SMAI-JCM
SMAI Journal of Computational Mathematics

A semi-exact primal-dual method for steady viscoplastic flows

François Bouchut & David Maltese

Volume 12 (2026), p. 135-170.
https://doi.org/10.5802/jcm.145

The SMAI Journal of Computational Mathematics is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/
Submissions at https://smai-jcm.centre-mersenne.org/ojs/submission
e-ISSN: 2426-8399

SMAI Journal of Computational Mathematics
Vol. 12, 135-170 (2026)

A semi-exact primal-dual method for steady viscoplastic flows

François Bouchut [1]
David Maltese [2]

[1] Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), CNRS, Univ. Gustave Eiffel, UPEC, F-77454, Marne-la-Vallée, France
E-mail address: francois.bouchut@univ-eiffel.fr
[2] Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), CNRS, Univ. Gustave Eiffel, UPEC, F-77454, Marne-la-Vallée, France
E-mail address: david.maltese@univ-eiffel.fr.

Abstract. Constitutive laws for viscoplastic materials involve a multivalued nonlinearity. Duality methods for such equations are known to converge, but the convergence is slow, requiring a large number of iterations. We introduce here a new iterative method of implicit primal-dual type for a class of variational problems, with a particular asymmetrical form in terms of the primal and dual unknowns, with an exact resolution of one relation and an approximate resolution of the other (semi-exact method). Two fast algorithms are proposed. The first includes an adaptive choice of a variable parameter, and the second includes a Newton-like linearized correction procedure. Applying the semi-exact method to a steady viscoplastic problem leads at each iteration to the resolution of a Laplace type equation, thus equivalent in terms of cost per iteration to the well-known Augmented Lagrangian or dual FISTA methods. Numerical tests show that the semi-exact method achieves a faster convergence in terms of number of iterations, compared to the augmented Lagrangian method or to the FISTA* method with acceleration, for Bingham or Herschel–Bulkley laws. This is particularly true when a large zero order term is present, as is the case when solving a time dependent problem.

2020 Mathematics Subject Classification. 35J20, 76A05, 65K15, 74S05.

Keywords. Viscoplastic materials, semi-exact method, implicit primal-dual algorithm, adaptive parameter, Newton linearized correction, steady flows.

1. Introduction

Basic viscoplastic models can be written

$$\alpha u - \operatorname{div}\sigma = f \quad \text{in } \Omega, \tag{1.1}$$

$$\sigma \in \partial F(Du) \quad \text{in } \Omega, \tag{1.2}$$

where Ω is a bounded open subset of $\mathbb{R}^N$, $\alpha \geq 0$ is a constant, the unknown u has values in $\mathbb{R}^N$, $Du := (\nabla u + (\nabla u)^t)/2$, the dual unknown σ has values in $\mathcal{S}_N$ the space of square symmetric matrices of size $N \times N$, and $f \in L^2(\Omega)$. The problem must be completed with boundary conditions, we shall take here homogeneous Dirichlet conditions

$$u = 0 \quad \text{on } \partial\Omega. \tag{1.3}$$

The nonlinearity $F : \mathcal{S}_N \to (-\infty, \infty]$ is a proper (i.e. not identically $+\infty$) convex lower semi-continuous function. In (1.2) the notation ∂F stands for the Moreau subdifferential of F, that is for $\gamma \in \mathcal{S}_N$

$$\partial F(\gamma) = \Big\{\sigma \in \mathcal{S}_N; F(\overline{\gamma}) \geq F(\gamma) + \sigma : (\overline{\gamma} - \gamma) \text{ for all } \overline{\gamma} \in \mathcal{S}_N\Big\}, \tag{1.4}$$

where we use the Frobenius scalar product $\gamma : \overline{\gamma} = \sum_{ij} \gamma_{ij}\overline{\gamma}_{ij}$ and the associated norm $|\gamma| = (\gamma : \gamma)^{1/2}$, for $\gamma, \overline{\gamma} \in \mathcal{S}_N$. The relation (1.2) has to be understood in a pointwise sense, which means that it must

https://doi.org/10.5802/jcm.145

hold for a.e. $x \in \Omega$. We consider a nonlinearity F that is not everywhere differentiable, this being fundamental to describe physically relevant materials. In particular, it is expected that $\partial F(0)$ is a (convex) nonempty set that is not reduced to a single point.

The problem (1.1)–(1.3) is the simplest to describe non-Newtonian materials. Indeed $\alpha = 0$ corresponds to the physically steady case, whereas $\alpha > 0$ is involved to the transient problem

$$\partial_t u - \operatorname{div} \sigma = f \quad \text{in } \Omega, \qquad \sigma \in \partial F(Du) \quad \text{in } \Omega, \tag{1.5}$$

since an implicit time discretization leads to

$$\frac{u^{n+1} - u^n}{\delta t} - \operatorname{div} \sigma^{n+1} = f \quad \text{in } \Omega, \qquad \sigma^{n+1} \in \partial F(Du^{n+1}) \quad \text{in } \Omega, \tag{1.6}$$

thus reducing to (1.1)–(1.3) for the unknown u^{n+1}, with $\alpha = 1/\delta t > 0$ and a right-hand side $f + \frac{u^n}{\delta t}$.

Therefore (1.1)–(1.3) is of interest in the two following cases: $\alpha = 0$ for the physically steady case, or $\alpha = 1/\delta t$ is large (large zero order term) for the transient case.

Solutions to (1.1)–(1.3) are known to exist, when Ω is sufficiently smooth and F has some regularity and growth properties, see for example [14, 18, 19, 21, 23]. If $\alpha = 0$, F must also satisfy some coercivity property. For the time dependent problem (1.5) we refer to [9].

Many numerical methods of iterative type have been proposed for the problem (1.1)–(1.3), we refer in particular to the review paper [26]. The regularization method is popular, it consists in replacing the nonlinearity F by a smoothed one F_ϵ and then using a quite standard iteration. It runs quite fast, but often does not resolve well the solid zones where $Du = 0$. Duality methods do this correctly, they are proved to converge, but a lot of iterations can be necessary. Among them we can cite explicit methods where there is no linear system to solve; they are popular in the image processing community [3, 11, 12, 13]. Implicit duality methods are popular in the viscoplastic community, they are robust and efficient, the cost necessary for the resolution of a linear system being compensated by the smaller number of iterations. Among these methods, the most well-known is the augmented Lagrangian method [24]. A related one is the Bermudez–Moreno method [6, 16], it can be seen as the Arrow–Hurwicz explicit method [2] in which a part $\omega \operatorname{div} Du$ has been implicited. The dual FISTA method [27, 28] has been developed more recently, as well as the interior point method [1, 5]. Duality methods generally have a rate of convergence $\|u - u_{exact}\|_{L^2} \leq C/k$, where k is the number of iterations, see [4, 13] in the explicit case, and [20, 27] for applications to viscoplastic flows. The Newton method with damping has been implemented with faster convergence [25], but the linear problem to be solved is quite stiff and needs particularly efficient iterative algorithms.

Primal-dual methods formulate the problem with two unknowns, u and σ in our case, related to the two equations (1.1) and (1.2). These two equations are resolved approximately, and somehow symmetrically. In this paper we introduce a new iterative primal-dual method with an asymmetrical form in terms of the primal and dual unknowns, with an exact resolution of one relation and an approximate resolution of the other (semi-exact method). A linear system has to be solved at each iteration (implicit method). We prove the stability of the method under some conditions on the parameters, one main parameter r_k being possibly variable during the iteration process. In the previously used methods with parameter [3, 11, 27] the variable parameter is defined by an a priori rule and tends to infinity. Here we introduce an algorithm with adaptive choice of the main parameter r_k, that allows r_k to be bounded or not, depending on the computed errors. A second algorithm includes a Newton-like linearized correction procedure.

We apply our method to the viscoplastic problem (1.1)–(1.3), leading to the resolution of a single Laplace type problem at each iteration, similarly to the Augmented Lagrangian or FISTA* methods. Our numerical experiments show competitive convergence rates for our two semi-exact algorithms in comparison to the augmented Lagrangian method or to the FISTA* method with acceleration, on known test cases for a Bingham or Herschel–Bulkley nonlinearity. Moreover we show that the semi-exact method performs particularly well in the regime when the parameter α is large. The accuracy of the method is related to contraction properties in the linear regime, as stated in Proposition 2.2.

2. Primal-dual semi-exact iteration algorithm

Our semi-exact method can be formulated in abstract Hilbert spaces, see Appendix A, but we consider here the more concrete case of (1.1)–(1.3). We classically assume that the nonlinearity $F:\mathcal{S}_N\to(-\infty,\infty]$ is a proper (i.e. not identically $+\infty$) convex lower semi-continuous function. It is useful to assume moreover that $F''\geq\eta_c$ for some $\eta_c\geq 0$ (coercivity when $\eta_c>0$). The rigorous formulation of this is indeed that

$$\gamma\mapsto F(\gamma)-\eta_c\frac{|\gamma|^2}{2}\quad\text{is convex on }\mathcal{S}_N,\quad\text{for some }\eta_c\geq 0,\tag{2.1}$$

where as mentioned above $|\gamma|$ denotes the Frobenius norm of γ. We shall use several times that for $\overline{F}$ a convex proper l.s.c. function, one has for $\gamma_1,\gamma_2\in\mathcal{S}_N$, $\sigma_1\in\partial\overline{F}(\gamma_1)$, $\sigma_2\in\partial\overline{F}(\gamma_2)$ the monotonicity property $(\sigma_2-\sigma_1):(\gamma_2-\gamma_1)\geq 0$, that follows from the definition (1.4). For (1.1)–(1.3) to be well-posed it is necessary to assume

$$\eta_c>0\quad\text{or}\quad\alpha>0.\tag{2.2}$$

This assumption implies classically that if there is a solution (u,σ) to (1.1)–(1.3), then u is unique (but not σ).

2.1. Space continuous formulation

We consider the following iterative algorithm for the problem (1.1)–(1.3). The approximation is made through a couple of primal-dual unknowns u_k, σ_k, for $k=0,1,\dots$. We have first to define u_0 and σ_0. Then once (u_k,σ_k) is known (for $k=0,1,\dots$) the update is performed by resolving the system in the unknowns (u_{k+1},σ_{k+1})

$$\begin{cases}\alpha u_{k+1}-\operatorname{div}\sigma_{k+1}-f=0, & (2.3)\\ \sigma_{k+1}-\sigma_k=-(\overline{\eta}-\eta_c)(Du_{k+1}-Du_k)+r_k\Big(Du_{k+1}-S_{r_k-\eta_c}\big(\sigma_k+(r_k-\eta_c)Du_k\big)\Big), & (2.4)\end{cases}$$

where we have the parameters

$$\overline{\eta}\geq 0\quad\text{and}\quad r_k>0,\tag{2.5}$$

the latter eventually depending on k in a way that will be precised later on. The nonlinear function S_t for $t>-\eta_c$ and $\sigma\in\mathcal{S}_N$ is defined by

$$S_t(\sigma)=(\partial F+t\,\mathrm{Id})^{-1}(\sigma),\tag{2.6}$$

or in other words

$$S_t(\sigma)=\gamma\quad\text{iff}\quad\partial F(\gamma)+t\gamma\ni\sigma.\tag{2.7}$$

Because of the assumption (2.1), for any $\sigma\in\mathcal{S}_N$ there is a unique $\gamma\in\mathcal{S}_N$ solution to (2.7), hence S_t is well-defined, for $t>-\eta_c$. For this property and other essential features on convex functions one can refer to [14]. In particular, (2.1) implies that $S_{r_k-\eta_c}$ is Lipschitz continuous on $\mathcal{S}_N$ (see Lemma 2.8),

$$|S_{r_k-\eta_c}(\sigma_2)-S_{r_k-\eta_c}(\sigma_1)|\leq|\sigma_2-\sigma_1|/r_k\quad\text{for all }\sigma_1,\sigma_2\in\mathcal{S}_N.\tag{2.8}$$

Notice that in (2.4) the nonlinear function $S_{r_k-\eta_c}$ acts pointwise on $\sigma_k+(r_k-\eta_c)Du_k$.

The system (2.3), (2.4) can be solved by inserting the value of σ_{k+1} given by the second equation into the first one, giving

$$\alpha u_{k+1}-(r_k-\overline{\eta}+\eta_c)\operatorname{div}Du_{k+1}=f+\operatorname{div}\Big(\sigma_k+(\overline{\eta}-\eta_c)Du_k-r_kS_{r_k-\eta_c}\big(\sigma_k+(r_k-\eta_c)Du_k\big)\Big).\tag{2.9}$$

Together with the boundary condition (1.3), this Laplace type equation determines u_{k+1}, as long as $r_k - \overline{\eta} + \eta_c \geq 0$ and, in case of equality, $\alpha > 0$. Knowing (2.2) we shall assume the sufficient condition

$$r_k \geq \overline{\eta}. \tag{2.10}$$

Then once u_{k+1} is known, σ_{k+1} can be computed directly by (2.4).

Properties. We can state several properties of the algorithm (2.3), (2.4).

(1) Because of (2.7), the update (2.3), (2.4) has the property that it leaves the data invariant (i.e. $u_{k+1} = u_k$ and $\sigma_{k+1} = \sigma_k$) if and only if u_k, σ_k is a solution to (1.1), (1.2).

(2) Strikingly and in contrast with other primal-dual methods, the equation (1.1) is resolved exactly in (2.3), whereas as usual (1.2) is solved approximately, by (2.4).

Remark 2.1. The semi-exact method under the form (2.9), (2.4) is algebraically quite similar to the FISTA* method described in [27], see Appendix D. However in our method we do not put backward dependency ("acceleration"). In the semi-exact method the parametrized operator $S_{r_k-\eta_c}$ is involved, instead of just S_0 in FISTA*. The parameters in (2.4) are different than in FISTA*, and allow for an arbitrarily large viscosity coefficient $r_k - \overline{\eta} + \eta_c \geq \eta_c$ in (2.9), in contrast with a viscosity bounded by η_c in FISTA*. In the case $r_k = \overline{\eta} = \eta_c$, the semi-exact method indeed coincides with the FISTA* method without "acceleration" (i.e. with $t_k \equiv 1$). The case $r_k = \overline{\eta}$ is indeed related to the Arrow–Hurwicz algorithm, see Appendix C.

An important motivation for our algorithm is that it has particular accuracy properties. We shall say here that a sequence a_k converges linearly to a with rate $q \in (0,1)$ if $a_k - a = O(q^k)$.

Proposition 2.2 (Convergence rate for a purely viscous law)**.** *For a purely viscous linear law where $F(\gamma) = \eta|\gamma|^2/2$ with $\eta \geq \eta_c \geq 0$, one has*

(i) *In the case $\alpha = 0$, $\eta_c > 0$, for the choice $r_k = \overline{\eta} = \eta_c$, the algorithm* (2.3), (2.4) *converges in two steps: $u_2 = u_{exact}$, whatever are the initial data u_0, σ_0. The constraint σ converges linearly with rate $1 - \eta_c/\eta < 1$.*

(ii) *For arbitrary α, for any $\overline{\eta} \geq 0$ and for the choice $r_k = \eta + \overline{\eta} - \eta_c$ (assumed positive), the algorithm* (2.3), (2.4) *converges linearly with rate $1/2$ in both u and σ.*

Proof. For property (i), with this F and since $\alpha = 0$, the exact solution u_{exact} solves

$$-\operatorname{div}(\eta D u_{exact}) = f, \tag{2.11}$$

with the boundary condition (1.3). One has $S_t(\sigma) = \sigma/(\eta + t)$, thus since $r = \eta_c$,

$$S_{r-\eta_c}(\sigma) = \frac{\sigma}{\eta}. \tag{2.12}$$

Taking into account that $\overline{\eta} = \eta_c$, the equation (2.9) gives

$$-\eta_c \operatorname{div} D u_{k+1} = f + \operatorname{div}\Big(\sigma_k - \frac{\eta_c}{\eta}\sigma_k\Big). \tag{2.13}$$

For $k \geq 1$, since (2.3) has been resolved exactly at the previous step, one has $-\operatorname{div}\sigma_k = f$, thus (2.13) reduces to

$$-\eta_c \operatorname{div} D u_{k+1} = \frac{\eta_c}{\eta} f. \tag{2.14}$$

Comparing to (2.11) this means that $u_{k+1} = u_{exact}$ for $k \geq 1$, as claimed. Moreover we can write (2.4) as

$$\sigma_{k+1} - \sigma_k = \eta_c\Big(D u_{k+1} - \frac{\sigma_k}{\eta}\Big), \tag{2.15}$$

and using that $\sigma_{exact} = \eta D u_{exact} = \eta D u_{k+1}$ for $k \geq 1$, we can express then σ_{k+1} as

$$\sigma_{k+1} = (1 - \frac{\eta_c}{\eta})\sigma_k + \frac{\eta_c}{\eta}\sigma_{exact}. \tag{2.16}$$

This proves that the zero divergence part of σ_k converges linearly to that of σ_{exact} with a rate $1 - \eta_c/\eta < 1$.

For property (ii), the exact solution u_{exact} solves

$$\alpha u_{exact} - \operatorname{div}(\eta D u_{exact}) = f, \tag{2.17}$$

with the boundary condition (1.3), and $\sigma_{exact} = \eta D u_{exact}$. One has still $S_t(\sigma) = \sigma/(\eta + t)$, thus (2.3), (2.4) writes

$$\begin{cases} \alpha u_{k+1} - \operatorname{div} \sigma_{k+1} - f = 0, & (2.18) \\ \sigma_{k+1} - \sigma_k = -(\overline{\eta} - \eta_c)(Du_{k+1} - Du_k) + rDu_{k+1} - \dfrac{r}{\eta + r - \eta_c}(\sigma_k + (r - \eta_c)Du_k). & (2.19) \end{cases}$$

Taking into account the value $r = \eta + \overline{\eta} - \eta_c$, (2.19) gives

$$\sigma_{k+1} = \eta D u_{k+1} + \frac{\eta - \eta_c}{2(\eta - \eta_c) + \overline{\eta}}(\sigma_k - \eta D u_k). \tag{2.20}$$

Let us denote by

$$\lambda^* = \frac{\eta - \eta_c}{2(\eta - \eta_c) + \overline{\eta}}. \tag{2.21}$$

Plugging (2.20) into (2.18) and using that for $k \geq 1$ $\alpha u_k - \operatorname{div} \sigma_k - f = 0$ yields

$$\alpha u_{k+1} - \operatorname{div}(\eta D u_{k+1}) - f = \lambda^* \operatorname{div}(\sigma_k - \eta D u_k) = \lambda^* \big(\alpha u_k - \operatorname{div}(\eta D u_k) - f\big). \tag{2.22}$$

This proves that u_k converges linearly to u_{exact} with a ratio λ^*. Since by (2.20) $\sigma_k - \eta D u_k$ tends to zero with the same rate, we conclude that both u_k and σ_k tend to their exact value at rate λ^*. Since $0 \leq \lambda^* \leq 1/2$ this proves the claim. ∎

Remark 2.3. The rate λ^* in (2.21) tends to 0 as $\overline{\eta} \to \infty$ (or equivalently $r \to \infty$ since we impose $r = \eta + \overline{\eta} - \eta_c$), giving superlinear convergence. However, as we shall see in Theorem 2.6, nonlinear stability requires that $\overline{\eta}$ remains bounded, see in particular (2.33).

Remark 2.4. Taking $\eta_c = \eta$ in Proposition 2.2 (ii) gives a vanishing rate. Indeed according to (2.18) and (2.20), in this case the method converges in one step for both u and σ. Similarly, taking $\eta_c = \eta$ in Proposition 2.2 (i) gives convergence in two steps for both u and σ.

Remark 2.5. In the case of a general nonlinearity F, it is intuitive to think of it in a first approximation as a space dependent viscosity $\eta(x)$. Then η_c has to be taken $\eta_c = \inf_x \eta(x)$, thus the equality $\eta_c = \eta$ cannot be reached, in particular in plug regions where $\eta(x) = \infty$. The case (i) of Proposition 2.2 gives therefore a slow rate on σ in plug regions. The slow rate on σ contaminates to some extent the rate on u by nonlinear coupling, depending on the test case. The case (ii) always gives a rate 1/2, which is better. But the relation $r_k = \eta + \overline{\eta} - \eta_c$ is not reachable for a viscosity $\eta(x)$ since it would need a coefficient $r(x)$. Our adaptive algorithm SEAR defined in Section 2.4 takes for r_k a mean value in space consistent with this relation. It is even possible to go further in accuracy and replace the viscosity $r_k - \overline{\eta} + \eta_c$ in (2.9) by F'', as we propose for our algorithm SELC in Section 3.

2.2. **P1/P0 finite element approximation**

We consider now a finite element approximation of the problem (1.1)–(1.3) in a P1/P0 setting.

Consider a conforming mesh $\mathcal{T}_h$ of the open bounded domain $\Omega \subset \mathbb{R}^N$, of size h: $\mathcal{T}_h$ is a finite set of disjoint open simplices such that $\bigcup_{K \in \mathcal{T}_h} \overline{K} = \overline{\Omega}$, and h is the maximum value of the diameter of all $K \in \mathcal{T}_h$. The mesh is conforming in the sense that for two distinct elements K, L of $\mathcal{T}_h$, $\overline{K} \cap \overline{L}$ is either empty or a simplex included in an affine subset with dimension strictly lower than N, whose vertices are simultaneously vertices of K and L.

We define

$$V_h = \Big\{u \in C(\overline{\Omega}, \mathbb{R}^N) \ ;\ u_{|K} \text{ is affine } \ \forall\, K \in \mathcal{T}_h,\ u_{|\partial\Omega} = 0\Big\}, \tag{2.23}$$

$$M_h = \Big\{\sigma \in L^\infty(\Omega, \mathcal{S}_N) \ ;\ \sigma_{|K} \text{ is constant } \ \forall\, K \in \mathcal{T}_h\Big\}. \tag{2.24}$$

With these definitions, for $u \in V_h$ one has $Du \in M_h$. Therefore for $u \in V_h$ and $\sigma \in M_h$ the equations (1.2) or (2.4) make sense in M_h since the nonlinearity S_t leaves M_h invariant. However (1.1) with the boundary condition (1.3) has to be replaced, for an exact discrete solution $(\widehat{u}, \widehat{\sigma}) \in V_h \times M_h$, by

$$\begin{aligned} &\widehat{u} \in V_h, \\ &\alpha\, \langle \widehat{u}, v\rangle + \langle \widehat{\sigma}, Dv\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h, \end{aligned} \tag{2.25}$$

where $\langle , \rangle$ denotes the scalar product in $L^2(\Omega)$, either of two functions with vector values in $\mathbb{R}^N$ or of two functions with matrix values in $\mathcal{S}_N$. The equation (1.2) is unchanged, i.e.

$$\widehat{\sigma} \in \partial F(D\widehat{u}). \tag{2.26}$$

Our notation here is that the hat stands for “exact solution” (of the discrete problem). The problem (2.25), (2.26) can be proved to have a solution under weak assumptions, see Appendix B. Our discrete semi-exact algorithm is obtained by replacing the iterative equation (2.3) (with boundary condition (1.3)) by a variational formulation over V_h and keeping (2.4) unchanged, which gives

$$\begin{cases} u_{k+1} \in V_h, \\ \alpha\, \langle u_{k+1}, v\rangle + \langle \sigma_{k+1}, Dv\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h, & (2.27)\\ \sigma_{k+1} - \sigma_k = -(\overline{\eta} - \eta_c)(Du_{k+1} - Du_k) + r_k\Big(Du_{k+1} - S_{r_k - \eta_c}(\sigma_k + (r_k - \eta_c)Du_k)\Big). & (2.28) \end{cases}$$

As in the continuous case, inserting the value of σ_{k+1} given by (2.28) into (2.27) we get

$$\begin{aligned} &u_{k+1} \in V_h, \\ &\alpha\, \langle u_{k+1}, v\rangle + (r_k - \overline{\eta} + \eta_c)\, \langle Du_{k+1}, Dv\rangle \\ &\qquad + \langle \sigma_k + (\overline{\eta} - \eta_c)Du_k - r_k S_{r_k - \eta_c}(\sigma_k + (r_k - \eta_c)Du_k), Dv\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h. \end{aligned} \tag{2.29}$$

As long as (2.10) holds, this system (2.29) has a unique solution u_{k+1}. Once u_{k+1} is known, σ_{k+1} is computed by (2.28). The case $r_k - \overline{\eta} + \eta_c = 0$ is possible (if $r_k = \overline{\eta}$, $\eta_c = 0$, $\alpha > 0$), it leads to an explicit method if one uses mass lumping. As in the continuous case, the equation (2.27) is an exact resolution of (2.25). In particular, since for $k \geq 1$ the couple (u_k, σ_k) satisfies (2.27), making the difference with (2.29) we obtain the equivalent update formula

$$\begin{aligned} &u_{k+1} \in V_h, \\ &\alpha\, \langle u_{k+1} - u_k, v\rangle + (r_k - \overline{\eta} + \eta_c)\, \langle Du_{k+1} - Du_k, Dv\rangle \\ &\qquad + r_k\, \langle Du_k - S_{r_k - \eta_c}(\sigma_k + (r_k - \eta_c)Du_k), Dv\rangle = 0, \quad \forall\, v \in V_h. \end{aligned} \tag{2.30}$$

The second line represents here the consistency error with respect to (2.26).

2.3. **Stability and convergence**

Since the discrete solution $(\widehat{u}, \widehat{\sigma}) \in V_h \times M_h$ to (2.25), (2.26) is known to converge to the continuous solution to (1.1)–(1.3) as the mesh size tends to 0 under some assumptions on F, see [8, 19], we consider here only the issue of stability and convergence of the discrete iterative algorithm to the discrete solution.

We introduce a constant $\lambda_{max} > 0$ such that

$$\|Dv\|^2 \leq \lambda_{max}\|v\|^2, \quad \forall\, v \in V_h, \tag{2.31}$$

where $\|\cdot\|$ denotes the L^2 norm over Ω. We have also the Poincaré-Korn inequality

$$\|Dv\|^2 \geq \lambda_{min}\|v\|^2, \quad \forall\, v \in V_h, \tag{2.32}$$

for some $\lambda_{min} > 0$. The constant λ_{min} can be taken independent of the mesh, but λ_{max} is mesh dependent, with a generic blow up as $\lambda_{max} \sim C/h^2$ for a quasi-uniform mesh.

Theorem 2.6 (Stability and convergence of the P1/P0 semi-exact method)**.** *We assume that F satisfies the coercivity assumptions* (2.1), (2.2), *and suppose that there is a solution $(\widehat{u},\widehat{\sigma}) \in V_h \times M_h$ to* (2.25), (2.26)*. If the parameters $\overline{\eta}$, r_k satisfy* (2.5), (2.10) *and for any $k = 0, 1, \ldots$*

$$r_k \geq \frac{\overline{\eta}^2}{2(\eta_c + \alpha/\lambda_{max})}, \tag{2.33}$$

$$r_{k+1} \leq \overline{\eta} + \sqrt{r_k^2 - 2r_k\Big(\overline{\eta} - \eta_c - \frac{\alpha}{\lambda_{max}}\Big)}, \tag{2.34}$$

and either $r_k \to \infty$ or

$$\liminf_{k\to\infty}\bigg(r_k^2 - 2r_k\Big(\overline{\eta} - \eta_c - \frac{\alpha}{\lambda_{max}}\Big) - (r_{k+1} - \overline{\eta})^2\bigg) > 0, \tag{2.35}$$

then the sequence of approximate solutions $(u_k, \sigma_k)_{k=0,\ldots} \in V_h \times M_h$ defined by (2.27), (2.28) *satisfies $u_k \to \widehat{u}$ as $k \to \infty$, and σ_k is bounded in L^2.*

If moreover $\liminf r_k > 0$, then for any subsequence $k = k(p)$ such that $\sigma_{k(p)} \to \overline{\sigma} \in M_h$ as $p \to \infty$, one has that $(\widehat{u}, \overline{\sigma})$ is a solution to (2.25), (2.26)*.*

Remark 2.7. We notice that because of (2.33) one has $r_k^2 - 2r_k(\overline{\eta} - \eta_c - \alpha/\lambda_{max}) = (r_k - \overline{\eta})^2 + 2r_k(\eta_c + \alpha/\lambda_{max}) - \overline{\eta}^2 \geq 0$, thus the square root in (2.34) is well-defined. Because of the previous identity we deduce also that

$$r_k \leq \overline{\eta} + \sqrt{r_k^2 - 2r_k\Big(\overline{\eta} - \eta_c - \frac{\alpha}{\lambda_{max}}\Big)}. \tag{2.36}$$

Lemma 2.8. *When F satisfies the coercivity assumption* (2.1)*, for $t > -\eta_c$ one has for all $\sigma_1, \sigma_2 \in \mathcal{S}_N$*

$$|S_t(\sigma_2) - S_t(\sigma_1)|^2 \leq \frac{1}{t + \eta_c}(S_t(\sigma_2) - S_t(\sigma_1)) : (\sigma_2 - \sigma_1), \tag{2.37}$$

where S_t is defined by (2.6)*.*

In particular this lemma implies (2.8).

Proof. Let $\gamma_1 = S_t(\sigma_1)$ and $\gamma_2 = S_t(\sigma_2)$. According to (2.7), one has $\partial F(\gamma_1) \ni \sigma_1 - t\gamma_1$, $\partial F(\gamma_2) \ni \sigma_2 - t\gamma_2$. Consider now $\overline{F}(\gamma) = F(\gamma) - \eta_c|\gamma|^2/2$. According to (2.1) $\overline{F}$ is convex, and one has $\partial\overline{F}(\gamma_1) \ni \sigma_1 - (t + \eta_c)\gamma_1$, $\partial\overline{F}(\gamma_2) \ni \sigma_2 - (t + \eta_c)\gamma_2$. Because of the monotonicity property it follows that

$$(\sigma_2 - (t + \eta_c)\gamma_2 - \sigma_1 + (t + \eta_c)\gamma_1) : (\gamma_2 - \gamma_1) \geq 0, \tag{2.38}$$

or in other words

$$(\sigma_2 - (t + \eta_c)S_t(\sigma_2) - \sigma_1 + (t + \eta_c)S_t(\sigma_1)) : (S_t(\sigma_2) - S_t(\sigma_1)) \geq 0, \tag{2.39}$$

proving (2.37). ∎

Proof of Theorem 2.6. Let $(u_k, \sigma_k) \in V_h \times M_h$ be defined by (2.27), (2.28), and $(\widehat{u}, \widehat{\sigma}) \in V_h \times M_h$ a solution to (2.25), (2.26). We apply Lemma 2.8 with $t = r_k - \eta_c$, under the form (2.39). We take

$\sigma_1 = \widehat{\sigma} + tD\widehat{u}$, then because of (2.26) one has $S_t(\sigma_1) = D\widehat{u}$. We take $\sigma_2 = \sigma_k + tDu_k$, then because of (2.28) we have

$$S_t(\sigma_2) = \frac{\sigma_k - \sigma_{k+1}}{r_k} + Du_{k+1} - \frac{\overline{\eta} - \eta_c}{r_k}(Du_{k+1} - Du_k). \tag{2.40}$$

Writing (2.39) multiplied by $-r_k$ and integrating over Ω we obtain

$$\begin{aligned} 0 \geq \Big\langle & \sigma_k - \sigma_{k+1} - (\overline{\eta} - \eta_c)(Du_{k+1} - Du_k) + r_k(Du_{k+1} - D\widehat{u}) \\ & + \widehat{\sigma} - \sigma_k - (r_k - \eta_c)(Du_k - D\widehat{u}), \\ & \sigma_k - \sigma_{k+1} - (\overline{\eta} - \eta_c)(Du_{k+1} - Du_k) + r_k(Du_{k+1} - D\widehat{u})\Big\rangle \equiv R. \end{aligned} \tag{2.41}$$

This quadratic expression R can be rewritten (see (A.17)–(A.22)) as

$$\begin{aligned} R = & \frac{1}{2}\|\sigma_{k+1} - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Du_{k+1} - D\widehat{u})\|^2 - \frac{1}{2}\|\sigma_k - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Du_k - D\widehat{u})\|^2 \\ & + \frac{1}{2}\|\sigma_{k+1} - \sigma_k + (\overline{\eta} - \eta_c - r_k)(Du_{k+1} - D\widehat{u}) + (\eta_c + r_k - 2\overline{\eta})(Du_k - D\widehat{u})\|^2 \\ & + \Big(\frac{r_k^2}{2} - r_k(\overline{\eta} - \eta_c)\Big)\|Du_{k+1} - D\widehat{u}\|^2 - \frac{1}{2}(r_k - \overline{\eta})^2\|Du_k - D\widehat{u}\|^2 \\ & - r_k\Big\langle\sigma_{k+1} - \widehat{\sigma}, Du_{k+1} - D\widehat{u}\Big\rangle. \end{aligned} \tag{2.42}$$

Next, combining (2.27) and (2.25) we get

$$\alpha\,\langle u_{k+1} - \widehat{u}, v\rangle + \langle\sigma_{k+1} - \widehat{\sigma}, Dv\rangle = 0, \quad \forall\, v \in V_h. \tag{2.43}$$

Taking $v = u_{k+1} - \widehat{u}$ this gives

$$-\langle\sigma_{k+1} - \widehat{\sigma}, Du_{k+1} - D\widehat{u}\rangle = \alpha\|u_{k+1} - \widehat{u}\|^2. \tag{2.44}$$

For any $k \geq 0$ let us consider

$$a_k = \frac{1}{2}\|\sigma_k - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Du_k - D\widehat{u})\|^2 + \frac{1}{2}(r_k - \overline{\eta})^2\|Du_k - D\widehat{u}\|^2. \tag{2.45}$$

Using (2.42), (2.44) and (2.31), the inequality $R \leq 0$ in (2.41) leads to

$$\begin{aligned} a_k \geq a_{k+1} + & \frac{1}{2}b_k\|Du_{k+1} - D\widehat{u}\|^2 \\ & + \frac{1}{2}\|\sigma_{k+1} - \sigma_k + (\overline{\eta} - \eta_c - r_k)(Du_{k+1} - D\widehat{u}) + (\eta_c + r_k - 2\overline{\eta})(Du_k - D\widehat{u})\|^2, \end{aligned} \tag{2.46}$$

with

$$b_k = r_k^2 - 2r_k\Big(\overline{\eta} - \eta_c - \frac{\alpha}{\lambda_{max}}\Big) - (r_{k+1} - \overline{\eta})^2. \tag{2.47}$$

Knowing (2.10), the stability condition (2.34) gives that b_k is nonnegative. If follows with (2.46) that the sequence (a_k) is nonincreasing, and it is therefore bounded. Summing up (2.46) we deduce that

$$\sum_{k=0}^{\infty} b_k\|Du_{k+1} - D\widehat{u}\|^2 < \infty, \tag{2.48}$$

hence that $b_k\|Du_{k+1} - D\widehat{u}\|^2 \to 0$ as $k \to \infty$. Now, either $r_k \to \infty$ or (2.35) holds. If $r_k \to \infty$, the boundedness of a_k gives because of the second term in (2.45) that $(r_k - \overline{\eta})(Du_k - D\widehat{u})$ is bounded in L^2, and thus that $Du_k - D\widehat{u} \to 0$ in L^2. Otherwise if assumption (2.35) holds, or in other words $\liminf b_k > 0$, we deduce that

$$\|Du_{k+1} - D\widehat{u}\|^2 \to 0 \quad \text{as } k \to \infty. \tag{2.49}$$

Therefore in any case (2.49) holds. Using (2.32) this gives $u_k \to \widehat{u}$ in L^2. Next, since a_k is bounded we have using the first term in (2.45) that σ_k is bounded in L^2. The sum over k of the second line of (2.46) is also finite, thus this term of the second line tends to 0 as $k \to \infty$. Using (2.49) we deduce that

$$\sigma_{k+1} - \sigma_k - r_k(Du_{k+1} - Du_k) \to 0 \quad \text{in } L^2. \tag{2.50}$$

Using this in (2.28) yields that

$$r_k\Big(S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k) - Du_k\Big) \to 0 \quad \text{in } L^2. \tag{2.51}$$

Let us denote by $\gamma_k = S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k)$. Then we have

$$\partial F(\gamma_k) + (r_k - \eta_c)\gamma_k \ni \sigma_k + (r_k - \eta_c)Du_k, \tag{2.52}$$

or equivalently

$$\partial F(\gamma_k) \ni \sigma_k + (r_k - \eta_c)(Du_k - \gamma_k) \quad \text{a.e. in } \Omega. \tag{2.53}$$

By (2.51) and the assumption that $\liminf r_k > 0$, we have $\gamma_k \to D\widehat{u}$ in L^2, and by (2.51) again $r_k(\gamma_k - Du_k) \to 0$ in L^2. It follows that the term $(r_k - \eta_c)(Du_k - \gamma_k)$ in the right-hand side of (2.53) tends to 0 in L^2. The sequence σ_k being bounded in the finite-dimensional space M_h, consider a subsequence $k = k(p)$ such that $\sigma_{k(p)} \to \overline{\sigma} \in M_h$ as $p \to \infty$, which means in the L^2 sense and pointwise in each $K \in \mathcal{T}_h$. Passing to the limit in (2.53) with the definition (1.4) of the subdifferential we obtain $\partial F(D\widehat{u}) \ni \overline{\sigma}$ a.e., hence $(\widehat{u}, \overline{\sigma})$ verifies (2.26). We can obviously pass to the limit in (2.27) and obtain that $(\widehat{u}, \overline{\sigma})$ verifies (2.25). ■

Remark 2.9. In Theorem 2.6, the most simple choice is to take any $\overline{\eta} \geq 0$, and $r_k = cst = r$, such that $r \geq \overline{\eta}$, $r > \overline{\eta}^2/2(\eta_c + \alpha/\lambda_{max})$. Then the conditions (2.5), (2.10), (2.33), (2.34), (2.35) are satisfied, because of (2.36).

Remark 2.10. If $\eta_c > 0$ a way to satisfy (2.33), (2.34) is to use conditions slightly more restrictive,

$$r_k \geq \frac{\overline{\eta}^2}{2\eta_c}, \qquad r_{k+1} \leq \overline{\eta} + \sqrt{r_k^2 - 2r_k(\overline{\eta} - \eta_c)}. \tag{2.54}$$

One has indeed to impose strict inequalities in (2.54) (with a safety margin) in order to get (2.35). The conditions (2.54) have the advantage that they do not need the evaluation of λ_{max} satisfying (2.31). Indeed it corresponds to taking $\lambda_{max} = \infty$.

Remark 2.11. Two main choices for $\overline{\eta}$ are either $\overline{\eta} = \eta_c + \alpha/\lambda_{max}$, or $\overline{\eta} = \eta_c$ in the case $\lambda_{max} \to \infty$ of Remark 2.10.

Remark 2.12. In the case $\eta_c = 0$, $\alpha > 0$, a lot of iterations may be necessary, because of the factor λ_{max} which is large.

Remark 2.13. For the time dependent problem (1.5) we have to solve (1.6), thus $\alpha = 1/\delta t$ is large, and additionally we have to be accurate not only on the unknown $u = u^{n+1}$, but indeed on αu in order to be consistent on the ratio $(u^{n+1} - u^n)/\delta t$. This makes stiff the problem of α large. Indeed the stiffness can be measured by the dimensionless number $\alpha/\eta_c\lambda_{min}$.

Remark 2.14. Primal-dual methods with variable parameter similar to r_k have been designed by many authors, in particular in [3, 11] and in the FISTA* algorithm of [27]. These authors always take the maximal possible value of r_{k+1}. For us it means equality in (2.34), i.e. take any $\overline{\eta} \geq 0$, start with $r_0 \geq \overline{\eta}$ such that $r_0 > \overline{\eta}^2/2(\eta_c + \alpha/\lambda_{max})$, and update r_k with

$$r_{k+1} = \overline{\eta} + \sqrt{r_k^2 - 2r_k\Big(\overline{\eta} - \eta_c - \frac{\alpha}{\lambda_{max}}\Big)}. \tag{2.55}$$

Then, because of (2.36), r_k is increasing, and one can prove that $r_k \sim (\eta_c + \alpha/\lambda_{max})k$ as $k \to \infty$. Thus $r_k \to \infty$ and the assumptions of Theorem 2.6 are satisfied. The boundedness of a_k defined by (2.45) gives because of the second term that $(r_k - \overline{\eta})(Du_k - D\widehat{u})$ is bounded in L^2 and thus that

$$Du_k - D\widehat{u} = O\Big(\frac{1}{k}\Big). \tag{2.56}$$

This rate is the best known for nondifferentiable laws, and was obtained in [27, Theorem 3.10] for the FISTA* algorithm. However here our numerical computations show that to let r_k go to $+\infty$ rapidly is far from being the best choice (see algorithm SEMI on Figure 4.1). We prefer different rules exposed below for choosing the parameters, even if we have no proof of rate of convergence like (2.56).

Remark 2.15. If we assume that for some k one has

$$\|\sigma_{k+1} - \sigma_k\| \leq \varepsilon, \quad \text{and} \quad (r_k - \overline{\eta} + \eta_c)\|Du_{k+1} - Du_k\| \leq \varepsilon, \tag{2.57}$$

it follows from (2.28) that

$$r_k\Big\|Du_k - S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k)\Big\| \leq 2\varepsilon, \tag{2.58}$$

giving an estimate of the strong consistency with (2.26). Conversely if (2.58) holds, taking $v = u_{k+1} - u_k$ in (2.30) gives $(r_k - \overline{\eta} + \eta_c)\|Du_{k+1} - Du_k\| \leq 2\varepsilon$, and with (2.28) we get $\|\sigma_{k+1} - \sigma_k\| \leq 4\varepsilon$. Hence (2.57) and (2.58) are equivalent up to some multiplicative constant. If r_k is large (2.58) is somehow too strong, see (3.7), (3.8); one should prefer to estimate the consistency by

$$\|Du_k - S_0(\sigma_k)\| \leq \varepsilon. \tag{2.59}$$

We shall call this quantity the consistency error (it is called *residual* in [28]). It is defined as long as $\eta_c > 0$ for S_0 to be well-defined. As long as r_k remains bounded, the two consistency errors are equivalent, as proved in (F.2) in Proposition F.1. An error estimate in terms of consistency error is also provided in Appendix F.

2.4. Semi-exact method with adaptive choice of r_k

When applying the iterative algorithm (2.27), (2.28), one has to define appropriate values for r_k, so that the assumptions of Theorem 2.6 are satisfied. We give here an adaptive algorithm. When r_k is known satisfying $r_k \geq \max(\overline{\eta}^2/2(\eta_c + \alpha/\lambda_{max}), \overline{\eta})$ one has to choose r_{k+1} such that

$$\max\Big(\frac{\overline{\eta}^2}{2(\eta_c + \alpha/\lambda_{max})}, \overline{\eta}\Big) \leq r_{k+1} \leq \overline{\eta} + \sqrt{r_k^2 - 2r_k(\overline{\eta} - \eta_c - \alpha/\lambda_{max})}. \tag{2.60}$$

According to Remark 2.7 we know that the value $r_{k+1} = r_k$ is admissible (it lies in the interval required in (2.60)). In view of r_k not going to ∞ we have to satisfy (2.35), which means that r_{k+1} must not approach the upper bound in (2.60). Indeed if r_{k+1} approaches this upper bound, the update $(u_k, \sigma_k) \mapsto (u_{k+1}, \sigma_{k+1})$ is not much contracting, and this should be avoided in order to have the best convergence rate. Our choice is to use the formula

$$r_{k+1} = \max(r_{min}, \min(r_{max}^k, r_{og}^k)), \tag{2.61}$$

where r_{min} is some fixed value,

$$r_{min} \geq \overline{\eta}, \quad r_{min} > \frac{\overline{\eta}^2}{2(\eta_c + \alpha/\lambda_{max})}, \tag{2.62}$$

$$r_{max}^k = \frac{1}{4}r_k + \frac{3}{4}\Big(\overline{\eta} + \sqrt{r_k^2 - 2r_k(\overline{\eta} - \eta_c - \alpha/\lambda_{max})}\Big), \tag{2.63}$$

and r_{og}^k is an "optimal guess" value. The coefficient $3/4 < 1$ has been put so that with the choice (2.61)–(2.63), if r_k remains bounded then the condition (2.35) is satisfied. We propose to take

$$r_{og}^k = \overline{\eta} - \eta_c + \frac{\|\sigma_{k+1} - \sigma_k\|}{\|Du_{k+1} - Du_k\|}, \tag{2.64}$$

where the norm is the L^2 norm over Ω.

The idea of using (2.64) comes from the update formula (2.28) for σ_{k+1}, that can be written

$$\sigma_{k+1} - \sigma_k = (r_k - \overline{\eta} + \eta_c)(Du_{k+1} - Du_k) + r_k\Big(Du_k - S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k)\Big). \tag{2.65}$$

The term with r_k in factor on the right-hand side vanishes if and only if $\sigma_k \in \partial F(Du_k)$. Thus in order to enforce this consistency relation one has to choose r_k so that as much as possible

$$\sigma_{k+1} - \sigma_k \simeq (r_k - \overline{\eta} + \eta_c)(Du_{k+1} - Du_k). \tag{2.66}$$

The formula (2.64) gives an average value consistent with that, except that r_{og}^k is used in the definition (2.61) of r_{k+1}, and will therefore be involved at the next iterative step, in the computation of σ_{k+2} and u_{k+2}.

Another interpretation of (2.64) follows from Proposition 2.2 (ii). Writing a formal linearization leads to consider in Proposition 2.2 (ii) the value $\eta = F''(Du_{k+1})$ (assuming for a moment that it is a scalar matrix). It follows that a formally good value of r_{k+1} would be $r_{k+1} - \overline{\eta} + \eta_c = F''(Du_{k+1})$. If we have already a rough approximation $\sigma_k \simeq F'(Du_k)$, $\sigma_{k+1} \simeq F'(Du_{k+1})$, we can write formally $F''(Du_{k+1})(Du_{k+1} - Du_k) \simeq (\sigma_{k+1} - \sigma_k)$. Taking the norm we obtain the formula (2.64), which is thus a scalar type optimal value. This linearized analysis is further developed in Section 3 for non scalar matrices.

Another idea in (2.64) is that the ratio involved in it will try to balance the errors in σ and in Du, in order to optimize the computational cost. If $\sigma_{k+1} - \sigma_k$ is large with respect to $Du_{k+1} - Du_k$ we take r_k large, and in the converse case if $Du_{k+1} - Du_k$ is large with respect to $\sigma_{k+1} - \sigma_k$ we take r_k small.

In the degenerate regions (plug zones) where $F''(Du)$ is infinite, the above analysis shows that in order to get fast convergence we should take large values of r. But doing this would slow down the convergence in the smooth part. In practice the balance between the errors in σ and in Du is enough to get accurate values on σ, which implies accuracy in degenerate regions.

In order for the optimal guess value (2.64) to be reachable by r_{k+1} in (2.61), we need that $r_{og}^k \geq \max(\overline{\eta}^2/2(\eta_c + \alpha/\lambda_{max}), \overline{\eta})$. Since $F'' \geq \eta_c$, it leads to the sufficient condition $\overline{\eta} \geq \overline{\eta}^2/2(\eta_c + \alpha/\lambda_{max})$, hence $\overline{\eta} \leq 2(\eta_c + \alpha/\lambda_{max})$. We shall assume

$$\overline{\eta} < 2(\eta_c + \alpha/\lambda_{max}), \qquad r_{min} = \overline{\eta}, \tag{2.67}$$

so that (2.62) is satisfied. We can now define our adaptive algorithm explicitly.

Algorithm 1 (SEAR / Semi-Exact method with Adaptive r_k)**.** *Input: η_c such that* (2.1), (2.2) *hold, startup data u_0, σ_0, parameters $r_{min} = \overline{\eta} > 0$, $r_0 \geq r_{min}$ (see the possible choices below).*

(SEAR.0) *Initialize $k = 0$.*

(SEAR.1) *Solve* (2.29) *to get u_{k+1}.*

(SEAR.2) *Get σ_{k+1} by* (2.28).

(SEAR.3) *If a stopping criterion is satisfied, return u_{k+1}, σ_{k+1} and stop.*

(SEAR.4) *Define r_{k+1} by* (2.61)*, where r_{og}^k is defined by* (2.64)*, and r_{max}^k by* (2.63).

(SEAR.5) *Set $k \leftarrow k + 1$ and go to* (SEAR.1).

Two main choices for $\overline{\eta}$ are

$$\overline{\eta}=\eta_c+\frac{\alpha}{\lambda_{max}}\quad\text{or}\quad\overline{\eta}=\eta_c. \tag{2.68}$$

The latter choice is possible only in the case $\eta_c>0$. It is a priori less optimal, but enables to avoid the evaluation of λ_{max} by taking $\lambda_{max}=\infty$ (see Remark 2.10). It is relevant when $\alpha/\lambda_{max}<<\eta_c$.

In any of the two cases of (2.68), the update (2.63) then simplifies to

$$r^k_{max}=r_k+\frac{3}{4}\overline{\eta}. \tag{2.69}$$

The construction (2.61), (2.64), (2.63) verifies $\overline{\eta}\leq r_{k+1}\leq r^k_{max}$ and thus satisfies the assumptions (2.5), (2.10), (2.33), (2.34), (2.35) of Theorem 2.6. It follows that the algorithm SEAR is always convergent.

For the initial value r_0 we take an heuristic formula

$$r_0=\left(\eta_c+\frac{\alpha}{\lambda_{min}}\right)^{\theta}\left(\eta_c+\frac{\alpha}{\lambda_{max}}\right)^{1-\theta}, \tag{2.70}$$

for some $0\leq\theta\leq 1$. In our simulations, not mentioning θ means the default value $\theta=0$, giving $r_0=\overline{\eta}$. Notice that if $\alpha=0$, the value of θ does not matter.

2.5. **Stopping criterion**

When applying the iterative algorithm (2.27), (2.28), we have to use a stopping criterion. Since the solution we want to compute is $(u_{exact},\sigma_{exact})$ solution to (1.1)–(1.3), we have to take into account that (u_k,σ_k) tends to the solution $(\widehat{u},\widehat{\sigma})$ to the discrete problem (2.25), (2.26), that differs from $(u_{exact},\sigma_{exact})$. Indeed this difference is only due to the fact that in (2.25) the unknown and the test function are in V_h instead of the whole space. One could think to use a posteriori estimators, which evaluate the approximate solution directly over the given mesh. This is done in particular in [17] for a problem which is not far from ours. However a posteriori estimates are not yet known for problems with non unique stress as we have here. As a consequence we propose here a more rough criterion based on the mesh size. Since we have $u_{exact}\in H^1(\Omega)$, we expect that $\|\widehat{u}-u_{exact}\|\sim Ch$, with h the grid size. It follows that it looks useless to look for an error $\|u_k-\widehat{u}\|$ much smaller than $Ch/2$, see Figure 2.1. Hence an optimal stopping criterion should involve the grid size h. Also, as explained in Remark 2.13, for α large the error has to be measured as $\alpha\|u_k-\widehat{u}\|$.

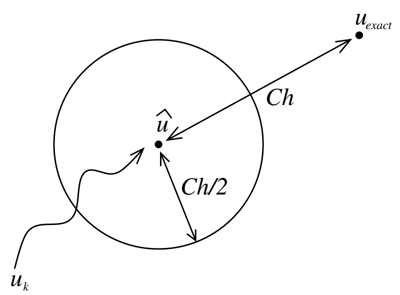

Figure 2.1. Convergence of u_k to the discrete solution $\widehat{u}$ depending on the grid. One has $\|\widehat{u}-u_{exact}\|\sim Ch$ with u_{exact} the exact solution to the continuous problem.

A first idea, proposed in [28], is to use the primal-dual gap. For an approximate solution (u,σ) to the discrete problem (2.25), (2.26) satisfying the momentum equation (2.25) exactly (semi-exact approximation), the latter reduces to

$$pdg(u,\sigma)=\int_\Omega\Big(F(Du)+F^*(\sigma)-\sigma:Du\Big), \tag{2.71}$$

where F^* is the convex conjugate function of F, see (A.3). Setting $\gamma=Du$, this can be computed by setting $\gamma_\sigma=S_0(\sigma)=(\partial F)^{-1}(\sigma)$, then $F^*(\sigma)=\sigma:\gamma_\sigma-F(\gamma_\sigma)$, which yields

$$F(\gamma)+F^*(\sigma)-\sigma:\gamma=F(\gamma)-F(\gamma_\sigma)-\sigma:(\gamma-\gamma_\sigma). \tag{2.72}$$

We notice that by (2.1) the quantity (2.72) is lower bounded by $\eta_c|\gamma-\gamma_\sigma|^2/2$, thus the primal-dual gap (2.71) is always larger than $\eta_c/2$ times the square of the L^2 consistency error $\|Du-S_0(\sigma)\|_{L^2}$, already considered in Remark 2.15. Moreover we have that the exact discrete solution $\widehat{u}$ minimizes the functional $J(v)=\int_\Omega(\alpha|v|^2/2+F(Dv)-f\cdot v)$ over V_h, and (A.3) implies that $pdg(u,\sigma)\geq J(u)-J(\widehat{u})$. It follows with (2.1) that

$$pdg(u,\sigma)\geq\int_\Omega\left(\alpha\frac{|u-\widehat{u}|^2}{2}+\eta_c\frac{|Du-D\widehat{u}|^2}{2}\right). \tag{2.73}$$

Thus as explained in [28], the primal-dual gap is good way to evaluate the error to the discrete solution. A stopping criterion could be then

$$\Big((\alpha+\eta_c\lambda_{min})pdg(u_{k+1},\sigma_{k+1})\Big)^{1/2}\leq Ch\lambda_{min}\|\sigma_{k+1}\|, \tag{2.74}$$

where $\|\sigma_{k+1}\|$ is put to make the inequality scale invariant, and C is a constant. Unfortunately our numerical experiments show that such criterion is a bit too strong, because of the gradient Du of u is involved in it. It could be used nevertheless, and our tests show a good value of $C=1.2$ for the problem without constraint, or $C=0.5$ for the problem with incompressible constraint.

We propose here another normalized stopping criterion

$$\max(E_{k+1},E_k)\left(r_k+\frac{\alpha}{\lambda_{min}}\right)\leq\|\sigma_{k+1}\|\frac{h}{8}, \tag{2.75}$$

where $E_k=\|u_k-u_{k-1}\|$ and the norms are all L^2 norms over Ω. In the right-hand side of (2.75), $\|\sigma_{k+1}\|$ is a normalizing factor that makes the inequality scale invariant, as well as λ_{min}, and $1/8$ is an empirical coefficient. The coefficient r_k in the left-hand side is put in order to scale as the error of σ in H^{-1}, taking into account that according to Section 2.4 we expect $\|\sigma_{k+1}-\sigma_k\|\sim r_k\|Du_{k+1}-Du_k\|$, and $\|\sigma_{k+1}-\sigma_k\|_{H^{-1}}\sim r_k\|u_{k+1}-u_k\|$. The maximum of E_{k+1} and E_k is taken because the criterion does not depend explicitly on $\|\sigma_{k+1}-\sigma_k\|_{H^{-1}}$ (computing this H^{-1} norm would be expensive). Indeed it can happen that $\|u_{k+1}-u_k\|$ is small but $\|\sigma_{k+1}-\sigma_k\|_{H^{-1}}$ is not small. However if both $\|u_{k+1}-u_k\|$ and $\|u_k-u_{k-1}\|$ are small, then it is likely that $\|\sigma_{k+1}-\sigma_k\|_{H^{-1}}$ is small also.

Instead of (2.75) one could think to measure the error in $r_k\|Du_{k+1}-Du_k\|$ and $\|\sigma_{k+1}-\sigma_k\|$, knowing that u_{exact} is expected to be H^2 when $\eta_c>0$ [7]. Our numerical runs show that this is too demanding (and it could be that $\sigma_{exact}\notin H^1$).

3. Semi-exact method with Newton-like linearized correction

We introduce here a second algorithm for our semi-exact primal-dual method. It uses a linearization in the spirit of the Newton method, leading to the resolution of an elliptic space dependent linear problem.

The starting idea is to observe that in our argument of setting the best parameter in Section 2.4, one would like to use F'' as in $r_{k+1}=\overline{\eta}-\eta_c+F''(Du_{k+1})$. But this is possible only for a scalar matrix F'', and we try here some kind of generalisation to arbitrary matrices. Indeed, in order to optimize simultaneously the smooth part and the plug zones, one should take a locally dependent parameter. Moreover since F'' can be infinite (plug zones), we have to apply some cutoff on F''.

Let us consider an approximate solution (u_{app}, σ_{app}) to (1.1)–(1.3), let us assume in particular that (1.1) is solved exactly,

$$\alpha u_{app} - \operatorname{div} \sigma_{app} = f. \tag{3.1}$$

Denote by (u_{ex}, σ_{ex}) an exact solution to (1.1)–(1.3). We make now a formal computation as if F were a C^2 functional. Forming the difference $u_{ex} - u_{app}$ we have

$$\alpha(u_{ex} - u_{app}) - \operatorname{div}\big(F'(Du_{ex}) - F'(Du_{app})\big) = \operatorname{div}\big(F'(Du_{app}) - \sigma_{app}\big). \tag{3.2}$$

Linearizing around u_{app} as in the Newton method, we are thus led to finding a corrected solution u_{cor} by

$$\alpha(u_{cor} - u_{app}) - \operatorname{div}\Big(F''(Du_{app})(Du_{cor} - Du_{app})\Big) = \operatorname{div}\big(F'(Du_{app}) - \sigma_{app}\big). \tag{3.3}$$

Then we introduce a cutoff factor $\big(\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r\big)^{-1}$, for some parameter $r > 0$. If r is large this factor is close to Id. We introduce this factor in (3.3), leading to

$$\begin{aligned}\alpha(u_{cor} - u_{app}) - \operatorname{div}\Big(\frac{F''}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}(Du_{cor} - Du_{app})\Big)&\\ = \operatorname{div}\Big(\frac{\operatorname{Id}}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}\big(F'(Du_{app}) - \sigma_{app}\big)\Big)&.\end{aligned} \tag{3.4}$$

We observe that

$$\frac{F''}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r} = r \operatorname{Id} - \frac{(r - \eta_c)\operatorname{Id}}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}. \tag{3.5}$$

We replace the right-hand side of (3.4) by the consistency error of the second line of (2.30),

$$\begin{aligned}\alpha(u_{cor} - u_{app}) - \operatorname{div}\Big(\frac{F''}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}(Du_{cor} - Du_{app})\Big)&\\ = \operatorname{div}\Big(rDu_{app} - rS_{r-\eta_c}\big(\sigma_{app} + (r - \eta_c)Du_{app}\big)\Big)&.\end{aligned} \tag{3.6}$$

Indeed the right-hand sides of (3.6) and (3.4) are close. This can be seen by using the identity $Du_{app} = S_{r-\eta_c}\big(F'(Du_{app}) + (r - \eta_c)Du_{app}\big)$, that implies that

$$\begin{aligned}rDu_{app} &- rS_{r-\eta_c}\big(\sigma_{app} + (r - \eta_c)Du_{app}\big)\\ &= rS_{r-\eta_c}\big(F'(Du_{app}) + (r - \eta_c)Du_{app}\big) - rS_{r-\eta_c}\big(\sigma_{app} + (r - \eta_c)Du_{app}\big)\\ &\simeq rS'_{r-\eta_c}() \times \big(F'(Du_{app}) - \sigma_{app}\big).\end{aligned} \tag{3.7}$$

According to (2.6) we have

$$S'_{r-\eta_c}(\sigma) = \big(F''(\gamma) + (r - \eta_c)\operatorname{Id}\big)^{-1}, \quad \text{with } \gamma = S_{r-\eta_c}(\sigma), \tag{3.8}$$

thus plugging this in (3.7) we obtain the right-hand side of (3.4).

In order to complete the determination of u_{cor} that is computed as the solution to the steady diffusion equation (3.6), we have to say how to compute the ratio (3.5), that occurs as diffusion matrix in (3.6). We compute it by setting

$$F'' = F''(\gamma), \qquad \gamma = S_{r-\eta_c}\big(\sigma_{app} + (r - \eta_c)Du_{app}\big). \tag{3.9}$$

Indeed we assume that F is piecewise C^2, and when $F''(\gamma)$ is not well-defined (which means formally that $F''(\gamma) = +\infty \operatorname{Id}$), we set the ratio (3.5) to $r \operatorname{Id}$. Assuming that F satisfies (2.1) with $\eta_c > 0$ and that $r \geq \eta_c$, we observe then that the diffusion matrix is symmetric, bounded by r and lower bounded by η_c. The problem (3.6) is thus a standard space dependent elliptic problem. Since u_{app} satisfies (3.1), adding it to (3.6) yields an equivalent equation

$$\alpha u_{cor} - \operatorname{div}\Big(\frac{F''}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}(Du_{cor} - Du_{app})\Big) \\ = f + \operatorname{div}\Big(\sigma_{app} + rDu_{app} - rS_{r-\eta_c}(\sigma_{app} + (r-\eta_c)Du_{app})\Big). \quad (3.10)$$

We have thus

$$\alpha u_{cor} - \operatorname{div}\sigma_{cor} = f, \quad (3.11)$$

with

$$\sigma_{cor} = \sigma_{app} + \frac{F''}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}(Du_{cor} - Du_{app}) + rDu_{app} - rS_{r-\eta_c}(\sigma_{app} + (r-\eta_c)Du_{app}). \quad (3.12)$$

The formula (3.12) is very much similar to (2.4), only the factor $r - \overline{\eta} + \eta_c$ has been replaced by the operator (3.5) which is a cutoff of F''.

To summarize, the linearized correction procedure consists in, given (u_{app}, σ_{app}) satisfying (3.1), computing the new approximation (u_{cor}, σ_{cor}) solving (3.11), (3.12) (or equivalently solve (3.10) to get u_{cor} and with (3.12) obtain σ_{cor}). When discretized, as in Section 2.2 we just replace the conservation equation (3.11) by

$$u_{cor} \in V_h, \\ \alpha \langle u_{cor}, v\rangle + \langle \sigma_{cor}, Dv\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h, \quad (3.13)$$

leading to the discrete analogue of (3.10)

$$u_{cor} \in V_h, \\ \alpha \langle u_{cor}, v\rangle + \Big\langle \frac{F''}{\operatorname{Id} + (F'' - \eta_c \operatorname{Id})/r}(Du_{cor} - Du_{app}), Dv\Big\rangle \\ + \Big\langle \sigma_{app} + rDu_{app} - rS_{r-\eta_c}(\sigma_{app} + (r-\eta_c)Du_{app}), Dv\Big\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h. \quad (3.14)$$

Since Du_{app} and σ_{app} belong to M_h the space of piecewise constant functions, the diffusion matrix also lies in this space with (3.9).

When iterating the linearized correction procedure, the larger r is, the faster the convergence is, because we are closer to the Newton method. However, there is no stability here, and practical simulations show that the larger r is, the more unstable the method is.

Our chosen algorithm is finally to mix the stable algorithm as stated in (2.3), (2.4), with constant parameter r, and the linearized correction procedure above, with the same r.

Algorithm 2 (SELC / Semi-Exact method with linearized correction)**.** *Input:* η_c *such that* (2.1), (2.2) *hold, startup data* u_0, σ_0, *parameters* $\overline{\eta} > 0$, $r \geq \overline{\eta}$, *integer* $m \geq 1$ *(see our choice below).*

(SELC.0) *Initialize* $k = 0$.

(SELC.1) *If* $k+1$ *is a multiple of* m, *set* $u_{app} = u_k$, $\sigma_{app} = \sigma_k$.

(SELC.2) *Solve* (3.14) *to get* u_{cor}.

(SELC.3) *Obtain* σ_{cor} *by* (3.12).

(SELC.4) *Reset* $u_k \leftarrow u_{cor}$, $\sigma_k \leftarrow \sigma_{cor}$.

(SELC.5) *Solve* (2.29) *to get* u_{k+1}.

(SELC.6) *Get* σ_{k+1} *by* (2.28).

(SELC.7) *If a stopping criterion is satisfied, return* u_{k+1}, σ_{k+1} *and stop.*

(SELC.8) *Set* $k \leftarrow k+1$ *and go to* (SELC.1).

Using $m = 1$ leads in practice to instability. In general taking a larger r gives faster convergence in terms of number of corrections since the algorithm is closer to a Newton method, but becomes more unstable thus needs to take a larger m to stabilize. Note that without linear correction ($m = \infty$) the algorithm is stable and convergent by Remark 2.9. After some testing, our empirical choice of parameters is

$$\overline{\eta} = \eta_c + \frac{\alpha}{\lambda_{max}}, \qquad r = \max(10\,\overline{\eta}, r_\theta), \qquad m = 3. \tag{3.15}$$

with r_θ defined by the right-hand side of (2.70) ($\theta = 0$ being the main choice).

Finally we can combine the adaptive algorithm SEAR and the linearly corrected algorithm SELC as the following algorithm SEARLC. We stop the linearized correction after some time, because our numerical tests show that it is useful only in the beginning of the iteration process.

Algorithm 3 (SEARLC / Semi-Exact method with adaptive r and linearized correction)**.** *Input:* η_c *such that* (2.1), (2.2) *hold, startup data* u_0, σ_0, *parameters* $r_{min} = \overline{\eta} > 0$, $r_0 \geq r_{min}$, *integers* $m \geq 1$, k_{stop} *(see our choice below).*

(SEARLC.0) *Initialize* $k = 0$.

(SEARLC.1) *If* $k + 1$ *is a multiple of* m *and* $k + 1 \leq k_{stop}$, *set* $u_{app} = u_k$, $\sigma_{app} = \sigma_k$, $r = r_k$ *and*

(SEARLC.2) *Solve* (3.14) *to get* u_{cor}.

(SEARLC.3) *Obtain* σ_{cor} *by* (3.12).

(SEARLC.4) *Reset* $u_k \leftarrow u_{cor}$, $\sigma_k \leftarrow \sigma_{cor}$.

(SEARLC.5) *Solve* (2.29) *to get* u_{k+1}.

(SEARLC.6) *Get* σ_{k+1} *by* (2.28).

(SEARLC.7) *If a stopping criterion is satisfied, return* u_{k+1}, σ_{k+1} *and stop.*

(SEARLC.8) *Define* r_{k+1} *by* (2.61), *where* r^k_{og} *is defined by* (2.64), *and* r^k_{max} *by* (2.63).

(SEARLC.9) *Set* $k \leftarrow k + 1$ *and go to* (SEARLC.1).

Our choice is

$$\overline{\eta} = \eta_c + \frac{\alpha}{\lambda_{max}}, \qquad m = 3, \qquad k_{stop} = 30, \tag{3.16}$$

$$r_0 = \max(10\,\overline{\eta}, r_\theta), \tag{3.17}$$

with r_θ defined by the right-hand side of (2.70), for some $0 \leq \theta \leq 1$.

4. Numerical tests

We consider several numerical tests. The first is a basic one with analytical solution, and the other ones are incompressible tests taken from [27]. Our nonlinearity is taken as a Herschel–Bulkley law with additional Newtonian viscosity

$$F(\gamma) = \sigma_{yield}|\gamma| + \frac{\eta}{2}|\gamma|^2 + \frac{\kappa}{n+1}|\gamma|^{n+1}, \tag{4.1}$$

with $\sigma_{yield} \geq 0$, $\eta \geq 0$, $\kappa \geq 0$, $n > 0$. For $\kappa = 0$ it is a Bingham law, and for $\eta = 0$ it is a Herschel–Bulkley law. The inversion formulas associated to the law (4.1) are provided in Appendix E. For our tests we take

$$\sigma_{yield} = \sqrt{2}, \quad \eta_c = \eta + n\kappa\gamma_{max}^{n-1}, \quad n = 1/2, \tag{4.2}$$

where γ_{max} is a suitable expected bound on $|Du|$. The iteration algorithms are initialized at 0 for u and σ. For a square of side L we use the approximation $\lambda_{min} = 18/L^2$, $\lambda_{max} = 32/h^2$, with $h = L/n_x$, n_x being the number of segments in which the boundary is subdivided. We use the choice $\overline{\eta} = \eta_c + \alpha/\lambda_{max}$.

The choice $\overline{\eta} = \eta_c$ ($\lambda_{max} = \infty$, see Remark 2.10) gives very similar results because in our tests we always have $\alpha/\lambda_{max} << \eta_c$.

We measure the L^2 error of the velocity with respect to an exact or reference solution, the L^2 consistency error defined by (2.59), and the primal-dual gap (2.71). Both last quantities measure the distance to the exact discrete solution, according to (2.73) and Proposition F.1.

We consider the semi-exact methods of constant r ($r = \overline{\eta}$), SEAR, SELC, SEARLC, Augmented Lagrangian (AL) as defined in Appendix D, and FISTA* also defined in Appendix D. For illustration we show also the SEMI algorithm (semi-exact with maximal increase) which is defined as the same as SEAR, except that (SEAR.4) is replaced by

(SEMI.4) Define $r_{k+1} = r_k + \overline{\eta}$.

4.1. **Test 1: analytical solution**

This test has been proposed in [23] without viscosity, and we extend it here to the case with viscosity. We consider the square $(x, y) \in \Omega = (-1, 1) \times (-1, 1)$, and an analytical radially symmetric solution to the problem (1.1)–(1.3). The solution is taken under the form $u(x, y) = \Phi(r)\left(\begin{smallmatrix}-y\\ x\end{smallmatrix}\right)$ where $r = \sqrt{x^2 + y^2}$, which gives

$$Du = \frac{\Phi'}{r} \begin{pmatrix} -xy & (x^2-y^2)/2 \\ (x^2-y^2)/2 & xy \end{pmatrix}, \tag{4.3}$$

where prime means differentiation with respect to r. Then $|Du|/\sqrt{2} = r|\Phi'|/2$ and we take $\sigma = \sigma_0 + \eta Du + \kappa|Du|^{n-1}Du$ with

$$\sigma_0 = \frac{2}{r^2}\operatorname{sgn}(\Phi') \begin{pmatrix} -xy & (x^2-y^2)/2 \\ (x^2-y^2)/2 & xy \end{pmatrix}, \tag{4.4}$$

where $\operatorname{sgn}(\Phi')$ is a function of r with values in $[-1, 1]$ which is 1 if $\Phi' > 0$ and -1 if $\Phi' < 0$. We consider a right-hand side $f = f_0 - \operatorname{div} f_1$, with $f_1 = \eta Du + \kappa|Du|^{n-1}Du$, so that the equation then writes

$$\left(\alpha\,\Phi - \frac{(\operatorname{sgn}\Phi')'}{r} - \frac{2\operatorname{sgn}\Phi'}{r^2}\right)\begin{pmatrix}-y\\ x\end{pmatrix} = f_0. \tag{4.5}$$

The function Φ must be continuous, as well as $\operatorname{sgn}\Phi'$. We take:
For $0 \le r \le 1/6$,

$$\operatorname{sgn}\Phi' = (12r - 36r^2)^2, \quad \Phi = 1, \quad f_0 = \Big(\alpha - 2\times 12^2(1-3r)(2-9r)\Big)\begin{pmatrix}-y\\ x\end{pmatrix}.$$

For $1/6 \le r \le 1/3$,

$$\operatorname{sgn}\Phi' = 1, \quad \Phi = 6r, \quad f_0 = \Big(6\alpha r - 2/r^2\Big)\begin{pmatrix}-y\\ x\end{pmatrix}.$$

For $1/3 \le r \le 1/2$,

$$\operatorname{sgn}\Phi' = \cos(\pi(6r-2)), \quad \Phi = 2, \quad f_0 = \left(2\alpha + \frac{6\pi\sin(\pi(6r-2))}{r} - \frac{2\cos(\pi(6r-2))}{r^2}\right)\begin{pmatrix}-y\\ x\end{pmatrix}.$$

For $1/2 \le r \le 5/6$,

$$\operatorname{sgn}\Phi' = -1, \quad \Phi = 5 - 6r, \quad f_0 = \Big(\alpha(5-6r) + 2/r^2\Big)\begin{pmatrix}-y\\ x\end{pmatrix}.$$

For $5/6 \le r \le 1$,

$$\operatorname{sgn}\Phi' = -\frac{1+\cos(\pi(6r-5))}{2}, \quad \Phi = 0, \quad f_0 = \left(\frac{-3\pi\sin(\pi(6r-5))}{r} + \frac{1+\cos(\pi(6r-5))}{r^2}\right)\begin{pmatrix}-y\\ x\end{pmatrix}.$$

α	constant r	SEAR	SEMI	SELC	AL	FISTA*
0	4	4	60	7	5	4
1	5	6	60	7	5	6
100	18	14	13	11	16	18
10000	60	38	34	25	31	60

Table 4.1. Test 1: number of iterations to reach the stopping criterion (2.75) in terms of α and the chosen method. The simulation is stopped anyway at $k = 60$.

For $1 \le r$,

$$\operatorname{sgn}\Phi' = 0, \quad \Phi = 0, \quad f_0 = 0.$$

The regularity is here only $u \in H^1$, because the source term is in H^{-1} if $\eta > 0$. We have three plug regions, $r < 1/6$, $1/3 < r < 1/2$ and $5/6 < r < 1$.

For our tests here we take $\eta = 1$, $\kappa = 1/2$, $\gamma_{max} = 3.5$. Our mesh is unstructured, made of 242818 triangles obtained from dividing the boundary as 320×320 segments. We use the (P1/P0) finite elements, and our code is written in FreeFem++ [22]. We use the direct solver UMFPACK.

We first consider $\alpha = 0$. On Figure 4.1 we plot the error with respect to the exact continuous solution u_{exact}, the consistency error, and the primal-dual gap. We observe on the first subfigure that Constant r, SEAR, SELC, FISTA* reach their optimal distance $\|u_k - u_{exact}\| \sim \|\widehat{u} - u_{exact}\|$ (see Figure 2.1) within $k \le 7$ iterations. The AL method takes a little more iterations, and SEMI a lot more. It is evidence here that the maximal increase of r_k is not the best choice. On the two last subfigures we observe that still SEMI and AL take more iterations, but also Constant r looks less good than SEAR, SELC, FISTA* (but this is not really true as seen on the first subfigure). SELC looks better, even if the slope at large k is lower then the one of SEAR and FISTA*. If we refer to the consistency error, the order of convergence is almost 2 for SEAR and FISTA* (i.e. an error in C/k^2), and is 1 for AL and Constant r (i.e. an error in C/k). In terms of primal-dual gap, the order of convergence is 1 for AL and Constant r, 2 for FISTA*, and 3 for SEAR.

Then we consider $\alpha = 10000$. Here the stiffness is $\alpha/(\eta_c\lambda_{min}) = 1960$. For the AL method we take $\theta = 1/4$, otherwise $\theta = 0$. On Figure 4.2 we plot the error with respect to the exact continuous solution, the consistency error, and the primal-dual gap. We observe on the first subfigure that SEAR, SEMI, SELC, AL almost reach their optimal distance to the exact solution within $k \le 30$ iterations. The FISTA* method takes more iterations, and constant r a lot more. On the two last subfigures we observe that still constant r takes really more iterations. SEAR and SEMI are more accurate on the primal-dual gap, but less accurate on the consistency error, in comparison to FISTA*. The methods AL and SELC are rather good, but have non optimal slopes on the primal-dual gap. If we refer to the consistency error, the order of convergence is 2 for SEAR, SEMI, SELC, AL, and 2.5 for FISTA*. In terms of primal-dual gap, the order of convergence is less than 2 for AL, SELC, more than 2 for FISTA*, and more than 3 for SEAR, SEMI.

On Table 4.1 we show the number of iterations to reach the stopping criterion (2.75), for the different methods and different values of α.

From the results of Test 1 we conclude that SEAR and SELC give good results in comparison to FISTA*, and especially in the case of large α. Indeed for $\alpha = 10000$ we observe that the methods that work well are those for which the iterative viscosity, i.e. $r_k - \overline{\eta} + \eta_c$ in the velocity equation (2.9) for the semi-exact method, or r in (D.2) for AL, can be taken large enough. This is not the case for FISTA* or constant r where the iterative viscosity coefficient is η_c.

In the iteration process, all methods show two distinct phases. In the first phase the smooth part of the solution is resolved, until reaching the optimal distance to the exact continuous solution, as shown

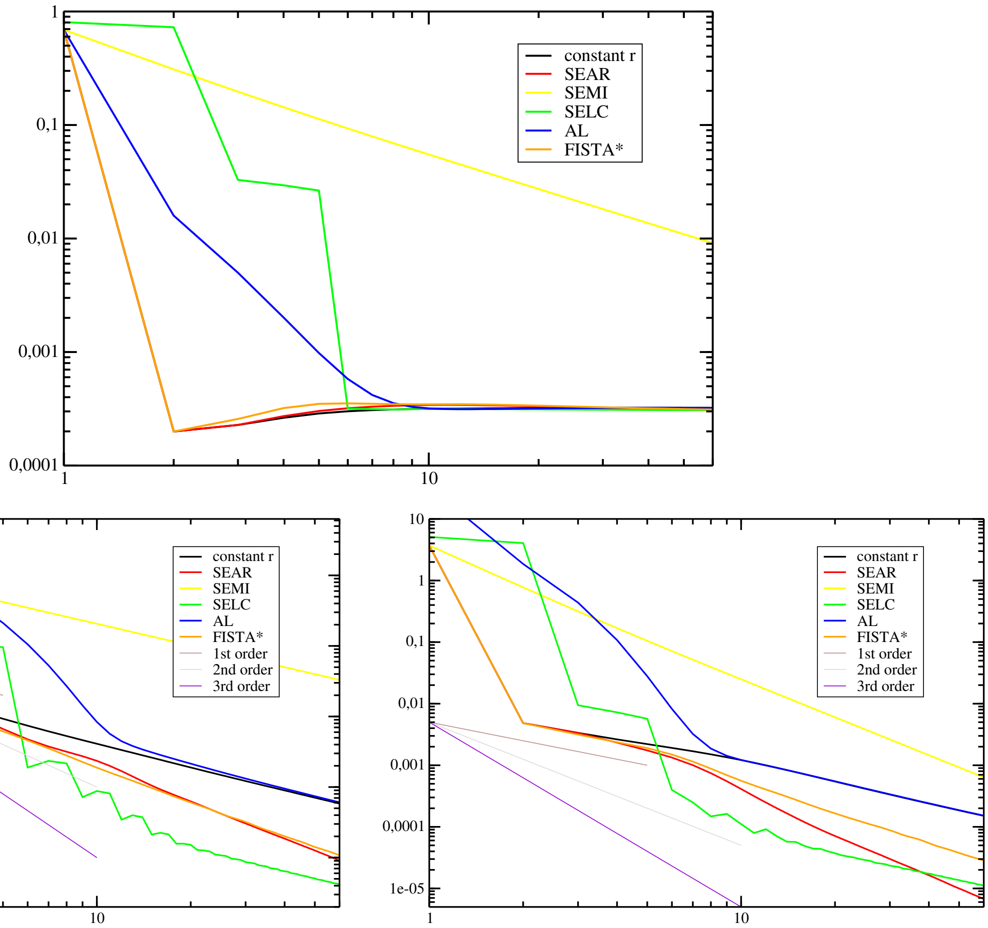

FIGURE 4.1. Test 1 with $\alpha = 0$. Above: Error $\|u_k - u_{exact}\|_{L^2}$ with u_{exact} the exact continuous solution, in terms of iteration number k in log-log scale for $k \leq 60$. Below left: same for the L^2 norm of the consistency error (2.59). Below right: same for the primal-dual gap (2.71). The three straight lines represent the slopes that a method should have to show a p-order convergence, i.e. $error \sim Ck^{-p}$, or equivalently $\log error \simeq -p \log k + \log C$.

on Figure 2.1. In this first phase r must not be too large, and the linearized correction procedure is particularly efficient. In the second phase, the singular part (plug regions) converges to the discrete solution $\widehat{u}$, in particular the detail of the stress σ, and this can take many iterations. For this second phase to be efficient, either r must become large enough (semi-exact method), or t_k must be large enough (FISTA*).

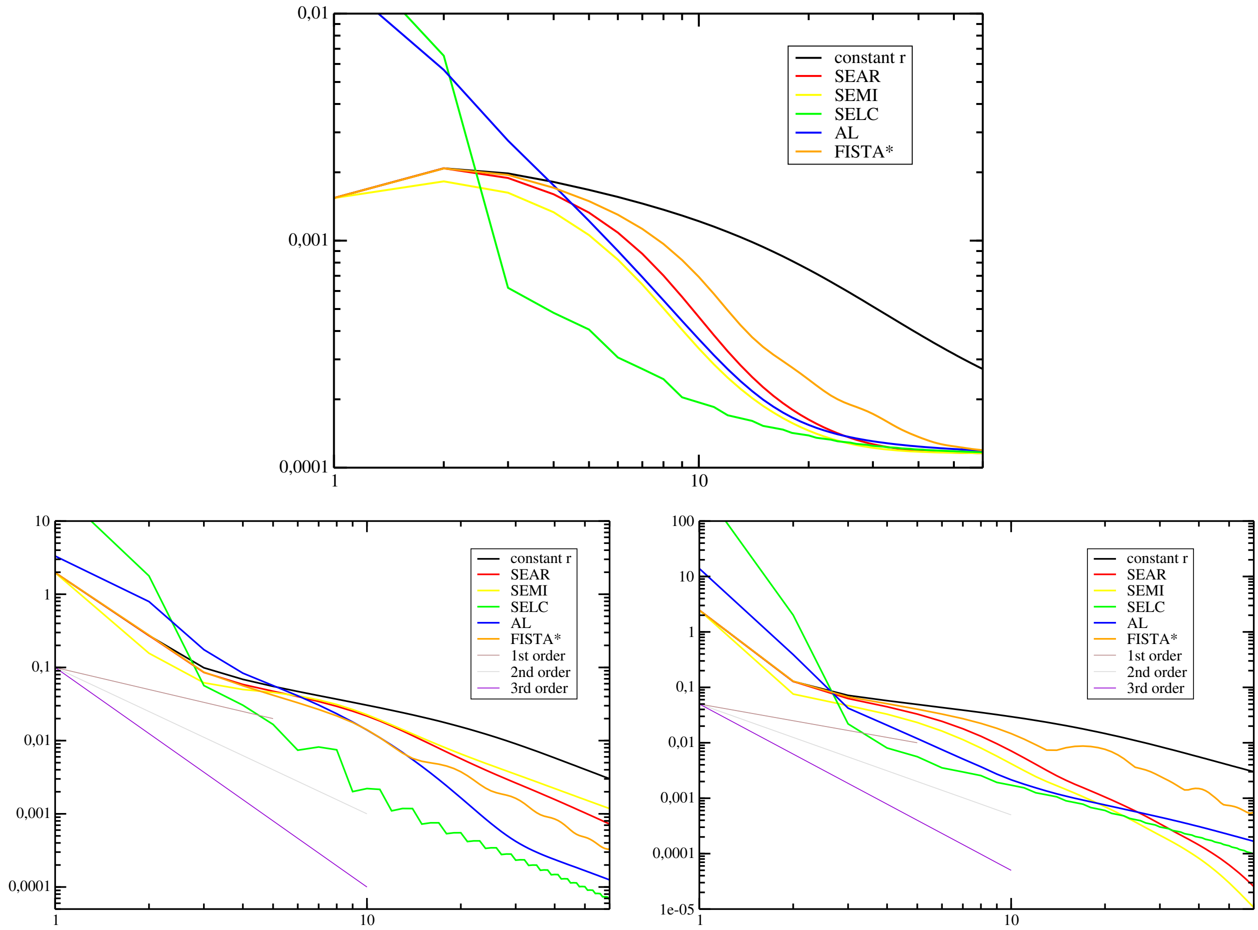

Figure 4.2. Test 1 for $\alpha = 10000$. Above: Error $\|u_k - u_{exact}\|_{L^2}$ with u_{exact} the exact continuous solution, in terms of iteration number k in log-log scale for $k \leq 60$. Below left: same for the L^2 norm of the consistency error (2.59). Below right: same for the primal-dual gap (2.71).

4.2. Test 2: the force driven cavity incompressible flow with Bingham law

We consider next the test cases of [27], that are incompressible. Thus we have to mention how to adapt our algorithm to that situation. Note that for time dependent incompressible problems another method is the Chorin–Temam projection method, that can be used in the viscoplastic setting as described in [10]. The (steady) incompressible problem consists in replacing (1.1) by

$$\alpha u - \operatorname{div} \sigma + \nabla p = f \quad \text{in } \Omega, \qquad \operatorname{div} u = 0 \quad \text{in } \Omega, \tag{4.6}$$

whereas (1.2), (1.3) are unchanged. The discrete problem is to find $(\widehat{u}, \widehat{p}, \widehat{\sigma}) \in V_h \times W_h \times M_h$ such that

$$\begin{cases} \alpha \langle \widehat{u}, v\rangle + \langle \widehat{\sigma}, Dv\rangle - \langle \widehat{p}, \operatorname{div} v\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h, & (4.7)\\ \langle q, \operatorname{div} \widehat{u}\rangle = 0, \quad \forall\, q \in W_h, & (4.8)\\ \widehat{\sigma} \in \partial F(D\widehat{u}), & (4.9)\end{cases}$$

where V_h, W_h, M_h are suitable finite element spaces. Then our algorithm (2.27), (2.28) is modified as searching for $u_{k+1} \in V_h$, $p_{k+1} \in W_h$, $\sigma_{k+1} \in M_h$ such that

$$\begin{cases} \alpha \langle u_{k+1}, v\rangle + \langle \sigma_{k+1}, Dv\rangle - \langle p_{k+1}, \operatorname{div} v\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h, & (4.10) \\ \langle q, \operatorname{div} u_{k+1}\rangle = 0, \quad \forall\, q \in W_h, & (4.11) \\ \langle \sigma_{k+1} - \sigma_k, s\rangle = \Big\langle -(\overline{\eta} - \eta_c)(Du_{k+1} - Du_k) \\ \qquad\qquad + r_k\Big(Du_{k+1} - S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k)\Big), s\Big\rangle, \quad \forall\, s \in M_h. & (4.12) \end{cases}$$

We assume that

$$\forall\, v \in V_h, \quad Dv \in M_h, \tag{4.13}$$

so that using (4.12) in (4.10) yields

$$\begin{cases} \alpha \langle u_{k+1}, v\rangle + (r_k - \overline{\eta} + \eta_c)\langle Du_{k+1}, Dv\rangle - \langle p_{k+1}, \operatorname{div} v\rangle - \langle f, v\rangle \\ \qquad + \langle \sigma_k + (\overline{\eta} - \eta_c)Du_k - r_k S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k), Dv\rangle = 0, \quad \forall\, v \in V_h, & (4.14) \\ \langle q, \operatorname{div} u_{k+1}\rangle = 0, \quad \forall\, q \in W_h. & (4.15) \end{cases}$$

As long as $r_k - \overline{\eta} + \eta_c > 0$ the Stokes system (4.14), (4.15) is solved classically with boundary conditions, and then σ_{k+1} is obtained by (4.12). Following [19, 27], our choice is the $P1 - iso - P2$ finite element for V_h. This means that we divide each triangle of the original mesh into four subtriangles by cutting the three edges by their middle, to obtain a finer mesh. Then V_h is $P1$ on this finer mesh, M_h is P0 on the finer mesh, and W_h is P1 on the original mesh. It follows that the inf-sup condition between V_h and W_h holds, the condition (4.13) holds, and the nonlinearities $S_t()$ leave invariant the space M_h. This last property enables to "remove" the test functions s in (4.12) and avoid a projection error. This ensures the existence of a solution to (4.7)–(4.9), see Appendix B, and the convergence of u_k to $\widehat{u}$ (the proof is identical to that of Theorem 2.6).

Another possible choice is to take $P1$ for V_h, $P0$ for M_h, and $W_h = \operatorname{div} V_h \subset P0$. Then the inf-sup condition is satisfied if the mesh is Powell–Sabin in 2d, or Worsey–Farin in 3d [15]. This choice ensures that the constraint $\langle q, \operatorname{div} v\rangle = 0$ for all $q \in W_h$ reduces to the exact constraint $\operatorname{div} v = 0$. However our simulations show that having this constraint satisfied exactly does not improve the accuracy (not shown here).

Our Test 2 is over $\Omega = (0, 1) \times (0, 1)$, the nonlinearity F is given by (4.1) with (4.2), and a Bingham fluid $\kappa = 0$, $\eta = 0.2$. The right-hand side is given by

$$f(x, y) = 30(y - 1/2, 1/2 - x). \tag{4.16}$$

It corresponds to the case $Bi = 10\sqrt{2}$ in [27]. The original mesh is unstructured, it is obtain by dividing the domain into 39146 triangles, obtained from dividing each side of the boundary into 128 segments. Since the triangles are subdivided for the $P1 - iso - P2$ space, the size h involved in the stopping criterion (2.75) is $1/(2 \times 128)$. We use a reference solution obtained with 1000 iterations.

We consider $\alpha = 0$. On Figure 4.3 we plot the error with respect to the reference solution, the consistency error, and the primal-dual gap. We observe that constant r and AL are only first order accurate. SEAR and FISTA* give quite similar results of second-order, FISTA* being a bit more accurate. SEAR gives a smaller primal-dual gap than FISTA*, whereas it is the converse for the consistency error. SELC and SEARLC are third order accurate in the first phase of the iteration process; then SELC becomes first-order, and SEARLC second-order. Level lines of $\|\sigma\|$ at the stopping criterion (2.75) for SEARLC is shown on Figure 4.4.

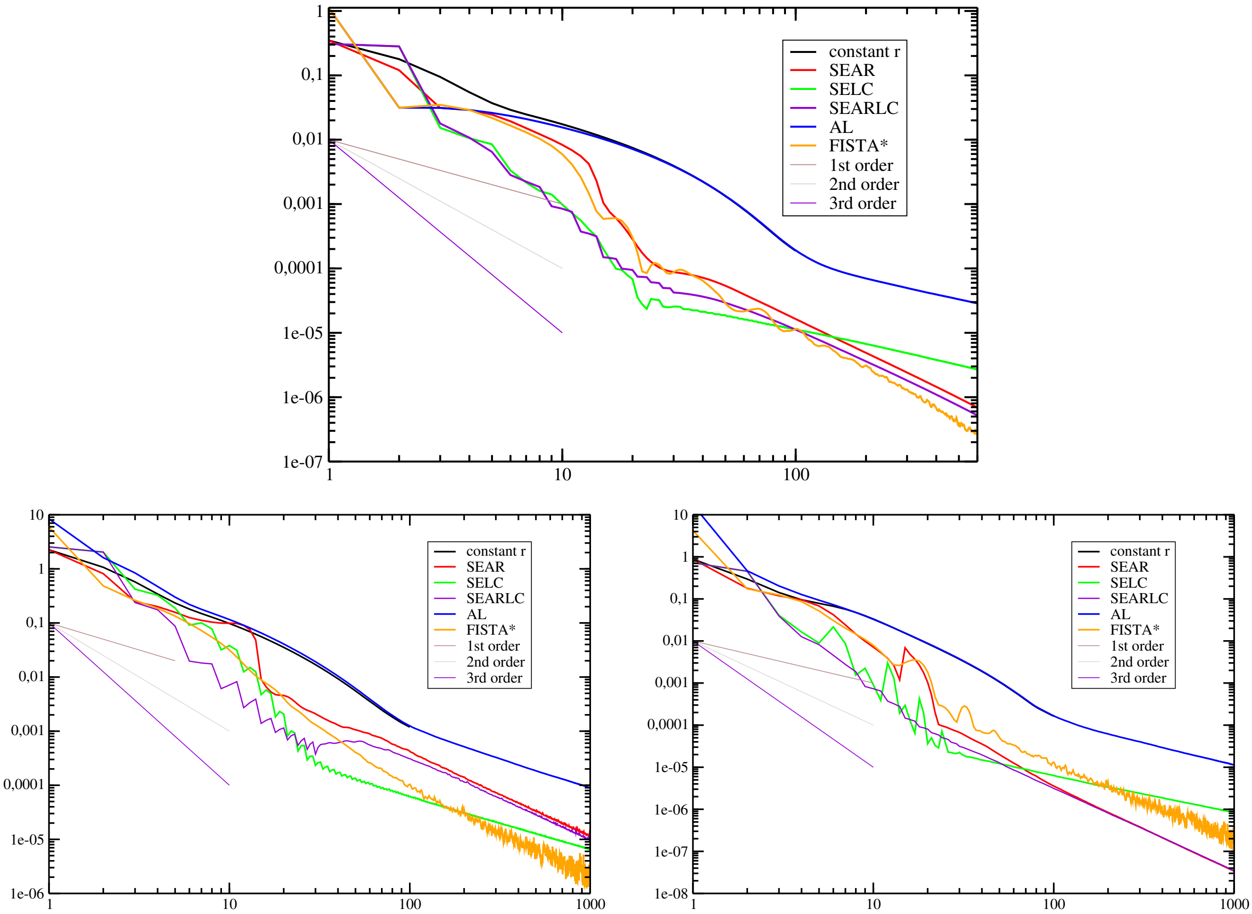

Figure 4.3. Test 2 with $\alpha = 0$. Above: Error $\|u_k - u_{ref}\|_{L^2}$ with u_{ref} the reference solution, in terms of iteration number k in log-log scale. Below left: same for the L^2 norm of the consistency error (2.59). Below right: same for the primal-dual gap (2.71). Note that since the reference solution is obtained for $k = 1000$, the value of the L^2 error $\|u_k - u_{ref}\|_{L^2}$ for k approaching 1000 could be misleading. This is why we cut the above error plot for $k \leq 600$.

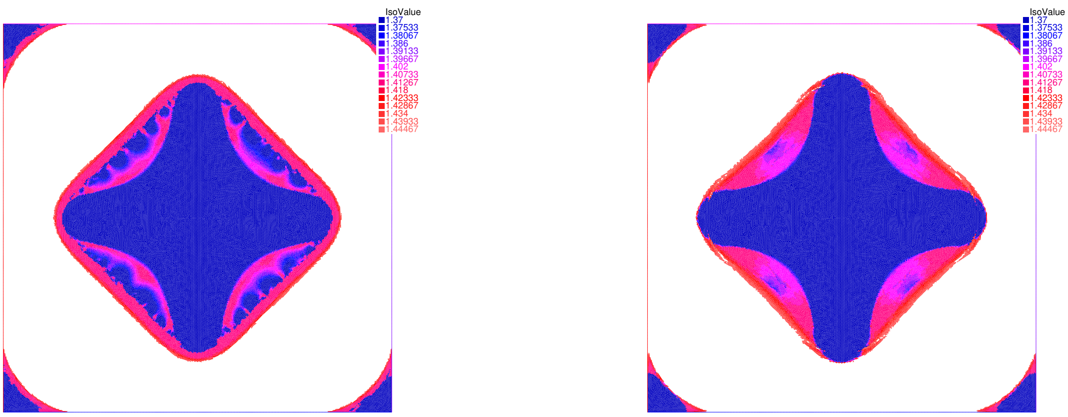

Figure 4.4. Test 2 with $\alpha = 0$. Level lines of $\|\sigma\|$ around the yield value $\sqrt{2}$. Left: reference solution ($k = 1000$). Right: SEARLC at the stopping criterion ($k = 7$). In the white domain one has $\|\sigma\| \geq 1.45$, and in the blue domain one has $\|\sigma\| \leq 1.37$.

Our conclusion for Test 2 is that SEAR or SEARLC are performing equally well with FISTA* for $\alpha = 0$. SEARLC improves the accuracy if one uses a limited number of iterations. We have performed Test 2 with $\alpha = 10000$, results are similar to those of Test 3, thus we omit to show them.

4.3. **Test 3: the force driven cavity incompressible flow with Herschel–Bulkley law**

Our Test 3 is incompressible, over $\Omega = (0,1) \times (0,1)$, the nonlinearity F is given by (4.1) with (4.2), and a Herschel–Bulkley fluid $\eta = 0$, $\kappa = \sqrt{2}/10$. The right-hand side is again (4.16), corresponding still to $Bi = 10\sqrt{2}$ in [27]. The mesh is the same as in Test 2.

We first consider $\alpha = 0$. We take $\gamma_{max} = 66$, twice the value of the L^2 norm of Du_{ref}. For FISTA* this value leads to a rapid blow up, thus in that case we take $\gamma_{max} = 200$. On Figure 4.5 we plot the error with respect to the reference solution, the consistency error, and the primal-dual gap. The methods AL, SEAR, SEARLC, FISTA* are second-order accurate on the L^2 error and primal-dual gap. Again SEAR gives a smaller primal-dual gap than FISTA*, whereas it is the converse for the consistency error. As previously SELC and SEARLC perform very well in the first phase of the iteration process, but it is preferable to stop the correction for the second phase (i.e. use SEARLC)

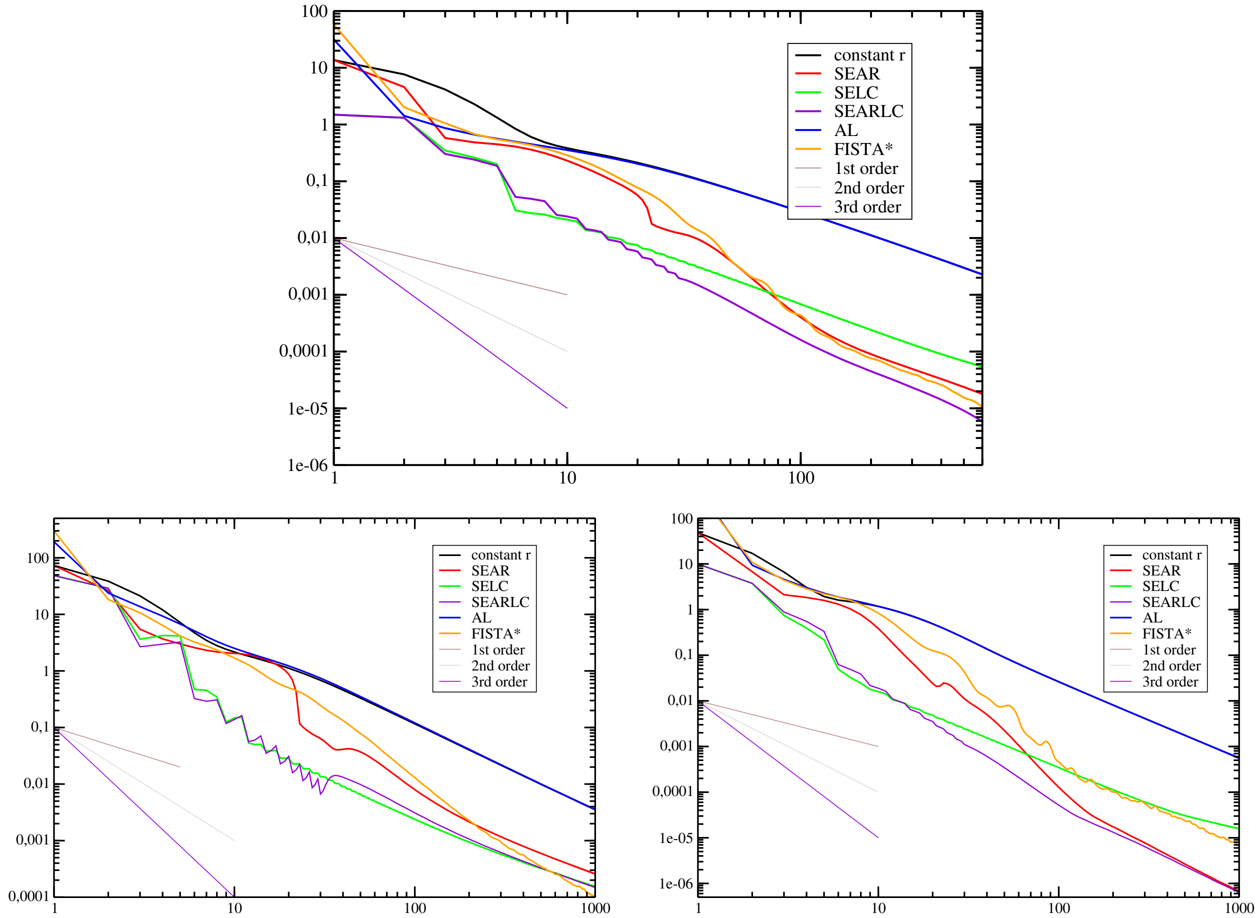

Figure 4.5. Test 3 with $\alpha = 0$. Above: Error $\|u_k - u_{ref}\|_{L^2}$ with u_{ref} the reference solution, in terms of iteration number k in log-log scale. Below left: same for the L^2 norm of the consistency error (2.59). Below right: same for the primal-dual gap (2.71).

α	constant r	SEAR	SELC	SEARLC	AL	FISTA*
0	10	20	7	7	7	7
10000	286	93		78	59	244

Table 4.2. Test 3: number of iterations to reach the stopping criterion (2.75) in terms of α and the chosen method.

Then we consider $\alpha = 10000$. The stiffness is $\alpha/(\eta_c\lambda_{min}) = 2484$. We take $\gamma_{max} = 0.1$. For the AL method we take $\theta = 0.75$, otherwise $\theta = 0$ (the default) or, when mentioned, $\theta = 0.75$. In order to make more iterations until $k = 5000$ without a too long run we take here a half-sized mesh with 64 segments on the boundary. On Figure 4.6 we plot the error with respect to the reference solution, the consistency error, and the primal-dual gap. The L^2 error and the primal-dual gap show a plateau for SEAR and SEARLC for $300 \leq k \leq 1800$. It corresponds to the time necessary for finding two passages though the central plug, that can be seen on Figure 4.7. Once this plateau is passed, the L^2 error and the primal-dual gap decay very fast. AL shows a more extended plateau, and for FISTA* it is not clear if there is such a plateau (more iterations would be necessary). FISTA* is much less accurate than SEAR and SEARLC. We have plotted on Figure 4.6 the FISTA* method with Nesterov sequence with $\alpha_N = 4$, as recommended in [28]. This does not change significantly the results in comparison with the original sequence. AL works rather nicely as long as the parameter r is well chosen (here with $\theta = 0.75$). Taking $r = \overline{\eta}$ (i.e. $\theta = 0$) for AL gives indeed a poor result similar to constant r (not shown). The results of SEAR and SEARLC are quite sensitive to the value of r_0 in the beginning of the iteration, but not for large k. The value $\theta = 0.75$ improves the first phase of iterations with respect to $\theta = 0$ for SEAR, but not so nicely for SEARLC. For $\theta = 0.75$, SEARLC is not as good as SEAR and AL, it seems that it is not a good idea to increase r_0 for SEARLC.

On Table 4.2 we show the number of iterations to reach the stopping criterion (2.75), for the different methods and different values of α. The criterion well indicates the separation between the two phases of the iteration, except for constant r and AL, and the criterion is excellent for $\alpha = 10000$.

Our conclusions for Test 3 are that first SEAR or SEARLC are performing equally well with FISTA* for small α, but largely outperform FISTA* for large α. Second, for the Herschel–Bulkley law involved Test 3, the choice of η_c in (4.2) via γ_{max} cannot be as sharp as for the Bingham law of Test2. We observe that FISTA* loses more accuracy and robustness because of this non sharp value than SEAR or SEARLC, and this is even more true for large α. We have tested changing the initial value r_0. It strongly affects the results of SEAR, SEARLC and AL. It can improve or make worse the accuracy of the method, but for SEAR and SEARLC it does not affect the fast convergence for large k. Thus taking $\theta \neq 0$ for SEAR or SEARLC is a bit risky, whereas it is somehow mandatory for AL in order to get a reasonable rate of convergence. We notice that the test $\alpha = 10000$ corresponds to solving the time dependent problem (1.5) with initial data $u^0 = 0$ over a timestep $\delta t = 1/\alpha \simeq h/80$. It is thus an important issue.

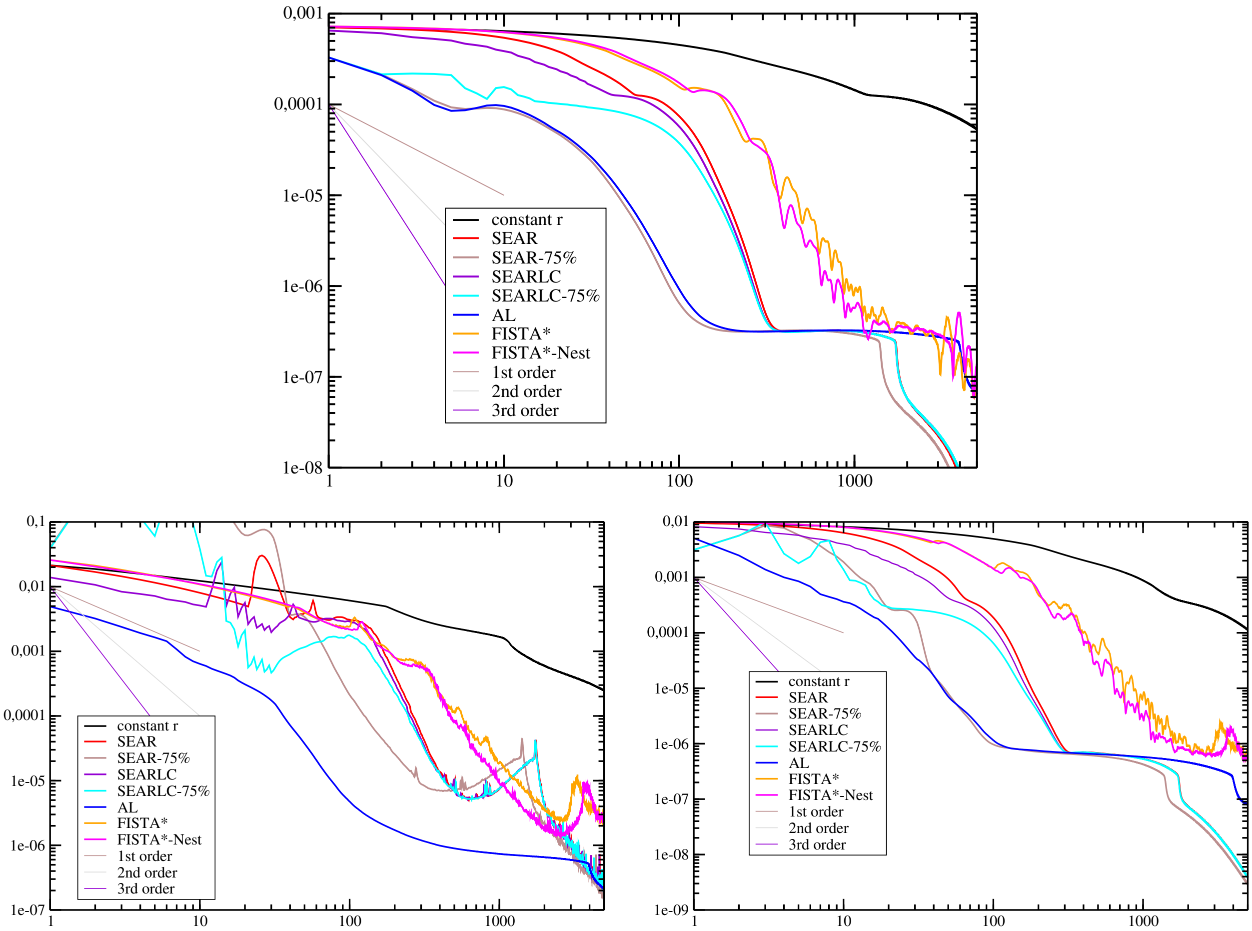

FIGURE 4.6. Test 3 with $\alpha = 10000$. Above: Error $\|u_k - u_{ref}\|_{L^2}$ with u_{ref} the reference solution, in terms of iteration number k in log-log scale. Below left: same for the L^2 norm of the consistency error (2.59). Below right: same for the primal-dual gap (2.71). The indication 75% means that $\theta = 0.75$.

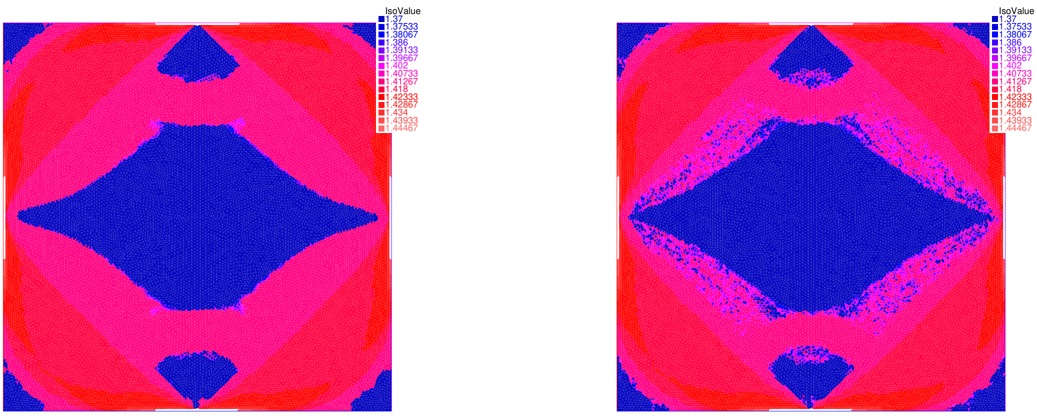

FIGURE 4.7. Test 3 with $\alpha = 10000$. Level lines of $\|\sigma\|$ around the yield value $\sqrt{2}$, at $k = 5000$. Left: SEAR. Right: FISTA*. In the blue domain one has $\|\sigma\| \leq 1.37$.

4.4. **Test 4: the lid driven cavity incompressible flow with Bingham law**

We consider another incompressible test from [27]. We have still $\Omega = (0,1)\times(0,1)$, F is given by (4.1) with (4.2), and a Bingham fluid $\eta = \sqrt{2}/10$, $\kappa = 0$ corresponding to $Bi = 20$ in [27]. The right-hand side is $f = 0$, the mesh is the same as in Test 2, and the boundary conditions are $u_y = 0$ on $\partial\Omega$, $u_x = 0$ on $\partial\Omega$ except on the upper boundary $y = 1$ where we set $u_x = 1$. We take $\alpha = 0$. Here $\alpha = 10000$ is not meaningful because a vanishing velocity does not satisfy the boundary conditions (a relevant way would be to add a right-hand side αu^0 with u^0 satisfying the boundary conditions).

The results are shown on Figure 4.8. SEAR, SEARLC and FISTA* are second-order accurate. Again FISTA* is more accurate on the consistency error, whereas SEAR and SEARLC are more accurate on the primal-dual gap. The number of iterations to reach the stopping criterion (2.75) is shown on Table 4.3. Level lines of $\|\sigma\|$ at the stopping for SEARLC is shown on Figure 4.9.

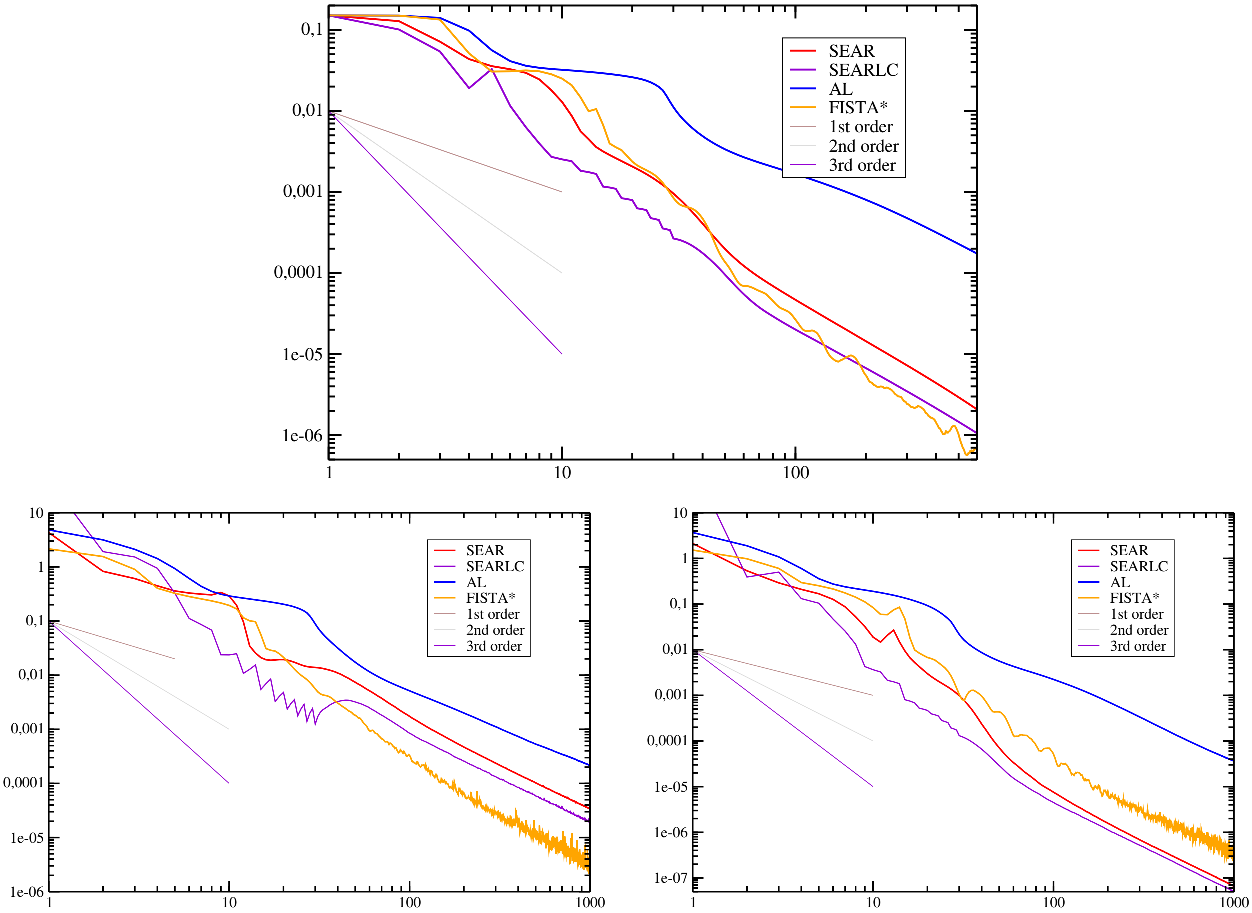

FIGURE 4.8. Test 4 with $\alpha = 0$. Above: Error $\|u_k - u_{ref}\|_{L^2}$ with u_{ref} the reference solution, in terms of iteration number k in log-log scale. Below left: same for the L^2 norm of the consistency error (2.59). Below right: same for the primal-dual gap (2.71).

Test 4 shows that the good properties of our semi-exact method are preserved on a test case with discontinuous boundary conditions, hence with a solution with limited smoothness.

α	SEAR	SEARLC	AL	FISTA*
0	14	10	9	16

TABLE 4.3. Test 4: number of iterations to reach the stopping criterion (2.75) in terms of the chosen method.

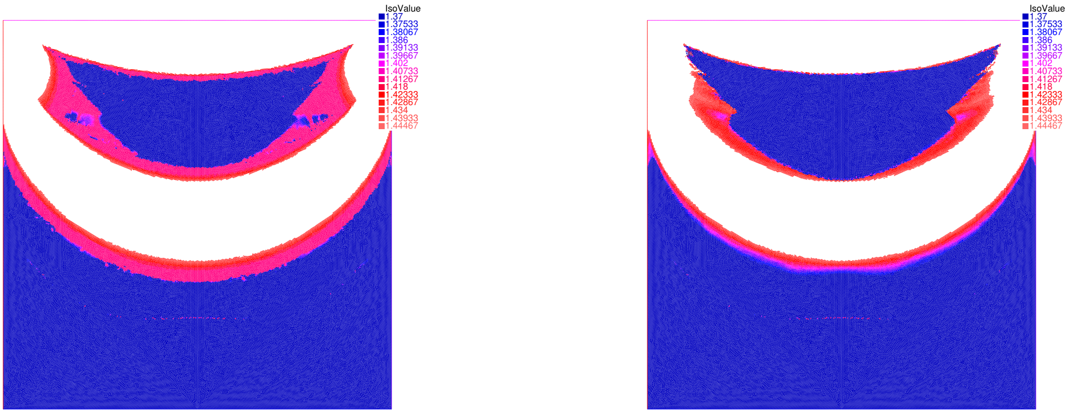

FIGURE 4.9. Test 4 with $\alpha = 0$. Level lines of $\|\sigma\|$ around the yield value $\sqrt{2}$. Left: reference solution ($k = 1000$). Right: SEARLC at the stopping criterion ($k = 10$).

5. Conclusion

The viscoplastic problem (1.1)–(1.3) is known to be hard to solve with a limited number of iterations when it involves plugs. The difficulty is increased when it is completed with an incompressibility constraint. Moreover the case of large α, that occurs under the form (1.6) when solving a time dependent problem (1.5), amplifies dramatically the number of iterations.

The semi-exact method, that we have introduced, resolves the momentum equation (1.1) exactly (at the discrete level), whereas the nonlinear relation (1.2) is solved approximately. The method involves a sequence (u_k, σ_k) of primal-dual unknowns that are updated through (2.3), (2.4) in the continuous formulation, or equivalently (2.9), (2.4). The particular form of the iteration scheme enables to have a parameter r_k that can be adapted, and that can become as large as necessary. The method is proved to converge under some sharp conditions stated in Theorem 2.6.

We have defined two main schemes, the Semi-Exact Adaptive r (SEAR) algorithm, and the Semi-Exact Adaptive r with Linearized Correction (SEARLC) algorithm. The first adjusts the value of r so as to balance as much as possible the velocity and stress errors, and the second additionally applies a linearized correction in the spirit of the Newton method, with a cutoff of the second derivative of the nonlinearity F.

Our numerical tests show that in a first phase of the iterations the smooth part of the solution is resolved. This needs a not too large value of the parameter r. In a second phase the singular part of the solution is resolved, including plugs. This needs a larger value of the parameter r. The accuracy of the first phase is improved when using the linearized correction, whereas this procedure does not improve in the second phase. For small α SEAR or SEARLC perform very similarly as the FISTA* method developed in [27]. For large α SEAR and SEARLC largely outstrips FISTA*. This is related to the property that the iterative viscosity $r_k - \overline{\eta} + \eta_c$ involved in the operator $\alpha\,\mathrm{Id} - (r_k - \overline{\eta} + \eta_c)\,\mathrm{div}\,D$ in (2.9) can become as large as necessary when α is large. In this situation of large α the Augmented Lagrangian method developed in [24] can perform well if its parameter r is well chosen, which is difficult in general. The semi-exact method has the ability to perform well in this case without having to find the best parameter. The accuracy of the semi-exact method can be explained by a linearized analysis developed in Proposition 2.2, in particular the rate in (ii) is independent of α, and in Section 3. However we have no proved convergence rate in the general case, that would justify this balance between the smooth and singular parts of the solution. Such rate should be at least as good as the proved rate for the FISTA* method, i.e. $O(1/k^2)$ for the primal-dual gap, or $O(1/k)$ for the error. Notice that this rate holds for the largest growth of r_k, cf. Remark 2.14, but this choice is not interesting since it is far from being the best in practice.

We have proposed a stopping criterion for the iterative process based on the principle of an optimal distance to the exact continuous solution related to the size of the mesh. It can be used for optimizing the numerical cost of iterations if one is interested in the distance to the exact solution. But more iterations are necessary if one is interested in the details of the plug regions.

The semi-exact method needs the existence of a coercivity constant η_c satisfying (2.1), (2.2). It is applicable if $\eta_c = 0$ and $\alpha > 0$, though we have only tested the case $\eta_c > 0$. It needs also in principle the spectral constants λ_{max} and λ_{min} satisfying (2.31) and (2.32). However for most practical cases we have $\alpha/\lambda_{max} << \eta_c$ thus one can neglect the (beneficial) effect of α/λ_{max} (therefore formally letting $\lambda_{max} \to \infty$). The exact value of λ_{min} is also not strictly necessary since we have never used quantitatively the inequality (2.32). Indeed λ_{min} has only been used as a scaling constant.

The main forms of SEAR and SEARLC (with $\theta = 0$ in (2.70), i.e. $r_0 = \overline{\eta}$) are robust and efficient. They do not need to tune any parameter, as is the case for AL or for regularization methods.

Appendix A. Semi-exact primal-dual method for a class of variational problems

We consider here the class of variational problems set in [14, III Remark 4.2], that gives somehow a nonlocal generalization of the viscoplastic problem (1.1)–(1.3). Assume that V, M are two Hilbert spaces with inner product denoted both by $\langle\cdot,\cdot\rangle$, with associated norms $\|\cdot\|$. Assume that $G : V \to M$ is a bounded linear operator, and denote by $G^* : M \to V$ its adjoint. We are looking for solutions $(u,\sigma) \in V \times M$ to the problem

$$\partial g(u) \ni -G^*\sigma, \qquad \partial h(Gu) \ni \sigma, \tag{A.1}$$

where g and h are two convex proper lower semi-continuous functions on V and M respectively. Solving the system (A.1) means to minimize $h(Gu) + g(u)$, or to find a saddle point to the Lagrangian

$$\mathcal{L}(u,\sigma) = \langle Gu, \sigma\rangle + g(u) - h^*(\sigma), \tag{A.2}$$

where h^* is the convex conjugate function of h. The primal-dual gap is

$$\begin{aligned} pdg(u,\sigma) &= \sup_{\overline{\sigma}\in M} \mathcal{L}(u,\overline{\sigma}) - \inf_{\overline{u}\in V} \mathcal{L}(\overline{u},\sigma) \\ &= h(Gu) + h^*(\sigma) - \langle Gu,\sigma\rangle + g(u) + g^*(-G^*\sigma) - \langle u, -G^*\sigma\rangle. \end{aligned} \tag{A.3}$$

The case of (1.1)–(1.3) corresponds to $Gu = Du$,

$$h(\gamma) = \int_\Omega F(\gamma), \qquad g(u) = \int_\Omega \alpha\frac{|u|^2}{2} - \int_\Omega f\cdot u. \tag{A.4}$$

Well-known primal-dual algorithms [2, 11, 12] to solve (A.1) involve the proximal operators $(\partial g + t\,\mathrm{Id})^{-1}$ and $(\partial h + t\,\mathrm{Id})^{-1}$ for $t > 0$, and both equations in (A.1) are resolved approximately at each iteration stage. We write down here our asymmetrical semi-exact method that resolves the first equation of (A.1) exactly, the second approximately, and involves the operators $(\partial h+t\,\mathrm{Id})^{-1}$ and $(\partial g+tG^*G)^{-1}$, as in the dual FISTA method [27].

We assume that h and g satisfy coercivity conditions,

$$\gamma\mapsto h(\gamma)-\eta_c\frac{\|\gamma\|^2}{2}\quad\text{is convex on } M,\quad\text{for some } \eta_c\geq 0,\tag{A.5}$$

$$u\mapsto g(u)-\beta\frac{\|Gu\|^2}{2}\quad\text{is convex on } V,\quad\text{for some } \beta\geq 0,\tag{A.6}$$

$$\eta_c>0\quad\text{or}\quad\beta>0.\tag{A.7}$$

Our iterative algorithm defines a sequence of primal-dual couples $(u_k,\sigma_k)\in V\times M$. We have first to define $u_0\in V$, $\sigma_0\in M$. Then once $(u_k,\sigma_k)\in V\times M$ is known (for $k=0,1,\dots$), the update is performed by solving the equations in the unknowns $(u_{k+1},\sigma_{k+1})\in V\times M$

$$\begin{cases}\partial g(u_{k+1})\ni -G^*\sigma_{k+1}, & \text{(A.8)}\\ \sigma_{k+1}-\sigma_k=-(\overline{\eta}-\eta_c)(Gu_{k+1}-Gu_k)\\ \qquad\qquad +r_k\Big(Gu_{k+1}-S_{r_k-\eta_c}(\sigma_k+(r_k-\eta_c)Gu_k)\Big), & \text{(A.9)}\end{cases}$$

where we have the parameters

$$\overline{\eta}\geq 0\quad\text{and}\quad r_k>0,\tag{A.10}$$

the latter eventually depending on k, and for $t>-\eta_c$

$$S_t(\sigma)=(\partial h+t\,\mathrm{Id})^{-1}(\sigma).\tag{A.11}$$

We observe that taking the value of σ_{k+1} given by (A.9) and plugging it into (A.8) we get

$$\partial g(u_{k+1})+(r_k-\overline{\eta}+\eta_c)G^*Gu_{k+1}\ni -G^*\Big(\sigma_k+(\overline{\eta}-\eta_c)Gu_k-r_kS_{r_k-\eta_c}(\sigma_k+(r_k-\eta_c)Gu_k)\Big).\tag{A.12}$$

This equation has a unique solution $u_{k+1}\in V$ as soon as $r_k\geq\overline{\eta}$ and G is strongly injective, i.e. $\|Gu\|\geq c\|u\|$ for some $c>0$. Once u_{k+1} is known we get σ_{k+1} by (A.9).

Theorem A.1 (Stability and convergence of the abstract semi-exact method)**.** *We assume that h, g satisfy the coercivity assumptions* (A.5)–(A.7)*, and that G is strongly injective. We suppose that there is a solution $(\widehat{u},\widehat{\sigma})\in V\times M$ to* (A.1)*. If the parameters $\overline{\eta}$, r_k satisfy* (A.10)*, $r_k\geq\overline{\eta}$ and for any $k=0,1,\dots$*

$$r_k\geq\frac{\overline{\eta}^2}{2(\eta_c+\beta)},\tag{A.13}$$

$$r_{k+1}\leq\overline{\eta}+\sqrt{r_k^2-2r_k(\overline{\eta}-\eta_c-\beta)},\tag{A.14}$$

and either $r_k\to\infty$ or

$$\liminf_{k\to\infty}\Big(r_k^2-2r_k(\overline{\eta}-\eta_c-\beta)-(r_{k+1}-\overline{\eta})^2\Big)\ >0,\tag{A.15}$$

then the sequence of approximate solutions $(u_k,\sigma_k)_{k=0,\dots}\in V\times M$ defined by (A.8), (A.9) *satisfies $u_k\to\widehat{u}$ in V as $k\to\infty$, and σ_k is bounded in M.*

If moreover $\liminf r_k>0$, then for any subsequence $k=k(p)$ such that $\sigma_{k(p)}\rightharpoonup\overline{\sigma}$ as $p\to\infty$ weakly in M, one has that $(\widehat{u},\overline{\sigma})$ is a solution to (A.1).

Remark A.2. Remark 2.7 remains valid here, as well as Remarks 2.9, 2.10, 2.11, 2.14, 2.15 in which α/λ_{max} has to be replaced by β and D by G. The algorithm SEAR can also be applied in the context of Theorem A.1.

Proof of Theorem A.1. Let $(u_k, \sigma_k) \in V \times M$ be defined by (A.8), (A.9), and $(\widehat{u}, \widehat{\sigma}) \in V \times M$ a solution to (A.1). We apply Lemma 2.8 (with F replaced by h) with $t = r_k - \eta_c$, under the form (2.39). We take $\sigma_1 = \widehat{\sigma} + tG\widehat{u}$, then because of (A.1) one has $S_t(\sigma_1) = G\widehat{u}$. We take $\sigma_2 = \sigma_k + tGu_k$, then because of (A.9) we have

$$S_t(\sigma_2) = \frac{\sigma_k - \sigma_{k+1}}{r_k} + Gu_{k+1} - \frac{\overline{\eta} - \eta_c}{r_k}(Gu_{k+1} - Gu_k). \tag{A.16}$$

Writing (2.39) multiplied by $-r_k$ we obtain

$$\begin{aligned} 0 \geq \Big\langle &\sigma_k - \sigma_{k+1} - (\overline{\eta} - \eta_c)(Gu_{k+1} - Gu_k) + r_k(Gu_{k+1} - G\widehat{u}) \\ &+ \widehat{\sigma} - \sigma_k - (r_k - \eta_c)(Gu_k - G\widehat{u}), \\ &\sigma_k - \sigma_{k+1} - (\overline{\eta} - \eta_c)(Gu_{k+1} - Gu_k) + r_k(Gu_{k+1} - G\widehat{u})\Big\rangle \equiv R. \end{aligned} \tag{A.17}$$

In order to have lighter expressions let us denote

$$w_k = Gu_k - G\widehat{u}, \qquad s_k = \sigma_k - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Gu_k - G\widehat{u}). \tag{A.18}$$

Then

$$R = \Big\langle (\overline{\eta} - r_k)w_k + r_k w_{k+1} - s_{k+1}, s_k + r_k w_{k+1} - s_{k+1}\Big\rangle. \tag{A.19}$$

We use the classical identity

$$\langle s_{k+1}, s_{k+1} - s_k\rangle = \frac{1}{2}\|s_{k+1}\|^2 - \frac{1}{2}\|s_k\|^2 + \frac{1}{2}\|s_{k+1} - s_k\|^2, \tag{A.20}$$

so that we develop (A.19) as

$$\begin{aligned} R &= \frac{1}{2}\|s_{k+1}\|^2 - \frac{1}{2}\|s_k\|^2 + \frac{1}{2}\|s_{k+1} - s_k\|^2 - \langle s_{k+1}, r_k w_{k+1}\rangle \\ &\quad + \langle s_k - s_{k+1}, (\overline{\eta} - r_k)w_k + r_k w_{k+1}\rangle + \langle r_k w_{k+1}, (\overline{\eta} - r_k)w_k + r_k w_{k+1}\rangle \\ &= \frac{1}{2}\|s_{k+1}\|^2 - \frac{1}{2}\|s_k\|^2 + \frac{1}{2}\|s_{k+1} - s_k - r_k w_{k+1} + (r_k - \overline{\eta})w_k\|^2 \\ &\quad - \frac{1}{2}\|r_k w_{k+1} + (\overline{\eta} - r_k)w_k\|^2 + \langle r_k w_{k+1}, (\overline{\eta} - r_k)w_k + r_k w_{k+1}\rangle \\ &\quad - \langle s_{k+1}, r_k w_{k+1}\rangle \\ &= \frac{1}{2}\|s_{k+1}\|^2 - \frac{1}{2}\|s_k\|^2 + \frac{1}{2}\|s_{k+1} - s_k - r_k w_{k+1} + (r_k - \overline{\eta})w_k\|^2 \\ &\quad + \frac{1}{2}\|r_k w_{k+1}\|^2 - \frac{1}{2}\|(\overline{\eta} - r_k)w_k\|^2 - \langle s_{k+1}, r_k w_{k+1}\rangle. \end{aligned} \tag{A.21}$$

Replacing w_k and s_k by their values (A.18) gives

$$\begin{aligned} R = &\frac{1}{2}\|\sigma_{k+1} - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Gu_{k+1} - G\widehat{u})\|^2 \\ &- \frac{1}{2}\|\sigma_k - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Gu_k - G\widehat{u})\|^2 \\ &+ \frac{1}{2}\|\sigma_{k+1} - \sigma_k + (\overline{\eta} - \eta_c - r_k)(Gu_{k+1} - G\widehat{u}) + (\eta_c + r_k - 2\overline{\eta})(Gu_k - G\widehat{u})\|^2 \\ &+ \frac{r_k^2}{2}\|Gu_{k+1} - G\widehat{u}\|^2 - \frac{1}{2}(r_k - \overline{\eta})^2\|Gu_k - G\widehat{u}\|^2 \\ &- r_k\langle\sigma_{k+1} - \widehat{\sigma}, Gu_{k+1} - G\widehat{u}\rangle - r_k(\overline{\eta} - \eta_c)\|Gu_{k+1} - G\widehat{u}\|^2. \end{aligned} \tag{A.22}$$

Next, since $(\widehat{u}, \widehat{\sigma})$ solves (A.1) and (u_{k+1}, σ_{k+1}) verifies (A.8), we have

$$\partial g(\widehat{u}) \ni -G^*\widehat{\sigma}, \qquad \partial g(u_{k+1}) \ni -G^*\sigma_{k+1}. \tag{A.23}$$

Using (A.6) the monotonicity property yields

$$\langle -G^*(\sigma_{k+1} - \widehat{\sigma}), u_{k+1} - \widehat{u}\rangle \geq \beta\|G(u_{k+1} - \widehat{u})\|^2. \tag{A.24}$$

According to (A.17) we have $R \leq 0$, thus with the expression (A.22) and (A.24) this gives

$$\begin{aligned} 0 \geq\ & \frac{1}{2}\|\sigma_{k+1} - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Gu_{k+1} - G\widehat{u})\|^2 - \frac{1}{2}\|\sigma_k - \widehat{\sigma} + (\overline{\eta} - \eta_c)(Gu_k - G\widehat{u})\|^2 \\ & + \frac{1}{2}\|\sigma_{k+1} - \sigma_k + (\overline{\eta} - \eta_c - r_k)(Gu_{k+1} - G\widehat{u}) + (\eta_c + r_k - 2\overline{\eta})(Gu_k - G\widehat{u})\|^2 \\ & + \left(\frac{r_k^2}{2} - r_k(\overline{\eta} - \eta_c - \beta)\right)\|Gu_{k+1} - G\widehat{u}\|^2 - \frac{1}{2}(r_k - \overline{\eta})^2\|Gu_k - G\widehat{u}\|^2. \end{aligned} \tag{A.25}$$

The remainder of the proof is identical to that of Theorem 2.6, replacing α/λ_{max} by β, D by G and ∂F by ∂h. ■

Appendix B. Existence of a solution to the discrete problems

We state here basic existence results under weak assumptions for the solutions to the discrete problems, that is assumed in Theorem 2.6.

Proposition B.1 (Existence of the P1/P0 discrete solution)**.** *Assume that F satisfies the coercivity assumptions* (2.1), (2.2)*, and that F is finite everywhere and lower bounded. Then there exist a solution $(\widehat{u}, \widehat{\sigma}) \in V_h \times M_h$ to the P1/P0 discrete problem* (2.25), (2.26)*. Moreover $\widehat{u}$ is unique and achieves the minimum over V_h of the functional $J(v) = \int_\Omega (\alpha|v|^2/2 + F(Dv) - f \cdot v)$.*

Proposition B.2 (Existence of the incompressible discrete solution)**.** *Assume that F satisfies the coercivity assumptions* (2.1), (2.2)*, and that F is finite everywhere and lower bounded. Then there exist a solution $(\widehat{u}, \widehat{p}, \widehat{\sigma}) \in V_h \times W_h \times M_h$ to the discrete problem* (4.7), (4.8), (4.9)*, where V_h, W_h, M_h are as stated in Section 4.2. Moreover $\widehat{u}$ is unique and achieves the minimum of the functional $J(v) = \int_\Omega (\alpha|v|^2/2 + F(Dv) - f \cdot v)$ over V_h under the constraint that $\langle q, \operatorname{div} v\rangle = 0$ for all $q \in W_h$.*

The proofs can be done in a classical way by approximating F by the Moreau envelope F_ϵ of F, see [23, Proposition 4.2.4].

Appendix C. Connection to the Arrow–Hurwicz algorithm

We give here a connection between the semi-exact method (A.8), (A.9) for solving (A.1) in the case $r_k = \overline{\eta}$, and the Arrow–Hurwicz algorithm [2]. In view of assumption (A.5), we can define

$$h_1(\gamma) = h(\gamma) - \eta_c \frac{\|\gamma\|^2}{2}, \tag{C.1}$$

which is a convex proper lower semi-continuous function on M. Then we have $\partial h_1(\gamma) = \partial h(\gamma) - \eta_c \gamma$. We define also

$$g_1(u) = g(u) + \eta_c \frac{\|Gu\|^2}{2}, \tag{C.2}$$

so that $\partial g_1(u) = \partial g(u) + \eta_c G^* G u$, $h(Gu) + g(u) = h_1(Gu) + g_1(u)$, and the problem (A.1) can be written equivalently

$$\partial g_1(u) \ni -G^* \sigma^{(1)}, \qquad \partial h_1(Gu) \ni \sigma^{(1)}, \tag{C.3}$$

with $\sigma^{(1)} = \sigma - \eta_c G u$.

Assuming $r_k = \overline{\eta}$, the equation (A.12) can be written

$$\partial g_1(u_{k+1}) \ni -G^* \Big(\sigma_k^{(1)} + rGu_k - rS_{r-\eta_c}(\sigma_k^{(1)} + rGu_k)\Big), \tag{C.4}$$

with $\sigma_k^{(1)} = \sigma_k - \eta_c G u_k$, whereas (A.9) can be written

$$\sigma_{k+1}^{(1)} - \sigma_k^{(1)} = r\Big(Gu_k - S_{r-\eta_c}(\sigma_k^{(1)} + rGu_k)\Big). \tag{C.5}$$

Next, according to the Moreau identity one has

$$\sigma - rS_{r-\eta_c}(\sigma) = (\mathrm{Id} + r\partial h_1^*)^{-1}(\sigma). \tag{C.6}$$

Therefore, (C.4)–(C.5) can be written equivalently

$$\begin{cases} \sigma_{k+1}^{(1)} = (\mathrm{Id} + r\partial h_1^*)^{-1}(\sigma_k^{(1)} + rGu_k), & \text{(C.7)} \\ \partial g_1(u_{k+1}) \ni -G^*(\sigma_{k+1}^{(1)}). & \text{(C.8)} \end{cases}$$

This is the Arrow–Hurwicz method for (C.3) in which the parameter τ has been sent to infinity, recovering the first relation of (C.3) exactly. According to Theorem A.1 the method is convergent for $0 < r < 2(\eta_c + \beta)$.

Remark C.1. In the case of the problem (1.1)–(1.3), according to Proposition 2.2 (i), if $\alpha = 0$ a particularly good value for r in (C.7) is $r = \eta_c$. In this case the method coincides with the FISTA* method without "acceleration".

Appendix D. The Augmented Lagrangian (AL) and FISTA* algorithms

We formulate here (in the continuous framework) the formulas we have used for these two algorithms.

The so called Augmented Lagrangian algorithm, as settled in [24], can be written as follows. It is initialized with (u_0, σ_0). Then for an integer $k \geq 0$, the values of u_k, σ_k being known, it sets

$$\sigma_{k+1} = (\mathrm{Id} + r\partial F^*)^{-1}(\sigma_k + rDu_k), \tag{D.1}$$

$$\alpha u_{k+1} - \operatorname{div}\big(rDu_{k+1} - rDu_k + 2\sigma_{k+1} - \sigma_k\big) - f = 0, \tag{D.2}$$

where $r > 0$ is a parameter. Notice that according to Moreau's identity one has $(\mathrm{Id} + r\partial F^*)^{-1} = \mathrm{Id} - rS_r$, with S_r defined by (2.6). Here we have eliminated $\gamma_{k+1} = (\sigma_k + rDu_k - \sigma_{k+1})/r = S_r(\sigma_k + rDu_k)$ from the formulas of [24]. It is not easy to choose a good value for r. In our simulations we take the rule (2.70) for some $0 \leq \theta \leq 1$. Notice that AL is not a semi-exact method: $Du_k = S_0(\sigma_k)$ for some k does not imply that the momentum equation $\alpha u_k - \operatorname{div}\sigma_k - f = 0$ holds.

The most simple form of the dual FISTA algorithm, as formulated in [27], is as follows. It starts with an initial guess σ_0 for the stress, and we set $\sigma_{-1} = \sigma_0$. Then for an integer $k \geq 0$, the values of σ_k and σ_{k-1} being known, it sets

$$\widetilde{\sigma}_k = \sigma_k + \frac{t_{k-1} - 1}{t_k}(\sigma_k - \sigma_{k-1}), \tag{D.3}$$

$$\alpha u_{k+1} - \operatorname{div}\big(\tau Du_{k+1} + \widetilde{\sigma}_k - \tau S_0(\widetilde{\sigma}_k)\big) - f = 0, \tag{D.4}$$

$$\sigma_{k+1} = \widetilde{\sigma}_k + \tau Du_{k+1} - \tau S_0(\widetilde{\sigma}_k), \tag{D.5}$$

where S_0 is defined by (2.6), i.e. $S_0 = (\partial F)^{-1}$, and $\tau > 0$ is a parameter satisfying $\tau \leq \eta_c$ (we have to assume that $\eta_c > 0$). For our tests we take $\tau = \eta_c$. The sequence t_k is defined as $t_{-1} = t_0 = 1$, and for $k \geq 0$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}. \tag{D.6}$$

This sequence t_k is called the "acceleration scheme", it makes the algorithm backward dependent by (D.3). Without acceleration it is $t_k \equiv 1$, then the scheme is identical to our semi-exact algorithm with $r_k = \overline{\eta} = \eta_c$. A variant is to use instead of $(t_{k-1}-1)/t_k$ in (D.3) a Nesterov sequence $k/(k+\alpha_N)$, with $\alpha_N > 3$.

We notice that the couple (u_{k+1}, σ_{k+1}) of approximate velocity and stress satisfies $\alpha u_{k+1} - \operatorname{div}\sigma_{k+1} - f = 0$ exactly, thus in the most simple form (D.3)–(D.5), FISTA* is a semi-exact method.

Appendix E. Formulas for a Bingham/Herschel–Bulkley law

We give here the inversion formulas for the composite Bingham/Herschel–Bulkley law (4.1), as well as the formulas to compute the diffusion matrix (3.5) that involves F'', which is necessary to apply the algorithm SELC.

When computing S_t defined in (2.6), given $\sigma \in \mathcal{S}_N$ we have to find $\gamma \in \mathcal{S}_N$ so that $S_t(\sigma) = \gamma$, which means according to (2.7) that $\partial F(\gamma) + t\gamma \ni \sigma$. Taking into account (4.1) this can be written

$$\sigma \in \sigma_{yield}\frac{\gamma}{|\gamma|} + (\eta + t)\gamma + \kappa|\gamma|^{n-1}\gamma. \tag{E.1}$$

If $|\sigma| \le \sigma_{yield}$, the solution is $\gamma = 0$.

Otherwise if $|\sigma| > \sigma_{yield}$, the solution is $\gamma = |\gamma|\frac{\sigma}{|\sigma|}$, with $|\gamma| > 0$ solving

$$\sigma_{yield} + (\eta + t)|\gamma| + \kappa|\gamma|^n = |\sigma|. \tag{E.2}$$

The value of $|\gamma|$ is obtained by the Newton method starting from any positive value, as

$$|\gamma|_{l+1} = |\gamma|_l - \frac{\sigma_{yield} - |\sigma| + (\eta+t)|\gamma|_l + \kappa|\gamma|_l^n}{\eta + t + \kappa n|\gamma|_l^{n-1}}. \tag{E.3}$$

If $\kappa = 0$ or $n = 1$, the solution is indeed given by $|\gamma| = (|\sigma| - \sigma_{yield})/(\eta + t + \kappa)$.

If $n = 1/2$ the solution is indeed given directly by

$$|\gamma| = \frac{4(|\sigma| - \sigma_{yield})^2}{\left(\sqrt{\kappa^2 + 4(\eta+t)(|\sigma| - \sigma_{yield})} + \kappa\right)^2}. \tag{E.4}$$

Next we can compute $F''(\gamma)$ when $\gamma \neq 0$,

$$F''(\gamma) = \left(\eta + \frac{\sigma_{yield}}{|\gamma|} + \kappa|\gamma|^{n-1}\right)\mathrm{Id} + \left(-\frac{\sigma_{yield}}{|\gamma|} + (n-1)\kappa|\gamma|^{n-1}\right)\frac{\gamma\otimes\gamma}{|\gamma|^2}, \tag{E.5}$$

where Id denotes the linear operator on $\mathcal{S}_N : \overline{\gamma} \mapsto \overline{\gamma}$, and $\gamma\otimes\gamma$ the operator $\overline{\gamma} \mapsto (\gamma : \overline{\gamma})\gamma$. Using the formulas

$$\left(a\,\mathrm{Id} + b\frac{\gamma\otimes\gamma}{|\gamma|^2}\right)\left(c\,\mathrm{Id} + d\frac{\gamma\otimes\gamma}{|\gamma|^2}\right) = ac\,\mathrm{Id} + (ad + bc + bd)\frac{\gamma\otimes\gamma}{|\gamma|^2}, \tag{E.6}$$

and for $a > 0$, $a + b > 0$

$$\left(a\,\mathrm{Id} + b\frac{\gamma\otimes\gamma}{|\gamma|^2}\right)^{-1} = \frac{1}{a}\left(\mathrm{Id} - \frac{b}{a+b}\frac{\gamma\otimes\gamma}{|\gamma|^2}\right), \tag{E.7}$$

we obtain

$$\begin{aligned}&\left(\mathrm{Id} + \frac{F''(\gamma) - \eta_c\,\mathrm{Id}}{r}\right)^{-1}\\ &\qquad = \frac{r}{(r - \eta_c + \eta + \kappa|\gamma|^{n-1})|\gamma| + \sigma_{yield}}\left(|\gamma|\,\mathrm{Id} - \frac{(n-1)\kappa|\gamma|^n - \sigma_{yield}}{r - \eta_c + \eta + n\kappa|\gamma|^{n-1}}\frac{\gamma\otimes\gamma}{|\gamma|^2}\right).\end{aligned} \tag{E.8}$$

To obtain the diffusion matrix we use the formula in the right-hand side of (3.5). We remark that this diffusion matrix tends to $r\,\mathrm{Id}$ as $\gamma \to 0$ in the case $0 < n < 1$ and $\kappa > 0$.

Appendix F. Consistency error and strong error

We provide here estimates showing that the consistency error (2.59) controls the distance in norm between the approximate solution u_k and the exact discrete solution $\widehat{u}$, as does the primal-dual gap in the estimate (2.73).

Let us define

$$\overline{\gamma}_k = S_0(\sigma_k), \qquad \gamma_k = S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k), \tag{F.1}$$

where the nonlinear functions S_t are defined by (2.6).

Proposition F.1 (Consistency error and strong error)**.** *We assume that F satisfies the coercivity assumptions* (2.1), (2.2)*, and suppose that there is a solution $(\widehat{u}, \widehat{\sigma}) \in V_h \times M_h$ to* (2.25), (2.26)*. If the parameters $\overline{\eta}$, r_k satisfy* (2.5), (2.10) *then the sequence of approximate solutions $(u_k, \sigma_k)_{k=0,\dots} \in V_h \times M_h$ defined by* (2.27), (2.28) *satisfy the estimates*

$$\begin{cases} \|Du_k - \overline{\gamma}_k\| \leq \dfrac{r_k}{\eta_c}\|Du_k - \gamma_k\|, \\ \|Du_k - \gamma_k\| \leq \left(2 - \dfrac{\eta_c}{r_k}\right)\|Du_k - \overline{\gamma}_k\|, \end{cases} \tag{F.2}$$

$$\alpha\|u_k - \widehat{u}\|^2 + \frac{\eta_c}{2}\|Du_k - D\widehat{u}\|^2 \leq \left(\frac{r_k^2}{2\eta_c} + \frac{\eta_c}{2}\right)\|\gamma_k - Du_k\|^2 + \|\sigma_k - \widehat{\sigma}\|\|\gamma_k - Du_k\|. \tag{F.3}$$

Remark F.2. The consistency error in (2.59) is $\|Du_k - \overline{\gamma}_k\|$. The term $\|Du_k - \gamma_k\|$ is the one of (2.58) without the factor r_k, it is another type of consistency error more adapted to the definition of the iterative scheme (2.28). The estimate (F.2) shows that these two quantities are comparable, except that the ratio of their respective magnitude may vary significantly if r_k is large. The estimate (F.3) proves that $u_k - \widehat{u}$ tends to zero as $r_k\|Du_k - \gamma_k\|$ tends to zero (this quantity is exactly the one of (2.58)), or equivalently if $\|\sigma_{k+1} - \sigma_k\| \to 0$ and $r_k\|Du_{k+1} - Du_k\| \to 0$ (see (2.57)). It holds also if the consistency error $\|Du_k - \overline{\gamma}_k\|$ tends to zero if r_k is not too large.

Proof of Proposition F.1. For (F.2), we write that according to (F.1) and (2.7)

$$\partial F(\gamma_k) \ni \sigma_k + (r_k - \eta_c)(Du_k - \gamma_k), \tag{F.4}$$

which implies that

$$\gamma_k = S_0(\sigma_k + (r_k - \eta_c)(Du_k - \gamma_k)). \tag{F.5}$$

Since $\overline{\gamma}_k = S_0(\sigma_k)$ and by Lemma 2.8, S_0 is $1/\eta_c$ Lipschitz continuous, we get

$$\|\gamma_k - \overline{\gamma}_k\| \leq \frac{r_k - \eta_c}{\eta_c}\|Du_k - \gamma_k\|. \tag{F.6}$$

Then we write $\|Du_k - \overline{\gamma}_k\| \leq \|Du_k - \gamma_k\| + \|\gamma_k - \overline{\gamma}_k\|$, which yields the first estimate of (F.2). Next, we can write symmetrically $\partial F(\overline{\gamma}_k) \ni \sigma_k$, which gives $\partial F(\overline{\gamma}_k) + (r_k - \eta_c)\overline{\gamma}_k \ni \sigma_k + (r_k - \eta_c)\overline{\gamma}_k$, thus

$$\overline{\gamma}_k = S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)\overline{\gamma}_k). \tag{F.7}$$

Since $\gamma_k = S_{r_k-\eta_c}(\sigma_k + (r_k - \eta_c)Du_k)$ and $S_{r_k-\eta_c}$ is $1/r_k$ Lipschitz continuous, we get

$$\|\overline{\gamma}_k - \gamma_k\| \leq \frac{r_k - \eta_c}{r_k}\|\overline{\gamma}_k - Du_k\|. \tag{F.8}$$

Writing $\|Du_k - \gamma_k\| \leq \|Du_k - \overline{\gamma}_k\| + \|\overline{\gamma}_k - \gamma_k\|$, we obtain the second estimate of (F.2).
To prove (F.3), we write the exact conservation for u_k and $\widehat{u}$ according to (2.27) and (2.25): $u_k, \widehat{u} \in V_h$,

$$\begin{aligned} \alpha\,\langle u_k, v\rangle + \langle \sigma_k, Dv\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h, \\ \alpha\,\langle \widehat{u}, v\rangle + \langle \widehat{\sigma}, Dv\rangle - \langle f, v\rangle = 0, \quad \forall\, v \in V_h. \end{aligned} \tag{F.9}$$

Taking the difference we obtain

$$\alpha\,\langle u_k - \widehat{u}, v\rangle + \Big\langle(\sigma_k + (r_k - \eta_c)Du_k) - (\widehat{\sigma} + (r_k - \eta_c)D\widehat{u}), Dv\Big\rangle = (r_k - \eta_c)\,\langle Du_k - D\widehat{u}, Dv\rangle\,. \tag{F.10}$$

We write

$$\begin{aligned} \partial F(\gamma_k) + (r_k - \eta_c)\gamma_k \ni \sigma_k + (r_k - \eta_c) Du_k, \\ \partial F(D\widehat{u}) + (r_k - \eta_c) D\widehat{u} \ni \widehat{\sigma} + (r_k - \eta_c) D\widehat{u}, \end{aligned} \tag{F.11}$$

which leads by monotonicity of $\partial F - \eta_c \,\mathrm{Id}$ to

$$\Big\langle (\sigma_k + (r_k - \eta_c) Du_k) - (\widehat{\sigma} + (r_k - \eta_c) D\widehat{u}), \gamma_k - D\widehat{u} \Big\rangle \geq r_k \|\gamma_k - D\widehat{u}\|^2. \tag{F.12}$$

Choosing $v = u_k - \widehat{u}$ in (F.10) and taking into account (F.12) we get

$$\begin{aligned} \alpha\|u_k - \widehat{u}\|^2 + r_k\|\gamma_k - D\widehat{u}\|^2 + \Big\langle (\sigma_k + (r_k - \eta_c) Du_k) - (\widehat{\sigma} + (r_k - \eta_c) D\widehat{u}), Du_k - \gamma_k \Big\rangle \\ \leq (r_k - \eta_c)\|Du_k - D\widehat{u}\|^2. \end{aligned} \tag{F.13}$$

Writing $\gamma_k - D\widehat{u} = (Du_k - D\widehat{u}) + (\gamma_k - Du_k)$, one has

$$\|\gamma_k - D\widehat{u}\|^2 = \|Du_k - D\widehat{u}\|^2 + 2\left\langle Du_k - D\widehat{u}, \gamma_k - Du_k \right\rangle + \|\gamma_k - Du_k\|^2. \tag{F.14}$$

Multiplying this by r_k and using the result in (F.13) yields

$$\begin{aligned} \alpha\|u_k - \widehat{u}\|^2 + \eta_c\|Du_k - D\widehat{u}\|^2 + 2r_k \left\langle Du_k - D\widehat{u}, \gamma_k - Du_k \right\rangle \\ + \Big\langle (\sigma_k + (r_k - \eta_c) Du_k) - (\widehat{\sigma} + (r_k - \eta_c) D\widehat{u}), Du_k - \gamma_k \Big\rangle \leq -r_k\|\gamma_k - Du_k\|^2, \end{aligned} \tag{F.15}$$

thus

$$\begin{aligned} \alpha\|u_k - \widehat{u}\|^2 + \eta_c\|Du_k - D\widehat{u}\|^2 \\ \leq \left\langle \sigma_k - \widehat{\sigma}, \gamma_k - Du_k \right\rangle - (\eta_c + r_k)\left\langle Du_k - D\widehat{u}, \gamma_k - Du_k \right\rangle - r_k\|\gamma_k - Du_k\|^2. \end{aligned} \tag{F.16}$$

We write the Hölder inequality

$$(\eta_c + r_k)\Big|\left\langle Du_k - D\widehat{u}, \gamma_k - Du_k \right\rangle\Big| \leq \frac{\eta_c}{2}\|Du_k - D\widehat{u}\|^2 + \frac{(\eta_c + r_k)^2}{2\eta_c}\|\gamma_k - Du_k\|^2, \tag{F.17}$$

which in (F.16) yields (F.3) and concludes the proof. ∎

References

[1] Vincent Acary, Paul Armand, Hoang Minh Nguyen, and Maksym Shpakovych. Second-order cone programming for frictional contact mechanics using interior point algorithm. *Optim. Methods Softw.*, 39(3):634–663, 2024.

[2] Kenneth J. Arrow, Leonid Hurwicz, and Hirofumi Uzawa. *Studies in linear and non-linear programming*, volume 2 of *Stanford Mathematical Studies in the Social Sciences*. Stanford University Press, 1958. with contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow.

[3] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[4] Amir Beck and Marc Teboulle. A fast dual proximal gradient algorithm for convex minimization and applications. *Oper. Res. Lett.*, 42(1):1–6, 2014.

[5] Clément Berger. *Fluides à seuil : interactions modèles et données.* PhD thesis, Ecole Normale Supérieure de Lyon, France, 2024.

[6] Alfredo Bermudez and Carlos Moreno. Duality methods for solving variational inequalities. *Comput. Math. Appl.*, 7:43–58, 1981.

[7] François Bouchut, Carsten Carstensen, and Alexandre Ern. H^1 regularity of the minimizers for the inviscid total variation and Bingham fluid problems for H^1 data. *Nonlinear Anal., Theory Methods Appl.*, 258: article no. 113809 (21 pages), 2025.

[8] François Bouchut, Robert Eymard, and Alain Prignet. Convergence of conforming approximations for inviscid incompressible Bingham fluid flows and related problems. *J. Evol. Equ.*, 14(3):635–669, 2014.

[9] Miroslav Bulíček, Erika Maringová, and Josef Málek. On nonlinear problems of parabolic type with implicit constitutive equations involving flux. *Math. Models Methods Appl. Sci.*, 31(10):2039–2090, 2021.

[10] Rénald Chalayer, Laurent Chupin, and Thierry Dubois. A bi-projection method for incompressible Bingham flows with variable density, viscosity, and yield stress. *SIAM J. Numer. Anal.*, 56(4):2461–2483, 2018.

[11] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1):120–145, 2011.

[12] Yunmei Chen, Guanghui Lan, and Yuyuan Ouyang. Optimal primal-dual methods for a class of saddle point problems. *SIAM J. Optim.*, 24(4):1779–1814, 2014.

[13] Wei Deng and Wotao Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, 66(3):889–916, 2016.

[14] Ivar Ekeland and Roger Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1999. unabridged, corrected republication of the 1976 English original.

[15] Maurice Fabien, Johnny Guzmán, Michael Neilan, and Ahmed Zytoon. Low-order divergence-free approximations for the Stokes problem on Worsey–Farin and Powell–Sabin splits. *Comput. Methods Appl. Mech. Eng.*, 390: article no. 114444 (21 pages), 2022.

[16] Enrique D. Fernández-Nieto, José M. Gallardo, and Paul Vigneaux. Efficient numerical schemes for viscoplastic avalanches. Part 2: The 2D case. *J. Comput. Phys.*, 353:460–490, 2018.

[17] François Févotte, Ari Rappaport, and Martin Vohralík. Adaptive regularization, discretization, and linearization for nonsmooth problems based on primal-dual gap estimators. *Comput. Methods Appl. Mech. Eng.*, 418B: article no. 116558 (33 pages), 2024.

[18] Martin Fuchs and Gregory Seregin. *Variational methods for problems from plasticity theory and for generalized Newtonian fluids*, volume 1749 of *Lecture Notes in Mathematics*. Springer, 2000.

[19] Roland Glowinski. *Numerical methods for nonlinear variational problems*. Scientific Computation. Springer, 2008.

[20] Roland Glowinski. On alternating direction methods of multipliers: a historical perspective. In *Modeling, Simulation and Optimization for Science and Technology*, volume 34 of *Computational Methods in Applied Sciences*, pages 59–82. Springer, 2014.

[21] Roland Glowinski, Jacques-Louis Lions, and Raymond Trémolières. *Numerical analysis of variational inequalities*, volume 8 of *Studies in Mathematics and its Applications*. North-Holland, 1981.

[22] Friedrich Hecht. New development in FreeFem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[23] Duc Nguyen Hoai. *Analysis and approximation of compressible viscoplastic models with general nonlinearity for granular flows*. PhD thesis, Univ. Gustave Eiffel, 2020. `https://theses.hal.science/tel-03078670`.

[24] Nicolas Roquet and Pierre Saramito. An adaptive finite element method for Bingham fluid flows around a cylinder. *Comput. Methods Appl. Mech. Eng.*, 192(31-32):3317–3341, 2003.

[25] Pierre Saramito. A damped Newton algorithm for computing viscoplastic fluid flows. *J. Non-Newton. Fluid Mech.*, 238:6–15, 2016.

[26] Pierre Saramito and Anthony Wachs. Progress in numerical simulation of yield stress fluid flows. *Rheol. Acta*, 56:211–230, 2017.

[27] Timm Treskatis, Miguel A. Moyers-González, and Chris J. Price. An accelerated dual proximal gradient method for applications in viscoplasticity. *J. Non-Newton. Fluid Mech.*, 238:115–130, 2016.

[28] Timm Treskatis, Ali Roustaei, Ian Frigaard, and Anthony 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-Newton. Fluid Mech.*, 262:149–164, 2018.