ACCURACY AND STABILITY OF INTEGRATION ALGORITHMS FOR ELASTOPLASTIC CONSTITUTIVE RELATIONS

M. ORTIZ
Division of Engineering, Brown University, Providence, Rhode Island, U.S.A.

E. P. POPOV
Department of Civil Engineering, University of California, Berkeley, California, U.S.A.

SUMMARY
An analysis of accuracy and stability of algorithms for the integration of elastoplastic constitutive relations is carried out in this paper. Reference is made to a very general internal variable formulation of plasticity and to two families of algorithms that generalize the well-known trapezoidal and midpoint rules to fit the present context. Other integration schemes such as the radial return, mean normal and closest point procedures are particular cases of this general formulation. The meaning of first and second-order accuracy in the presence of the plastic consistency condition is examined in detail, and the criteria derived are used to identify two second-order accurate members of the proposed algorithms. A general methodology is also derived whereby the numerical stability properties of integration schemes can be systematically assessed. With the aid of this methodology, the generalized midpoint rule is seen to have far better stability properties than the generalized trapezoidal rule. Finally, numerical examples are presented that illustrate the performance of the algorithms.

1. INTRODUCTION
In any numerical scheme employed for the analysis of elastoplastic problems it eventually becomes necessary to integrate the constitutive equations governing material behaviour. Whereas accuracy in the computation of the tangent stiffness matrix can be substantially relaxed at the expense of convergence speed, the precision with which constitutive relations are integrated has a direct impact on the overall accuracy of the analysis. The importance of understanding and controlling this source of numerical error has motivated a number of recent studies.\textsuperscript{1-5} In spite of these advances, some obscure points remain. In the context of inviscid plasticity, a second-order accurate integration algorithm appears to be as yet unavailable and the difficult but crucial aspect of numerical stability has received little attention. Even more importantly, there appear to be no general methodologies that would permit a systematic analysis of accuracy and stability for arbitrary constitutive laws.

In this paper, two families of algorithms for the integration of elastoplastic constitutive equations are proposed. To enhance the applicability of the method, reference is made to a general internal variable formulation of plasticity capable of characterizing a large class of elastoplastic materials. The proposed integration schemes generalize the well-known trapezoidal and midpoint rules in a way that facilitates satisfaction of the consistency condition of inviscid plasticity, whereby states of stress are required to be confined to the elastic domain. The radial return,\textsuperscript{6} mean normal\textsuperscript{7} and closest point\textsuperscript{8} procedures are contained in the proposed families of algorithms as particular cases.

0029–5981/85/091561–16$01.60
© 1985 by John Wiley & Sons, Ltd.

Received 24 July 1984
Revised 17 October 1984
In Section 4, the meaning of consistency and second-order accuracy of an integration scheme for inviscid elastoplastic relations is examined in detail and the criteria derived are used to demonstrate the second-order accuracy of two members of the proposed families of algorithms. In addition, a general methodology is developed that permits a systematic stability analysis of arbitrary integration schemes. This method of analysis is novel in that it is based on some concepts from Riemannian geometry. The aim here is to determine conditions under which an algorithm is contractive with respect to some suitable Riemannian metric defined over the yield surface. This notion of stability, to be termed in what follows large-scale stability, in conjunction with consistency or first-order accuracy, has been shown to guarantee convergence of the numerical solution towards the exact one as the step size tends to zero. Furthermore, in Section 4 it is shown that large-scale stability is equivalent to the simpler concept of small-scale stability, or stability with respect to infinitesimal perturbations in the initial conditions. This has the effect of greatly simplifying the stability analysis since infinitesimal perturbations are propagated by a linearized algorithm to which the familiar stability analysis tools for linear problems can be conveniently applied. These observations would appear to have a potential value in connection with other transient problems in computational mechanics such as nonlinear heat conduction and structural dynamics.

The stability analysis methodology derived in Section 4 is subsequently applied to the proposed families of algorithms and the unconditionally and conditionally stable subclasses are indentified. For the latter, stability criteria are given. The analysis shows that the stability range of the generalized trapezoidal rule is very sensitive to the degree of distortion of the loading surface. By contrast, the stability range of the generalized midpoint rule is fixed regardless of the shape of the loading surface. This would appear to point to the midpoint rule as preferable to the trapezoidal rule. Finally, in Section 5 numerical examples are presented that illustrate the accuracy of the proposed algorithms.

2. CONSTITUTIVE RELATIONS FOR INFINITESIMAL PLASTICITY

For several decades, the physical foundations of classical plasticity as well as the mathematical properties of related boundary value problems have been the subject of very active research. As a result, today classical plasticity is a sizable and well-established subfield of solid mechanics. The success of plasticity theories as a basis for material modelling stems largely from their ability to approximate in a relatively simple manner the quasistatic constitutive behaviour of a broad class of materials, ranging from metals to concrete and soils. Excellent reviews on these and other related topics may be found in References 9 and 10.

A wide range of elastoplastic materials can be characterized by means of a set of constitutive relations of the general form

\[ e_{ij} = E_{ij}^e + e_{ij}^p \]  
(1a)

\[ \sigma_{ij} = D_{ijkl} e_{kl}^e \]  
(1b)

\[ \dot{e}_{ij}^p = \lambda r_{ij}(\sigma, q) \]  
(1c)

\[ \dot{q}_a = \lambda h_a(\sigma, q) \]  
(1d)

where, following standard notation, \( e_{ij}, e_{ij}^e \) and \( e_{ij}^p \) denote the total, elastic and plastic strain tensors, \( \sigma_{ij} \) the Cauchy stress tensor and \( q_a \) signifies some suitable set of internal variables. Equation (1a) expresses the commonly assumed additive decomposition of infinitesimal strain into elastic and plastic parts. On the other hand, generalized Hooke's law (1b) linearly relates stresses and elastic strains through a stiffness tensor \( D_{ijkl} \). The symmetry relations \( D_{ijkl} = D_{klij} \).
inherent to hyperelasticity are assumed. Equation (1c) expresses a generally non-associated flow rule for the plastic strain rates and (1d) is some suitable set of hardening laws which govern the evolution of the plastic variables. In these equations, \( r_{ij} \) is the plastic flow direction, \( h_0 \) the plastic moduli and \( \dot{\lambda} \) is a plastic parameter to be determined with the aid of the loading–unloading criterion. This can be expressed in Kuhn–Tucker form as

\[
\begin{align*}
\phi(\sigma, q) & \leq 0 \\
\dot{\lambda} & \geq 0 \\
\phi \dot{\lambda} & = 0
\end{align*}
\]  
(2a-b-c)

Here, \( \phi(\sigma, q) \) signifies the yield function of the material and (2a) characterizes the corresponding elastic domain, which is assumed to be convex. Along any process of loading, conditions (2) must hold simultaneously. For \( \phi < 0 \), (2c) yields \( \dot{\lambda} = 0 \), i.e. elastic behaviour, while plastic flow is characterized by \( \dot{\lambda} > 0 \) which, in view of (2c), necessitates satisfaction of the yield criterion \( \phi = 0 \). From this latter constraint, along a process of plastic loading the so-called plastic consistency condition is obtained

\[
\dot{\phi} = \frac{\partial \phi}{\partial \sigma_{ij}} \dot{\sigma}_{ij} + \frac{\partial \phi}{\partial q_s} \dot{q}_s = \eta_{ij} \dot{\sigma}_{ij} + \xi_s \dot{q}_s = 0
\]  
(3)

which has the effect of confining the stress trajectory to the yield surface. In (3) one writes, for simplicity of notation,

\[
\eta_{ij}(\sigma, q) = \frac{\partial \phi}{\partial \sigma_{ij}}, \quad \xi_s(\sigma, q) = \frac{\partial \phi}{\partial q_s}
\]  
(4)

Note that \( \eta_{ij} \) and \( \xi_s \) are the normals to the yield surface in stress and plastic variable spaces, respectively.

A particular form of flow rule can be obtained by identifying

\[ r_{ij} = \eta_{ij} \]  
(5)

i.e. by postulating that plastic strain rates are normal to the yield surface. Assumption (5) is known as the normality rule and the resulting flow laws are termed associated. The normality rule is known to hold for metals but is violated by numerous materials of engineering interest, such as concrete and soils. Therefore, in order to keep the formulation general, non-associated plasticity will be considered throughout.

3. TWO FAMILIES OF ALGORITHMS FOR THE INTEGRATION OF ELASTOPLASTIC CONSTITUTIVE RELATIONS

In the preceding section, constitutive equations capable of characterizing a wide variety of elastoplastic materials have been briefly summarized. The problem considered here is to devise accurate and efficient algorithms for the integration of such constitutive relations. In the context of finite element analysis using isoparametric elements, the integration of constitutive equations is carried out at the Gauss points for given deformation increments. The unknowns to be found are the updated stresses and plastic variables.

An acceptable algorithm should satisfy three basic requirements: (a) consistency with the
constitutive relations to be integrated or first-order accuracy; (b) numerical stability; (c) incremental plastic consistency. A non-required but nevertheless desirable feature to add to the above list is: (d) second-order accuracy.

Conditions (a) and (b) are needed for attaining convergence of the numerical solution as the time step becomes vanishingly small. Condition (c) is the algorithmic counterpart of the plastic consistency condition and requires that states of stress computed from the algorithm be contained within the elastic domain.

**Generalized trapezoidal rule.** A class of algorithms for the integration of relations (1)–(3) with a potential for satisfying conditions (a)–(d) is given by

\begin{align}
\sigma_{n+1} &= D : (\varepsilon_{n+1} - \varepsilon_n) \\
\varepsilon_{n+1}^p &= \varepsilon_n^p + \lambda [ (1 - \alpha) r_n + \alpha r_{n+1} ] \\
q_{n+1} &= q_n + \lambda [ (1 - \alpha) h_n + \alpha h_{n+1} ] \\
\phi_{n+1} &= 0
\end{align}

where, for simplicity of notation, bold-face symbols are used for variables of tensorial character and the symbol (:) signifies doubly contracted tensor product, e.g. \( D : \varepsilon \)\( \equiv D_{ijkl} \varepsilon_{ij} \).

In (6), \( \varepsilon_n, \varepsilon_n^p, \sigma_n \) and \( q_n \) are the known strains, plastic strains, stress and plastic variables at time \( t_n \), whereas \( \varepsilon_{n+1}^p, \sigma_{n+1} \) and \( q_{n+1} \) are the corresponding unknown variables at time \( t_{n+1} = t_n + h, h \) being the time step. The updated strains \( \varepsilon_{n+1} \) are assumed given and one writes

\begin{align}
r_n &= r(\sigma_n, q_n); \quad r_{n+1} = r(\sigma_{n+1}, q_{n+1}) \\
h_n &= h(\sigma_n, q_n); \quad h_{n+1} = h(\sigma_{n+1}, q_{n+1})
\end{align}

In (6), \( \lambda \) may be regarded as an incremental plastic parameter to be determined with the aid of the incremental plastic consistency condition (6d). Finally, the algorithmic parameter \( \alpha \) ranges from 0 to 1.

**Remarks**

(1) Algorithm (6) extends the trapezoidal rule in a manner suitable for application to elastoplastic constitutive relations. In particular, incremental plastic consistency follows automatically from the enforcement of condition (6d).

(2) A revealing geometric interpretation of algorithm (6) can be given by rephrasing it as

\begin{align}
\sigma_{n+1} &= \sigma_n + 2 D : [(1 - \alpha) r_n + \alpha r_{n+1}] \\
q_{n+1} &= q_n + \lambda [(1 - \alpha) h_n + \alpha h_{n+1}] \\
\phi_{n+1} &= 0
\end{align}

where one writes

\[ \sigma_{n+1}^* = \sigma_n + D : (\varepsilon_{n+1} - \varepsilon_n) \]

for the elastically updated stress predictor. The situation expressed in (8) is graphically shown in Figure 1, which illustrates how stresses are updated in two steps. The first one takes \( \sigma_n \) into an elastic predictor \( \sigma_{n+1}^* \), obtained from integration of the elastic stress–strain relations. The elastic predictor so determined is subsequently mapped onto a suitably updated yield surface, thus restoring plastic consistency. Such return mapping can be in turn divided into two substeps: first, stresses are projected along the initial plastic flow direction \( r_n \); the stresses so obtained are then projected onto the yield surface along the final plastic flow direction \( r_{n+1} \).
(3) It is apparent from the above remark that equations (6)–(9) generalize the concept of return mapping introduced by Wilkins for von Mises yield criterion and further refined in References 1–4, 7, 12 and 13. In the present setting, the algorithm allows for non-associated plasticity and arbitrary hardening rules. Furthermore, the definition of the return mapping is improved with respect to previous work by introducing the adjustable parameter $\alpha$. This determines the relative weight given to the initial and final plastic flow directions $r_n$ and $r_{n+1}$, respectively. For $\alpha > 0$ the algorithm is implicit. For $\alpha = \frac{1}{2}$ and the particular case of a von Mises yield criterion with associated perfect plasticity, the proposed algorithm coincides with the mean-normal procedure proposed by Rice and Tracy. For $\alpha = 1$ and associated plasticity, the closest point projection algorithm is obtained. This particular choice of return mapping was proposed in Reference 4 as a generalization of the well-known radial return algorithm to plasticity models other than linearly hardening von Mises, to which early formulations were confined. The closest point algorithm is known to be unconditionally stable provided the elastic region is convex, but it is only first-order accurate. This can be a limitation in many cases of practical interest. As shown in Section 5, a suitable choice of the adjustable parameter $\alpha$ may result in improved accuracy with respect to the closest point procedure.

Generalized midpoint rule. An alternative family of algorithms may be obtained from a midpoint-type rule of the following general form:

$$
\sigma_{n+1} = D : (\varepsilon_{n+1} - \varepsilon^p_{n+1}) \tag{10a}
$$

$$
\varepsilon^p_{n+1} = \varepsilon_n^p + \lambda r_{n+2} \tag{10b}
$$

$$
q_{n+1} = q_n + \lambda h_{n+2} \tag{10c}
$$

$$
\phi_{n+1} = 0 \tag{10d}
$$

where one writes

$$
r_{n+2} = r((1-\alpha)\sigma_n + \alpha \sigma_{n+1}, (1-\alpha)q_n + \alpha q_{n+1}) \tag{11}
$$

$$
h_{n+2} = h((1-\alpha)\sigma_n + \alpha \sigma_{n+1}, (1-\alpha)q_n + \alpha q_{n+1})
$$

As in the case of the generalized trapezoidal rule, (10)–(11) are a set of nonlinear algebraic equations to be solved for the unknowns $\sigma_{n+1}, \varepsilon^p_{n+1}, q_{n+1}$ and $\lambda$.

---

1The case $\alpha = 0$ corresponding to explicit integration has also been extensively used in computation, particularly in conjunction with sub-incrementation (c.e. see References 15–17)
Remarks

(4) As before, incremental plastic consistency follows from the enforcement of condition (10d).

(5) Figure 2 depicts a geometric interpretation of algorithm (10)–(11). In it is seen from this figure that the generalized midpoint rule may be also regarded as a return mapping algorithm in which the elastic predictor \( \sigma^*_{n+1} \) is projected onto a suitably updated yield surface along a flow direction evaluated at the midpoint \((\sigma_{n+2}, q_{n+2})\).

(6) For the particular case of von Mises yield criterion with associated linearly hardening plasticity the generalized trapezoidal and midpoint rules coincide.

4. ALGORITHM ANALYSIS

A critical appraisal of any time integration algorithm requires examination of its accuracy and stability characteristics. In the present context, an accuracy analysis is complicated by the presence of the plastic consistency constraint (2)–(3). The analysis that follows overcomes this difficulty and shows that both the generalized trapezoidal and midpoint rules contain a second-order accurate member, \( \alpha = \frac{1}{2} \), as well as an unconditionally stable subclass. The sense in which this numerical stability is attained is made precise in Section 4.2.

4.1. Accuracy analysis

In the context of finite element analysis, as mentioned before, integration of constitutive relations is carried out for given strain increments. As a result, the updated strains \( \varepsilon_{n+1} = \varepsilon(t_{n+1}) = \varepsilon(t_n + h) \) may be viewed as known functions of the step size \( h \). The remaining updated variables \( \sigma_{n+1}, \varepsilon_{n+1}^p \) and \( q_{n+1} \), as well as the incremental plastic parameter \( \lambda \), also become functions of \( h \) implicitly defined through relations (6) or (10). It is clear from these relations that as \( h \to 0 \) and, hence, \( \varepsilon_{n+1} \to \varepsilon_n \) the limiting values

\[
\sigma_{n+1} \to \sigma_n; \quad \varepsilon_{n+1}^p \to \varepsilon_n^p; \quad q_{n+1} \to q_n; \quad \lambda \to 0
\]  

(12)

are obtained. Furthermore, by the implicit function theorem \( \sigma_{n+1}, \varepsilon_{n+1}^p, q_{n+1} \) and \( \lambda \) are differentiable functions of \( h \) provided the functions \( r, h \) and \( \phi \) are sufficiently smooth, which will be assumed as needed.

First-order accuracy or consistency of the algorithms (6) and (10) with constitutive relations (1) necessitates that the numerically integrated variables \( \sigma_{n+1}, \varepsilon_{n+1}^p \) and \( q_{n+1} \) agree with the their
exact values $\sigma(t_{n+1}), \epsilon_p(t_{n+1})$ and $q(t_{n+1})$ to within second-order terms in $h$. An alternative statement requires that

$$\frac{d}{dh}(\sigma_{n+1})_{h=0} = \hat{\sigma}_n = D: (\hat{\epsilon}_n - \hat{\epsilon}_n)$$  \hspace{1cm} (13a)$$

$$\frac{d}{dh}(\epsilon_{p,n+1})_{h=0} = \hat{\epsilon}_n^{p} = \hat{\lambda}_n^{-1} r_n$$  \hspace{1cm} (13b)$$

$$\frac{d}{dh}(q_{n+1})_{h=0} = \hat{q}_n = \hat{\lambda}_n h_n$$  \hspace{1cm} (13c)$$

$$\frac{d}{dh}(\hat{\lambda})_{h=0} = \hat{\lambda}_n$$  \hspace{1cm} (13d)$$

where the plastic parameter $\hat{\lambda}_n$ is determined with the aid of the plastic consistency condition

$$\phi_n = \eta_n: \sigma_n + \xi_n: q_n = 0$$  \hspace{1cm} (14)$$

In this latter expression one writes $\eta_n = \eta(\sigma_n, q_n)$, $\xi_n = \xi(\sigma_n, q_n)$ and $\hat{\epsilon}_n$ in (13a) is assumed given.

It remains to be seen, therefore, whether the generalized trapezoidal and midpoint rules satisfy the consistency requirement (13). As for the former, differentiating (6) with respect $h$ leads to

$$\frac{d}{dh} \sigma_{n+1} = D: \left( \frac{d}{dh} \epsilon_{n+1} - \frac{d}{dh} \epsilon_{p,n+1} \right)$$

$$\frac{d}{dh} \epsilon_{p,n+1} = \frac{d}{dh} \left[ (1 - \alpha) r_n + \alpha r_{n+1} \right] + \alpha \lambda \left[ \left( \frac{\partial r}{\partial \sigma} \right)_{n+1} \frac{d}{dh} \sigma_{n+1} + \left( \frac{\partial r}{\partial q} \right)_{n+1} \frac{d}{dh} q_{n+1} \right]$$

$$\frac{d}{dh} q_{n+1} = \frac{d}{dh} \left[ (1 - \alpha) h_n + \alpha h_{n+1} \right] + \alpha \lambda \left[ \left( \frac{\partial h}{\partial \sigma} \right)_{n+1} \frac{d}{dh} \sigma_{n+1} + \left( \frac{\partial h}{\partial q} \right)_{n+1} \frac{d}{dh} q_{n+1} \right]$$

$$0 = \frac{d}{dh} \phi_{n+1} = \eta_n: \sigma_{n+1} + \xi_{n+1}: q_{n+1}$$  \hspace{1cm} (15)$$

Taking the limit $h \to 0$ in these expressions and recalling (12) it is readily found

$$\frac{d}{dh} (\sigma_{n+1})_{h=0} = D: \left[ \dot{\epsilon}_n - \frac{d}{dh} (\epsilon_{p,n+1})_{h=0} \right]$$

$$\frac{d}{dh} (\epsilon_{p,n+1})_{h=0} = \left( \frac{d}{dh} \right)_{h=0} r_n$$  \hspace{1cm} (16a)$$

$$\frac{d}{dh} (q_{n+1})_{h=0} = \left( \frac{d}{dh} \right)_{h=0} h_n$$

$$0 = \frac{d}{dh} (\phi_{n+1})_{h=0} = \eta_n: \frac{d}{dh} (\sigma_{n+1})_{h=0} + \xi_n: \frac{d}{dh} (q_{n+1})_{h=0}$$

Comparing (16) with (13) it becomes apparent that $(d\lambda/dh)_{h=0} = \hat{\lambda}_n$ and, consequently, the remaining consistency conditions (13a)–(13c) are also satisfied.

Taking the accuracy analysis a step further, second order accuracy of algorithm (6) requires that the numerically integrated variables agree with their exact values to within third-order terms in $h$. In
other words, in addition to (13) it is now required

\[
\frac{d^2}{dh^2}(\sigma_{n+1})_{h=0} = \ddot{\sigma}_n = D: (\ddot{\varepsilon}_n - \ddot{\varepsilon}_n^p)
\]

\[
\frac{d^2}{dh^2}(\varepsilon_{n+1}^p)_{h=0} = \ddot{\varepsilon}_n^p = \dot{\lambda}_n r_n + \dot{\lambda}_n \left[ \left( \frac{\partial r}{\partial \sigma} \right)_n \dot{\sigma}_n + \left( \frac{\partial r}{\partial q} \right)_n \dot{q}_n \right]
\]

\[
\frac{d^2}{dh^2}(q_{n+1})_{h=0} = \ddot{q}_n = \dot{\lambda}_n h_n + \dot{\lambda}_n \left[ \left( \frac{\partial h}{\partial \sigma} \right)_n \dot{\sigma}_n + \left( \frac{\partial h}{\partial q} \right)_n \dot{q}_n \right]
\]

\[
\frac{d^2}{dh^2}(\dot{\lambda})_{h=0} = \ddot{\lambda}_n
\]

where \(\dot{\lambda}_n\) is to be determined from imposing a second-order oscillatory satisfaction of the plastic consistency condition \(\phi = 0\), i.e.

\[
0 = \ddot{\phi}_n = \dot{\eta}_n: \dot{\sigma}_n + \dot{\eta}_n : \ddot{\sigma}_n + \dot{\xi}_n : \ddot{q}_n + \xi_n \dddot{q}_n
\]

(18)

To check whether this condition is satisfied, differentiating (15) with respect to \(h\) and taking the limit \(h \to 0\) it is obtained

\[
\frac{d^2}{dh^2}(\sigma_{n+1})_{h=0} = D: \left( \ddot{\varepsilon}_n - \frac{d^2}{dh^2}(\varepsilon_{n+1}^p)_{h=0} \right)
\]

\[
\frac{d^2}{dh^2}(\varepsilon_{n+1}^p)_{h=0} = \left( \frac{d^2 \dot{\lambda}}{dh^2} \right)_{h=0} r_n + 2\dot{\lambda}_n \left[ \left( \frac{\partial r}{\partial \sigma} \right)_n \dot{\sigma}_n + \left( \frac{\partial r}{\partial q} \right)_n \dot{q}_n \right]
\]

(19)

\[
\frac{d^2}{dh^2}(q_{n+1})_{h=0} = \left( \frac{d^2 \dot{\lambda}}{dh^2} \right)_{h=0} h_n + 2\dot{\lambda}_n \left[ \left( \frac{\partial h}{\partial \sigma} \right)_n \dot{\sigma}_n + \left( \frac{\partial h}{\partial q} \right)_n \dot{q}_n \right]
\]

\[
0 = \frac{d^2}{dh^2}(\phi_{n+1})_{h=0} = \dot{\eta}_n: \ddot{\sigma}_n + \dot{\eta}_n: \frac{d^2}{dh^2}(\sigma_{n+1})_{h=0} + \dot{\xi}_n \dddot{q}_n
\]

where use has been made of (12) and the assumed consistency of the algorithm as in (13).

Comparison of (19) with (17) and (18) immediately leads to the conclusion that \((d^2 \dot{\lambda}/dh^2)_{h=0} = \dot{\lambda}_n\) and that second-order accuracy is achieved provided one chooses \(2\dot{\lambda} = 1\), i.e. \(\dot{\lambda} = \frac{1}{2}\).

A similar analysis leads to the same conclusions for the generalized midpoint rule, i.e. it is consistent for all \(\dot{\lambda}\) and second-order accurate for \(\dot{\lambda} = \frac{1}{2}\).

From these results it is concluded that the choice of \(\dot{\lambda} = \frac{1}{2}\) leads to optimal accuracy for small strain increments. This is confirmed by the numerical examples presented in Section 5. By contrast, this numerical evidence and the stability analysis carried out below indicate that larger values of \(\dot{\lambda}\) may prove advantageous in terms of accuracy and even necessary for stability in the realm of large time steps.

4.2. Numerical stability

The concept of numerical stability plays a central role in approximation theory for initial value problems. Its relevance stems largely from the fact that consistency and stability are necessary and sufficient conditions for convergence as the time step size is allowed to tend to
In the context of linear analysis, the numerical stability of a time stepping algorithm can be characterized by way of a spectral decomposition or, equivalently, with the aid of suitably defined energy norms. By contrast, when the initial value problem under consideration is nonlinear it is not always clear from the structure of the governing equations how to characterize numerical stability. In the computational literature, stability analyses have by and large been confined to the linear case or else based on a limited number of numerical tests.

In the remaining of this section a new methodology is proposed whereby the stability properties of integration algorithms for elastoplastic constitutive relations can be readily established. For simplicity, attention is confined to perfect plasticity and smooth yield surfaces. The conclusions derived, nevertheless, can be extended to the hardening case without conceptual changes.

Large-scale and small-scale stability. The purpose of the stability analysis that follows is to determine under what conditions finite perturbations in the initial stresses are attenuated by the algorithm, i.e.

\[ d(\sigma_{n+1}^{(2)}, \sigma_{n+1}^{(1)}) \leq d(\sigma_n^{(2)}, \sigma_n^{(1)}) \]  

(20)

where \( d(\cdot, \cdot) \) is some suitable distance to be defined on the yield surface and \( \sigma_n^{(1)} \) and \( \sigma_n^{(2)} \) are two sets of updated stresses corresponding to arbitrary initial values \( \sigma_n^{(1)} \) and \( \sigma_n^{(2)} \), respectively, all of which are assumed to lie on the yield surface. Stability in the sense (20) will be henceforth referred to as 'large-scale' stability. It has been shown in Reference 8 that for nonlinear initial value problems defined on Banach manifolds consistency and large-scale stability with respect to a complete metric are sufficient for convergence.

In spite of the conceptual appeal of large-scale stability, the task of directly establishing estimates of the type (20) may indeed pose insurmountable difficulties. The stability analysis is significantly simplified, however, once it is recognized that attention can be confined to infinitesimal perturbations in the initial conditions of the type \( \sigma_n \rightarrow \sigma_n + d\sigma_n \). This rests on the fact that attenuation by the algorithm of infinitesimal perturbations,

\[ \|d\sigma_{n+1}\| \leq \|d\sigma_n\| \]  

(21)

with respect to some suitable norm \( \|\cdot\| \), of 'small-scale' stability, implies large-scale stability in the sense (20). To demonstrate this, let \( \|\cdot\| \) be the energy norm

\[ \|\sigma\|^2 = \sigma_{ij}C_{ijkl}\sigma_{kl} \]  

(22)

where one writes \( C \equiv D^{-1} \), and let the distance \( d(\cdot, \cdot) \) on the yield surface be defined as

\[ d(\sigma^{(1)}, \sigma^{(2)}) = \inf_{\gamma} \int_{\gamma} \|\sigma'(s)\| \, ds \]  

(23)

where the infimum is taken over all stress paths \( \gamma \) on the yield surface joining \( \sigma^{(1)} \) and \( \sigma^{(2)} \). Then, it follows from the theory of Riemannian manifolds (Reference 14, Chapter 1) that, for a smooth yield surface, (23) does indeed define a so-called geodesic distance which endows the yield surface with a complete metric structure.

Consider next any two initial states of stress \( \sigma_n^{(1)} \) and \( \sigma_n^{(2)} \) and let \( \sigma_{n+1}^{(1)} \) and \( \sigma_{n+1}^{(2)} \) be the corresponding updated values, respectively, all of which are assumed to lie on the yield surface. Then, there exists a unique geodesic curve joining \( \sigma_n^{(1)} \) and \( \sigma_n^{(2)} \) for which the infimum in (23) is attained (Reference 14, Chapter 1). If \( \gamma_n \) is such curve, then by definition

\[ d(\sigma_n^{(1)}, \sigma_n^{(2)}) = \int_{\gamma_n} \|\sigma'(s)\| \, ds \]  

(24)
Let $\gamma_{n+1}$ be the transform of $\gamma_n$ by the algorithm. By construction, $\gamma_{n+1}$ lies on the yield surface and joins $\sigma_{n+1}^{(1)}$ and $\sigma_{n+1}^{(2)}$. Then, by the definition of geodesic distance, it follows that

$$d(\sigma_{n+1}^{(1)}, \sigma_{n+1}^{(2)}) \leq \int_{\gamma_{n+1}} ||\sigma'(s)|| ds$$  \hspace{1cm} (25)

But under the assumption of small-scale stability of the algorithm one has $||\sigma'(s_{n+1})||ds = ||d\sigma(s_{n+1})|| = ||\sigma'(s_n)||ds$ for every pair of corresponding points $s_n$ and $s_{n+1}$ on $\gamma_n$ and $\gamma_{n+1}$, respectively, and hence

$$\int_{\gamma_{n+1}} ||\sigma'(s)|| ds \leq \int_{\gamma_n} ||\sigma'(s)|| ds$$  \hspace{1cm} (26)

Combining (24), (25) and (26) it is finally concluded that

$$d(\sigma_{n+1}^{(1)}, \sigma_{n+1}^{(2)}) \leq d(\sigma_n^{(1)}, \sigma_n^{(2)})$$  \hspace{1cm} (27)

which proves large-scale stability.

The main conclusion of the above argument may be stated as follows: small-scale stability in the energy norm is equivalent to large-scale stability in the associated geodesic distance.

This result is of practical importance since it shows that stability analysis may be confined to the assessment of small-scale stability. This, in turn, has the effect of rendering the analysis tractable since infinitesimal perturbations are propagated by a linearized algorithm to which the familiar linear analysis techniques can be applied.

**Generalized trapezoidal rule.** To carry out a small-scale stability analysis of the generalized trapezoidal rule it is necessary to determine first how it propagates infinitesimal perturbations in the initial conditions. To this end, differentiating (6) yields

$$d\sigma_{n+1} = -D:d\varepsilon_{n+1}^{a}; \quad d\sigma_n = -D:d\varepsilon_n^{a}$$

$$d\varepsilon_{n+1}^{a} - d\varepsilon_n^{a} = d\lambda[(1-\alpha)r_n + \alpha r_{n+1}] + \lambda \left[ (1-\alpha)\left( \frac{\partial r}{\partial \sigma}_n \right):d\sigma_n + \alpha \left( \frac{\partial r}{\partial \sigma}_{n+1} \right):d\sigma_{n+1} \right]$$  \hspace{1cm} (28)

$$0 = d\delta_{n+1} = \eta_{n+1}:d\sigma_{n+1}$$

Introducing the notation

$$B = \partial r/\partial \sigma; \quad \tilde{r}_{n+1} = (1-\alpha)r_n + \alpha r_{n+1}$$  \hspace{1cm} (29)

equations (28) can be simplified to read

$$(C + \alpha\lambda B_n):d\sigma_{n+1} = d\lambda \tilde{r}_{n+1} + (C - (1-\alpha)\lambda B_n):d\sigma_n$$  \hspace{1cm} (30a)

$$0 = \eta_{n+1}:d\sigma_{n+1}$$  \hspace{1cm} (30b)

Substituting (30a) into (30b), the value of $d\lambda$ may be solved for, leading to

$$d\lambda = \eta_{n+1} : (C + \alpha\lambda B_{n+1})^{-1} : (C - (1-\alpha)\lambda B_n):d\sigma_n$$  \hspace{1cm} (31)

Solving in (30) for $d\sigma_{n+1}$ and making use of (31) one obtains

$$d\sigma_{n+1} = P : (C + \alpha\lambda B_{n+1})^{-1} : (C - (1-\alpha)\lambda B_n):d\sigma_n$$  \hspace{1cm} (32)
where
\[ \mathbf{P} = \mathbf{I} - \frac{\hat{r}_{n+2} \otimes \eta_{n+1}}{\hat{r}_{n+2} : \eta_{n+1}} \]
\[ \hat{r}_{n+2} = (\mathbf{C} + \alpha \lambda \mathbf{B}_{n+1})^{-1} \cdot \hat{r}_{n+2} \]  

To derive an estimate of type (21) from (32) it proves convenient to define the energy norm of a matrix in the usual fashion:
\[ \| \mathbf{A} \| = \sup_{\sigma} \| \mathbf{A} : \sigma \| \quad \| \sigma \| \]  

Taking the energy norm of (32) and recalling the well-known inequalities \( \| \mathbf{A} : \sigma \| \leq \| \mathbf{A} \| \| \sigma \| \) and \( \| \mathbf{A}_1 : \mathbf{A}_2 \| \leq \| \mathbf{A}_1 \| \| \mathbf{A}_2 \| \) pertaining to matrix multiplication it is found that
\[ \| d\sigma_{n+1} \| \leq \| \mathbf{P} \| \| (\mathbf{C} + \alpha \lambda \mathbf{B}_{n+1})^{-1} : (\mathbf{C} - (1 - \alpha) \lambda \mathbf{B}_n) \| \| d\sigma_n \| \]  

It remains to estimate, therefore, the matrix norms involved in (35). Concerning the norm of \( \mathbf{P} \) it suffices to note that \( \mathbf{P} \) defines an oblique projection along the direction \( \hat{r}_{n+2} \) onto the hyperplane orthogonal to \( \eta_{n+1} \), i.e. \( \mathbf{P} : \hat{r}_{n+2} = 0 \) and \( \mathbf{P} : \sigma = \sigma \) for every \( \sigma \) orthogonal to \( \eta_{n+1} \). From these properties of \( \mathbf{P} \) and definition (34) it readily follows that \( \| \mathbf{P} \| = 1 \) and hence \( \| \mathbf{P} \| \) drops from inequality (35).

In what follows, it is assumed that the tensor field \( \mathbf{B} = \partial \mathbf{r} / \partial \sigma \) is symmetric and positive definite everywhere on the yield surface. This is tantamount to assuming that the flow direction \( \mathbf{r} \) derives from a convex potential or loading function, a feature which is common to many plastic models. Under these assumptions, straightforward considerations show that
\[ \| (\mathbf{C} + \alpha \lambda \mathbf{B}_{n+1})^{-1} : (\mathbf{C} - (1 - \alpha) \lambda \mathbf{B}_n) \| = \frac{\| \delta_{\text{max}} : (\mathbf{C} - (1 - \alpha) \lambda \mathbf{B}_n) : \delta_{\text{max}} \|}{\| \delta_{\text{max}} : (\mathbf{C} + \alpha \lambda \mathbf{B}_{n+1}) : \delta_{\text{max}} \|} \]  

where \( \delta_{\text{max}} \) is the eigenvector corresponding to the maximum eigenvalue of the eigenproblem
\[ [(\mathbf{C} - (1 - \alpha) \lambda \mathbf{B}_n) - \mu(\mathbf{C} + \alpha \lambda \mathbf{B}_{n+1})] \cdot \delta = 0 \]  

which may be normalized to satisfy
\[ \| \delta_{\text{max}} \|^2 = \delta_{\text{max}} : \mathbf{C} : \delta_{\text{max}} = 1 \]  

Denoting
\[ \beta_n = \delta_{\text{max}} : \mathbf{B}_n ; \delta_{\text{max}} ; \quad \beta_{n+1} = \delta_{\text{max}} : \mathbf{B}_{n+1} ; \delta_{\text{max}} \]  

and making use of (38), (36) reduces to
\[ \| (\mathbf{C} + \alpha \lambda \mathbf{B}_{n+1})^{-1} : (\mathbf{C} - (1 - \alpha) \lambda \mathbf{B}_n) \| = \frac{1 - (1 - \alpha) \lambda \beta_n}{1 + \alpha \lambda \beta_{n+1}} \]  

which substituted into (35) leads to
\[ \| d\sigma_{n+1} \| \leq \frac{1 - (1 - \alpha) \lambda \beta_n}{1 + \alpha \lambda \beta_{n+1}} \| d\sigma_n \| \]  

From the assumed symmetry and positive definiteness of the tensor field \( \mathbf{B} \) it follows that both \( \beta_n \) and \( \beta_{n+1} \) are positive scalars. Under this condition, it follows by simple inspection that
\[
\frac{1 - (1 - \alpha)\lambda \beta_n}{1 + \alpha \lambda \beta_{n+1}} \leq \frac{1 - \alpha}{\alpha} \frac{\beta_n}{\beta_{n+1}}, \quad 0 < \alpha \leq 1
\] (42)

To majorize this bound, one may define the quantity
\[
s = \sup_{\delta : B_1 : \delta} \frac{\delta : B_2 : \delta}{\delta : B_2 : \delta}
\] (43)

where \( B_1 \) and \( B_2 \) are evaluated at arbitrary points \( \sigma_1 \) and \( \sigma_2 \) on the yield surface and the supremum is taken over all incremental stress \( \delta \) and stress points \( \sigma_1 \) and \( \sigma_2 \). For reasons that are clarified below \( s \) will be referred to as the distortion index of the loading surface. From definition (43) it follows that
\[
\frac{\beta_n}{\beta_{n+1}} \leq s
\] (44)

which in combination with (41) and (42) yields
\[
\| d\sigma_{n+1} \| \leq \frac{1 - \alpha}{\alpha} s \| d\sigma_n \| = c \| d\sigma_n \| ; \quad 0 < \alpha \leq 1
\] (45)

where one writes
\[
c \equiv \frac{1 - \alpha}{\alpha} s
\] (46)

Unconditional stability requires \( c \leq 1 \) which in view of (46) necessitates
\[
\alpha \geq \frac{s}{1 + s} \equiv \alpha_{\text{min}}
\] (47)

As discussed below, it follows from definition (43) that the distortion index \( s \) ranges from 1 for loading functions of constant curvature such as von Mises to \( \infty \) for loading surfaces with corners. As a consequence, the minimum value \( \alpha_{\text{min}} \) of \( \lambda \) to achieve unconditional stability ranges from \( \frac{1}{s} \) for \( s = 1 \) to 1 for \( s = \infty \).

For \( \alpha < \alpha_{\text{min}} \) stability is only conditional. A little reflection soon leads to the conclusion that
\[
\frac{1 - (1 - \alpha)\lambda \beta_n}{1 + \alpha \lambda \beta_{n+1}} \leq 1
\] (48)

in (41) if and only if \( \lambda \) is confined to be
\[
\lambda \leq \frac{2/\beta_{\text{max}}}{s - (1 + s)\alpha} = \lambda_c, \quad \alpha < \alpha_{\text{min}}
\] (49)

where \( \beta_{\text{max}} \) is the maximum eigenvalue among all tensors \( B \) over the yield surface. For \( \alpha = \alpha_{\text{min}} \), the critical value \( \lambda_c \) of the incremental plastic parameter becomes infinity and unconditional stability is recovered.

**Remarks**

1. Under the assumption of a loading function \( \psi \) acting as a potential for the plastic flow direction \( r = \partial \psi / \partial \sigma \), one has \( B = \partial^2 \psi / \partial \sigma^2 \), i.e. tensor \( B \) coincides with the hessian of \( \psi \). Thus, \( B \) determines how the normal \( r \) to the loading surfaces \( \psi = \text{const} \) varies over neighbouring points, i.e. it is a measure of the curvature of the loading surface.
(2) Let $B_1$ and $B_2$ be evaluated at two arbitrary points $\sigma_1$ and $\sigma_2$ on the yield surface. Then, the scalar

$$s(\sigma_1, \sigma_2) = \sup_{\delta} \frac{\delta : B_1 : \delta}{\delta : B_2 : \delta}$$

(50)

where the supremum is taken over all incremental stresses $\delta$, is a measure of the curvature of the loading surface at point $\sigma_1$ relative to that at point $\sigma_2$. From this perspective, the scalar

$$s = \sup_{\sigma_1, \sigma_2} s(\sigma_1, \sigma_2)$$

(51)

where the supremum is taken over all pairs of stress point $\sigma_1$ and $\sigma_2$ on the yield surface, gives an idea of the extent of variation of the curvature of the loading surface over points of the yield surface. This provides the reason for calling $s$ the distortion index of the loading surface. If, for instance, one takes $\psi = \frac{1}{2} s_{ij} \delta_{ij}$, with $s_{ij} = \sigma_{ij} - (\sigma_{kk}/3)\delta_{ij}$, corresponding to the Prandtl–Reuss flow rule, it follows that $B_{ijkl} = \partial^2 \psi / \partial \sigma_{ij} \partial \sigma_{kl} = \delta_{ik} \delta_{jl}$, which is constant over the associated von Mises yield surface $\phi = \psi = k^2$. Hence, from (50) and (51) it follows that the distortion index takes the value $s = 1$. In other words, the value $s = 1$ corresponds to a loading surface of constant curvature, such as that associated with the Prandtl–Reuss flow rule. If, on the contrary, the loading surface exhibits corners an infinite distortion index $s = \infty$ is obtained.

(3) It is interesting to note how the shape of the loading surface strongly influences the stability properties of the generalized trapezoidal rule. Whereas for linear dissipative equations of evolution, such as the heat equation, the trapezoidal rule is known to be unconditionally stable for $\alpha \geq \frac{1}{2}$, in the present context a high degree of distortion of the loading surface significantly reduces the unconditional stability range, requiring values of $\alpha$ close to 1. In particular, it follows from (47) that for loading surfaces with corners, for which case $s = \infty$, the only value of $\alpha$ leading to unconditional stability is $\alpha = 1$. This choice corresponds to the closest point return mapping algorithm proposed in Reference 4. Also, the critical value of the incremental plastic parameter $\lambda$, defined in (49) is seen to be highly sensitive to the degree of distortion of the loading surface. In this sense, the von Mises model, for which $s = 1$, is a best possible case for the trapezoidal rule since unconditional stability follows for $\alpha \geq \frac{1}{2}$ as in the linear setting.

**Generalized midpoint rule.** An entirely similar analysis of the generalized midpoint rule leads to the following estimate:

$$\| d\sigma_{n+1} \| \leq \frac{1 - (1 - \alpha) \lambda \beta_{n+2}}{1 + \alpha \lambda \beta_{n+2}} \| d\sigma_n \|$$

(52)

where now $\beta_{n+2}$ is the maximum eigenvalue of the tensor $B_{n+2}$. Making use of inequality (42), (52) reduces to

$$\| d\sigma_{n+1} \| \leq \left| \frac{1 - \alpha}{\alpha} \right| \| d\sigma_n \| = c \| d\sigma_n \|$$

(53)

where one writes

$$c = \left| \frac{1 - \alpha}{\alpha} \right|$$

(54)

Unconditional stability requires $c \leq 1$ which in view of (54) necessitates

$$\alpha \geq \alpha_{\text{min}} = \frac{1}{2}$$

(55)
In conclusion, the generalized midpoint rule is unconditionally stable for $\alpha \geq \frac{1}{2}$, regardless of the choice of loading potential. This is in sharp contrast with the generalized trapezoidal rule for which stability has been shown above to be strongly dependent on the shape of the loading surface.

For $\alpha < \frac{1}{2}$ the stability of the generalized midpoint rule is only conditional. Making use of (48) and (49) the following stability condition is obtained:

$$\lambda \leq \frac{2/\beta_{\text{max}}}{1 - 2\alpha} = \lambda_c, \quad \alpha < \frac{1}{2}$$

which, unlike (49), is independent of the choice of loading potential. For $\alpha = \frac{1}{2}$, the critical value $\lambda_c$ becomes infinity and unconditional stability is recovered.

5. NUMERICAL EXAMPLES

In this section, the accuracy analysis of the proposed algorithm is completed by way of numerical testing. For this purpose, the perfectly plastic von Mises model is considered. In this simple case, the generalized trapezoidal and midpoint rules coincide. The position of the stress points on the yield surface may be characterized by an angle $\theta$, as indicated in Figure 3. Furthermore, let $\Delta e$ denote the prescribed deviatoric strain increment. Figure 3 displays a generic $\Delta e$ resolved into its radial component $\Delta e_r$ and tangential component $\Delta e_t$.

Figures 4(a–f) show the isoeffect maps for the angle $\theta$ computed numerically as a function of the prescribed radial and tangential strain increments and the algorithmic parameter $\alpha$. The exact solution for $\theta$ is given in Reference 1. It is observed that for small strain increments optimal accuracy is obtained for $\alpha = \frac{1}{2}$, for which choice of $\alpha$ the integration rules have been shown to be second-order accurate. By contrast, when large strain increments are present higher values of $\alpha$ consistently lead to better accuracy. For the limiting case of $|\Delta e| \to \infty$, for instance, the exact value $\theta = \pi/2$ is only matched exactly by the radial return algorithm, $\alpha = 1$, whereas the mean normal procedure $\alpha = \frac{1}{2}$ yields the highest error.

In conclusion, the optimal choice of $\alpha$ may depend on the nature of the problem under consideration. When large strain increments are anticipated the closest point algorithm $\alpha = 1$ is probably optimal. On the other hand, in problems where strain increments are kept to a small size the choice $\alpha = \frac{1}{2}$ may yield improved accuracy with respect to the closest point procedure.

6. SUMMARY AND CONCLUSIONS

Two families of algorithms have been presented which generalize the trapezoidal and midpoint rules in a manner that facilitates satisfaction of the plastic consistency condition. The algorithms are applicable to arbitrary non-associated hardening plasticity models and contain
Figure 4. Isoerror maps for the generalized trapezoidal and midpoint rules plotted as a function of the tangential and radial deviatoric strain increments. Results correspond to the perfectly plastic von Mises model and to various choices of the algorithmic parameter $\alpha$ within the region of unconditional stability. Errors are measured in terms of the angle $\theta$, see Figure 3.
well-known integration schemes such as the radial return, mean normal and closest point procedures as particular cases.

As in the linear case, it shown that the choice \( \alpha = \frac{1}{2} \) of the algorithmic parameter results in second-order accuracy. Numerical examples demonstrate, however, that in the presence of large strain increments larger values of \( \alpha \) may result in improved accuracy.

A new methodology has been also proposed which allows a systematic assessment of the numerical stability properties of integration rules for elastoplastic constitutive relations. A salient feature of such methodology is the fact that it requires only consideration of infinitesimal perturbations in the initial conditions. This renders the stability analysis linear and, hence, readily tractable. With the aid of this method, it is observed that the stability properties of the generalized trapezoidal rule are very sensitive to the degree of distortion of the loading surface. In particular, in the presence of corners stability of the trapezoidal rule requires \( \alpha = 1 \), which corresponds to the closest point procedure. By contrast, the generalized midpoint rule is unconditionally stable for \( \alpha \geq \frac{1}{2} \), regardless of the choice of loading surface. This remarkable fact would appear to point to the generalized midpoint rule as preferable to the generalized trapezoidal rule, apart from simple cases such as the von Mises model for which both integration rules coincide.

REFERENCES