A family of predictor-corrector schemes for a class of variational inequalities

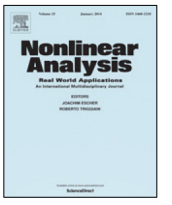
J K Djoko, Jonas Koko, B. Daya Reddy

Nonlinear Analysis: Real World Applications 73 (2023) 103881

Contents lists available at ScienceDirect

Nonlinear Analysis: Real World Applications

www.elsevier.com/locate/nonrwa



A family of predictor–corrector schemes for a class of variational inequalities

J.K. Djoko [a], J. Koko [b], B.D. Reddy [c,*]

[a] *African Peer Review Mechanism, Johannesburg, South Africa*
[b] *LIMOS, Université Clermont-Auvergne – CNRS UMR 6158, B.P 10448, 63000 Clermont-Ferrand, France*
[c] *Department of Mathematics and Applied Mathematics, University of Cape Town, 7701 Rondebosch, South Africa*

ARTICLE INFO

Article history:
Received 21 November 2022
Received in revised form 22 February 2023
Accepted 6 March 2023
Available online xxxx

Dedicated to Weimin Han on the occasion of his sixtieth birthday

Keywords:
Predictor-corrector scheme
Variational inequalities
Steepest descent scheme
A Newton-like method
Convergence

ABSTRACT

This work is concerned with an abstract problem in the form of a variational inequality, or equivalently a minimization problem involving a non-differential functional. The problem is inspired by a formulation of the initial–boundary value problem of elastoplasticity. The objective of this work is to revisit the predictor–corrector algorithms that are commonly used in computational applications, and to establish conditions under which these are convergent or, at least, under which they lead to decreasing sequences of the functional for the problem. The focus is on the predictor step, given that the corrector step by definition leads to a decrease in the functional. The predictor step may be formulated as a minimization problem. Attention is given to the tangent predictor, a line search approach, the method of steepest descent, and a Newton-like method. These are all shown to lead to decreasing sequences.

1. Introduction

The theory of elastoplastic solids has yielded an abundance of interesting and important mathematical problems, and the various formulations of the initial–boundary value problem have been the subject of a range of corresponding computational approaches and associated algorithms (see for example [1,2] and the references in these works). The elastoplasticity problem takes the form of a variational inequality, of one of two kinds, depending on the manner in which the flow relation is treated. Whatever the formulation, the nonlinear nature of the problem at its simplest is a result of behaviour being either elastic and reversible, or plastic and irreversible, with the relevant branch not being known a priori.

* Corresponding author.
E-mail addresses: jules.djokokamdem@gmail.com (J.K. Djoko), jonas.koko@uca.fr (J. Koko), daya.reddy@uct.ac.za (B.D. Reddy).

https://doi.org/10.1016/j.nonrwa.2023.103881
1468-1218/

Consequently, a challenge lies in constructing solution algorithms which are efficient and convergent. The most popular algorithms are of predictor–corrector form, with the predictor step entailing an approximate solution of the equilibrium equation and the corrector step a solution to the variational inequality associated with the flow relation. Approximations in the first step lead to elastic, secant or tangent predictors. The corrector step is known in the computational literature as the return map, and entails either the solution of a variational inequality of the second kind, if formulated in what is known as the primal form [1], that is, in terms of the dissipation or support function, or alternatively the dual form, in which this step is formulated using the yield function or, in the language of convex analysis, the indicator function of the set defining the elastic region [2,3].

While the algorithms for the abstract problem corresponding to an elastic or secant predictor have been shown to be convergent [4], the same has not been established for the tangent predictor, which is known from computational experiments to converge most rapidly. Modifications of this predictor in the form of a perturbation or line search procedure, on the other hand, lead to guaranteed convergence.

This work focuses on an abstract version of the primal problem of elastoplasticity, in the form of a minimization problem involving a non-differential functional or, equivalently, a variational inequality in two variables. The objective here is to explore some of the algorithms in use, to review the conditions under which these lead to decreasing sequences of the functional, or indeed to convergence. In particular, existing results for the tangent predictor are complemented by those establishing the conditions under which line search, steepest descent, or Newton-like methods lead to minimizing sequences, in the sense that the functional value reduces with each iteration. Full convergence studies for these latter methods remain open, and will be the subject of a subsequent work.

The structure of the rest of this work is as follows. In Section 2 the abstract variational problem is formulated, and conditions for its well-posedness presented. Section 3 is concerned with properties of the predictor–corrector algorithm, in particular, the conditions for solution sequences to be decreasing. In Section 4 attention is turned first to the line search procedure, which is shown to lead to a decreasing sequence. Corresponding results follow for the steepest descent and Newton-like approaches.

2. Description of the problem and predictor–corrector algorithm

2.1. Formulation of the abstract problem

We begin by introducing various spaces, functionals and assumptions.

(h_1) Let V and Λ be two Hilbert spaces, and set $W = V \times \Lambda$.

(h_2) Let $b : V \times V \to \mathbb{R}$, $c : V \times \Lambda \to \mathbb{R}$ and $d : \Lambda \times \Lambda \to \mathbb{R}$ be continuous bilinear forms, with $b(\cdot,\cdot)$ and $d(\cdot,\cdot)$ symmetric, and assume the existence of positive constants $c_0, \widetilde{c}$ and k_0 such that for all $u, v \in V$ and $\lambda, \mu \in \Lambda$,

$$\begin{aligned} b(u,v) \leq & \widetilde{c}\|u\|_V\|v\|_V\,, \quad b(v,v) \geq c_0\|v\|_V^2\,, \\ d(\lambda,\lambda) \geq & (c_0+\kappa_0)\|\lambda\|_\Lambda^2, \quad d(\lambda,\mu) \leq \widetilde{c}\|\lambda\|_\Lambda\|\mu\|_\Lambda, \\ c(v,\mu) \leq & \widetilde{c}\|v\|_V\|\mu\|_\Lambda\,. \end{aligned} \tag{2.1}$$

(h_3) Let $\ell_1 : V \to \mathbb{R}$ and $\ell_2 : \Lambda \to \mathbb{R}$ be continuous linear functionals.

(h_4) Let $j : \Lambda \to \mathbb{R}$ be a non-negative convex functional, twice Gateaux-differentiable except at zero.

We consider the following abstract problem: Find $(u, \lambda) \in V \times \Lambda$ such that

$$b(u,v) - c(\lambda, v) = \ell_1(v) \quad \text{for all } v \in V, \tag{2.2a}$$

$$d(\lambda, \mu - \lambda) - c(\mu - \lambda, u) + j(\mu) - j(\lambda) \geq \ell_2(\mu - \lambda) \quad \text{for all } \mu \in \Lambda\,. \tag{2.2b}$$

It is noted that, with assumptions $(h_1) - (h_4)$, the variational problem (2.2) has a unique solution $w := (u, \lambda) \in W$ [5] (see also [1,6] for details of the proof).

For $w = (u, \lambda)$ and $z = (v, \mu)$ elements of W, we define

$$\begin{cases} a(w, z) = b(u, v) - c(\lambda, v) - c(\mu, u) + d(\lambda, \mu), \\ \ell(z) = \ell_1(v) + \ell_2(\mu), \\ j(z) = j(\mu)\,. \end{cases} \tag{2.3}$$

Here, for convenience, we write $j(z)$ for the functional $\bar{j} : W \to \mathbb{R}$ defined by $\bar{j}(z) = j(\mu)$. We note that the pair (2.2) is equivalent to the variational inequality

$$\begin{cases} \text{Find } w = (u, \lambda) \in W \text{ such that for all } z = (v, \mu) \in W\,, \\ a(w, z - w) + j(z) - j(w) - \ell(z - w) \geq 0\,. \end{cases} \tag{2.4}$$

Now since the bilinear form $a(\cdot, \cdot)$ is symmetric, (2.4) is equivalent to the minimization problem

$$\begin{cases} \text{Find } w = (u, \lambda) \in W \text{ such that for all } z = (v, \mu) \in W\,, \\ \mathcal{L}(w) \leq \mathcal{L}(z), \quad \text{with} \\ \mathcal{L}(z) = \frac{1}{2} a(z, z) + j(z) - \ell(z)\,. \end{cases} \tag{2.5}$$

2.2. A predictor–corrector solution strategy

We will later explore specific forms of a predictor–corrector approach to the solution of (2.4) or (2.5); the approach has been motivated by problems in elastoplasticity (see [1] and the references cited therein). Assuming an algorithm in which steps are indexed by i $(i = 1, 2, \ldots)$, the objective is to generate a sequence of solutions $w^i := (u^i, \lambda^i)$, given $w^{i-1} = (u^{i-1}, \lambda^{i-1})$, such that $w^i \to w$ as $i \to \infty$.

For the predictor step we need the quadratic functional $\mathcal{L}^{(i)}$ defined by

$$\mathcal{L}^{(i)}(z) = \tfrac{1}{2} a(z, z) + j^{(i)}(z) - \ell(z) \tag{2.6}$$

where $j^{(i)}$ is the Taylor approximation of order two of j at w^{i-1} (assuming that $j(\mu)$ is twice differentiable for $\mu \neq 0$); that is,

$$j^{(i)}(z) = j(w^{i-1}) + (j'(w^{i-1}), z - z^{i-1}) + \tfrac{1}{2} j''(w^{i-1})(z - w^{i-1}, z - w^{i-1})\,, \tag{2.7}$$

where $(\ )'$ and $(\ )''$ denote first and second Gateaux derivatives. The first derivative is treated as a member of W, in view of the Riesz isometry between this product space and its dual. Furthermore, here and in what follows the inner product and norm on $V \times \Lambda$ are denoted by $(\cdot, \cdot)$ and $\|\cdot\|$, respectively. The second Gateaux derivative is a bilinear operator on the product space.

The functional $\mathcal{L}^{(i)}$ is quadratic and twice Gateaux differentiable, with first and second Gateaux derivatives given by

$$\begin{aligned} (\mathcal{L}^{(i)\prime}(z), w) &= a(z, w) + (j'(w^{i-1}), w) + j''(w^{i-1})(w, z - z^{i-1}) - \ell(w)\,, \\ \mathcal{L}^{(i)\prime\prime}(z)(w, e) &= a(e, w) + j''(w^{i-1})(w, e)\,. \end{aligned} \tag{2.8}$$

The predictor–corrector scheme is as follows. Choose $(u^0, \lambda^0) \in V \times \Lambda$. For $i = 1, \ldots$ and given iterates $(u^{i-1}, \lambda^{i-1}) \in V \times \Lambda$,

Predictor step: Find $(u^i, \lambda^{*i}) \in W$ the solution of

$$\mathcal{L}^{(i)}(u^i, \lambda^{*i}) = \inf_{(v,\mu) \in W} \mathcal{L}^{(i)}(v, \mu)\,. \tag{2.9}$$

Corrector step: Knowing u^i, find $\lambda^i \in \Lambda$ the solution of

$$\mathcal{L}(u^i, \lambda) = \inf_{\mu \in \Lambda} \mathcal{L}(u^i, \mu) \tag{2.10}$$

where $\mathcal{L}$ is given in (2.5).

Equivalently, using the optimality condition, (2.9) and (2.10) may be written as follows: Find $(u^i, \lambda^{*i}) \in W$ and $\lambda^i \in \Lambda$ such that

$$\begin{aligned} &\mathcal{L}^{(i)\prime}(u^i, \lambda^{*i})(v, \mu) = 0 \quad \text{for all } (v, \mu) \in W\,, \\ &0 \in \partial_\mu \mathcal{L}(u^i, \lambda^i)\,, \end{aligned} \tag{2.11}$$

where $\partial_\mu \mathcal{L}$ denotes the subdifferential of $\mathcal{L}$ with respect to the second variable. Using (2.3) and (2.5), the algorithm is as follows:

Predictor step: Find $(u^i, \lambda^{*i}) \in V \times \Lambda$ that solve

$$\begin{cases} b(u^i, v) - c(\lambda^{*i}, v) = \ell_1(v) \quad \text{for all } v \in V\,, \\ -c(\mu, u^i) + d(\lambda^{*i}, \mu) + j''(\lambda^{i-1})(\lambda^{*i}, \mu) = \ell_2(\mu) \quad \text{for all } \mu \in \Lambda\,. \end{cases} \tag{2.12}$$

Corrector step: Find $\lambda^i \in \Lambda$ that solves

$$d(\lambda^i, \mu - \lambda^i) - c(\mu - \lambda^i, u^i) + j(\mu) - j(\lambda^i) \geq \ell_2(\mu - \lambda^i) \quad \text{for all } \mu \in \Lambda\,. \tag{2.13}$$

Remark 2.1. See [1, Section 12.2] for concrete examples of the above abstract formulation, for both conventional and strain-gradient plasticity.

Remark 2.2. The optimality condition $(2.11)_2$ or (2.13) can be approximated by a variational equation in which the functional j is replaced by a differentiable functional j_ε with ε a small parameter such that $j_\varepsilon \to j$ as $\varepsilon \to 0$. Then the approximation of (2.10) is

$$\mathcal{L}(u^i, \lambda_\varepsilon) = \inf_{\mu} \mathcal{L}_\varepsilon(u^i, \mu)\,,$$

in which

$$\mathcal{L}_\varepsilon(z) = \tfrac{1}{2} a(z, z) + j_\varepsilon(z) - \ell(z).$$

This approach has been used with success in various applications, for example, [7] for problems in plasticity and [6] for Bingham flow (see also [1, sect. 12.4], for various examples of regularization and their convergence properties).

Remark 2.3. The pair of Eqs. (2.12) can be reduced to a single equation for the unknown variable u, as is conventionally done computationally. To do this, define the set of operators

$$\begin{aligned} &\mathcal{B} : V \to \widehat{V}, && \langle \mathcal{B}u, v\rangle = b(u, v) \\ &\mathcal{C} : V \to \widehat{\Lambda} \text{ and } \widehat{\mathcal{C}} : \Lambda \to \widehat{V}, && \langle \mathcal{C}v, \mu\rangle = \langle \widehat{\mathcal{C}}\mu, v\rangle = c(\mu, v)\,, \\ &\mathcal{D} : \Lambda \to \widehat{\Lambda}, && \langle \mathcal{D}\lambda, \mu\rangle = d(\lambda, \mu)\,, \\ &\mathcal{G} : \Lambda \to \widehat{\Lambda}, && \langle \mathcal{G}\lambda, \mu\rangle = j''(\lambda^{i-1})(\lambda, \mu)\,. \end{aligned}$$

Here $\widehat{V}$ and $\widehat{\Lambda}$ denote respectively the topological duals of V and Λ, and $\widehat{\mathcal{C}}$ is the dual operator of $\mathcal{C}$. From the properties (2.1) the operators $\mathcal{B}$, $\mathcal{C}, \mathcal{D}$ and $\mathcal{G}$ are bounded, and $\mathcal{B}$, $\mathcal{D}$ and $\mathcal{G}$ are bounded below, and hence invertible. With these definitions Eqs. (2.12) for the predictor step can be rewritten as

$$\mathcal{B}u^i - \widehat{\mathcal{C}}\lambda^{*i} = \ell_1\,, \tag{2.14a}$$

$$-\mathcal{C}u^i + \mathcal{D}\lambda^{*i} + \mathcal{G}\lambda^{*i} = \ell_2. \tag{2.14b}$$

Elimination of λ^{*i} gives the equation

$$[\mathcal{B} - \widehat{\mathcal{C}}(\mathcal{D}+\mathcal{G})^{-1}\mathcal{C}]u^i = \ell_1 + \widehat{\mathcal{C}}(\mathcal{D}+\mathcal{B})^{-1}\ell_2 \tag{2.15}$$

for u^i. The discrete analogue of the operator on the left hand side of (2.15) is referred to as the consistent tangent.

Remark 2.4. Discrete and computational approaches in plasticity conventionally formulate the flow relation (2.12) or (2.14b) in dual form. To obtain this form, (2.14b) is rewritten as

$$j^{(i)\prime}(\lambda^{*i}) = \mathcal{C}u^i - \mathcal{D}\lambda^{*i} + \ell_2 := \tau \in \widehat{\Lambda}\,.$$

The dual form of this expression is

$$\lambda^{*i} = \widehat{j^{(i)}}'(\tau)\,, \tag{2.16}$$

where $\widehat{j^{(i)}}$ denotes the Legendre–Fenchel convex conjugate of $j^{(i)}$. Given the quadratic form of $j^{(i)}$, its conjugate is found in a straightforward manner and is given by

$$\widehat{j^{(i)}}(\tau) = \tfrac{1}{2}\langle \tau - j'(\lambda^{i-1}), \mathcal{G}^{-1}(\tau - j'(\lambda^{i-1}))\rangle - j(\lambda^{i-1}).$$

Then Eq. (2.16) gives

$$\lambda^{*i} = \mathcal{G}^{-1}(\tau - j'(\lambda^{i-1}))\,. \tag{2.17}$$

As with the version (2.15), the expression (2.17) may be used to eliminate λ^{*i} from (2.14a).

3. Convergence: Preliminary approach

The goal here is to study the conditions under which $(\mathcal{L}(u^i,\lambda^i))_i$ is a decreasing sequence. Indeed, the sequence $(\mathcal{L}(u^i,\lambda^i))_i$ is bounded from below by $\mathcal{L}(u,\lambda)$. These two properties imply that $(\mathcal{L}(u^i,\lambda^i))_i$ is convergent to some limit $(\geq \mathcal{L}(u,\lambda))$.

We first revisit the result presented in [1, Section 12.2].

Let $(u^i,\lambda^i)_i$ be the sequence given by the relations (2.12), (2.13). Then

$$\mathcal{L}(u^i,\lambda^i) - \mathcal{L}(u^{i-1},\lambda^{i-1}) = \mathcal{L}(u^i,\lambda^i) - \mathcal{L}(u^i,\lambda^{*i}) + \mathcal{L}(u^i,\lambda^{*i}) - \mathcal{L}(u^{i-1},\lambda^{i-1})\,. \tag{3.1}$$

Clearly from (2.10), $\mathcal{L}(u^i,\lambda^i) - \mathcal{L}(u^i,\lambda^{*i})$ is non-positive, and hence (3.1) implies that

$$\mathcal{L}(u^i,\lambda^i) - \mathcal{L}(u^{i-1},\lambda^{i-1}) \leq \mathcal{L}(u^i,\lambda^{*i}) - \mathcal{L}(u^{i-1},\lambda^{i-1})\,.$$

From the definition of $\mathcal{L}$ and $\mathcal{L}^{(i)}$ we have

$$\mathcal{L}(u^i,\lambda^{*i}) - \mathcal{L}(u^{i-1},\lambda^{i-1}) = \mathcal{L}^{(i)}(u^i,\lambda^{*i}) - \mathcal{L}^{(i)}(u^{i-1},\lambda^{i-1}) + j(\lambda^{*i}) - j^{(i)}(\lambda^{*i})\,. \tag{3.2}$$

Now using (2.9), one has $\mathcal{L}^{(i)}(u^i,\lambda^{*i}) - \mathcal{L}^{(i)}(u^{i-1},\lambda^{i-1}) \leq 0$. Hence (3.2) implies that

$$\mathcal{L}(u^i,\lambda^{*i}) - \mathcal{L}(u^{i-1},\lambda^{i-1}) \leq j(\lambda^{*i}) - j^{(i)}(\lambda^{*i})\,. \tag{3.3}$$

Considering the definition of $j^{(i)}(\lambda^{*i})$ (see (2.7)), and setting $\Delta\lambda^{*i} := \lambda^{*i} - \lambda^{i-1}$, one obtains

$$j(\lambda^{*i}) - j^{(i)}(\lambda^{*i}) = \underbrace{j(\lambda^{*i}) - j(\lambda^{i-1}) - (j'(\lambda^{i-1}), \Delta\lambda^{*i})}_{\geq 0} - \underbrace{\tfrac{1}{2}j''(\lambda^{i-1})(\Delta\lambda^{*i},\Delta\lambda^{*i})}_{\geq 0}\,, \tag{3.4}$$

the signs of the two sets of terms on the right hand side following from the convexity of j.

Thus, the right hand side of (3.4) is non-positive if

$$j(\lambda^{*i}) - j(\lambda^{i-1}) - (j'(\lambda^{i-1}), \Delta\lambda^{*i}) \leq \tfrac{1}{2} j''(\lambda^{i-1})(\Delta\lambda^{*i}, \Delta\lambda^{*i})\,. \tag{3.5}$$

The requirement (3.5) is very restrictive even if the functional j is convex and twice Gateaux-differentiable at λ^{i-1}. Also, the fact that λ^{*i} is an unknown makes the fulfilment of (3.5) difficult. Thus, the sign of $j(\lambda^{*i}) - j^{(i)}(\lambda^{*i})$ is unclear and the tangent predictor in its original form (2.12) might not produce a decreasing sequence.

As an alternative, one may make use of a perturbed Taylor expansion of j about λ^{i-1}, of the form [1, Section 12.3]

$$j_\varepsilon^{(i)}(\lambda) = j(\lambda^{i-1}) + j'(\lambda^{i-1})(\lambda - \lambda^{i-1}) + \tfrac{1}{2} j''(\lambda^{i-1})(\lambda - \lambda^{i-1}, \lambda - \lambda^{i-1}) + \varepsilon\|\lambda - \lambda^{i-1}\|^2 \tag{3.6}$$

with ε chosen so that

$$j(\mu) \leq j_\varepsilon^{(i)}(\mu) \quad \text{for all}\, \mu \in \Lambda \tag{3.7}$$

holds, at least in a neighbourhood of λ^{i-1}. It is clear that with (3.6), and (3.7) enforced locally, (3.3) implies that

$$\mathcal{L}(u^i, \lambda^{*i}) - \mathcal{L}(u^{i-1}, \lambda^{i-1}) \leq \quad j(\lambda^{*i}) - j_\varepsilon^{(i)}(\lambda^{*i}) \leq 0.$$

Remark 3.1. If $j''(\lambda^{i-1}) = 0$, then (2.12) becomes

$$\begin{cases} b(u^i, v) - c(\lambda^{*i}, v) = \ell_1(v) \quad \text{for all}\, v \in V\,, \\ -c(\mu, u^i) + d(\lambda^{*i}, \mu) = \ell_2(\mu) \quad \text{for all}\, \mu \in \Lambda\,, \end{cases}$$

which is known as the conventional tangent predictor. The relation (3.3) remains valid, and (3.4) becomes

$$j(\lambda^{*i}) - j^{(i)}(\lambda^{*i}) = j(\lambda^{*i}) - j(\lambda^{i-1}) - (j'(\lambda^{i-1}), \Delta\lambda^{*i}) \tag{3.8}$$

which is non-negative because of the convexity and Gateaux-differentiability of j. Hence we cannot conclude that the sequence $\mathcal{L}(u^i, \lambda^i)$ is decreasing.

A convergence result. The following result was proved in [4,8] (see also [1], Section 12.3).

Theorem 3.1. *Under the assumptions (2.1), for $j^{(i)}$ chosen such that*

$$j^{(i)}(\lambda^{i-1}) = j(\lambda^{i-1})\,, \quad j'^{(i)}(\lambda^{i-1}) = j'(\lambda^{i-1})\,, \quad j(\mu) \leq j^{(i)}(\mu) \;\; \forall \mu \in \Lambda\,, \tag{3.9}$$

and with the condition $r := \tilde{c}^2/[c_0(c_0 + \kappa_0)] < \frac{1}{3}$, where the constants are given in (2.1), the algorithm (2.12), (2.13) converges.

Thus in particular the algorithm with the perturbed tangent predictor (3.6) converges.

We explore next various approaches based on descent methods, and establish conditions under which the algorithms lead to decreasing sequences.

4. Descent methods

The descent method or line search approach to the predictor step (see (2.9)) (see for example [9], Sections 10.4 and 10.5) entails the generation of a sequence of iterates $w^{*i} := \{u^i, \lambda^{*i}\}$ with

$$w^{*i} = w^{i-1} - \alpha_i \hat{w}^i\,, \tag{4.1}$$

where α_i is non-negative and represents the step size, and $\hat{w}^i$ is the search direction. Different choices of α_i and $\hat{w}^i$ will determine different methods of descent. It is clear that in each iteration we have two stages:

(i) choice of a descent search direction $\hat{w}^i$;
(ii) choice of the step size α_i along the search direction $\hat{w}^i$.

We now consider some specific examples of descent methods.

4.1. Line search

The line search approach proposed in [10] was analysed for the discrete case in [11], and is presented here in the present Hilbert space context. The main idea is to treat the solution to (2.9) for the tangent predictor as an intermediate one, then to seek an improvement in the form given by (4.1).

We show now that it is possible to find α satisfying $0 < \alpha < 1$ and a direction $\hat{w}^i$ such that

$$\mathcal{L}(w^{*i}) \leq \mathcal{L}(w^{i-1})\,. \tag{4.2}$$

Thus the predictor (2.9), when combined with the corrector step, leads to $\{\mathcal{L}(w^i)\}_i$ being a decreasing sequence.

Consider the functionals $\mathcal{L}(w^{*i})$ and $\mathcal{L}^{(i)}(w^{*i})$ as functions of α and set

$$\mathcal{J}(\alpha) := \mathcal{L}(w^{i-1} - \alpha\hat{w}^i)\,, \quad \mathcal{J}^{(i)}(\alpha) := \mathcal{L}^{(i)}(w^{i-1} - \alpha\hat{w}^i).$$

For any integer $i \geq 1$ and knowing w^{i-1}, we want to find $\alpha > 0$ that solves

$$\mathcal{L}(w^{i-1} - \alpha_i\hat{w}^i) = \inf_{\alpha} \mathcal{L}(w^{i-1} - \alpha\hat{w}^i)\,. \tag{4.3}$$

Since $\mathcal{J}$ is not quadratic it is not straightforward to find the minimizer α_i. Instead, we approach the problem by seeking an approximation, that is, by minimizing $\mathcal{J}^{(i)}$.

The value at which the quadratic function $\mathcal{J}^{(i)}$ achieves a minimum is found from

$$0 = \frac{d\mathcal{J}^{(i)}}{d\alpha} = \alpha[a(\hat{w}^i, \hat{w}^i) + j''(w^{i-1})(\hat{w}^i, \hat{w}^i)] - a(w^{i-1}, \hat{w}^i) - (j'(w^{i-1}), \hat{w}^i) + \ell(\hat{w}^i). \tag{4.4}$$

Thus,

$$\alpha_{\min} = \frac{a(w^{i-1}, \hat{w}^i) + (j'(w^{i-1}), \hat{w}^i) - \ell(\hat{w}^i)}{a(\hat{w}^i, \hat{w}^i) + j''(w^{i-1})(\hat{w}^i, \hat{w}^i)}\,. \tag{4.5}$$

Now from (2.14b) the solution w^{*i} satisfies $(\mathcal{L}^{(i)\prime}(w^{i-1} - \alpha\hat{w}^i), z) = 0$, that is,

$$a(w^{i-1} - \alpha\hat{w}^i, z) + (j'(w^{i-1}), z) + j''(w^{i-1})(\hat{w}, z) = \ell(z)\,. \tag{4.6}$$

Setting $z = \hat{w}^i$ and rearranging, we find that the denominator in (4.5) is given by

$$a(\hat{w}^i, \hat{w}^i) + j''(w^{i-1})(\hat{w}^i, \hat{w}^i) = a(w^{i-1}, \hat{w}^i) + (j'(w^{i-1}), \hat{w}^i) - \ell(\hat{w}^i)\,. \tag{4.7}$$

Thus $\alpha_{\min} = 1$. At this stage we need to show that there is a value $\alpha' \in (0, \alpha_{\min})$ approximating the solution of (4.3).

We note that

$$\begin{aligned}[\mathcal{J}(\alpha) + \alpha^2\mathcal{L}^{(i)\prime\prime}(\hat{w}^i, \hat{w}^i)]\Big|_{\alpha=0} &= \mathcal{J}(0)\\ &= \mathcal{J}^{(i)}(0) - j^{(i)}(0) + j(0)\\ &= \mathcal{J}^{(i)}(0).\end{aligned}$$

Secondly,

$$\begin{aligned}\frac{d}{d\alpha}\left[\mathcal{J}(\alpha)+\alpha^2\mathcal{L}^{(i)\prime\prime}(\hat{w}^i,\hat{w}^i)\right]\Big|_{\alpha=0} &= \frac{d}{d\alpha}\mathcal{J}^{(i)}(0)\\ &= -(\mathcal{L}^{(i)\prime}(w^{i-1}),\hat{w}^i)\\ &= -\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i,\hat{w}^i)<0\,. \end{aligned} \tag{4.8}$$

Thus $\frac{d}{d\alpha}[\mathcal{J}(\alpha)+\alpha^2\mathcal{L}^{(i)\prime\prime}(\hat{w}^i,\hat{w}^i)]$ will be non-positive for some α in the range $0<\alpha<1$, and hence

$$\begin{aligned}\mathcal{L}(w^{*i}) &= \mathcal{J}(\alpha)\\ &\le \mathcal{J}(\alpha)+\alpha^2\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i,\hat{w}^i)\\ &\le \mathcal{J}(0)\quad(0<\alpha<1)\\ &= \mathcal{L}(w^{i-1})\,. \end{aligned} \tag{4.9}$$

It follows that with the appropriate choice of α, the line search procedure for the predictor step together with the conventional corrector leads to a minimizing sequence for $\mathcal{L}(w^i)$.

4.2. Steepest descent method

We take $w^i=(u^i,\lambda^i)$ and $w^{*i}=(u^i,\lambda^{*i})$.

A family of descent methods is constructed with the search direction satisfying

$$(\mathcal{L}^{(i)\prime}(w^{i-1}),\hat{w}^i)>0\,. \tag{4.10}$$

The method of steepest descent corresponds to a gradient approach with the direction $\hat{w}^i$ given by

$$\hat{w}^i=\mathcal{L}^{(i)\prime}(w^{i-1})\,. \tag{4.11}$$

Clearly we have (4.10) with (4.11). Next, we compute $\alpha_i>0$ as the solution to the minimization problem

$$\mathcal{L}^{(i)}(w^{i-1}-\alpha_i\hat{w}^i)=\inf_{\alpha}\mathcal{L}^{(i)}(w^{i-1}-\alpha\hat{w}^i)\,; \tag{4.12}$$

that is,

$$(\mathcal{L}^{(i)\prime}(w^{i-1}-\alpha_i\hat{w}^i),\hat{w}^i)=0\,. \tag{4.13}$$

Since $\mathcal{L}^{(i)}$ is quadratic one obtains

$$\alpha_i=\frac{(\mathcal{L}^{(i)\prime}(w^{i-1}),\hat{w}^i)}{\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i,\hat{w}^i)}=\frac{(\mathcal{L}^{(i)\prime}(w^{i-1}),\mathcal{L}^{(i)\prime}(w^{i-1}))}{\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i,\hat{w}^i)}.$$

The predictor step corresponding to the steepest descent method is thus as follows.

Predictor (steepest descent): Given $w^{i-1}\in W$, find $\hat{w}^i$ and w^{*i} in W from

$$\hat{w}^i=\mathcal{L}^{(i)\prime}(w^{i-1})\,, \tag{4.14a}$$

$$\alpha_i=\frac{\|\mathcal{L}^{(i)\prime}(w^{i-1})\|^2}{\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i,\hat{w}^i)}\,, \tag{4.14b}$$

$$w^{*i}=w^{i-1}-\alpha_i\hat{w}^i\,. \tag{4.14c}$$

The corrector step is as in (2.10) or (2.13).

Remark 4.1. From the definition (4.12) and (4.13),

$$\mathcal{L}^{(i)}(w^{*i}) \leq \mathcal{L}^{(i)}(w^{i-1}) \quad \text{and} \quad (\mathcal{L}^{(i)\prime}(w^{*i}), \mathcal{L}^{(i)\prime}(w^{i-1})) = 0\,. \tag{4.15}$$

Hence the steepest descent approach is guaranteed to make at least some progress in finding a minimizer of $\mathcal{L}^{(i)}$, and the search directions are independent from one iteration to the next.

Remark 4.2. Given $j^{(i)}$, $\mathcal{L}$ and $\mathcal{L}^{(i)}(z) = \mathcal{L}(z) + j^{(i)}(z) - j(z)$, it follows that

$$\mathcal{L}^{(i)}(w^{i-1}) = \mathcal{L}(w^{i-1}), \quad \mathcal{L}^{(i)\prime}(w^{i-1}) = \mathcal{L}'(w^{i-1}), \quad \mathcal{L}^{(i)\prime\prime}(w^{i-1}) = \mathcal{L}''(w^{i-1})\,, \tag{4.16}$$

and

$$\alpha_i = \frac{\|\mathcal{L}'(w^{i-1})\|^2}{\mathcal{L}''(w^{i-1})(\hat{w}^i, \hat{w}^i)}.$$

Theorem 4.1. *Let (u^i, λ^i) given by (4.14a)–(2.10) and (4.14c). Then $(\mathcal{L}(u^i, \lambda^i))_i$ is a decreasing sequence for α_i in a neighbourhood of zero.*

This result is easily obtained by adapting the proof used for the line search approach. It suffices to show that there exists a neighbourhood of α_i such that the sequence $\{\mathcal{L}(w^i)\}_i$ is decreasing. We present a different proof here.

Proof. We have the decomposition

$$\mathcal{L}(w^i) - \mathcal{L}(w^{i-1}) = \underbrace{\mathcal{L}(w^i) - \mathcal{L}(w^{*i})}_{\text{corrector step}} + \underbrace{\mathcal{L}(w^{*i}) - \mathcal{L}(w^{i-1})}_{\text{predictor step}}\,. \tag{4.17}$$

We start with the corrector step, and note that $\mathcal{L}(u^i, \lambda^i) = \inf_{\mu \in \Lambda} \mathcal{L}(u^i, \mu)$. Thus

$$\mathcal{L}(w^i) - \mathcal{L}(w^{*i}) \leq 0\,. \tag{4.18}$$

We note that

$$\mathcal{L}''(w^{i-1})(\hat{w}^i, \hat{w}^i) = a(\hat{w}^i, \hat{w}^i) + j''(w^{i-1})(\hat{w}^i, \hat{w}^i) \geq (c+m)\|\mathcal{L}'(w^{i-1})\|^2\,,$$

where m is the smallest eigenvalue of $j''(w^{i-1})$.

Secondly, in the predictor step, with (4.14c) and Taylor's expansion (recall that $\mathcal{L}^{(i)}$ is quadratic), we have

$$\begin{aligned}\mathcal{L}(w^{*i}) &= \mathcal{L}^{(i)}(w^{*i}) + j(w^{*i}) - j^{(i)}(w^{*i}) \\ &= \mathcal{L}(w^{i-1}) - \alpha_i[(\mathcal{L}^{(i)\prime}(w^{i-1}), \hat{w}^i) - \tfrac{\alpha_i}{2}\mathcal{L}^{(i)\prime\prime}(w^i)(\hat{w}^{*i}, \hat{w}^{*i})] + j(w^{*i}) - j^{(i)}(w^{*i})\,,\end{aligned}$$

which after substitution for α_i gives

$$\mathcal{L}(w^{*i}) - \mathcal{L}(w^{i-1}) = -\frac{\|\mathcal{L}^{(i)\prime}(w^{i-1})\|^4}{2\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i, \hat{w}^i)} + j(w^{*i}) - j^{(i)}(w^{*i})\,. \tag{4.19}$$

We return to (4.17); using (4.18) and (4.19) we have

$$\mathcal{L}(w^i) - \mathcal{L}(w^{i-1}) \leq -\frac{\|\mathcal{L}^{(i)\prime}(w^{i-1})\|^4}{2\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i, \hat{w}^i)} + j(w^{*i}) - j^{(i)}(w^{*i})\,. \tag{4.20}$$

By definition, $j^{(i)}$ is the second order Taylor's approximation of j. Denoting by $j^{(k)}$ the k-derivative of j, we have

$$j(w^{*i}) - j^{(i)}(w^{*i}) = \sum_{3 \leq k \leq n} (-1)^k \frac{\alpha_i^k}{k!} j^{(k)}(w^{i-1})(\underbrace{\hat{w}^i, \ldots, \hat{w}^k}_{k \text{ terms}})\,.$$

We deduce from (4.20) that there exists a neighbourhood of α_i for which $\mathcal{L}(u^i, \lambda^i) - \mathcal{L}(u^{i-1}, \lambda^{i-1})$ is non-positive, which completes the proof of the theorem. □

4.3. A Newton-like method

In this approach the search direction at each iteration is defined to be $\hat{w}^i$, the solution of

$$\mathcal{H} : V \times \Lambda \to V \times \Lambda, \quad \mathcal{H}(\hat{w}^i) = \mathcal{L}^{(i)\prime}(w^{i-1}), \tag{4.21}$$

where $\mathcal{H}$ is an approximation of $\mathcal{L}^{(i)\prime\prime}(w^{i-1})$. The functional $\mathcal{L}^{(i)}$ is convex and twice differentiable, hence its second derivative is symmetric and positive definite. So, we have (4.10) with (4.21). Next, from (4.12) and (4.21),

$$\alpha_i = \frac{\mathcal{L}^{(i)\prime}(w^{i-1}) \cdot \mathcal{H}^{-1}\mathcal{L}^{(i)\prime}(w^{i-1})}{\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i, \hat{w}^i)}. \tag{4.22}$$

The predictor step corresponding to the Newton-like method is as follows.

Predictor (Newton-like): Given $w^{i-1} = (u^{i-1}, \lambda^{i-1})$ in W, find $\hat{w}^i$, α_i and $w^{*i} \in W$ from

$$\begin{aligned} &\mathcal{H}(\hat{w}^i) = \mathcal{L}^{(i)\prime}(w^{i-1}), \\ &\alpha_i = \frac{\mathcal{L}^{(i)\prime}(w^{i-1}) \cdot \mathcal{H}^{-1}\mathcal{L}^{(i)\prime}(w^{i-1})}{\mathcal{L}^{(i)\prime\prime}(w^{i-1})(\hat{w}^i, \hat{w}^i)}, \\ &w^{*i} = w^{i-1} - \alpha_i \hat{w}^i. \end{aligned} \tag{4.23a}$$

The corrector step is as in (2.10) or (2.13).

Remark 4.3. From the construction of α_i, the iterates $w^{*i} = (u^i, \lambda^{*i})$ are such that

$$\mathcal{L}(w^{*i}) \le \mathcal{L}(w^{i-1}) \quad \text{and} \quad \mathcal{L}'(w^{*i}) \cdot \mathcal{H}^{-1}\mathcal{L}'(w^{i-1}) = 0.$$

We have the following result

Theorem 4.2. *Let $w^i = (u^i, \lambda^i)$ given by (4.23a). Then $(\mathcal{L}(w^i))_i$ is decreasing in a neighbourhood of α_i.*

Proof. We proceed as in the proof of Theorem 4.1. □

References

[1] W. Han, B.D. Reddy, Plasticity Mathematical Theory and Numerical Analysis, second ed., in: Interdisciplinary Applied Mathematics, vol. 9, Springer, New York, 2013.
[2] J.C. Simo, T.J. Hughes, Computational Inelasticity, Interdisciplinary Applied Mathematics, vol. 7, Springer, New York, 1998.
[3] J.C. Simo, R.L. Taylor, Consistent tangent operators for rate independent elastoplasticity, Comput. Methods Appl. Mech. Engrg. 48 (1985) 101–118.
[4] J.K. Djoko, F. Ebobisse, A.T. McBride, B.D. Reddy, A discontinuous Galerkin formulation for classical and gradient plasticity. Part 1: Formulation and analysis, Comput. Methods Appl. Mech. Engrg. 196 (2007) 3881–3897.
[5] J.-L. Lions, G. Stampacchia, Variational inequalities, Comm. Pure Appl. Math. 20 (1967) 493–519.
[6] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, in: Springer Series in Computational Physics, Springer-Verlag, 2008.
[7] B.D. Reddy, T.B. Griffin, Variational principles and convergence of finite element approximations of a holonomic elastic–plastic problem, Numer. Math. 52 (1988) 101–117.
[8] J.K. Djoko, F. Ebobisse, A.T. McBride, B.D. Reddy, A discontinuous Galerkin formulation for classical and gradient plasticity. Part 2: Algorithms and numerical results, Comput. Methods Appl. Mech. Engrg. 197 (2007) 1–21.
[9] D.G. Luenberger, Optimization by Vector Space Methods, John Wiley & Sons, New York, 1969.
[10] M.A. Crisfield, Accelerating and damping the modified Newton–Raphson method, Comput. Struct. 18 (1984) 395–407.
[11] J.B. Martin, S. Caddemi, Sufficient conditions for convergence of the Newton–Raphson iterative algorithm in incremental elastic–plastic analysis, Eur. J. Mech. A Solids 13 (1994) 351–365.