε			
			1×10^{-3}	1×10^{-4}	5×10^{-5}	2×10^{-5}
10^3	1	2.7×10^4	12	17.5	26.6	*
	2	1.0×10^5	11.3	16.25	17	23
	3	4.3×10^5	12.88	13.67	14.5	13
	4	1.7×10^6	6.48	*	*	*
10^5	1	2.7×10^4	1.78	2.16	1.75	1.2
	2	1.0×10^5	1.6	1.3	1.16	1.07
	3	4.3×10^5	1.78	1.17	1.05	1
	4	1.7×10^6	1.31	1.13	1.03	1

TABLE 9. Average number of Krylov iterations per Newton step as ε decreases for Problem (Q2) with $k = 2$, $\mathrm{Pe} = 10$, $\theta_H = 10$, $\mathrm{Br} = 0$. The symbol * means that the maxmimum permitted number of nonlinear iterations was reached.



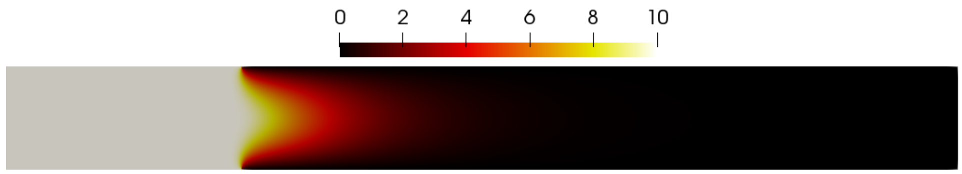
(A) Temperature



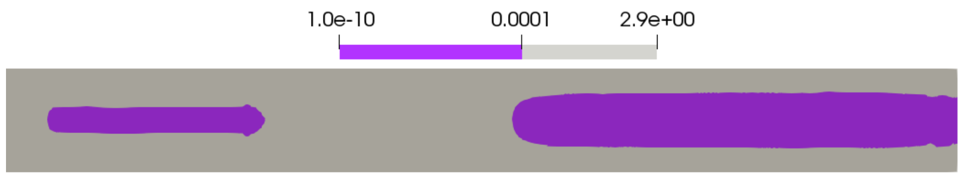
(B) Magnitude of the symmetric velocity gradient

FIGURE 6. Temperature field and yielded regions for the Bingham flow on a cooling channel (Problem (Q1)), with $\mathrm{Pe} = 10$, $\theta_H = 10$, $\mathrm{Br} = 0.1$.

6. J. Blechta, J. Málek, and K. R. Rajagopal. On the classification of incompressible fluids and a mathematical analysis of the equations that govern their motion. *SIAM J. Math. Anal.*, 52(2):1232–1289, 2020.
7. D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*. Springer, 2013.
8. J. Boussinesq. *Théorie Analytique de la Chaleur*. Gauthier–Villars, Paris, 1903.
9. M. Bulíček, E. Feireisl, and J. Málek. A Navier-Stokes-Fourier system for incompressible fluids with temperature dependent material coefficients. *Nonlinear Anal. Real World Appl.*, 10:992–1015, 2009.
10. M. Bulíček, P. Gwiazda, J. Málek, and A. Świerczewska-Gwiazda. On steady flows of incompressible fluids with implicit power-law-like rheology. *Adv. Calc. Var.*, 2:109–136, 2009.



(A) Temperature



(B) Magnitude of the symmetric velocity gradient.

FIGURE 7. Temperature field and yielded regions for Bingham flow on a cooling channel (Problem (Q2)), with $\mathrm{Pe} = 10$, $\theta_H = 10$, $\mathrm{Br} = 0$.

11. M. Bulíček, P. Gwiazda, J. Málek, and A. Świerczewska-Gwiazda. On unsteady flows of implicitly constituted incompressible fluids. *SIAM J. Math. Anal.*, 44(4):2756–2801, 2012.
12. M. Bulíček, J. Málek, and K. R. Rajagopal. Mathematical analysis of unsteady flows of fluids with pressure, shear-rate, and temperature dependent material moduli that slip at solid boundaries. *SIAM J. Math. Anal.*, 41(2):665–707, 2009.
13. E. Burman and A. Linke. Stabilized finite element schemes for incompressible flow using Scott–Vogelius elements. *Appl. Numer. Math.*, 58(11):1704–1719, 2008.
14. J. Casado-Díaz, T. Chacón-Rebollo, V. Girault, M. Gómez-Mármol, and F. Murat. Finite elements approximation of second order linear elliptic equations in divergence form with right-hand side in L^1. *Numer. Math.*, 105:337–374, 2007.
15. M. Crouzeix and P. A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *ESAIM: M2AN*, pages 33–75, 1973.
16. L. Diening, D. Kreuzer, and E. Süli. Finite element approximation of steady flows of incompressible fluids with implicit power-law-like rheology. *SIAM J. Numer. Anal.*, 51(2):984–1015, 2013.
17. J. Douglas and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. *Computing Methods in Applied Sciences. Lecture Notes in Physics*, vol. 58, 1976.
18. H. Elman, V. E. Howle, J. Shadid, R. Shuttleworth, and R. Tuminaro. Block preconditioners based on approximate commutators. *SIAM J. Sci. Comput.*, 27(5):1651–1668, 2006.
19. 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, second edition, 2014.
20. A. F. Emery and J. W. Lee. The effects of property variations on natural convection in a square enclosure. *J. Heat Transfer*, 121:57–62, 1999.
21. P. E. Farrell and P. A. Gazca-Orozco. Augmented Lagrangian preconditioner for implicitly-constituted non-Newtonian incompressible flow. *ArXiv Preprint: 2005.03150*, 2020.
22. P. E. Farrell, P. A. Gazca-Orozco, and E. Süli. Numerical analysis of unsteady implicitly constituted incompressible fluids: 3-field formulation. *SIAM J. Numer. Anal.*, 58(1):757–787, 2020.
23. P. E. Farrell, M. G. Knepley, L. E. Mitchell, and F. Wechsung. PCPATCH: software for the topological construction of multigrid relaxation methods. *ArXiv Preprint: 1912.08516*, 2019. In review.

24. P. E. Farrell, L. Mitchell, L. R. Scott, and F. Wechsung. A Reynolds-robust preconditioner for the Reynolds-robust Scott–Vogelius discretization of the stationary incompressible Navier–Stokes equations. *ArXiv Preprint: 2004.09398*, 2020.
25. P. E. Farrell, L. Mitchell, L. R. Scott, and F. Wechsung. Robust multigrid methods for nearly incompressible elasticity using macro elements. *ArXiv Preprint: 2002.02051*, 2020.
26. P. E. Farrell, L. Mitchell, and F. Wechsung. An augmented Lagrangian preconditioner for the 3D stationary incompressible Navier–Stokes equations at high Reynolds number. *SIAM J. Sci. Comput.*, 41(5):A3073–A3096, 2019.
27. S. Ferro and G. Gnavi. Effects of temperature-dependent viscosity in channels with porous walls. *Phys. Fluids*, 14(2):839–849, 2002.
28. V. Girault, R. H. Nochetto, and L. R. Scott. Max-norm estimates for Stokes and Navier–Stokes approximations in convex polyhedra. *Numer. Math.*, 131(4):771–882, 2015.
29. V. Girault and P. A. Raviart. *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*. Springer Verlag, 1986.
30. P. P. Grinevich and M. A. Olshanskii. An iterative method for the Stokes-type problem with variable viscosity. *SIAM J. Sci. Comput.*, 31(5):3959–3978, 2009.
31. F. M. Hewitt, D. P. McKenzie, and N. O. Weiss. Dissipative heating in convective flows. *J. Fluid Mech.*, 68(4):721–738, 1975.
32. V. E. Howle and R. C. Kirby. Block preconditioners for finite element discretization of incompressible flow with thermal convection. *Numer. Linear Algebra Appl.*, 19(2):427–440, 2012.
33. V. John, A. Linke, C. Merdon, M. Neilan, and L. G. Rebholz. On the Divergence Constraint in Mixed Finite Element Methods for Incompressible Flows. *SIAM Rev.*, 59(3):492–544, 2017.
34. Y. Kagei, M. Růžička, and G. Thäter. Natural convection with dissipative heating. *Commun. Math. Phys.*, 214:287–313, 2000.
35. D. Kay, D. Loghin, and A. Wathen. A preconditioner for the steady-state Navier–Stokes equations. *SIAM J. Sci. Comput.*, 24(1):237–256, 2002.
36. G. Ke, E. Aulisa, G. Bornia, and V. Howle. Block triangular preconditioners for linearization schemes of the Rayleigh–Bénard convection problem. *Numer. Linear Algebra Appl.*, 24(5):e2096, 2017.
37. G. Ke, E. Aulisa, G. Dillon, and V. Howle. Augmented Lagrangian-based preconditioners for steady buoyancy driven flow. *Appl. Math. Lett.*, 82:1–7, 2018.
38. C. Kreuzer and E. Süli. Adaptive finite element approximation of steady flows of incompressible fluids with implicit power-law-like rheology. *ESAIM: M2AN*, 50(5):1333–1369, 2016.
39. P. L. Lederer, A. Linke, and J. Merdon, C. Schöberl. Divergence-free reconstruction operators for pressure-robust Stokes discretizations with continuous pressure finite elements. *SIAM J. Numer. Anal.*, 55(3):1291–1314, 2017.
40. Y. J. Lee, J. Wu, J. Xu, and L. Zikatanov. Robust subspace correction methods for nearly singular systems. *Math. Models Methods Appl. Sci.*, 17(11):1937–196, 2007.
41. A. Linke. On the role of the Helmholtz decomposition in mixed methods for incompressible flows and a new variational crime. *Comput. Methods Appl. Mech. Engrg.*, 268:782–800, 2014.
42. A. Linke, G. Matthies, and L. Tobiska. Robust arbitrary order mixed finite element methods for the incompressible Stokes equations with pressure independent velocity errors. *ESAIM: M2AN*, 50(1):289–309, 2016.
43. A. Linke and C. Merdon. Pressure-robustness and discrete Helmholtz projectors in mixed finite element methods for the incompressible Navier–Stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 311:304–326, 2016.
44. A. Linke, C. Merdon, and W. Wollner. Optimal L^2 velocity error estimate for a modified pressure-robust Crouzeix–Raviart Stokes element. *IMA J. Numer. Anal.*, 37:354–374, 2017.
45. K. -A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebr.*, 18(1):1–40, 2011.
46. E. Maringová and J. Žabenský. On a Navier–Stokes–Fourier-like system capturing transitions between viscous and inviscid fluid regimes and between no-slip and perfect-slip boundary conditions. *Nonlinear Anal. Real World Appl.*, 41:152–178, 2018.
47. J. Nečas and T. Roubiček. Buoyancy-driven viscous flow with L^1-data. *Nonlinear Anal.*, 46:737–755, 2001.
48. A. Oberbeck. Über die Wärmeleitung der Flüssigkeiten bei der Berücksichtigung der Strömungen infolge von Temperaturdifferenzen. *Ann. Phys.*, 243(6):271–292, 1879.

49. S. Ostrach. *Internal viscous flows with body forces*, pages 185–208. Grenzschictsforschung. Springer Verlag, 1958.
50. J. Qin. *On the convergence of some low order mixed finite elements for incompressible fluids*. PhD thesis, Pennsylvania State University, 1994.
51. K. R. Rajagopal. On implicit constitutive theories. *Appl. Math.*, 48(4):279–319, 2003.
52. K. R. Rajagopal. On implicit constitutive theories for fluids. *J. Fluid Mech.*, 550:243–249, March 2006.
53. K. R. Rajagopal and A. R. Srinivasa. On the thermodynamics of fluids defined by implicit constitutive relations. *Z. Angew. Math. Phys*, 59:715–729, July 2008.
54. F. Rathgeber, D. A. Ham, L. Mitchell, M. Lange, F. Luporini, A. T. T. Mcrae, G-T. Bercea, G. R. Markall, and P. H. J. Kelly. Firedrake: automating the finite element method by composing abstractions. *ACM Trans. Math. Softw.*, 43(3), 2016.
55. T. Roubiček. Steady-state buoyancy-driven viscous flow with measure data. *Math. Bohem.*, 126(2):493–504, 2001.
56. J. Schöberl. Multigrid methods for a parameter dependent problem in primal variables. *Numer. Math.*, 84(1):97119, 1999.
57. J. Schöberl. *Robust Multigrid Methods for Parameter Dependent Problems*. PhD thesis, Johannes Kepler Universität Linz, 1999.
58. L. R. Scott and M. Vogelius. Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials. *Math. Modelling Numer. Anal.*, 19(1):111–143, 1985.
59. D. Silvester and A. Wathen. Fast iterative solution of stabilised Stokes systems. Part II: Using general block preconditioners. *SIAM J. Numer. Anal.*, 31(5):1352–1367, 1994.
60. E. Süli and T. Tscherpel. Fully discrete finite element approximation of unsteady flows of implicitly constituted incompressible fluids. *IMA J. Numer. Anal.*, dry097, 2019.
61. T. Tscherpel. *FEM for the Unsteady Flow of Implicitly Constituted Incompressible Fluids*. PhD thesis, University of Oxford, 2018.
62. D. L. Turcotte, A. T. Hsui, K. E. Torrance, and G. Schubert. Influence of viscous dissipation on Bénard convection. *J. Fluid Mech.*, 64(2):369–374, 1974.
63. M. G. Velarde and R. Pérez-Cordon. On the (non-linear) foundations of Boussinesq approximation applicable to a thin layer of fluid II: Viscous dissipation and large cell gap effects. *J. de Physique*, 37(3):178–182, 1976.
64. G. Vinay, A. Wachs, and J-F. Agassant. Numerical simulation of non-isothermal viscoplastic waxy crude oil flows. *J. Non-Newtonian Fluid Mech.*, 128:144–162, 2005.
65. J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, 1992.
66. J. Xu. The method of subspace corrections. *J. Comput. Appl. Math.*, 128(1):335–362, 2001.
67. S. Zhang. A new family of stable mixed finite elements for the 3D Stokes equations. *Math. Comput.*, 74(250):543–554, 2005.

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
Email address: patrick.farrell@maths.ox.ac.uk

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
Email address: gazcaorozco@maths.ox.ac.uk

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
Email address: endre.suli@maths.ox.ac.uk