A SEMISMOOTH NEWTON METHOD FOR IMPLICITLY CONSTITUTED NON-NEWTONIAN FLUIDS AND ITS APPLICATION TO THE NUMERICAL APPROXIMATION OF BINGHAM FLOW

PABLO ALEXEI GAZCA-OROZCO*

Abstract. We propose a semismooth Newton method for non-Newtonian models of incompressible flow where the constitutive relation between the shear stress and the symmetric velocity gradient is given implicitly; this class of constitutive relations captures for instance the models of Bingham and Herschel–Bulkley. The proposed method avoids the use of variational inequalities and is based on a particularly simple regularisation for which the (weak) convergence of the approximate stresses is known to hold. The system is analysed at the function space level and results in mesh-independent behaviour of the nonlinear iterations.

1. INTRODUCTION

Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz polygonal/polyhedral domain for $d \in \{2, 3\}$ and consider the following system for the velocity $\boldsymbol{u} \colon \Omega \to \mathbb{R}^d$, shear stress tensor $\mathsf{S} \colon \Omega \to \mathbb{R}^{d \times d}_{\text{sym}}$ and pressure $p \colon \Omega \to \mathbb{R}$ of an incompressible fluid:

$$
\begin{aligned}
\alpha \boldsymbol{u} - \operatorname{div} \mathsf{S} + \operatorname{div}(\boldsymbol{u} \otimes \boldsymbol{u}) + \nabla p &= \boldsymbol{f} && \text{in } \Omega, \\
\operatorname{div} \boldsymbol{u} &= 0 && \text{in } \Omega, \\
\boldsymbol{u} &= \boldsymbol{0} && \text{on } \partial\Omega,
\end{aligned}
\tag{1.1a}
$$

where $\boldsymbol{f} \colon \Omega \to \mathbb{R}^d$ is given, and $\alpha \geq 0$ is *e.g.* a parameter that arises from an implicit time discretisation ($\alpha = 0$ for the steady problem). In this work we will close the system above with an implicit constitutive relation of the form

$$
\mathsf{G}(\mathsf{S}, \mathsf{D}(\boldsymbol{u})) = \boldsymbol{0} \qquad \textit{a.e. in } \Omega,
\tag{1.1b}
$$

where $\mathsf{D}(\boldsymbol{u}) := \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\top)$ is the symmetric velocity gradient and $\mathsf{G} \colon \mathbb{R}^{d \times d}_{\text{sym}} \times \mathbb{R}^{d \times d}_{\text{sym}} \to \mathbb{R}^{d \times d}_{\text{sym}}$ defines a monotone graph on the space of symmetric matrices (the precise assumptions on the function G will be introduced later). The theory of implicitly constituted fluids, introduced in [37, 38], provides a framework that captures a wide range of models and allows for a thermodynamically consistent analysis (see *e.g.* [39]). For example, the usual

---

Department of Mathematics, FAU Erlangen-Nürnberg, 91058 Erlangen, Germany.
*Corresponding author: alexei.gazca@math.fau.de; gazcaorozco@gmail.com

Newtonian and power-law models can be written in the form (1.1b):

$$\mathbf{G}(\mathbf{S}, \mathbf{D}(\boldsymbol{u})) := \mathbf{S} - 2\nu\mathbf{D}(\boldsymbol{u}) \qquad\qquad \nu > 0, \qquad\qquad (1.2a)$$

$$\mathbf{G}(\mathbf{S}, \mathbf{D}(\boldsymbol{u})) := \mathbf{S} - K|\mathbf{D}(\boldsymbol{u})|^{r-2}\mathbf{D}(\boldsymbol{u}) \qquad\qquad K > 0,\, r > 1. \qquad\qquad (1.2b)$$

As a motivating example, in this work we will place special emphasis on the Bingham constitutive relation for viscoplastic fluids, defined by the dichotomy:

$$\begin{cases} |\mathbf{S}| \leq \tau_* \iff \mathbf{D}(\boldsymbol{u}) = \mathbf{0}, \\ |\mathbf{S}| > \tau_* \iff \mathbf{S} = 2\nu\mathbf{D}(\boldsymbol{u}) + \tau_* \dfrac{\mathbf{D}(\boldsymbol{u})}{|\mathbf{D}(\boldsymbol{u})|}, \end{cases} \qquad (1.3)$$

where $\tau_* \geq 0$ is the yield stress, and $\nu > 0$ is the viscosity; the closely related Herschel–Bulkley constitutive relation can be obtained *e.g.* by letting $\nu = \nu(\mathbf{D}(\boldsymbol{u})) := K|\mathbf{D}(\boldsymbol{u})|^{r-2}$, with $K > 0$ and $r > 1$. Such models can be used to describe waxy crude oil, toothpaste, paint, pastes, drilling muds, and mango jam, among other things [5, 7, 26, 29]; see [4] for a nice survey on various aspects of the modeling and simulation of viscoplastic fluids. Note that, in contrast to the constitutive relations (1.2), the relation (1.3) cannot be written in terms of a (single-valued) function $\mathbf{S} = \mathcal{S}(\mathbf{D}(\boldsymbol{u}))$. However, within the framework of implicitly constituted fluids it can be very naturally written, for instance, using either of the two following expressions:

$$\mathbf{G}(\mathbf{S}, \mathbf{D}(\boldsymbol{u})) := (|\mathbf{S}| - \tau_*)^+ \mathbf{S} - 2\nu\Big(\tau_* + (|\mathbf{S}| - \tau_*)^+\Big)\mathbf{D}(\boldsymbol{u}) = \mathbf{0}, \qquad (1.4a)$$

$$\mathbf{G}(\mathbf{S}, \mathbf{D}(\boldsymbol{u})) := |\mathbf{D}(\boldsymbol{u})|\mathbf{S} - (\tau_* + 2\nu|\mathbf{D}(\boldsymbol{u})|)\mathbf{D}(\boldsymbol{u}) = \mathbf{0}, \qquad (1.4b)$$

where $p^+ := \max\{p, 0\}$ is the positive part of p. For a rigorous mathematical analysis of models describing implicitly constituted incompressible fluids, the reader is referred to [10, 11]; in particular existence of weak solutions for large data is known to hold both for the steady and unsteady problems, under appropriate assumptions for $\mathbf{G}$ (see also [8] for a classification of various models that can be described within this framework). Regarding the numerical analysis, in [18, 44] the authors established (weak) convergence of the finite element approximations to a weak solution of the system for the steady and unsteady problems, respectively. This was later extended to formulations including the shear stress $\mathbf{S}$ in [21], and an augmented Lagrangian preconditioner was proposed in [20].

In previous works dealing with the numerical analysis of implicitly constituted fluids a fixed point argument is employed to obtain existence of discrete solutions, and then the emphasis is placed on proving that convergence to a weak solution of the problem holds. However, there is barely any discussion on how to solve such discrete nonlinear systems. For smooth constitutive relations, such as (1.2), this is not an issue since the standard Newton's method can be applied, but it is an important matter when dealing *e.g.* with the Bingham constitutive relation. In practice, it is common to circumvent the non-differentiability by regularising; a popular regularisation is for example the one proposed in [6]:

$$\mathbf{S}_\varepsilon = \mathcal{S}_\varepsilon(\mathbf{D}(\boldsymbol{u}_\varepsilon)) := 2\nu\mathbf{D}(\boldsymbol{u}_\varepsilon) + \tau_* \frac{\mathbf{D}(\boldsymbol{u}_\varepsilon)}{\sqrt{\varepsilon^2 + |\mathbf{D}(\boldsymbol{u}_\varepsilon)|^2}} \qquad \varepsilon > 0. \qquad (1.5)$$

Regularisations such as (1.5) are very popular owing to their simplicity. However, a major drawback of this approach is that by and large there are no results that guarantee the convergence of the shear stress $\mathbf{S}_\varepsilon \to \mathbf{S}$, as $\varepsilon \to 0$, and previous works suggest that this convergence might not actually hold in general (see *e.g.* [22, 33, 42]); one exception is the regularisation of the Bingham model introduced in [16], for which the authors prove convergence of the approximate solutions to a solution of the original problem. The main alternative, whenever an accurate approximation of the yield surfaces (*i.e.* the region of transition between the two regimes described by (1.3)) is essential, is given by the augmented Lagrangian method, which is based on a variational inequality formulation of the problem (see [26]). The main drawback of this method is that it requires the use of more

sophisticated mathematical tools to be analysed and convergence can be slow; the search for ways to accelerate the algorithm is still a very active area of research (see *e.g.* [19]).

Here we propose the use of an alternative regularisation, introduced in [13], for which the convergence $\mathsf{S}_\varepsilon \rightharpoonup \mathsf{S}$ can in fact be established. This regularisation takes the very simple form

$$\mathsf{G}_{\varepsilon_2}^{\varepsilon_1}(\mathsf{S}, \mathsf{D}(\boldsymbol{u})) := \mathsf{G}(\mathsf{S} - \varepsilon_1 \mathsf{D}(\boldsymbol{u}), \mathsf{D}(\boldsymbol{u}) - \varepsilon_2 \mathsf{S}) \qquad \varepsilon_1, \varepsilon_2 > 0. \tag{1.6}$$

The approach proposed here is conceptually very simple and avoids the use of formulations based on variational inequalities. Note that the regularised function $\mathsf{G}_{\varepsilon_2}^{\varepsilon_1}$ inherits the smoothness of the original constitutive relation, and so in general it will not be differentiable (it brings however other advantages that will be discussed later). For this reason a semismooth version of Newton's method will be required.

In the classical version of Newton's method for solving the equation $F(z) = 0$, where $F\colon Z \to X$ is a function defined between two Banach spaces Z and X, one applies iteratively the following steps until a convergence criterion is satisfied:

(1) Solve the linear problem: $\mathrm{DF}(z^k)d^k = -F(z^k)$;
(2) Update the solution: $z^{k+1} := z^k + d^k$.

With the appropriate hypotheses it is well known that the iterates produced by this method converge (locally) quadratically to the solution of the problem. In the semismooth version of the algorithm one has to replace $\mathrm{DF}(z^k)$ in step 1 above by an element of the generalised Jacobian of F. Then, assuming some regularity conditions, it is possible to prove that the iterates converge at least (locally) superlinearly to the solution (the required hypotheses will be described in detail in subsequent sections).

After introducing the relevant weak and finite element formulations in Section 2, we will proceed in Section 3 to write the problem (1.1a) with the regularised constitutive relation (1.6) in terms of a function F between appropriate Banach spaces and formulate a semismooth Newton algorithm that, with appropriate assumptions, can be shown to fall within the framework of semismooth Newton methods in function spaces of [49]; in particular, we will see that some expressions describing the Bingham constitutive relation are more advantageous. The analysis is carried out at the function space level in order to avoid mesh-dependent behaviour. Numerical experiments focusing on the Bingham constitutive relation that support this result are carried out in Section 4; these experiments employ the finite element method and a stress-velocity-pressure formulation.

Semismooth Newton methods have been applied before in the context of viscoplastic flow; semismooth Newton methods were analysed in [16, 17] for the solution of certain variational inequalities related to the description of Bingham flow, and so the approach is distinct to the one presented here (as mentioned above, one of the advantages the framework presented here is precisely that it avoids the formalism of variational inequalities). A non-smooth Newton method was developed for a different formulation describing Herschel–Bulkley fluids with $r \in (1, 2)$ in [41], but, as opposed to the present work, some mesh-dependence was observed in the solver. In addition, the analysis from [41] is closely tied to the particular structure of the Herschel–Bulkley relation, which is a constraint not present here: the framework employed here captures a wider variety of models; for instance, given the implicit nature of the constitutive relation, one could very easily consider stress-dependent viscosities or even swap the roles of the shear stress S and the symmetric velocity gradient $\mathsf{D}(\boldsymbol{u})$, and obtain models describing inviscid (*i.e.* Euler) fluids for low values of $|\mathsf{D}(\boldsymbol{u})|$, and viscous otherwise (see [8] for a more detailed description of related models).

2. Weak and finite element formulations

2.1. Implicitly constituted fluids

Throughout this work we will employ standard notation for Sobolev and Lebesgue spaces (*e.g.* $(W^{1,r}(\Omega), \|\cdot\|_{W^{1,r}(\Omega)})$). We define the space $W_0^{k,r}(\Omega)$, with $k \in \mathbb{N}$, $r > 1$, to be the closure of the space of

compactly supported smooth functions $C_0^\infty(\Omega)$ with respect to the norm $\|\cdot\|_{W^{k,r}(\Omega)}$. In addition, for $r \in (1, \infty)$ we define the following subspaces:

$$L_0^r(\Omega) := \left\{ q \in L^r(\Omega) : \int_\Omega q = 0 \right\},$$

$$W_{0,\mathrm{div}}^{1,r}(\Omega)^d := \overline{\{ \boldsymbol{v} \in C_0^\infty(\Omega)^d : \mathrm{div}\, \boldsymbol{v} = 0 \}}^{\|\cdot\|_{W^{1,r}(\Omega)}},$$

$$L_{\mathrm{tr}}^r(Q)^{d\times d} := \left\{ \boldsymbol{\tau} \in L^r(Q)^{d\times d} : \mathrm{tr}(\boldsymbol{\tau}) = 0 \right\},$$

$$L_{\mathrm{sym}}^r(Q)^{d\times d} := \left\{ \boldsymbol{\tau} \in L^r(Q)^{d\times d} : \boldsymbol{\tau}^\top = \boldsymbol{\tau} \right\},$$

$$L_{\mathrm{sym,tr}}^r(Q)^{d\times d} := L_{\mathrm{sym}}^r(Q)^{d\times d} \cap L_{\mathrm{tr}}^r(Q)^{d\times d}.$$

The operator $\mathrm{tr}(\boldsymbol{\tau})$ above denotes the usual trace of the matrix function $\boldsymbol{\tau}$. The Frobenius inner product between matrices will be written as $\boldsymbol{\tau} : \boldsymbol{\sigma} := \mathrm{tr}(\boldsymbol{\tau}\boldsymbol{\sigma}^\top)$.

Suppose that $\boldsymbol{f} \in L^{r'}(\Omega)^d$. In order to focus solely on the difficulties brought by the nonlinear constitutive relation, we will neglect the convective term in the subsequent analysis. Therefore, in the weak formulation of the problem we look for $(\mathsf{S}, \boldsymbol{u}, p) \in L_{\mathrm{sym,tr}}^{r'}(\Omega)^{d\times d} \times V^r \times L_0^{r'}(\Omega)$, where $V^r := W_{0,\mathrm{div}}^{1,r}(\Omega)^d \cap L^2(\Omega)^d$ and r' is the Hölder conjugate of r, such that

$$\alpha \int_\Omega \boldsymbol{u} \cdot \boldsymbol{v} + \int_\Omega \mathsf{S} : \mathsf{D}(\boldsymbol{v}) - \int_\Omega p\, \mathrm{div}\, \boldsymbol{v} = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v} \qquad \forall\, \boldsymbol{v} \in C_0^\infty(\Omega)^d, \qquad (2.1\mathrm{a})$$

$$\mathsf{G}(\mathsf{S}, \mathsf{D}(\boldsymbol{u})) = \boldsymbol{0} \qquad a.e. \text{ in } \Omega, \qquad (2.1\mathrm{b})$$

$$-\int_\Omega q\, \mathrm{div}\, \boldsymbol{u} = 0 \qquad \forall\, q \in C_0^\infty(\Omega). \qquad (2.1\mathrm{c})$$

In the existence analysis of systems such as (2.1), the assumptions on the constitutive relation are traditionally written in terms not of the function G, but of its induced graph on $\mathbb{R}_{\mathrm{sym}}^{d\times d} \times \mathbb{R}_{\mathrm{sym}}^{d\times d}$. In this work we will prefer to work directly with the function G, following the approach first proposed in [13], which leads to conditions that are easier to check in practice.

Assumption 2.1 (Properties of G). *We have for some $r > 1$:*

(G1) G *is locally Lipschitz and* $\mathsf{G}(\boldsymbol{0}, \boldsymbol{0}) = \boldsymbol{0}$;
(G2) *We have for any* $\boldsymbol{\tau}, \boldsymbol{\sigma} \in \mathbb{R}_{\mathrm{sym}}^{d\times d}$:

$$\partial_{\boldsymbol{\sigma}} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \geq \boldsymbol{0} \qquad\qquad \partial_{\boldsymbol{\tau}} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \leq \boldsymbol{0},$$

$$\partial_{\boldsymbol{\sigma}} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \partial_{\boldsymbol{\tau}} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) > \boldsymbol{0} \qquad\qquad \partial_{\boldsymbol{\tau}}(\boldsymbol{\sigma}, \boldsymbol{\tau})(\partial_{\boldsymbol{\sigma}} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}))^\top \leq \boldsymbol{0};$$

(G3) *Either*

$$\forall\, \boldsymbol{\tau} \in \mathbb{R}_{\mathrm{sym}}^{d\times d} \quad \liminf_{\boldsymbol{\sigma} \to \infty} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) : \boldsymbol{\sigma} > 0,$$

or

$$\forall\, \boldsymbol{\sigma} \in \mathbb{R}_{\mathrm{sym}}^{d\times d} \quad \liminf_{\boldsymbol{\tau} \to \infty} \mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) : \boldsymbol{\tau} < 0;$$

(G4) *For any* $(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathbb{R}_{\mathrm{sym}}^{d\times d} \times \mathbb{R}_{\mathrm{sym}}^{d\times d}$ *such that* $\mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \boldsymbol{0}$:

$$\boldsymbol{\sigma} : \boldsymbol{\tau} \geq c_1 \left(|\boldsymbol{\sigma}|^{r'} + |\boldsymbol{\tau}|^r \right) - c_2,$$

for some $c_1, c_2 > 0$.

The assumptions (G1)–(G4) imply that G induces a maximal monotone r-coercive graph on $\mathbb{R}^{d\times d}_{\mathrm{sym}}\times\mathbb{R}^{d\times d}_{\mathrm{sym}}$ (see [13], Def. 3.1) and so they guarantee the existence of a weak solution to the problem ([13], Thm. 2.2). It can be seen that the Bingham relation (1.4) and relations of power-law type like (1.2) satisfy the assumptions above (see *e.g.* [13, 48]).

The existence result obtained in [13] relies on an approximation scheme based on solving the problem associated to the approximate constitutive relation (1.6) and then passing to the limit $(\varepsilon_1,\varepsilon_2)\to 0$; in particular, one has that the approximate solutions $(\mathsf{S}^{\varepsilon_1,\varepsilon_2},\boldsymbol{u}^{\varepsilon_1,\varepsilon_2},p^{\varepsilon_1,\varepsilon_2})$ converge weakly to a weak solution $(\mathsf{S},\boldsymbol{u},p)$ of the system. From a theoretical point of view the regularisation (1.6) is advantageous because the resulting approximate graph is strictly monotone and 2-coercive, thus leading to an approximate problem with a Hilbert space structure. Figure 1 shows a plot of the regularised constitutive relation (1.6) for the Bingham and Herschel–Bulkley models for different values of $\varepsilon=\varepsilon_1=\varepsilon_2$.

Remark 2.2. In the theoretical analysis carried out in [13], the parameters ε_1 and ε_2 were taken to be equal:

$$\mathsf{G}_\varepsilon(\mathsf{S},\mathsf{D}(\boldsymbol{u})):=\mathsf{G}^\varepsilon_\varepsilon(\mathsf{S},\mathsf{D}(\boldsymbol{u}))=\mathsf{G}(\mathsf{S}-\varepsilon\mathsf{D}(\boldsymbol{u}),\mathsf{D}(\boldsymbol{u})-\varepsilon\mathsf{S})=0 \qquad \varepsilon>0. \tag{2.2}$$

This choice was made because having two distinct parameters did not bring any advantages to the analysis ([13], Rem. 4.2). Here we prefer to introduce the regularisation in the form (1.6) to emphasise the fact that the two parameters have different units: ε_1 has units of viscosity and ε_2 has units of fluidity; otherwise, starting directly from (2.2) can give the impression that one is commiting a "dimensional crime". Furthermore, when performing numerical computations it is not clear whether different choices for ε_1 and ε_2 could result in better behaviour. However, for the sake of simplicity we will employ the form (2.2) in the rest of this work.

Lemma 2.3 ([13]). *Suppose that G satisfies Assumptions (G1)–(G4). Then the graph induced by G_ε (recall (2.2)) is uniformly monotone and 2-coercive, i.e. for any $(\boldsymbol{\sigma}_i,\boldsymbol{\tau}_i)\in\mathbb{R}^{d\times d}_{\mathrm{sym}}\times\mathbb{R}^{d\times d}_{\mathrm{sym}}$ such that $\mathsf{G}_\varepsilon(\boldsymbol{\sigma}_i,\boldsymbol{\tau}_i)=\mathbf{0}$, with $i\in\{1,2\}$, we have that*

$$\boldsymbol{\sigma}_1:\boldsymbol{\tau}_1\geq c\big(|\boldsymbol{\sigma}_1|^2+|\boldsymbol{\tau}_1|^2\big)-\tilde{c},$$
$$(\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2):(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)\geq c_\varepsilon\big(|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2|^2+|\boldsymbol{\tau}_1-\boldsymbol{\tau}_2|^2\big), \tag{2.3}$$

where $c,\tilde{c}>0$ and $c_\varepsilon=\frac{\varepsilon}{1+\varepsilon^2}>0$.

2.2. Finite element approximation

We will now introduce the necessary notation to be able to define the finite element formulation of the system. Since the constitutive relation is implicit, a natural approach is to work with a 3-field formulation that includes the shear stress S (such a formulation was analysed in [21]). Let $\{\mathcal{T}_n\}_{n\in\mathbb{N}}$ be a family of shape-regular triangulations such that the mesh size $h_n:=\max_{K\in\mathcal{T}_n}h_K$ vanishes as $n\to\infty$. We then define the following conforming families of finite element spaces:

$$\Sigma^n:=\big\{\boldsymbol{\sigma}\in L^\infty_{\mathrm{sym}}(\Omega)^{d\times d}:\boldsymbol{\sigma}|_K\in\mathbb{P}_{\mathbb{S}}(K)^{d\times d},\,K\in\mathcal{T}_n\big\},$$
$$V^n:=\big\{\boldsymbol{v}\in W^{1,\infty}_0(\Omega)^d:\boldsymbol{v}|_K\in\mathbb{P}_{\mathbb{V}}(K)^d,\,K\in\mathcal{T}_n\big\},$$
$$M^n:=\big\{q\in L^\infty(\Omega):q|_K\in\mathbb{P}_{\mathbb{M}}(K),\,K\in\mathcal{T}_n\big\},$$

where $\mathbb{P}_{\mathbb{S}}(K)$, $\mathbb{P}_{\mathbb{V}}(K)$, and $\mathbb{P}_{\mathbb{M}}(K)$ are spaces of polynomials defined on the element $K\in\mathcal{T}_n$. We will also introduce the following useful subspaces:

$$M^n_0:=M^n\cap L^2_0(\Omega), \qquad \Sigma^n_{\mathrm{tr}}:=\Sigma^n\cap L^2_{\mathrm{sym,tr}}(\Omega)^{d\times d},$$
$$V^n_{\mathrm{div}}:=\bigg\{\boldsymbol{v}\in V^n:\int_\Omega q\,\mathrm{div}\,\boldsymbol{v}=0 \quad \forall q\in M^n\bigg\}.$$
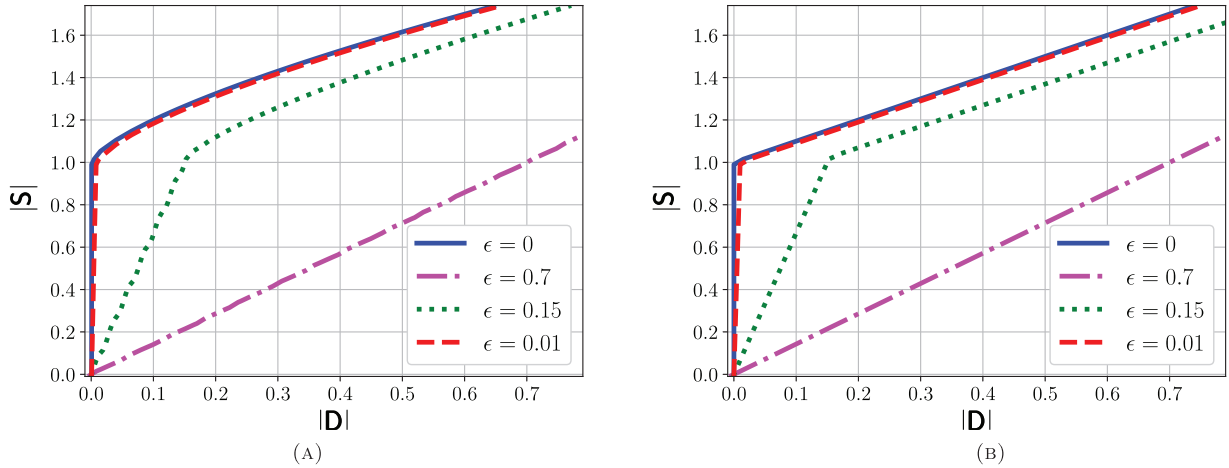
FIGURE 1. Regularisation (1.6) for the Bingham and Herschel–Bulkley constitutive relations with $\tau_* = 1$, taking the same value $\varepsilon = \varepsilon_1 = \varepsilon_2$. (A) Herschel–Bulkley relation ($r = 1.7$). (B) Bingham relation ($r = 2.0$).

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

$$\inf_{\overline{\boldsymbol{v}} \in V^n} \|\boldsymbol{v} - \overline{\boldsymbol{v}}\|_{W^{1,s}(\Omega)} \to 0 \qquad \text{as} \quad n \to \infty \quad \forall \, \boldsymbol{v} \in W_0^{1,s}(\Omega)^d,$$

$$\inf_{\overline{q} \in M^n} \|q - \overline{q}\|_{L^s(\Omega)} \to 0 \qquad \text{as} \quad n \to \infty \quad \forall \, q \in L^s(\Omega),$$

$$\inf_{\overline{\boldsymbol{\sigma}} \in \Sigma^n} \|\boldsymbol{\sigma} - \overline{\boldsymbol{\sigma}}\|_{L^s(\Omega)} \to 0 \qquad \text{as} \quad n \to \infty \quad \forall \, \boldsymbol{\sigma} \in L^s(\Omega)^{d \times d}.$$

Assumption 2.5 (Fortin Projector Π_Σ^n). *For each $n \in \mathbb{N}$ there is a linear projector $\Pi_\Sigma^n \colon L_{\mathrm{sym}}^1(\Omega)^{d \times d} \to \Sigma^n$ such that:*

– *(Preservation of divergence). For any $\boldsymbol{\sigma} \in L_{\mathrm{sym}}^1(\Omega)^{d \times d}$ we have that*

$$\int_\Omega \boldsymbol{\sigma} : \mathbf{D}(\boldsymbol{v}) = \int_\Omega \Pi_\Sigma^n(\boldsymbol{\sigma}) : \mathbf{D}(\boldsymbol{v}) \quad \forall \, \boldsymbol{v} \in V_{\mathrm{div}}^n;$$

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

$$\|\Pi_\Sigma^n \boldsymbol{\sigma}\|_{L^s(\Omega)} \le c \|\boldsymbol{\sigma}\|_{L^s(\Omega)} \qquad \forall \, \boldsymbol{\sigma} \in L_{\mathrm{sym}}^s(\Omega)^{d \times d}.$$

Assumption 2.6 (Fortin Projector Π_V^n). *For each $n \in \mathbb{N}$ there is a linear projector $\Pi_V^n \colon W_0^{1,1}(\Omega)^d \to V^n$ such that the following properties hold:*

– *(Preservation of divergence). For any $\boldsymbol{v} \in W_0^{1,1}(\Omega)^d$ we have that*

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

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

$$\|\Pi_V^n \boldsymbol{v}\|_{W^{1,s}(\Omega)} \le c \|\boldsymbol{v}\|_{W^{1,s}(\Omega)} \qquad \forall \, \boldsymbol{v} \in W_0^{1,s}(\Omega)^d.$$

Assumption 2.7 (Projector Π_M^n). *For each $n \in \mathbb{N}$ there is a linear projector $\Pi_M^n : L^1(\Omega) \to M^n$ such that for all $s \in (1, \infty)$ there is a constant $c > 0$, independent of n, such that:*

$$\|\Pi_M^n q\|_{L^s(\Omega)} \leq c \|q\|_{L^s(\Omega)} \qquad \forall q \in L^s(\Omega).$$

An important consequence of Assumptions 2.5 and 2.6 is inf-sup stability; *i.e.* we have that for any $s \in (1, \infty)$ there exist two constants $\beta_s, \gamma_s > 0$, that do not depend on n, such that

$$\inf_{q \in M^n \setminus \{0\}} \sup_{\boldsymbol{v} \in V^n \setminus \{0\}} \frac{\int_\Omega q \, \mathrm{div} \, \boldsymbol{v}}{\|\boldsymbol{v}\|_{W^{1,s}(\Omega)} \|q\|_{L^{s'}(\Omega)}} \geq \beta_s, \tag{2.4a}$$

$$\inf_{\boldsymbol{v} \in V_{\mathrm{div}}^n \setminus \{0\}} \sup_{\boldsymbol{\tau} \in \Sigma_{\mathrm{sym}}^n \setminus \{0\}} \frac{\int_\Omega \boldsymbol{\tau} : \mathbf{D}(\boldsymbol{v})}{\|\boldsymbol{\tau}\|_{L^{s'}(\Omega)} \|\boldsymbol{v}\|_{W^{1,s}(\Omega)}} \geq \gamma_s. \tag{2.4b}$$

There are several families of velocity and pressure pairs V^n–M^n that satisfy the assumptions above. We can for example mention the Taylor–Hood $\mathbb{P}_k$–$\mathbb{P}_{k-1}$ and MINI elements (see *e.g.* [9, 25]). The Scott–Vogelius element $\mathbb{P}_k$–$\mathbb{P}_{k-1}^{\mathrm{disc}}$ is also known to satisfy the assumptions, for instance, on barycentrically refined meshes [36]; this element has in addition the useful property that discretely divergence-free velocities are in fact also pointwise divergence-free. Regarding the shear stress, if the velocity element consists of globally continuous piecewise polynomials of degree $k \in \mathbb{N}$, then an example of a space satisfying Assumption 2.5 is given by (see *e.g.* [21]):

$$\Sigma^n = \left\{ \boldsymbol{\sigma} \in L_{\mathrm{sym}}^\infty(\Omega)^{d \times d} : \boldsymbol{\sigma}|_K \in \mathbb{P}_{k-1}(K)^{d \times d}, \text{ for all } K \in \mathcal{T}_n \right\}. \tag{2.5}$$

In the finite element formulation of the (regularised) problem we then look for functions $(\mathbf{S}^n, \boldsymbol{u}^n, p^n) \in \Sigma_{\mathrm{sym}}^n \times V^n \times M_0^n$ such that:

$$\int_\Omega \mathbf{G}_\varepsilon(\mathbf{S}^n, \mathbf{D}(\boldsymbol{u}^n)) : \boldsymbol{\tau} = 0 \qquad\qquad \forall \boldsymbol{\tau} \in \Sigma^n, \tag{2.6a}$$

$$\alpha \int_\Omega \boldsymbol{u}^n \cdot \boldsymbol{v} + \int_\Omega \mathbf{S}^n : \mathbf{D}(\boldsymbol{v}) - \int_\Omega p^n \, \mathrm{div} \, \boldsymbol{v} = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v} \qquad\qquad \forall \boldsymbol{v} \in V^n, \tag{2.6b}$$

$$- \int_\Omega q \, \mathrm{div} \, \boldsymbol{u}^n = 0 \qquad\qquad \forall q \in M^n. \tag{2.6c}$$

The Assumptions 2.4–2.7 and (G1)–(G4) guarantee that the discrete problem (2.6) has a solution, and that, as $n \to \infty$, the sequence of discrete solutions converges (weakly) to a solution of the continuous problem (see [21], Rem. 3.8). The rest of the paper will focus on how to solve the nonlinear problems described in this section by means of a semismooth Newton method.

3. Semismooth Newton method

3.1. Semismooth functions

Semismoothness is a concept originally introduced in [32] for the analysis of finite-dimensional nonsmooth optimisation problems (see also [34, 35]); it was helpful in the development of a generalised Newton method for functions that are not Fréchet-differentiable.

Suppose $F \colon \mathbb{R}^m \to \mathbb{R}^k$ is a locally Lipschitz function. Since F is (by Rademacher's theorem) differentiable everywhere except maybe on a set U_F of zero measure, one can define a generalised notion for the Jacobian of F at any point $z \in \mathbb{R}^m$, called Clarke's differential, as

$$\partial F(z) := \mathrm{co}\left\{ M \in \mathbb{R}^{k \times m} : \exists \{z_i\}_i \subset \mathbb{R}^m \setminus U_F \text{ with } z_i \to z, \, \nabla F(z_i) \to M \right\}, \tag{3.1}$$

where $\operatorname{co} A$ denotes the convex hull of the set A. It is known that Clarke's differential $\partial F(z)$ is nonempty, compact and convex, and that when F is continuously differentiable then $\partial F(z) = \{\nabla F(z)\}$ (for more details the reader is referred to [15]). We then say that F is *semismooth at z* if it is directionally differentiable and

$$\max_{M \in \partial F(z+h)} |f(z+h) - F(z) - Mh| = \mathrm{o}(|h|) \quad \text{as } h \to 0. \tag{3.2}$$

An important consequence of this definition is that if F is a locally Lipschitz and semismooth function in a neighborhood around $z_* \in \mathbb{R}^m$, where $F(z_*) = 0$, and $\partial F(z_*)$ is nonsingular, then the generalised Newton iteration

$$z^{j+1} = z^j - M_j^{-1} F(z^j), \tag{3.3}$$

where $M_j \in \partial F(z^j)$ is arbitrary, converges superlinearly to z_*, provided $|z^0 - z_*|$ is sufficiently small [14].

At the infinite dimensional level, where $F \colon Z \to X$ is now a function defined between two Banach spaces Z and X, Rademacher's theorem does not apply and therefore a characterisation of a generalised gradient such as (3.1) is not available. Thus, one has to work with a generally weaker and more abstract notion of generalised Jacobians. Let U be a neighborhood of $z \in Z$, and suppose $\partial F \colon U \subset Z \rightrightarrows \mathcal{L}(Z; X)$ is a set-valued function with $\partial F(\hat{z}) \neq \emptyset$ for every $\hat{z} \in U$. We say that F is *∂F-semismooth at z* (in the sense of Ulbrich [50]) if

$$\sup_{M \in \partial F(z+h)} \|F(z+h) - F(z) - Mh\|_X = \mathrm{o}(\|h\|_Z) \quad \text{as } h \to 0. \tag{3.4}$$

Similarly to the finite-dimensional case, one has that if F is continuously differentiable then it is $\{F'\}$-semismooth. Furthermore, as stated in the following proposition, the semismooth Newton algorithm also leads to superlinear convergence.

Proposition 3.1 ([49], Prop. 2.7). *Let X, Z be Banach spaces, and let U be a neighborhood of a point $z_* \in Z$. Suppose that the function $F \colon Z \to X$ is ∂F-semismooth at z_* and that $F(z_*) = 0$. Suppose that there exists a positive constant c, such that for every $z \in U$ and every $M \in \partial F(z)$, M is invertible with $\|M^{-1}\|_{\mathcal{L}(X;Z)} \leq c$. Then the generalised Newton–Kantorovich iteration defined by (3.3) converges superlinearly to z_* for all z^0 sufficiently close to z_*.*

We will now formulate a semismooth Newton method for the problem (1.1a). Define for $r \in (1, \infty)$ the mapping

$$F \colon Z := L^{r'}_{\mathrm{sym,tr}}(\Omega)^{d \times d} \times V^r \times L^{r'}_0(\Omega) \to X := L^q_{\mathrm{sym,tr}}(\Omega)^{d \times d} \times W^{-1,r'}(\Omega)^d \times L^r_0(\Omega),$$

$$F(\mathsf{S}, \boldsymbol{u}, p) := \begin{pmatrix} \overline{\mathsf{G}}(\mathsf{S}, \mathsf{D}(\boldsymbol{u})) \\ \alpha \boldsymbol{u} - \operatorname{div} \mathsf{S} + \nabla p - \boldsymbol{f} \\ \operatorname{div} \boldsymbol{u} \end{pmatrix}, \tag{3.5}$$

where $q > 1$ and $\overline{\mathsf{G}}$ denotes the Nemytskii operator associated to G. Observe that the second and third entries in the definition of F are given by linear and bounded operators, and so the semismoothness (or lack thereof) of F depends solely on that of $\overline{\mathsf{G}}$. This is advantageous because from this one immediately obtains a candidate for the multivalued function ∂F from Proposition 3.1. Namely, for $(\mathsf{S}, \boldsymbol{u}, p) \in Z$, the generalised Jacobian $\partial F(\mathsf{S}, \boldsymbol{u}, p)$ will consist of the linear operators $M \in \mathcal{L}(Z; X)$ of the form

$$M[\mathsf{T}, \boldsymbol{v}, q] = \begin{pmatrix} \boldsymbol{d}_1 \mathsf{T} + \boldsymbol{d}_2 \mathsf{D}(\boldsymbol{v}) \\ \alpha \boldsymbol{v} - \operatorname{div} \mathsf{T} + \nabla m \\ \operatorname{div} \boldsymbol{v} \end{pmatrix}, \tag{3.6}$$

where $[\boldsymbol{d}_1, \boldsymbol{d}_2]$ is a measurable selection of $\partial \mathsf{G}(\mathsf{S}(\cdot), \mathsf{D}(\boldsymbol{u}(\cdot)))$, with $\partial \mathsf{G}$ being the Clarke generalised gradient of G (*cf.* [50], Def. 3.40). The main algorithm can be summarised as follows:

Algorithm 3.2. Select an initial guess $(\mathbf{S}_0, \boldsymbol{u}_0, p_0) \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times V^r \times L_0^{r'}(\Omega)$, and for each $k = 0, 1, 2, \ldots$ perform the steps:

(1) If $F(\mathbf{S}_k, \boldsymbol{u}_k, p_k) = 0$, then terminate with $(\mathbf{S}_k, \boldsymbol{u}_k, p_k)$;
(2) Choose an arbitrary element $M_k \in \partial F$, where ∂F is defined through (3.6);
(3) Solve the system $M_k[\tilde{\mathbf{S}}_k, \tilde{\boldsymbol{u}}_k, \tilde{p}_k] = -F(\mathbf{S}_k, \boldsymbol{u}_k, p_k)$;
(4) Update $(\mathbf{S}_k, \boldsymbol{u}_k, p_k) \leftarrow (\mathbf{S}_k, \boldsymbol{u}_k, p_k) + (\tilde{\mathbf{S}}_k, \tilde{\boldsymbol{u}}_k, \tilde{p}_k)$.

The continuity of Nemytskii mappings between Lebesgue spaces has been well studied and it is known (see *e.g.* [27], Thm. 4) that if $\mathbf{G}$ satisfies, for some $q > 1$, the growth condition

$$|\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})| \leq c\left(|\boldsymbol{\sigma}|^{r'/q} + |\boldsymbol{\tau}|^{r/q}\right) + \tilde{c} \qquad \forall \boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{R}_{\mathrm{sym}}^{d \times d}, \tag{3.7}$$

where $c, \tilde{c} > 0$, then the Nemytskii map

$$(\mathbf{S}, \mathbf{T}) \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times L_{\mathrm{sym}}^{r}(\Omega)^{d \times d} \mapsto \overline{\mathbf{G}}(\mathbf{S}, \mathbf{T}) := \mathbf{G}(\mathbf{S}(\cdot), \mathbf{T}(\cdot)) \in L_{\mathrm{sym}}^{q}(\Omega)^{d \times d}, \tag{3.8}$$

is continuous. Suppose for a moment that the partial derivatives

$$\partial_{\boldsymbol{\sigma}} \mathbf{G}, \partial_{\boldsymbol{\tau}} \mathbf{G} \colon \mathbb{R}_{\mathrm{sym}}^{d \times d} \times \mathbb{R}_{\mathrm{sym}}^{d \times d} \to \mathcal{L}\left(\mathbb{R}_{\mathrm{sym}}^{d \times d}; \mathbb{R}_{\mathrm{sym}}^{d \times d}\right),$$

are continuous, and define

$$q := \min\{r, r'\} - \varepsilon \in [1, \min\{r, r'\}), \text{ with } \varepsilon \in (0, \min\{r, r'\} - 1]. \tag{3.9}$$

If we assume that the following growth conditions hold:

$$\begin{aligned}
\|\partial_{\boldsymbol{\sigma}} \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})\|_{\mathcal{L}(\mathbb{R}^{d \times d}; \mathbb{R}^{d \times d})} &\leq c_1 \left(|\boldsymbol{\sigma}|^{\frac{r'}{s}} + |\boldsymbol{\tau}|^{\frac{r}{s}}\right) + c_2, \\
\|\partial_{\boldsymbol{\tau}} \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})\|_{\mathcal{L}(\mathbb{R}^{d \times d}; \mathbb{R}^{d \times d})} &\leq \tilde{c}_1 \left(|\boldsymbol{\sigma}|^{\frac{r'}{p}} + |\boldsymbol{\tau}|^{\frac{r}{p}}\right) + \tilde{c}_2,
\end{aligned} \tag{3.10}$$

where $c_i, \tilde{c}_i$ are non-negative constants, and

$$s := \frac{qr'}{r' - q} \in [r, \infty) \qquad p := \frac{qr}{r - q} \in \left[\frac{r}{r - 1}, \frac{r}{r - 2}\right), \tag{3.11}$$

then the operators

$$(\mathbf{S}, \mathbf{T}) \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times L_{\mathrm{sym}}^{r}(\Omega)^{d \times d} \mapsto \partial_{\boldsymbol{\sigma}} \mathbf{G}(\mathbf{S}(\cdot), \mathbf{T}(\cdot)), \partial_{\boldsymbol{\tau}} \mathbf{G}(\mathbf{S}(\cdot), \mathbf{T}(\cdot)) \in L_{\mathrm{sym}}^{q}(\Omega)^{d \times d},$$

are continuous, which in turn implies that the corresponding Nemytskii maps are continuously differentiable (see *e.g.* [27], Thm. 7), and the derivatives

$$\begin{aligned}
\partial_1 \overline{\mathbf{G}} \colon L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times L_{\mathrm{sym}}^{r}(\Omega)^{d \times d} &\mapsto \mathcal{L}\left(L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d}; L_{\mathrm{sym}}^{q}(\Omega)^{d \times d}\right), \\
\partial_2 \overline{\mathbf{G}} \colon L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times L_{\mathrm{sym}}^{r}(\Omega)^{d \times d} &\mapsto \mathcal{L}\left(L_{\mathrm{sym}}^{r}(\Omega)^{d \times d}; L_{\mathrm{sym}}^{q}(\Omega)^{d \times d}\right),
\end{aligned} \tag{3.12}$$

are given by

$$\begin{aligned}
\left(\partial_1 \overline{\mathbf{G}}(\mathbf{S}, \mathbf{T}) \tilde{\mathbf{S}}\right)(\cdot) &= \partial_{\boldsymbol{\sigma}} \mathbf{G}(\mathbf{S}(\cdot), \mathbf{T}(\cdot)) \tilde{\mathbf{S}}(\cdot), \quad \tilde{\mathbf{S}} \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d}, \\
\left(\partial_2 \overline{\mathbf{G}}(\mathbf{S}, \mathbf{T}) \tilde{\mathbf{T}}\right)(\cdot) &= \partial_{\boldsymbol{\tau}} \mathbf{G}(\mathbf{S}(\cdot), \mathbf{T}(\cdot)) \tilde{\mathbf{T}}(\cdot), \quad \tilde{\mathbf{T}} \in L_{\mathrm{sym}}^{r}(\Omega)^{d \times d}.
\end{aligned} \tag{3.13}$$

In summary, if one assumes that the partial derivatives of G are continuous, and they satisfy the growth assumptions (3.10), then the Nemytskii map (3.8) is Fréchet-differentiable (and therefore so is F), which would make the use of a Newton–Kantorovich algorithm possible. Note that the number q defined (3.9) is smaller than both r and r'; this is an example of a norm gap, which is a phenomenon that appears in other contexts such as optimal control theory (see *e.g.* [47]).

As mentioned in the introduction, we wish here to work with Bingham-like constitutive relations that do not satisfy the differentiability requirement just described, and we therefore wish to have at our disposal a similar characterisation for semismoothness of superposition operators. The first results in this direction were obtained in [49] assuming Lipschitz continuity, which is too restrictive for our setting. Thankfully, this assumption was relaxed in [43], leading to conditions that guarantee semismoothness that are more widely applicable. The following lemma, consequence of Proposition A.1 from [43], summarises the required assumptions in the current setting.

Lemma 3.3. *Let $(\mathsf{S}, \mathsf{T}) \in L^{r'}_{\mathrm{sym}}(\Omega)^{d \times d} \times L^{r}_{\mathrm{sym}}(\Omega)^{d \times d}$, where $r \in (1, \infty)$ is arbitrary, and let $q \geq 1$ be defined by (3.9). Suppose that $\mathsf{G} \colon \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}} \to \mathbb{R}^{d \times d}_{\mathrm{sym}}$ is a locally Lipschitz function satisfying the growth assumption (3.7), with partial derivatives $\partial_{\boldsymbol{\sigma}} \mathsf{G}, \partial_{\boldsymbol{\tau}} \mathsf{G}$ satisfying (3.10). Suppose that $M := [\boldsymbol{d}_1 \; \boldsymbol{d}_2] \colon \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}} \to \mathcal{L}\left(\mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}}; \mathbb{R}^{d \times d}_{\mathrm{sym}}\right)$ satisfies*

$$
\begin{aligned}
\|\boldsymbol{d}_1(\boldsymbol{\sigma}, \boldsymbol{\tau})\|_{\mathcal{L}\left(\mathbb{R}^{d \times d}_{\mathrm{sym}}; \mathbb{R}^{d \times d}_{\mathrm{sym}}\right)} &\leq c_1 \left(|\boldsymbol{\sigma}|^{\frac{r'}{s}} + |\boldsymbol{\tau}|^{\frac{r}{s}}\right) + c_2, \\
\|\boldsymbol{d}_2(\boldsymbol{\sigma}, \boldsymbol{\tau})\|_{\mathcal{L}\left(\mathbb{R}^{d \times d}_{\mathrm{sym}}; \mathbb{R}^{d \times d}_{\mathrm{sym}}\right)} &\leq c_1 \left(|\boldsymbol{\sigma}|^{\frac{r'}{p}} + |\boldsymbol{\tau}|^{\frac{r}{p}}\right) + c_2,
\end{aligned}
\tag{3.14}
$$

where c_1, c_2 are positive constants, and s, p are defined by (3.11); assume further that M can be written as the a.e. pointwise limit of continuous functions. Then, if G is semismooth, the associated Nemytskii operator $\overline{\mathsf{G}}$ defined by (3.8) will satisfy

$$
\left\|\overline{\mathsf{G}}\left(\mathsf{S} + \tilde{\mathsf{S}}, \mathsf{T} + \tilde{\mathsf{T}}\right) - \overline{\mathsf{G}}(\mathsf{S}, \mathsf{T}) - M\left(\mathsf{S} + \tilde{\mathsf{S}}, \mathsf{T} + \tilde{\mathsf{T}}\right)\left(\tilde{\mathsf{S}}, \tilde{\mathsf{T}}\right)\right\|_{L^q(\Omega)} = \mathrm{o}\left(\left\|\left(\tilde{\mathsf{S}}, \tilde{\mathsf{T}}\right)\right\|_{L^{r'}(\Omega) \times L^r(\Omega)}\right) \quad as \; \left(\tilde{\mathsf{S}}, \tilde{\mathsf{T}}\right) \to \mathbf{0}.
\tag{3.15}
$$

Proof. First note that the existence of the partial derivatives $\partial_{\boldsymbol{\sigma}} \mathsf{G}, \partial_{\boldsymbol{\tau}} \mathsf{G}$ is guaranteed by Rademacher's theorem. Define now the local Lipschitz constants of G as

$$
L_1(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \sup_{|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}| \leq 1} \frac{|\mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \mathsf{G}(\tilde{\boldsymbol{\sigma}}, \boldsymbol{\tau})|}{|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}|}, \qquad L_2(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \sup_{|\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}| \leq 1} \frac{|\mathsf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \mathsf{G}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\tau}})|}{|\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}|},
\tag{3.16}
$$

where $\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}, \boldsymbol{\tau}, \tilde{\boldsymbol{\tau}} \in \mathbb{R}^{d \times d}_{\mathrm{sym}}$; then the growth assumption (3.10) can be reframed as

$$
\begin{aligned}
|L_1(\boldsymbol{\sigma}, \boldsymbol{\tau})| &\leq c_1 \left(|\boldsymbol{\sigma}|^{\frac{r'}{s}} + |\boldsymbol{\tau}|^{\frac{r}{s}}\right) + c_2, \\
|L_2(\boldsymbol{\sigma}, \boldsymbol{\tau})| &\leq c_1 \left(|\boldsymbol{\sigma}|^{\frac{r'}{p}} + |\boldsymbol{\tau}|^{\frac{r}{p}}\right) + c_2.
\end{aligned}
\tag{3.17}
$$

Since M is a pointwise limit of continuous functions (*i.e.* it is a Baire–Carathéodory function, but without an explicit space-dependence), Proposition A.1 of [43] implies that the function

$$
\begin{cases}
\frac{M(\mathsf{S}(x) + \boldsymbol{\sigma}, \mathsf{T}(x) + \boldsymbol{\tau})[\boldsymbol{\sigma}, \boldsymbol{\tau}] - (\mathsf{G}(\mathsf{S}(x) + \boldsymbol{\sigma}, \mathsf{T}(x) + \boldsymbol{\tau}) - \mathsf{G}(\mathsf{S}(x), \mathsf{T}(x)))}{|(\boldsymbol{\sigma}, \boldsymbol{\tau})|} & \Longleftrightarrow (\boldsymbol{\sigma}, \boldsymbol{\tau}) \neq \mathbf{0}, \\
\mathbf{0} & \Longleftrightarrow (\boldsymbol{\sigma}, \boldsymbol{\tau}) = \mathbf{0},
\end{cases}
\tag{3.18}
$$

defined for $(x, \boldsymbol{\sigma}, \boldsymbol{\tau}) \in \Omega \times \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}}$, induces a continuous Nemytskii operator, which in particular implies the desired condition (3.15). $\square$

In practice, the function M in Lemma 3.3 arises from derivatives of $\mathbf{G}$, and so the growth conditions (3.10) and (3.14) are redundant. This is the case for instance with the generalised Jacobian defined by (3.6). After having identified the necessary assumptions to guarantee semismoothness, it only remains to examine the solvability of the linearised system.

Assumption 3.4. *Let* $(\mathbf{S}, \boldsymbol{u}) \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times V^r$, *with* $r \in (1, \infty)$, *and define* q *through* (3.9). *Let* $\mathbf{G} \colon \mathbb{R}_{\mathrm{sym}}^{d \times d} \times \mathbb{R}_{\mathrm{sym}}^{d \times d} \to \mathbb{R}_{\mathrm{sym}}^{d \times d}$ *be locally Lipschitz. Given a triplet* $(\mathbf{H}, \boldsymbol{h}, h) \in L_{\mathrm{sym}}^q(\Omega)^{d \times d} \times L^{r'}(\Omega)^d \times L_0^r(\Omega)$, *there exists a solution* $(\mathbf{T}, \boldsymbol{v}, m) \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times V^r \times L_0^{r'}(\Omega)$ *of the linear system*

$$
\begin{aligned}
\boldsymbol{d}_1 \mathbf{T} + \boldsymbol{d}_2 \mathbf{D}(\boldsymbol{v}) &= \mathbf{H}, \\
\alpha \boldsymbol{v} - \operatorname{div} \mathbf{T} + \nabla m &= \boldsymbol{h}, \\
\operatorname{div} \boldsymbol{v} &= h,
\end{aligned}
\tag{3.19}
$$

where $[\boldsymbol{d}_1, \boldsymbol{d}_2]$ *is a measurable selection of* $\partial \mathbf{G}(\mathbf{S}(\cdot), \mathbf{D}\boldsymbol{u}(\cdot))$, *such that*

$$
\|\mathbf{T}\|_{L^{r'}(\Omega)} + \|\boldsymbol{v}\|_{W^{1,r}(\Omega)} + \|m\|_{L^{r'}(\Omega)} \le c \Big(\|\mathbf{H}\|_{L^q(\Omega)} + \|\boldsymbol{h}\|_{L^{r'}(\Omega)} + \|h\|_{L^r(\Omega)} \Big),
\tag{3.20}
$$

where c *is a positive constant.*

An immediate consequence of Proposition 3.1 is the following convergence result.

Theorem 3.5. *Let* $\mathbf{G} \colon \mathbb{R}_{\mathrm{sym}}^{d \times d} \times \mathbb{R}_{\mathrm{sym}}^{d \times d} \to \mathbb{R}_{\mathrm{sym}}^{d \times d}$ *be a function satisfying assumptions (G1)–(G4) for some* $r > 1$. *Suppose that the growth conditions* (3.7) *and* (3.10) *hold, and that Assumption 3.4 is satisfied. Let* $(\mathbf{S}, \boldsymbol{u}, p) \in L_{\mathrm{sym}}^{r'}(\Omega)^{d \times d} \times V^r \times L_0^{r'}(\Omega)$ *be such that* $F(\mathbf{S}, \boldsymbol{u}, p) = 0$, *where* F *is defined by* (3.5). *Then the iterates produced by Algorithm 3.2 converge superlinearly to* $(\mathbf{S}, \boldsymbol{u}, p)$, *for any initial guess* $(\mathbf{S}_0, \boldsymbol{u}_0, p_0)$ *sufficiently close to* $(\mathbf{S}, \boldsymbol{u}, p)$.

In some cases higher regularity results allow for less restrictive growth conditions for $\mathbf{G}$, meaning that a larger q could be selected in (3.5). For instance, the strict monotonicity (2.3) might result in higher regularity for the problem associated to the regularisation $\mathbf{G}_\varepsilon$ defined by (1.6), namely $(\mathbf{S}_\varepsilon, \boldsymbol{u}_\varepsilon) \in W_{\mathrm{sym}}^{1,\min\{2,r'\}}(\Omega)^{d \times d} \times W^{2,\min\{2,r\}}(\Omega)^d$; in that case one could *e.g.* guarantee the continuity of the Nemytskii operator $\overline{\mathbf{G}}_\varepsilon$ with the more general growth condition

$$
|\mathbf{G}_\varepsilon(\boldsymbol{\sigma}, \boldsymbol{\tau})| \le c \left(|\boldsymbol{\sigma}|^{\frac{\min\{2,r'\}^*}{q}} + |\boldsymbol{\tau}|^{\frac{\min\{2,r\}^*}{q}} \right) + \tilde{c} \qquad c, \tilde{c} > 0, \qquad \boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{R}_{\mathrm{sym}}^{d \times d},
\tag{3.21}
$$

where $p^* := \frac{pd}{d-p}$ denotes the Sobolev exponent of p (it is defined as infinity when $p = d$).

Without assuming any sort of strict monotonicity of the constitutive relation, it might prove difficult to solve the linearised system (3.19) from Assumption 3.4. In this case the use of the regularisation (1.6) also brings an advantage: the monotonicity condition (2.3) implies by the implicit function theorem that the partial derivatives of $\mathbf{G}_\varepsilon$ are uniformly positive definite, and so the matrix function $A_\varepsilon := -(\boldsymbol{d}_1^\varepsilon)^{-1} \boldsymbol{d}_2^\varepsilon$ is uniformly positive definite, which means that the system (3.19) can be rewritten as a Stokes-type system:

$$
\begin{aligned}
\alpha \boldsymbol{v}_\varepsilon - \operatorname{div}(A_\varepsilon \mathbf{D}(\boldsymbol{v}_\varepsilon)) + \nabla m_\varepsilon &= \boldsymbol{h} + \operatorname{div} \tilde{\mathbf{H}}, \\
\operatorname{div} \boldsymbol{v}_\varepsilon &= h,
\end{aligned}
\tag{3.22}
$$

where $\tilde{\mathbf{H}} := (\boldsymbol{d}_1^\varepsilon)^{-1} \mathbf{H} \in L^q(\Omega)^{d \times d}$. If $q \ge \max\{r, r'\}$, the system (3.22) can be solved by standard energy arguments, and the required estimate (3.20) is for instance a consequence of Theorem 1.4 from [12]. In the opposite case $q \in (1, \max\{r, r'\})$ the situation becomes slightly more delicate since the term $\operatorname{div} \tilde{\mathbf{H}}$, with $\mathbf{H} \in L_{\mathrm{sym}}^q(\Omega)^{d \times d}$ and q small, does not allow the use of standard energy estimates. The results from [12] still guarantee the existence of a solution, but the estimate takes the form

$$
\|\mathbf{T}_\varepsilon\|_{L^q(\Omega)} + \|\boldsymbol{v}_\varepsilon\|_{W^{1,q}(\Omega)} + \|m_\varepsilon\|_{L^q(\Omega)} \le c \Big(1 + \|\mathbf{H}\|_{L^q(\Omega)} + \|\boldsymbol{h}\|_{L^{r'}(\Omega)} + \|h\|_{L^r(\Omega)} \Big).
\tag{3.23}
$$

In this case, the addition of a smoothing step to Algorithm 3.2 that takes the solution back to $(\mathbf{T}_\varepsilon, \boldsymbol{v}_\varepsilon, m_\varepsilon) \in L^{r'}_{\mathrm{sym}}(\Omega)^{d \times d} \times W^{1,r}_0(\Omega)^d \times L^{r'}_0(\Omega)$ would be necessary (see *e.g.* [50] for more details on smoothing operators).

Corollary 3.6. *Let* $\mathbf{G} \colon \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}} \to \mathbb{R}^{d \times d}_{\mathrm{sym}}$ *be a function satisfying (G1)–(G4) for some* $r > 1$, *and let* $(\mathbf{S}_\varepsilon, \boldsymbol{u}_\varepsilon, p_\varepsilon) \in L^r_{\mathrm{sym}}(\Omega)^{d \times d} \times V^r \times L^r_0(\Omega)$ *be such that* $F_\varepsilon(\mathbf{S}_\varepsilon, \boldsymbol{u}_\varepsilon, p_\varepsilon) = 0$, *where* F_ε *is defined through* (3.5) *with* $\mathbf{G}$ *replaced by* $\mathbf{G}_\varepsilon$. *Then, if* $q \geq \max\{r, r'\}$ *(recall* (3.21)*), the iterates produced by Algorithm 3.2 (employing* F_ε *instead of* F*) converge superlinearly to* $(\mathbf{S}_\varepsilon, \boldsymbol{u}_\varepsilon, p_\varepsilon)$, *for any initial guess* $(\mathbf{S}^\varepsilon_0, \boldsymbol{u}^\varepsilon_0, p^\varepsilon_0)$ *sufficiently close to* $(\mathbf{S}_\varepsilon, \boldsymbol{u}_\varepsilon, p_\varepsilon)$. *If* $q < \max\{r, r'\}$ *the same result holds if a smoothing step is included in Algorithm 3.2.*

Remark 3.7. While the previous discussion dealt with the continuous problem (1.1), the exact same arguments regarding *e.g.* continuity could be applied to a weakly enforced form of the constitutive relation

$$\int_\Omega \mathbf{G}(\mathbf{S}, \mathbf{D}(\boldsymbol{u})) : \boldsymbol{\tau} = \mathbf{0} \qquad \forall\, \boldsymbol{\tau} \in C^\infty_0(\Omega)^{d \times d},$$

and so in turn also to the finite element discretisation (2.6). At the discrete level all norms are equivalent, which would make it straightforward to obtain estimate (3.20), without needing sophisticated results like the ones from [12]. However, it is likely that this would then result in mesh-dependent behaviour of the iterations; it is for this reason that we endeavoured to carry out the analysis at the function space level.

Remark 3.8. The convective term can be incorporated into the analysis in a similar manner as is done for the Newtonian problem. For example, if we employ a Picard-like linearisation of the form

$$\mathcal{B}[\boldsymbol{u}; \boldsymbol{v}, \boldsymbol{\phi}] := - \int_\Omega \boldsymbol{v} \otimes \boldsymbol{u} : \boldsymbol{\phi},$$

where $\boldsymbol{u} \in V^r$ is the current solution and $\boldsymbol{\phi} \in C^\infty_0(\Omega)^d$ is a test function, the solvability of the linearised system remains intact because of the skew-symmetry property $\mathcal{B}[\boldsymbol{u}; \boldsymbol{v}, \boldsymbol{v}] = 0$ for all $\boldsymbol{v} \in V^r$ (but note that one has to assume that r is large enough so that choosing $\boldsymbol{\phi} = \boldsymbol{v}$ is allowed), which is a consequence of the divergence-free condition $\operatorname{div} \boldsymbol{u} = 0$. If one wishes to employ instead the form of the convective term that would arise from Newton linearisation:

$$\mathcal{B}[\boldsymbol{u}; \boldsymbol{v}, \boldsymbol{\phi}] := - \int_\Omega (\boldsymbol{v} \otimes \boldsymbol{u} + \boldsymbol{u} \otimes \boldsymbol{v}) : \boldsymbol{\phi},$$

the situation is slightly more delicate, and even in the Newtonian case it is necessary to assume *e.g.* that the viscosity is large enough in order to guarantee that the linearised system can be solved (see *e.g.* [25], Lem. 3.2).

3.2. The Bingham constitutive relation

In this section we will look at some explicit examples of constitutive relations that fall into the framework described in the previous section. Namely, we will show that the Bingham constitutive relation can be described by means of a semismooth function $\mathbf{G}$ that also satisfies appropriate growth conditions.

First, we claim that the relation (1.4b) defining a Bingham fluid is semismooth. Note that, excepting the first term, all the terms appearing in the following function

$$\mathbf{G} \colon \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}} \to \mathbb{R}^{d \times d}_{\mathrm{sym}} \qquad \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) := |\boldsymbol{\tau}|\boldsymbol{\sigma} - (\tau_* + 2\nu|\boldsymbol{\tau}|)\boldsymbol{\tau} \qquad \tau_*, \nu > 0, \tag{3.24}$$

are smooth. Therefore we only need concern ourselves with the function

$$\boldsymbol{h} \colon \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}} \to \mathbb{R}^{d \times d}_{\mathrm{sym}} \qquad \boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{\tau}) := |\boldsymbol{\tau}|\boldsymbol{\sigma}. \tag{3.25}$$

The function $\boldsymbol{h}$ is locally Lipschitz, and its (Clarke) generalised Jacobian is given by

$$\partial \boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \begin{cases} \left\{ \left[|\boldsymbol{\tau}|\mathbf{I}, \boldsymbol{\sigma} \otimes \frac{\boldsymbol{\tau}}{|\boldsymbol{\tau}|} \right] \right\}, & \text{if } |\boldsymbol{\tau}| \neq 0, \\ \left\{ [\mathbf{0}, \boldsymbol{\sigma} \otimes \boldsymbol{\phi}] : \boldsymbol{\phi} \in \mathbb{R}^{d \times d}_{\mathrm{sym}}, |\boldsymbol{\phi}| \leq 1 \right\}, & \text{if } |\boldsymbol{\tau}| = 0, \end{cases} \tag{3.26}$$

where $\mathbf{I}$ denotes the fourth order identity tensor. Since for $\boldsymbol{\tau} \neq \mathbf{0}$ the function $\boldsymbol{h}$ is actually smooth, we only need to prove semismoothness at $\boldsymbol{\tau} = \mathbf{0}$:

$$|\boldsymbol{h}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}}) - \boldsymbol{h}(\boldsymbol{\sigma}, \mathbf{0}) - |\tilde{\boldsymbol{\tau}}|\tilde{\boldsymbol{\sigma}} - |\tilde{\boldsymbol{\tau}}|(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}})| = |\tilde{\boldsymbol{\tau}}||\tilde{\boldsymbol{\sigma}}||(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}})| = \mathrm{o}(|(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}})|), \text{ as } |(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}})| \to 0.$$

The function (3.24) is therefore semismooth and its generalised Jacobian is given by

$$\partial \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \begin{cases} \left\{ \left[|\boldsymbol{\tau}|\mathbf{I}, \boldsymbol{\sigma} \otimes \frac{\boldsymbol{\tau}}{|\boldsymbol{\tau}|} - (\tau_* + |\boldsymbol{\tau}|)\mathbf{I} - \frac{\boldsymbol{\tau} \otimes \boldsymbol{\tau}}{|\boldsymbol{\tau}|} \right] \right\}, & \text{if } |\boldsymbol{\tau}| \neq 0, \\ \left\{ [\mathbf{0}, \boldsymbol{\sigma} \otimes \boldsymbol{\phi} - \tau_*\mathbf{I}] : \boldsymbol{\phi} \in \mathbb{R}^{d \times d}_{\mathrm{sym}}, |\boldsymbol{\phi}| \leq 1 \right\}, & \text{if } |\boldsymbol{\tau}| = 0. \end{cases} \tag{3.27}$$

For the Bingham constitutive relation we have $r = 2$ and so the solution to the problem belongs to $(\mathbf{S}, \boldsymbol{u}, p) \in L^2_{\mathrm{sym}}(\Omega)^{d \times d} \times W^{1,2}_0(\Omega)^d \times L^2_0(\Omega)$. On the other hand, a simple application of Young's inequality yields the growth conditions for $\mathbf{G}$ and its derivatives:

$$\begin{aligned} |\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})| &\leq c(|\boldsymbol{\sigma}|^2 + |\boldsymbol{\tau}|^2) + \tilde{c}, \\ |\partial_{\boldsymbol{\sigma}}\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})| &\leq c(|\boldsymbol{\sigma}| + |\boldsymbol{\tau}|) + \tilde{c}, \\ |\partial_{\boldsymbol{\tau}}\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})| &\leq c(|\boldsymbol{\sigma}| + |\boldsymbol{\tau}|) + \tilde{c}, \end{aligned} \tag{3.28}$$

where $c, \tilde{c} > 0$. The growth conditions (3.28) alone do not suffice to be able to apply the results from the previous section, since they only imply that the Nemytskii operator (3.8) is continuous into $L^1_{\mathrm{sym}}(\Omega)^{d \times d}$, which brings great difficulties since it is unlikely that one can solve the linearised system (3.19) with a uniform bound on the solution, if the linearised system is at all solvable (recall the lack of monotonicity).

Thankfully, at this point we can make use of regularity results that guarantee the boundedness of gradients for Bingham and Herschel–Bulkley models [23,24]. Noting also that the mapping $\mathbf{D}(\boldsymbol{u}_\varepsilon) \mapsto \mathbf{S}_\varepsilon$ defined by the relation (1.6) is Lipschitz continuous ([30], Lem. 4.5), we obtain also boundedness of the stresses, which implies that the corresponding Nemytskii operator $\overline{\mathbf{G}}_\varepsilon$ will be continuous into $L^2_{\mathrm{sym}}(\Omega)^{d \times d}$. Moreover, thanks to the uniform monotonicity condition (2.3), it is possible to solve the linearised system in a straightforward manner, which implies that Corollary 3.6 can be applied to obtain convergence of Algorithm 3.2 (note that $\mathbf{G}_\varepsilon$ is semismooth, because it is a composition of a linear transformation with $\mathbf{G}$).

Remark 3.9. Strictly speaking, the expression (3.24) does not exactly represent the constitutive relation, since when $|\boldsymbol{\tau}| = 0$, the equation $\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \mathbf{0}$ is satisfied for arbitrary $\boldsymbol{\sigma} \in \mathbb{R}^{d \times d}_{\mathrm{sym}}$, and not just $|\boldsymbol{\sigma}| \leq \tau_*$ as required by the Bingham relation (1.3). However, the regularised relation $\mathbf{G}_\varepsilon$ does not exhibit this unwanted behaviour; *e.g.* Figure 1 was obtained using this expression. Numerical experiments in Section 4 will also show that the semismooth Newton algorithm will perform well using the regularised version of (3.24).

The alternative expression (1.4a) used to define the same Bingham constitutive relation satisfies the same growth conditions (3.28), but is not semismooth. To see this, note that its Clarke Jacobian is given by

$$\partial \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \begin{cases} \left\{ \left[(|\boldsymbol{\sigma}| - \tau_*)^+\mathbf{I} + \mathbb{1}_{\{|\boldsymbol{\sigma}| > \tau_*\}} \frac{(\boldsymbol{\sigma} - 2\nu\boldsymbol{\tau}) \otimes \boldsymbol{\sigma}}{|\boldsymbol{\sigma}|}, -2\nu(\tau_* + (|\boldsymbol{\sigma}| - \tau_*)^+)\mathbf{I} \right] \right\}, & \text{if } |\boldsymbol{\sigma}| \neq \tau_*, \\ \left\{ [(\boldsymbol{\sigma} - 2\nu\boldsymbol{\tau}) \otimes \boldsymbol{\phi}, -2\nu\tau_*\mathbf{I}] : \boldsymbol{\phi} \in \mathbb{R}^{d \times d}_{\mathrm{sym}}, |\boldsymbol{\phi}| \leq 1 \right\}, & \text{if } |\boldsymbol{\sigma}| = \tau_*. \end{cases}$$

Take $(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathbb{R}^{d \times d}_{\mathrm{sym}} \times \mathbb{R}^{d \times d}_{\mathrm{sym}}$ with $|\boldsymbol{\sigma}| = \tau_*$ and $\boldsymbol{\tau}$ arbitrary. To see that $\mathbf{G}$ is not semismooth, take a sequence $(\boldsymbol{\sigma}_k, \boldsymbol{\tau}_k) \to (\mathbf{0}, \mathbf{0})$ such that $|\boldsymbol{\sigma} + \boldsymbol{\sigma}_k| = \tau_*$. Then for any $M_{\boldsymbol{\phi}} \in \partial \mathbf{G}(\boldsymbol{\sigma} + \boldsymbol{\sigma}_k, \boldsymbol{\tau} + \boldsymbol{\tau}_k)$ we have that

$$|\mathbf{G}(\boldsymbol{\sigma} + \boldsymbol{\sigma}_k, \boldsymbol{\tau} + \boldsymbol{\tau}_k) - \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - M_{\boldsymbol{\phi}}[\boldsymbol{\sigma}_k, \boldsymbol{\tau}_k]| = |\boldsymbol{\sigma}_k : \boldsymbol{\phi}||2\nu(\boldsymbol{\tau} + \boldsymbol{\tau}_k) - (\boldsymbol{\sigma} + \boldsymbol{\sigma}_k)|,$$

which is not $\mathrm{o}(|(\boldsymbol{\sigma}_k, \boldsymbol{\tau}_k)|)$ as $k \to \infty$. However, it is possible to prove that

$$|\mathbf{G}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}, \boldsymbol{\tau} + \tilde{\boldsymbol{\tau}}) - \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - M[\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}}]| = \mathrm{o}(|(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}})|^\alpha), \text{ as } |(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\tau}})| \to 0,$$

for $(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathbb{R}_{\mathrm{sym}}^{d \times d} \times \mathbb{R}_{\mathrm{sym}}^{d \times d}$ with $|\boldsymbol{\sigma}| = \tau_*$, where $M \in \partial \mathbf{G}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}, \boldsymbol{\tau} + \tilde{\boldsymbol{\tau}})$, and $\alpha \in (0, 1)$ is arbitrary. This suggests that Algorithm 3.2 would still be convergent, albeit at a slower rate.

The following is yet another expression that can be used to describe the Bingham relation:

$$\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) := (|\boldsymbol{\sigma}| - \tau_*)^+ \frac{\boldsymbol{\sigma}}{|\boldsymbol{\sigma}|} - 2\nu \boldsymbol{\tau}. \tag{3.29}$$

The expression (3.29) in fact satisfies a less restrictive growth condition

$$|\mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})| \leq c(|\boldsymbol{\sigma}| + |\boldsymbol{\tau}|) + \tilde{c} \qquad c, \tilde{c} > 0, \tag{3.30}$$

which *e.g.* guarantees the continuity of the Nemytskii operator $\overline{\mathbf{G}}$ into $L_{\mathrm{sym}}^2(\Omega)^{d \times d}$ without the need for regularity results, but unfortunately it also fails to be semismooth. This shows that different expressions for the same constitutive can relation result in different behaviour, and so it would be worthwhile in future research to look for the most advantageous ways of writing the constitutive relation.

Remark 3.10. The Herschel–Bulkley relation can be analysed in a similar manner. For instance, we could use the following expression:

$$\mathbf{G} \colon \mathbb{R}_{\mathrm{sym}}^{d \times d} \times \mathbb{R}_{\mathrm{sym}}^{d \times d} \to \mathbb{R}_{\mathrm{sym}}^{d \times d} \qquad \mathbf{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) := |\boldsymbol{\tau}| \boldsymbol{\sigma} - \left(\tau_* + 2\nu |\boldsymbol{\tau}|^{r-1} \right) \boldsymbol{\tau} \qquad \tau_*, \nu > 0, r > 1. \tag{3.31}$$

Since $r > 1$, the function $\boldsymbol{\tau} \mapsto |\boldsymbol{\tau}|^{r-1} \boldsymbol{\tau}$ is actually differentiable, and so by the same arguments employed above, the function (3.31) is semismooth. As for the growth condition, one can *e.g.* see that the regularised expression satisfies

$$|\mathbf{G}_\varepsilon(\boldsymbol{\sigma}, \boldsymbol{\tau})| \leq c \left(|\boldsymbol{\sigma}|^{\max\{r,2\}} + |\boldsymbol{\tau}|^{\max\{r,2\}} \right) + \tilde{c} \qquad c, \tilde{c} > 0. \tag{3.32}$$

Then, for instance if we look at the shear-thinning case $r < 2$, the natural higher integrability $(\mathbf{S}, \mathbf{D}(\boldsymbol{u})) \in L_{\mathrm{sym}}^{2^*}(\Omega)^{d \times d} \times L_{\mathrm{sym}}^{r^*}(\Omega)^{d \times d}$ would suggest we rewrite (3.32) as

$$|\mathbf{G}_\varepsilon(\boldsymbol{\sigma}, \boldsymbol{\tau})| \leq c \left(|\boldsymbol{\sigma}|^{\frac{2^*}{2^*/2}} + |\boldsymbol{\tau}|^{\frac{r^*}{r^*/2}} \right) + \tilde{c} \qquad c, \tilde{c} > 0. \tag{3.33}$$

Therefore, if we assume *e.g.* that $r \geq \frac{3d}{d+2}$ (which ensures that $\frac{r^*}{2} \geq r'$), Corollary 3.6 implies the convergence of Algorithm 3.2 without the need for a smoothing step. We note that the condition $r \geq \frac{3d}{d+2}$ is advantageous when analysing the convergence of finite element approximations for the system including the convective term (see *e.g.* [18]).

Remark 3.11. The method presented in this work has the advantage that it works directly with the basic formulation conventionally derived in continuum mechanics: balance laws in the form of partial differential equations (1.1a) supplemented with a constitutive relation in the form of a pointwise algebraic constraint (1.1b). This leads to a lot of flexibility in the choice of constitutive relations; for instance, one could easily capture with the general form (1.1b) constitutive relations of the type

$$\begin{cases} |\mathbf{S}| \leq \tau_* \iff \mathbf{D}(\boldsymbol{u}) = \mathbf{0}, \\ |\mathbf{S}| > \tau_* \iff \mathbf{S} = \tau_* \dfrac{\mathbf{D}(\boldsymbol{u})}{|\mathbf{D}(\boldsymbol{u})|} + \mathbf{S}_2, \quad \text{where } \mathbf{D}(\boldsymbol{u}) = \alpha(|\mathbf{S}|)\mathbf{S}. \end{cases} \tag{3.34}$$

In contrast, the variational inequality formulation makes heavy use of the form of the potential, *e.g.* for the Bingham model, $\int_\Omega (\nu |\mathbf{D}(\boldsymbol{u})|^2 + \tau_* |\mathbf{D}(\boldsymbol{u})|)$ and it is not immediately clear how one should modify the formulation to capture a relation such as (3.34). In addition, the variational inequality formulation is itself derived as a *consequence* of the laws (1.1), and in some cases it is not completely obvious that a solution of the variational inequality is also a solution of the weak form of (1.1), *e.g.* in the three-dimensional unsteady Bingham model including inertia, where testing with the velocity $\boldsymbol{u}$ is not allowed.

4. Numerical examples

In this section we implement Algorithm 3.2 with the regularisation (1.6) applied to the Bingham constitutive relation expressed *via* (1.4b). All the examples were implemented using Firedrake [40]. The semismooth Newton iterations are carried out until the Euclidean norm of the residual falls below 1×10^{-9}, unless specified otherwise; The linear systems were solved with the sparse direct solver from MUMPS [2]. Firedrake makes use of the Unified Form Language [1], which automatically chooses an element of the generalised derivative of the positive part f_+ of a function f as:

$$\nabla f_+ = \begin{cases} \nabla f & \text{if } f > 0, \\ 0 & \text{if } f \leq 0. \end{cases} \tag{4.1}$$

4.1. Flow between two plates

Consider the problem posed on the domain $\Omega = (0,4) \times (-1,1)$ with $\boldsymbol{f} = \boldsymbol{0}$. The following function solves the Bingham problem exactly (*cf.* [20, 28]):

$$\boldsymbol{u}_e(\boldsymbol{x}) := \begin{cases} \left(\frac{\sqrt{2}\tau_*}{4}, 0\right)^{\top}, & \text{if } |x_2| \leq \frac{1}{2}, \\ \left(\sqrt{2}\tau_*\left(|x_2| - x_2^2\right), 0\right)^{\top}, & \text{if } |x_2| \geq \frac{1}{2}, \end{cases} \tag{4.2}$$

$$p_e(\boldsymbol{x}) := \sqrt{2}\tau_*(16 - x_1). \tag{4.3}$$

The boundary data are then chosen so as to match the given expression above.

Even though there are many families of finite element methods satisfying Assumptions 2.4–2.7 (recall the discussion in Sect. 2.2), which in theory lead to a convergent scheme for a big class of implicitly constituted fluids, in the particular case of viscoplastic flow it has been observed in practice that schemes for which the shear stress S is approximated by piecewise constant functions are more appropriate (see *e.g.* [46]). For instance, when using the Taylor–Hood $\mathbb{P}_2$–$\mathbb{P}_1$ element, oscillations that spoil the overall convergence of the nonlinear iterations might appear; Figure 2 shows *e.g.* such oscillations in the solution obtained using a mesh with 3.6×10^4 degrees of freedom. For this reason, in the rest of this work we will employ instead a stabilised $\mathbb{P}_1$–$\mathbb{P}_1$ element for the velocity and pressure; *i.e.* the finite element spaces will be chosen as (recall (2.5)):

$$\Sigma^n := \left\{\boldsymbol{\sigma} \in L^{\infty}_{\text{sym}}(\Omega)^{d \times d} \colon \boldsymbol{\sigma}|_K \in \mathbb{P}_0(K)^{d \times d}, \ K \in \mathcal{T}_n\right\},$$
$$V^n := \left\{\boldsymbol{v} \in C(\overline{\Omega})^d \colon \boldsymbol{v}|_{\partial\Omega} = \boldsymbol{u}_e, \ \boldsymbol{v}|_K \in \mathbb{P}_1(K)^d, \ K \in \mathcal{T}_n\right\},$$
$$M^n := \left\{q \in L^{\infty}_0(\Omega) \colon q|_K \in \mathbb{P}_1(K), \ K \in \mathcal{T}_n\right\},$$

and the following stabilisation term is added to the right-hand-side of the discrete equation (2.6c):

$$0.2h^2 \int_{\Omega} \nabla p^n \cdot \nabla q. \tag{4.4}$$

In all the numerical experiments, the phrase "degrees of freedom" (or "#dofs") refers to the total number of unknowns associated to the stress, velocity, and pressure.

In this example we set $\tau_* = 1$, and continuation was employed to obtain better initial guesses for the nonlinear iterations; in particular, the problem was solved for the values $\varepsilon \in \{0.5, 0.0166, 0.001, 0.0001\}$, with each solution being used as the initial guess in the iterations, for the next smaller value of ε. Figures 3 and 4 show the L^2 and H^1 error of the velocity as ε and the mesh size decrease, respectively; it can be observed that for values less than $\varepsilon = 0.001$, it is the discretisation error that dominates. Figures 5 and 6 show the decrease in the residual norm as the number of semismooth Newton iterations increases. The different values of ε and the mesh size result in very similar behaviour for the residual, and in particular we see that the solver does not seem to exhibit any mesh-dependence.
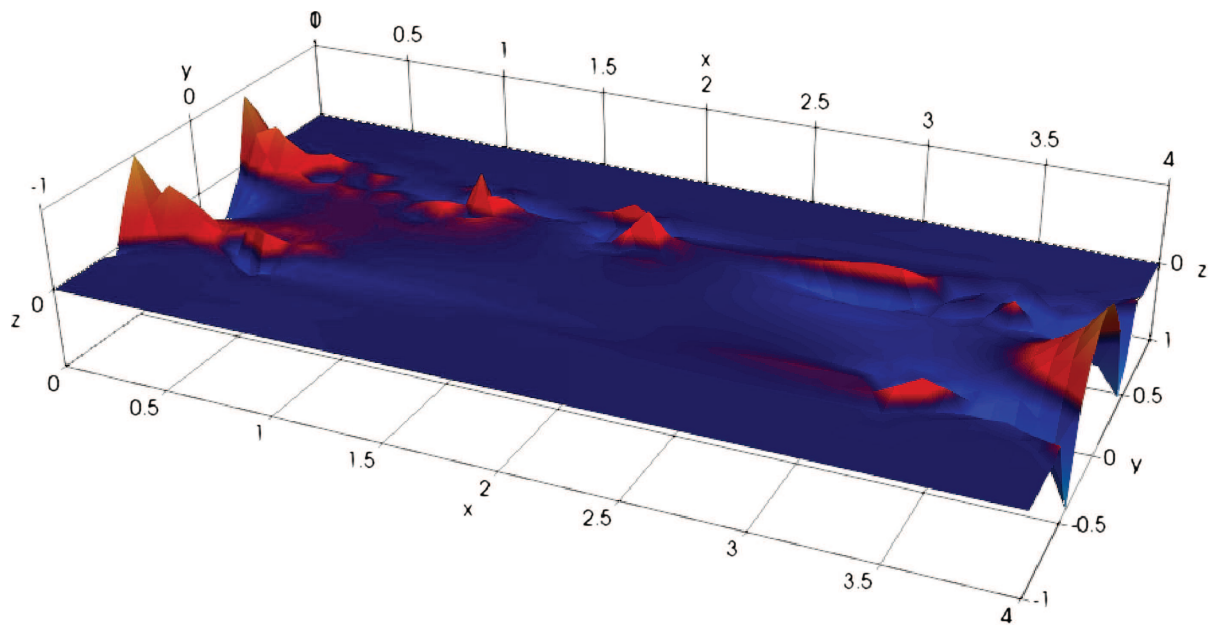
(A)
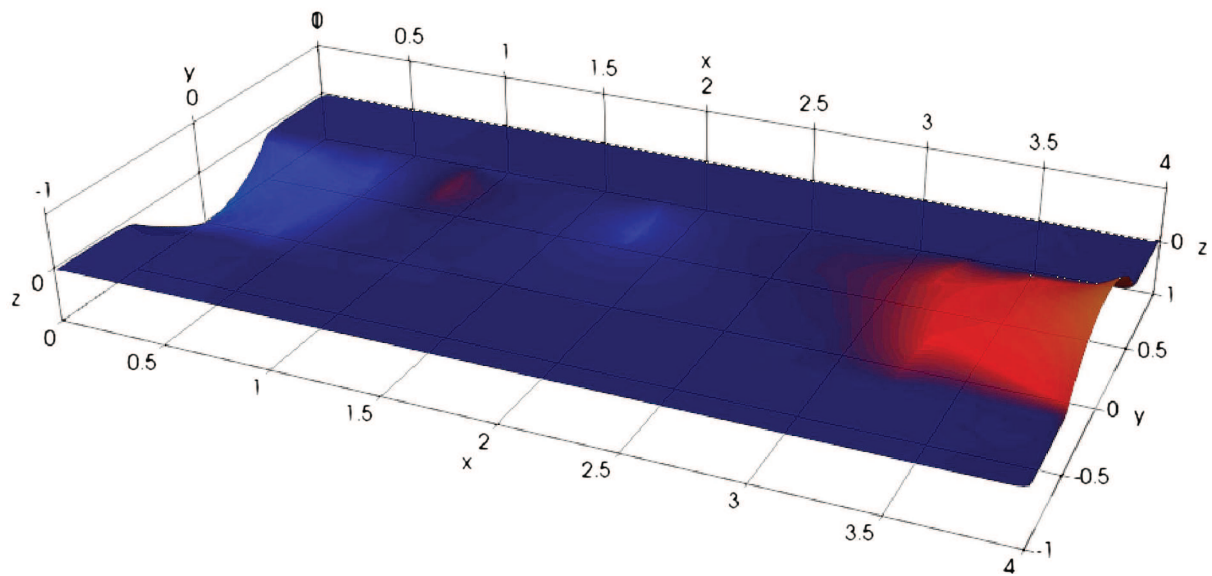
(B)

FIGURE 2. Pressure for the flow between two plates with $\varepsilon = 0.001$ and 3.6×10^4 dofs. (A) Pressure obtained with the Taylor–Hood $\mathbb{P}_2$–$\mathbb{P}_1$ element. (B) Pressure obtained with the stabilised $\mathbb{P}_1$–$\mathbb{P}_1$ element.
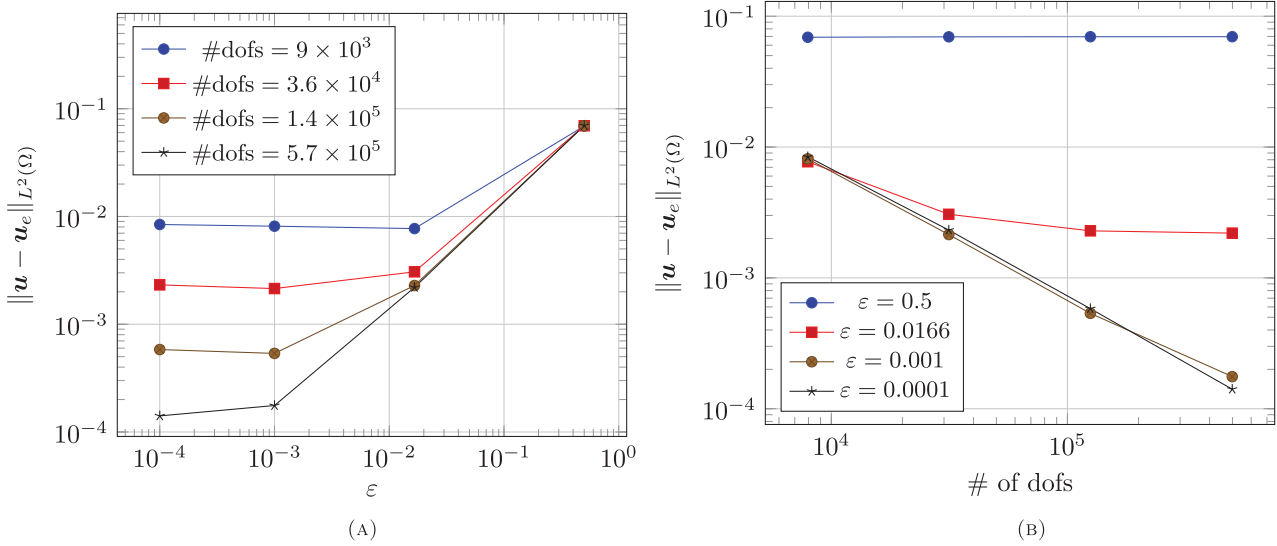
FIGURE 3. L^2 error of the velocity for the Bingham flow between two plates. (A) Velocity error as ε decreases. (B) Velocity error as the mesh size decreases.
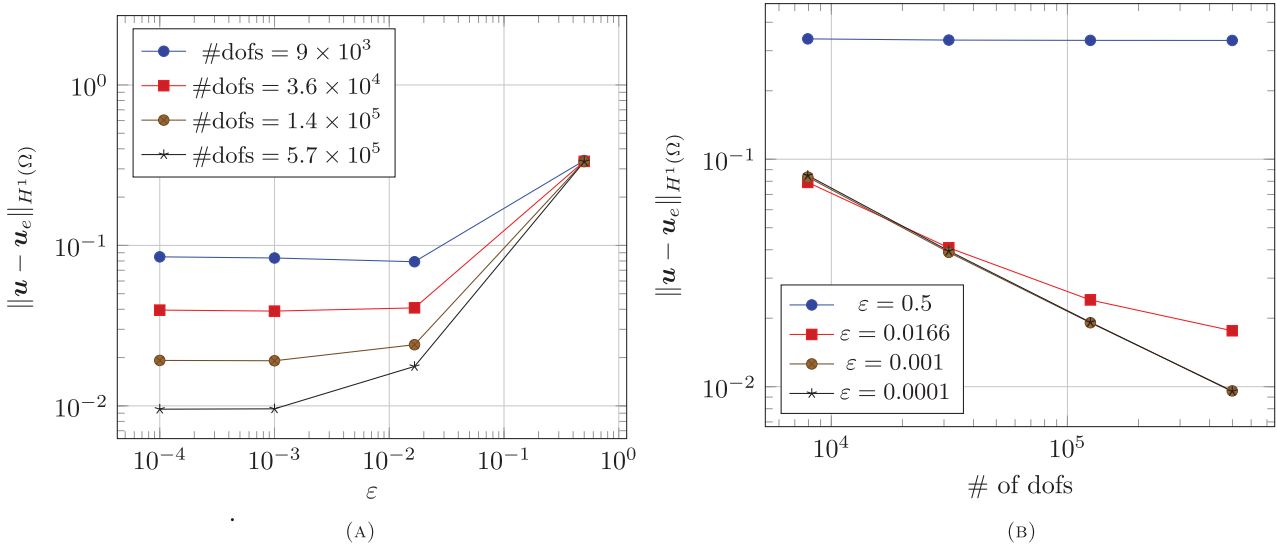

FIGURE 4. H^1 error of the velocity for the Bingham flow between two plates. (A) Velocity error as ε decreases. (B) Velocity error as the mesh size decreases.

For the sake of comparison, Figure 7 shows the decrease in the residual norm when using the alternative expression (1.4a) for the constitutive relation; recall that, as seen in Section 3.2, this expression does not define a semismooth function. To ensure convergence of the nonlinear iterations, in this case we had to employ a continuation scheme with respect to ε: given two previously computed solutions $\boldsymbol{z}_1, \boldsymbol{z}_2$, corresponding to the
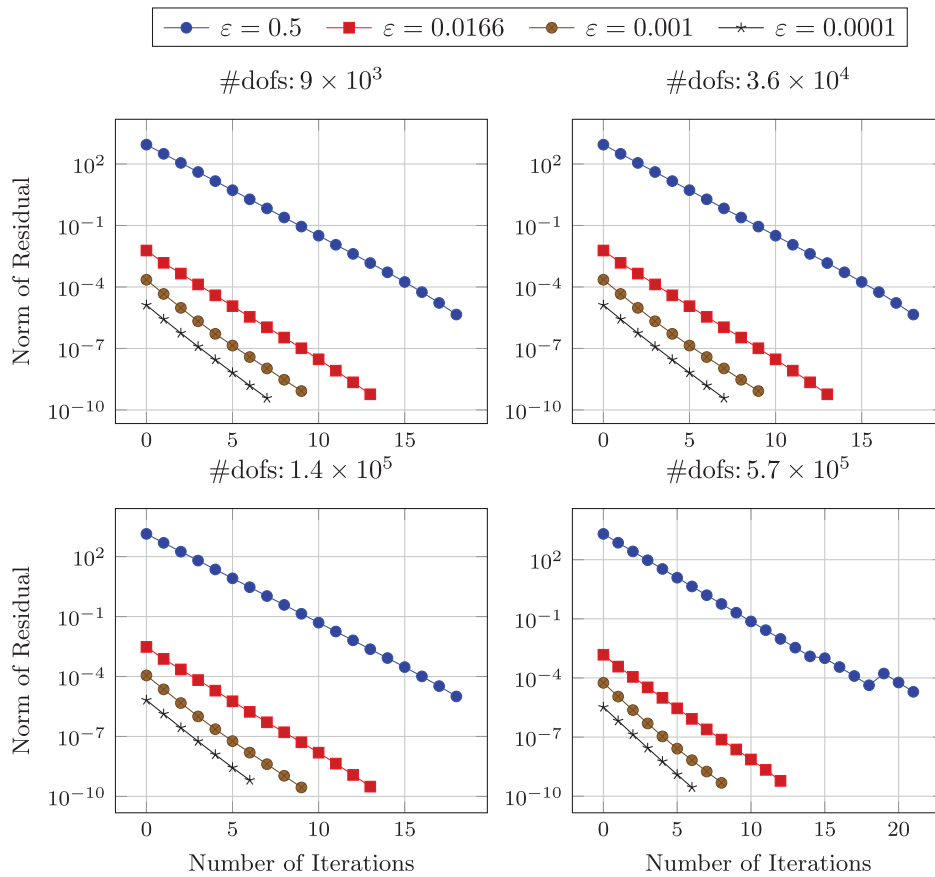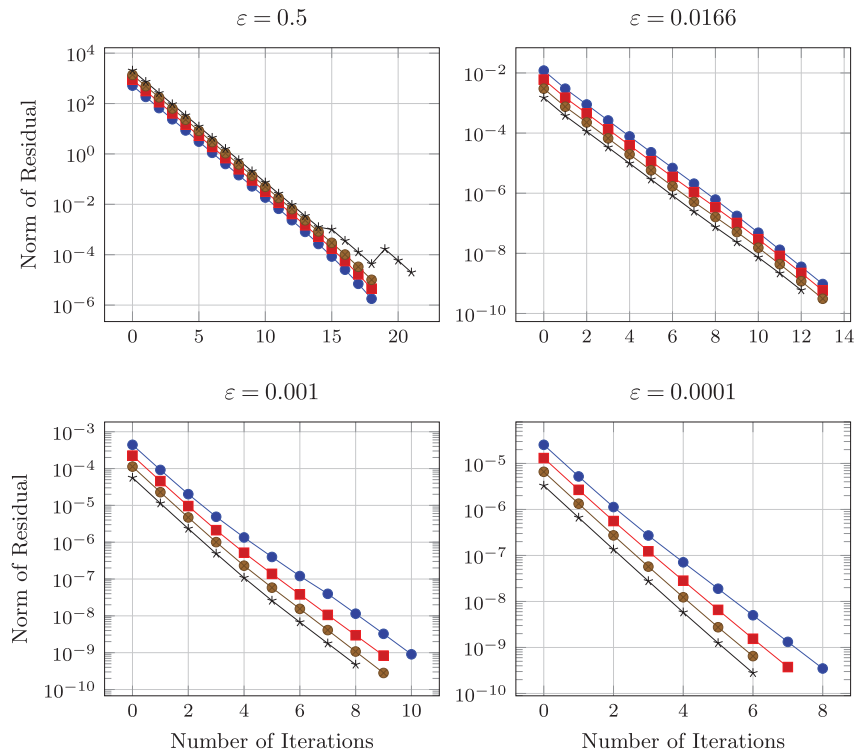
FIGURE 5. Residual norm *vs.* number of (semismooth) Newton iterations for the Bingham flow between two plates, for different mesh sizes.

values $\varepsilon_1, \varepsilon_2$, respectively, the initial guess for the semismooth Newton iteration for ε is chosen as

$$\frac{\varepsilon - \varepsilon_2}{\varepsilon_2 - \varepsilon_1}(\boldsymbol{z}_2 - \boldsymbol{z}_1) + \boldsymbol{z}_2. \tag{4.5}$$

Although convergence is achieved, the behaviour is somewhat more erratic in comparison to that obtained using (1.4b), and it takes a higher number of iterations and continuation steps for the solver to converge. This supports the results from Section 3.2, which suggest that the expression (1.4b) should be more advantageous.

4.2. Lid-driven cavity

The lid-driven cavity problem has been used as a test for many numerical schemes throughout the years (see *e.g.* [3, 45]). In this example we solve the unsteady problem on the domain defined by the square $\Omega = (0, 1)^2$; the following boundary conditions are imposed on the velocity field for all times $t \in [0, T]$:

$$\begin{aligned}
\partial\Omega_1 &:= (0, 1) \times \{1\}, & \partial\Omega_2 &:= \partial\Omega \setminus \partial\Omega_1, \\
\boldsymbol{u} &= \boldsymbol{0} & &\text{on } \partial\Omega_2, \\
\boldsymbol{u} &= (1, 0)^{\mathrm{T}} & &\text{on } \partial\Omega_1.
\end{aligned} \tag{4.6}$$

FIGURE 6. Residual norm *vs.* number of (semismooth) Newton iterations for the Bingham flow between two plates, for different values of ε.
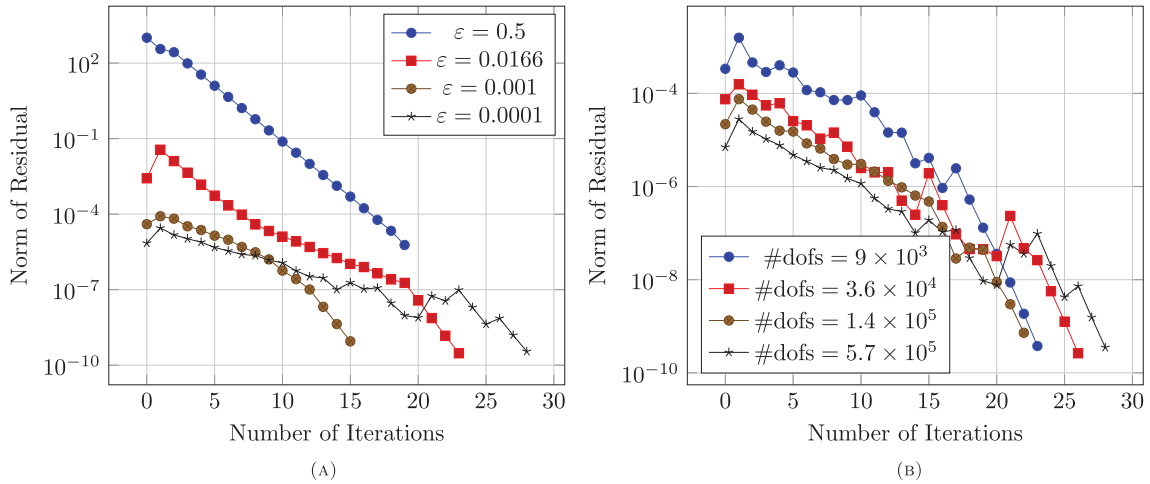
FIGURE 7. Residual norm *vs.* number of (semismooth) Newton iterations for the Bingham flow between two plates using (1.4a). (A) Decreasing regularisation parameter ε. (B) Decreasing mesh size h_n.
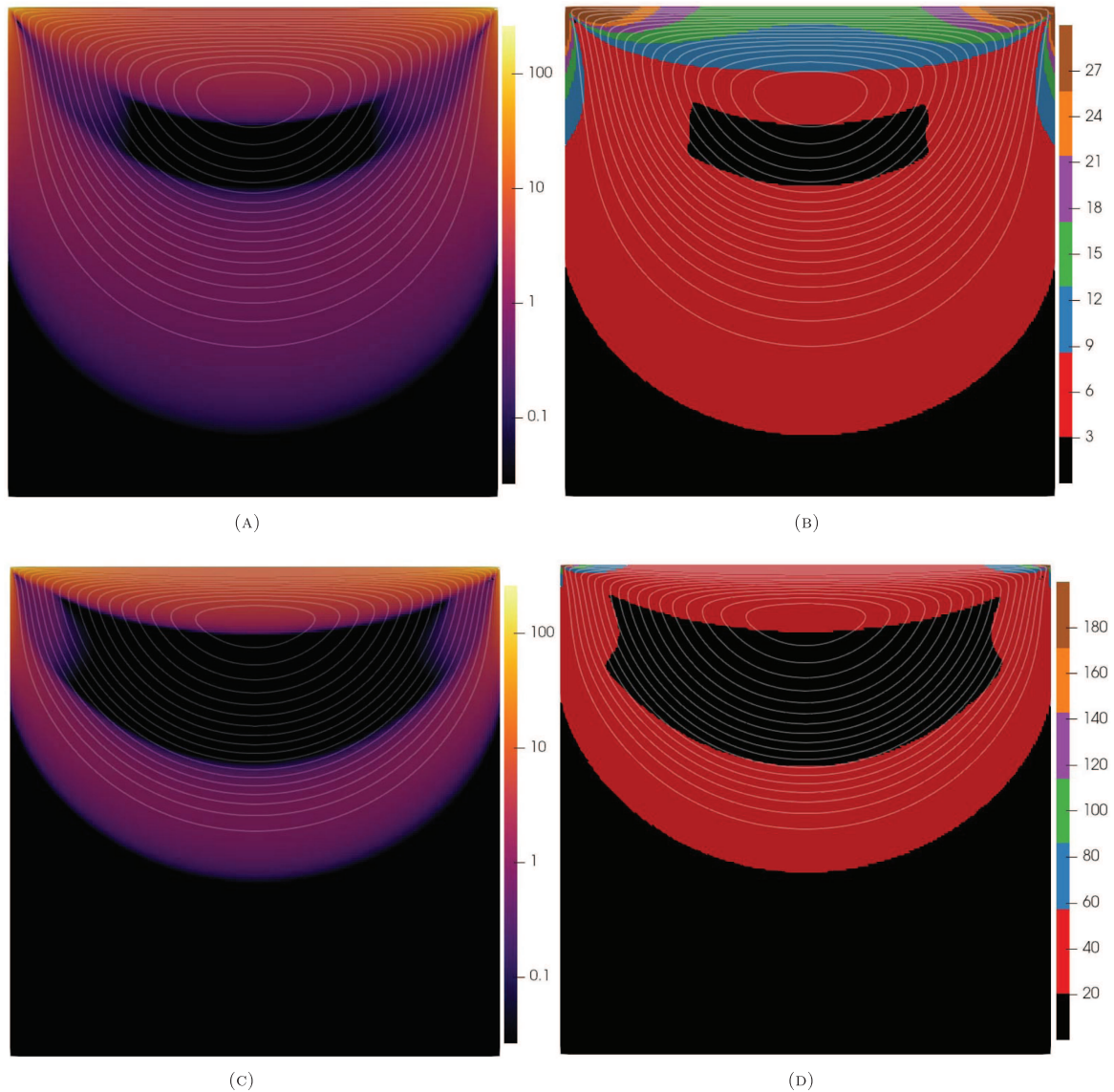
FIGURE 8. Magnitude of $\mathbf{D}(\boldsymbol{u})$ (*left*) and $\mathbf{S}$ (*right*) for the Bingham cavity with $\varepsilon = 10^{-4}$ and $\tau_* \in \{3, 20\}$. (A) $\tau_* = 3$; $\Delta t = 0.0005$. (B) $\tau_* = 3$; $\Delta t = 0.0005$. (C) $\tau_* = 20$; $\Delta t = 0.0001$. (D) $\tau_* = 20$; $\Delta t = 0.0001$.

The time derivative is discretised using implicit Euler's method with a time step Δt, and the goal is to compute the steady state of the system. In this example we use an absolute tolerance for the nonlinear iterations of 10^{-7}, and we deem the solution to have reached a steady state when the difference between two consecutive solutions $\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|_{L^2(\Omega)}$ is less than 10^{-6}.

Figure 8 shows comparative plots of the magnitude of the symmetric velocity gradient $\mathbf{D}(\boldsymbol{u})$ and the stress $\mathbf{S}$ of the steady state, for two different values of the yield stress, obtained using a discretisation with 4.5×10^5 degrees of freedom. The scheme yields a solution that exhibits the expected behaviour, *i.e.* a plug region appears in the upper half of the cavity, and it increases in size as τ_* increases. Table 1 shows, for several values of the

TABLE 1. Strength and vertical coordinate of the central vortex for the lid-driven cavity problem.

τ_*	0.5	3	5	10	20	40
$\psi_{\max}$	0.09472	0.07716	0.06916	0.05734	0.04554	0.03473
y_c	0.7768	0.8170	0.8348	0.8616	0.8884	0.9107
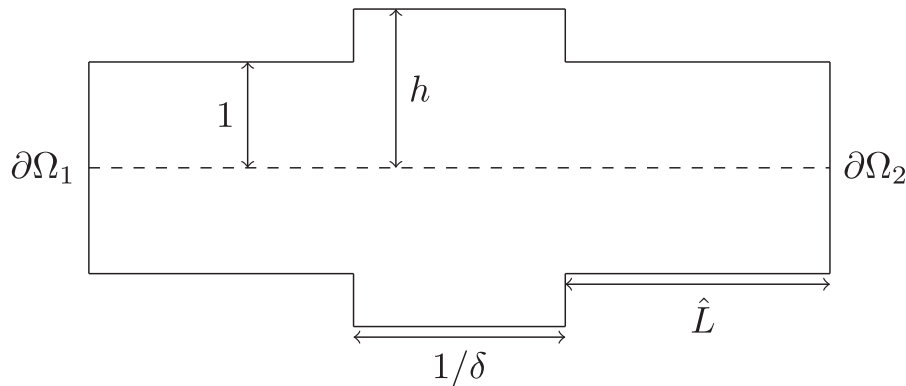
FIGURE 9. Geometry of the expansion-contraction problem.

yield stress τ_*, the maximum of the stream function $\psi_{\max}$, and the vertical coordinate of the center of the vortex that appears in the middle of the cavity. The obtained values are in good agreement with [45].

4.3. Expansion-contraction channel

In this section we consider the flow of a Bingham fluid on a channel whose width expands and then contracts; the geometry of the problem is depicted in Figure 9. The problem has already been non-dimensionalised, and in the current setting the Bingham number Bn takes the role of the yield stress; for details about the non-dimensionalisation procedure see [31], where this problem was solved employing the augmented Lagrangian method and a finite difference discretisation.

The boundary condition for the velocity at the inflow and outflow boundaries $\partial\Omega_1 \cup \partial\Omega_2$ is chosen to be a fully developed Poiseuille flow (*cf.* Sect. 4.1). On the rest of the boundary $\partial\Omega \setminus (\partial\Omega_1 \cup \partial\Omega_2)$ we impose a no-slip condition $\boldsymbol{u} = 0$. Figure 10 shows the magnitude of $\mathbf{D}(\boldsymbol{u})$ (using a logarithmic scale for the color map) for the solution obtained using a mesh with 330 819 degrees of freedom, and the values $\hat{L} = 4$, $\delta = \frac{1}{2}$, $h = 2$. Figure 11 then shows the variation of the pressure in the horizontal direction; the pressure exhibits a linear variation before and after the cavity, as expected from a Poiseuille profile.

As the height h gets closer to 1, the plug in the cavity breaks into two dead zones; Figure 12 shows two different solutions that exhibit this feature. Table 2 shows the (non-dimensionalised) length of the dead zone for several values of the Bingham number. Although the interfaces between yielded and unyielded regions is not as sharp as those obtained with the augmented Lagrangian method in [31], we find that the proposed algorithm is still capable of producing a behaviour consistent with [31], including *e.g.* the small plug regions that appear as the channel expands or contracts (*cf.* Fig. 12B).
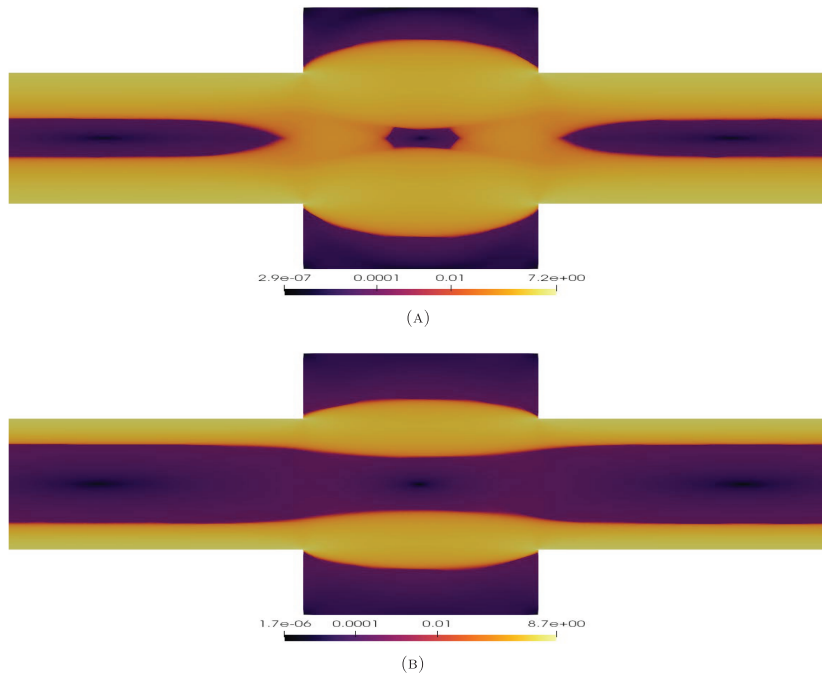
(A)

(B)

FIGURE 10. Magnitude of $|\mathbf{D}(\boldsymbol{u})|$ for the expansion-contraction problem with $\varepsilon = 10^{-4}$, $\hat{L} = 4$, $\delta = \frac{1}{2}$, $h = 2$. (A) Bn = 2. (B) Bn = 10.
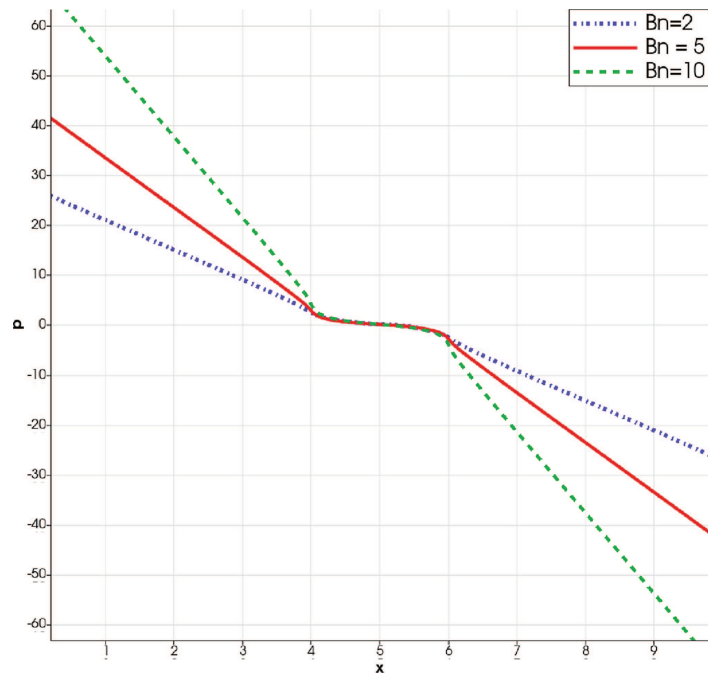

FIGURE 11. Pressure variation along a horizontal line with a height halfway between the boundary of the channel and the beginning of the Poiseuille plug.
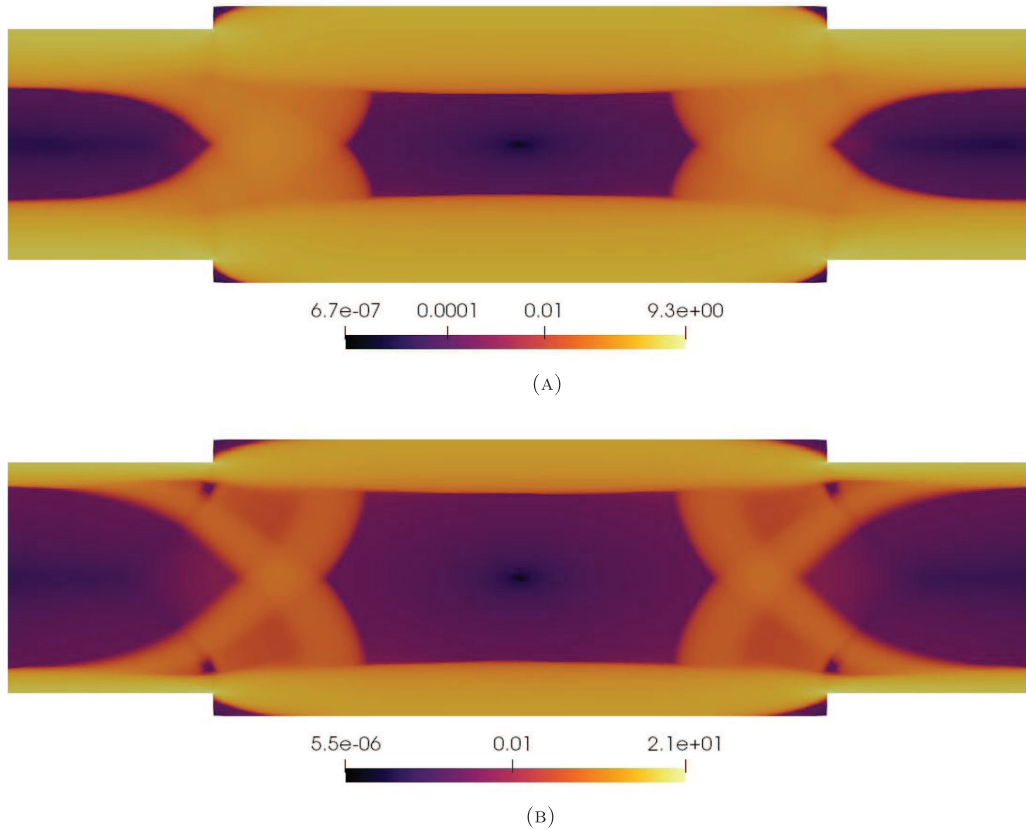
(A)

(B)

FIGURE 12. Magnitude of $|\mathbf{D}(\boldsymbol{u})|$ for the expansion-contraction problem with $\varepsilon = 10^{-4}$, $\hat{L} = 3$, $\delta = \frac{1}{5}$, $h = \frac{6}{5}$. (A) Bn = 5. (B) Bn = 50.

TABLE 2. Length of the dead zone for the expansion-contraction problem with $\hat{L} = 3$, $\delta = \frac{1}{5}$, $h = \frac{6}{5}$, for several values of the Bingham number.

Bn	2	5	10	20	30	40	50
L_d	0.0469	0.06453	0.07813	0.09688	0.10626	0.11876	0.12500

REFERENCES

[1] M.S. Alnaes, A. Logg, K.B. Ølgaard, M.E. Rognes and G.N. Wells, Unified form language: a domain-specific language for weak formulations of partial differential equations. *ACM Trans. Math. Softw.* **40** (2014) 1–37.

[2] P.R. Amestoy, I.S. Duff, J. Koster and J.Y. L'Excellent, A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.* **23** (2001) 15–41.

[3] A. Aposporidis, E. Haber, M.A. Olshanskii and A. Veneziani, A mixed formulation of the Bingham fluid flow problem: analysis and numerical solution. *Comput. Methods Appl. Mech. Eng.* **200** (2011) 2434–2446.

[4] N.J. Balmforth, I.A. Frigaard and G. Ovarlez, Yielding to stress: recent developments in viscoplastic fluid mechanics. *Ann. Rev. Fluid Mech.* **46** (2014) 121–146.

[5] S. Basu and U.S. Shivhare, Rheological, textural, microstructural, and sensory properties of sorbitol-substituted mango jam. *Food Bioprocess. Technol.* **6** (2013) 1401–1413.

[6] M. Bercovier and M. Engelman, A finite-element method for incompressible non-Newtonian flows. *J. Comput. Phys.* **36** (1980) 313–326.

[7] R.B. Bird, G.C. Dal and B.J. Yarusso, The rheology and flow of viscoplastic materials. *Rev. Chem. Eng.* **1** (1983) 1–70.

[8] 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** (2020) 1232–1289.

[9] D. Boffi, F. Brezzi and M. Fortin, Mixed Finite Element Methods and Applications. Springer (2013).

[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** (2009) 109–136.

[11] M. Bulíček, P. Gwiazda, J. Málek and A. Świerczewska-Gwiazda, On unsteady flows of implicitly constituted incompressible fluids. *SIAM J. Math. Anal.* **44** (2012) 2756–2801.

[12] M. Bulíček, J. Burczak and S. Schwarzacher, A unified theory for some non-Newtonian fluids under singular focing. *SIAM J. Math. Anal.* **48** (2016) 4241–4267.

[13] M. Bulíček, J. Málek and E. Maringová, On nonlinear problems of parabolic type with implicit constitutive equations involving flux. *Math. Models Methods Appl. Sci.* (2021) 1–52. DOI: 10.1142/S0218202521500457.

[14] X. Chen, Z. Nashed and L. Qi, Smoothing methods and semismooth methods for non-differentiable operator equations. *SIAM J. Numer. Anal.* **38** (2000) 1200–1216.

[15] F.H. Clarke, Optimization and Nonsmooth Analysis. Wiley, New York (1983).

[16] J.C. De los Reyes and S.G. Andrade, Numerical simulation of two-dimensional Bingham fluid flow by semismooth Newton methods. *J. Comput. Appl. Math.* **235** (2010) 11–32.

[17] J.C. De los Reyes and M. Hintermüller, A duality based semismooth Newton framework for solving variational inequalities of the second kind. *Interfaces Free Boundaries* **13** (2011) 437–462.

[18] L. Diening, C. Kreuzer and E. Süli, Finite element approximation of steady flows of incompressible fluids with implicit power-law-like rheology. *SIAM J. Numer. Anal.* **51** (2013) 984–1015.

[19] Y. Dimakopoulos, G. Makrigiorgos, G.C. Georgiou and J. Tsamopoulos, The PAL (Penalized Augmented Lagrangian) method for computing viscoplastic flows: a new fast converging scheme. *J. Non-Newt. Fluid Mech.* **256** (2018) 23–41.

[20] P.E. Farrell and P.A. Gazca-Orozco, Augmented Lagrangian preconditioner for implicitly-constituted non-Newtonian incompressible flow. *SIAM J. Sci. Comput.* **42** (2020) B1329–B1349.

[21] 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** (2020) 757–787.

[22] I.A. Frigaard and C. Nouar, On the usage of viscosity regularisation methods for viscoplastic fluid flow computation. *J. Non-Newt. Fluid Mech.* **127** (2005) 1–26.

[23] M. Fuchs and G. Seregin, Variational methods for fluids of Prandtl–Eyring type and plastic materials with logarithmic hardening. *Math. Methods Appl. Sci.* **22** (1999) 317–351.

[24] M. Fuchs and G. Seregin, Variational Methods for Problems from Plasticity Theory and for Generalized Newtonian Fluids. Lecture Notes in Math. 1749. Springer Verlag (2000).

[25] V. Girault and P.A. Raviart, Finite Element Methods for Navier–Stokes Equations: Theory and Algorithms. Springer Verlag (1986).

[26] R. Glowinski and A. Wachs, On the numerical simulation of viscoplastic fluid flow. In: Vol. XVI of Handbook of Numerical Analysis. *Numerical Methods for Non-Newtonian Fluids*. North Holland (2010) 483–718.

[27] H. Goldberg, W. Kampowsky and F. Tröltzsch, On Nemytskij operators in L_p-spaces of abstract functions. *Math. Nachr.* **155** (1992) 127–140.

[28] P.P. Grinevich and M.A. Olshanskii, An iterative method for the Stokes-type problem with variable viscosity. *SIAM J. Sci. Comput.* **31** (2009) 3959–3978.

[29] L.H. Luu and Y. Forterre, Drop impact of yield-stress fluids. *J. Fluid Mech.* **632** (2009) 301–327.

[30] 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** (2018) 152–178.

[31] A. Marly and P. Vigneaux, Augmented Lagrangian simulations study of yield-stress fluid flows in expansion-contraction and comparisons with physical experiments. *J. Non-Newtonian Fluid Mech.* **239** (2017) 35–52.

[32] R. Mifflin, Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.* **15** (1977) 959–972.

[33] A. Putz, I.A. Frigaard and D.M. Martínez, On the lubrication paradox and the use of regularisation methods for lubrication flows. *J. Non-Newt. Fluid Mech.* **163** (2009) 62–77.

[34] L.Q. Qi, Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.* **18** (1993) 227–244.

[35] L.Q. Qi and J. Sun, A nonsmooth version of Newton's method. *Math. Program.* **58** (1993) 353–367.

[36] J. Qin, *On the convergence of some low order mixed finite elements for incompressible fluids*. Ph.D. thesis, Pennsylvania State University (1994).

[37] K.R. Rajagopal, On implicit constitutive theories. *Appl. Math.* **48** (2003) 279–319.

[38] K.R. Rajagopal, On implicit constitutive theories for fluids. *J. Fluid Mech.* **550** (2006) 243–249.

[39] K.R. Rajagopal and A.R. Srinivasa, On the thermodynamics of fluids defined by implicit constitutive relations. *Z. Angew. Math. Phys.* **59** (2008) 715–729.

[40] 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** (2016) 1–27.

[41] P. Saramito, A damped Newton algorithm for computing viscoplastic fluid flows. *J. Non-Newt. Fluid Mech.* **238** (2016) 6–15.

[42] P. Saramito and A. Wachs, Progress in numerical simulation of yield stress fluid flows. *Rheol. Acta* **56** (2017) 211–230.

[43] A. Schiela, A simplified approach to semismooth Newton methods in function space. *SIAM J. Optim.* **19** (2008) 1417–1432.

[44] E. Süli and T. Tscherpel, Fully discrete finite element approximation of unsteady flows of implicitly constituted incompressible fluids. *IMA J. Numer. Anal.* **40** (2021) 801–849.

[45] A. Syrakos, G.C. Georgiou and A.N. Alexandrou, Solution of the square lid-driven cavity flow of a Bingham plastic using the finite volume method. *J. Non-Newtonian Fluid Mech.* **195** (2013) 19–31.

[46] T. Treskatis, A. Roustaei, I. Frigaard and A. Wachs, Practical guidelines for fast, efficient and robust simulations of yield-stress flows without regularisation: a study of accelerated proximal gradient and augmented Lagrangian methods. *J. Non-Newtonian Fluid Mech.* **262** (2018) 149–164.

[47] F. Tröltzsch, Optimal Control of Partial Differential Equations: Theory, Methods, and Applications. American Mathematical Soc. Vol. 112. (2010).

[48] T. Tscherpel, *FEM for the unsteady flow of implicitly constituted incompressible fluids*. Ph.D. thesis, University of Oxford (2018).

[49] M. Ulbrich, Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.* **13** (2003) 805–841.

[50] M. Ulbrich, Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. *MOS-SIAM Series on Optimization*. SIAM (2011).