SYMMETRY-PRESERVING RETURN MAPPING ALGORITHMS AND INCREMENTALLY EXTREMAL PATHS: A UNIFICATION OF CONCEPTS

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

JOHN B. MARTIN
Department of Civil Engineering, University of Cape Town, Rondebosch, Cape Town, South Africa

SUMMARY
In this work we seek to characterize the conditions under which an elastic–plastic stress update algorithm preserves the symmetries inherent to the material response. From a numerical standpoint, the aim is to determine under what conditions a stress update algorithm produces symmetric consistent tangents when applied to materials obeying normality. For the ideally plastic solid we show that only the fully implicit or closest point return mapping algorithm is symmetry preserving. For hardening plasticity, symmetry cannot be preserved in general unless suitable restrictions are imposed on the constitutive equations. We show that these restrictions amount to the existence of a pseudo-internal energy function acting as a joint potential for both the direction of plastic flow and the hardening moduli. In view of the fact that holonomic methods based on incrementally extremal paths also result in update rules possessing a potential structure and, hence, in symmetric tangents, we address the question of whether any connections exist between the two approaches. We show that holonomic methods and the fully implicit algorithm may indeed be brought into correspondence.

1. INTRODUCTION
Algorithms for the integration of constitutive relations play a central role in computational plasticity. Some recently proposed implicit formulations\textsuperscript{5,25} naturally lead to the notion of consistent tangent moduli, obtained by differentiation of the update algorithm with respect to the incremental strains. An undesirable feature of the consistent tangents is that they may not preserve constitutive symmetries, e.g. the tangents may not be symmetric for associated plasticity. The symmetry of the tangent operator is the cornerstone of numerous regularity properties and bounding theorems of plasticity.\textsuperscript{10} It also has a profound influence on questions of material stability and shear banding.\textsuperscript{22} It is therefore of critical importance that the numerical procedure preserve the symmetry of the problem. Furthermore, non-symmetric matrices add to the factorization cost of the analysis. Thus, it is of theoretical and practical interest to determine conditions under which the numerical tangents possess the symmetries inherent to the constitutive relations.

Two seemingly unrelated classes of algorithms are found to preserve symmetry. One is the fully implicit or backward Euler method, which in turn coincides with the closest point return mapping algorithm. The second class is furnished by holonomic approximations constructed from


Received 13 June 1988
Revised 28 December 1988
incrementally extremal paths. In the latter case, the rate equations of plasticity are replaced by a deformation theory based on the incremental strains. The existence of an incremental strain energy potential insures symmetry of the consistent tangents. It is shown in this paper that, under suitable restrictions on the constitutive equations, the fully implicit algorithm and holonomic methods based on extremal paths can in fact be brought into correspondence. Thus, the fully implicit algorithm is seen to possess a potential structure resulting in symmetry of the tangents. A further conclusion is that the fully implicit algorithm integrates the rate equations of plasticity exactly along the extremal paths, i.e. strain paths which minimize the work of deformation.

2. STRESS UPDATE ALGORITHMS AND CONSISTENT TANGENTS

Throughout this paper, attention is confined to the small strain plasticity problem. This section establishes our notation and reviews some concepts which lie at the basis of subsequent discussions. We focus on the generalized midpoint rule as a representative class of stress update algorithms. In addition, we briefly review some aspects of an implicit finite element formulation which leads to the notion of consistent elastoplastic tangents.

2.1. Constitutive equations

The constitutive equations governing the behaviour of an elastoplastic solid may be expressed in the form

\[ \sigma = D^e \varepsilon + D^p \varepsilon^p \]  
\[ \dot{\varepsilon}^p = \dot{\gamma}^p \tau(\sigma, q) \]  
\[ \dot{q} = \dot{\gamma}^p h(\sigma, q) \]  
\[ \dot{\gamma} = \phi(\sigma, q)/\eta \]

where \( \sigma, \varepsilon, \varepsilon^p \) are the stress, strain and plastic strain tensors, respectively, \( q \) is some set of plastic variables, \( \tau \) is the plastic flow direction, \( h \) are the plastic moduli, \( \gamma \) may be regarded as an effective plastic strain, \( \phi \) is an effective driving stress, \( D^e \) are the elastic moduli and \( \eta \) a scalar viscosity parameter. Equation (1) is assumed to govern the elastic response of the material, equation (2) is the flow rule and equations (3) and (4) are the hardening and viscosity laws, respectively.

The rate-independent limit can be attained by letting \( \eta \to 0 \), i.e. as the inviscid limit, in which case the requirement that \( \dot{\gamma} < \infty \) necessitates

\[ \phi = 0 \]

during plastic loading. This is the yield criterion of inviscid plasticity, which replaces the viscosity law (4) of viscoplasticity. The yield criterion and the loading-unloading conditions can be expressed in Kuhn–Tucker form as the requirement that the constraints

\[ \phi \leq 0 \]
\[ \dot{\gamma} \geq 0 \]
\[ \phi \dot{\gamma} = 0 \]

be simultaneously satisfied at all times.
For rate-independent plasticity and under conditions of plastic loading, a standard argument yields the following incremental relation between stress and strain:

\[
\dot{\sigma} = D^p : \epsilon
\]

(7)

where

\[
D^p = D^e - \left( \frac{(D^e : r)}{v : D^e : r - \xi : h} \right) \epsilon
\]

(8)

are the elastic-plastic moduli, and one writes

\[
v = \frac{\partial \phi}{\partial \sigma}, \quad \xi = \frac{\partial \phi}{\partial q}
\]

(9)

No such moduli exist in the case of rate-dependent plasticity.

For rate-independent plasticity, the flow rule is said to be associative if

\[
r = \frac{\partial \phi}{\partial \sigma} = v
\]

(10)

This condition is also referred to as the normality rule, in as much as it requires that the direction of incremental plastic flow \( r \) point in the direction of the normal \( v \) to the yield surface. As may be directly appreciated from equation (8), normality results in the symmetry of the elastic-plastic moduli. For the rate-dependent solid, by contrast, no set of moduli exist other that the elastic and an analogous concept of normality cannot be introduced from considerations of symmetry alone.

2.2 Stress-update algorithms

Regardless of the solution procedure utilized in the analysis, at some point during the computations it becomes necessary to integrate the constitutive equations (1) to (4). In a displacement finite element formulation, this process is strain-driven, i.e. the given of the problem are the strain increment \( \Delta \epsilon \) and the initial conditions \( \sigma_n, \epsilon, q_n, \gamma_n \), henceforth collectively denoted as \( A_n \), and the sought outcome is the updated state \( A_{n+1} \). Here and subsequently, the subindices \( n \) and \( n+1 \) are used to identify the value of the various state variables at times \( t_n \) and \( t_{n+1} = t_n + \Delta t \).

An example of a stress-update algorithm is furnished by the generalized midpoint rule, which in the present context takes the form

\[
\epsilon_{n+1} = \epsilon_n + \Delta \epsilon
\]

\[
\sigma_{n+1} = D^e : (\epsilon_{n+1} - \epsilon_n)
\]

\[
\epsilon_{n+1}^p = \epsilon_n^p + \Delta \gamma r_{n+1}
\]

\[
q_{n+1} = q_n + \Delta \gamma h_{n+1}
\]

\[
\Delta \gamma = \Delta t \phi_{n+1}/\eta
\]

(11)

where one writes

\[
r_{n+1} = r(\sigma_{n+1}, q_{n+1})
\]

\[
h_{n+1} = h(\sigma_{n+1}, q_{n+1})
\]

\[
\phi_{n+1} = \phi(\sigma_{n+1+1}, q_{n+1+1})
\]

\[
\sigma_{n+1} = (1 - \alpha) \sigma_n + \alpha \sigma_{n+1}
\]

\[
q_{n+1} = (1 - \alpha) q_n + \alpha q_{n+1}
\]

(12)
and the algorithmic parameter $\alpha$ takes values in the interval $[0, 1]$. Note the requirement that the viscosity law be satisfied in a fully implicit manner regardless of the value of $\alpha$. In this fashion, in the inviscid limit $\eta \to 0$ the last equations (11) reduces to

$$\phi_{n+1} = 0$$

(13)

which necessitates compliance with the yield criterion at time $t_{n+1}$.

Some particular cases of this algorithm are noteworthy. For $\alpha = 1$ and rate-independent plasticity, one recovers the closest point return algorithm, whereby the elastic predictor is relaxed into the closest point on the updated yield surface. Note that closedness in stress space is here understood in the sense of the metric defined by the inverse elastic moduli $D^{-1}$. If, on the other extreme, one takes $\alpha = 0$ and linearizes with respect to $\Delta y$ one obtains the rate tangent procedure of Pierce et al.\textsuperscript{16, 17} For $\alpha = 1/2$ and $J_2$ flow theory, the algorithm reduces to Rice and Tracy's mean normal method.\textsuperscript{23} An analysis given by Ortiz and Popov\textsuperscript{15} shows that the algorithm is second-order accurate for $\alpha = 1/2$ and unconditionally stable for $\alpha \geq 1/2$, regardless of the shape of the yield surface. This latter property is in contrast to the generalized trapezoidal rule for which the range of stability of the parameter $\alpha$ is very sensitive to the degree of distortion of the yield surface, being significantly reduced by the presence of regions of high curvature. It is also noteworthy that the second-order accuracy which is accomplished for $\alpha = 1/2$ only results in increased accuracy in the limit of small strain increments. When large steps are to be expected, the choice $\alpha = 1$ results in the highest accuracy.\textsuperscript{6, 15, 24}

Expressions (11) define a set of non-linear equations to be solved for $A_{n+1}$. For complex material models, the effort involved in a Newton–Raphson solution of the equations can be rather substantial. An exception is the semi-implicit method $\alpha = 0$, for which the only unknown defined implicitly is the scalar $\Delta y$. The semi-implicit algorithm, however, lacks the robustness of its fully implicit counterpart and, as it will be demonstrated subsequently, does not preserve constitutive symmetries.

2.3. Linearization of momentum balance

In a finite element setting, the equilibrium equations are satisfied weakly. This results in a system of algebraic constraints

$$\sum_{e} \int_{\Omega_e} B^T \sigma(t) d\Omega_e = f(t)$$

(14)

where $\{ \Omega_e, e = 1, \ldots, Nel \}$ signifies a finite element partition of the domain of analysis, $B$ is the discrete symmetric gradient operator and $f$ are the applied nodal forces. Enforcing satisfaction of (14) at time $t_{n+1}$ we obtain

$$\sum_{e} \int_{\Omega_e} B^T \sigma_{n+1} d\Omega_e = f_{n+1}$$

(15)

Next we note that the net effect of a stress-update algorithm is to define, possibly implicitly, a relation of the type

$$\sigma_{n+1} = \bar{\sigma}(\varepsilon_{n+1}; A_n, \Delta t)$$

(16)

where the parametric dependence on the initial conditions $A_n$ and on the time increment is made explicit for emphasis. Thus, the function $\bar{\sigma}$ represents a general stress-update algorithm. The dependence on $\Delta t$ expressed in (16) drops out in the rate-independent limit, rendering the incremental solution independent of the choice of time scale.

Using (16) we have

$$\sum_{e} \int_{\Omega_e} B^T \bar{\sigma}(\varepsilon_{n+1}; A_n, \Delta t) d\Omega_e = \sum_{e} \int_{\Omega_e} B^T \sigma_{n+1} d\Omega_e = f_{n+1}$$

Finally, the displacement increment is

$$u_{n+1} = u_{n+1}$$

which defines a coupled algorithm. It is equally well possible to express the velocity in a weak form; however, the procedure is more cumbersome and, as is shown in the earlier articles of this series, the coupled approach will include the dependency on $\varepsilon_{n+1}$ on the right side of the constitutive equation, which is appealing in practice as this requires no iteration at all.

Note that in the variational setting, in particular, in the case of the stress update, the linearization take the form

$$\sum_{e} \int_{\Omega_e} B^T \sigma_{n+1} d\Omega_e = f_{n+1}$$

where $B$ is the symmetric gradient operator.

Note that these equations are strongly coupled.

In the rate-independent limit, the stress update is elastic–plastic and linear in the time step, i.e., the stress update is an implicit algorithm. For non-linear problems, however, the procedure requires that

$$\sigma_{n+1} = \bar{\sigma}(\varepsilon_{n+1}; A_n, \Delta t)$$

Take a particular point

$$\sigma_{n+1} = \bar{\sigma}(\varepsilon_{n+1}; A_n, \Delta t)$$

for some given $\varepsilon_{n+1}$ and $A_n$. Clearly, the left side of this equation is non-linear in $\varepsilon_{n+1}$ and $A_n$. Thus, the left side must be evaluated at some particular value of $\varepsilon_{n+1}$ and $A_n$. In the rate-independent case, the linear constitutive equation guarantees that

$$\sigma_{n+1} = \bar{\sigma}(\varepsilon_{n+1}; A_n, \Delta t)$$

where $\bar{\sigma}$ is the rate-independent stress function, and $\varepsilon_{n+1}$ and $A_n$ are given. Therefore, a coupled algorithm in the rate-independent limit is absolutely consistent.
Using (16), equation (15) can be recast in terms of the incremental strains to read

$$
\sum_e \int_{\Omega_e} B^T \delta(\varepsilon_{n+1}; \Lambda_n, \Delta t) d\Omega_e = f_{n+1}
$$

(17)

Finally, the discretized compatibility equations can be expressed as

$$
\varepsilon_{n+1} = Bu_{n+1}
$$

(18)

where $u_{n+1}$ are the updated nodal displacements. Combining (17) and (18) we find

$$
\sum_e \int_{\Omega_e} B^T \delta(Bu_{n+1}; \Lambda_n, \Delta t) d\Omega_e = f_{n+1}
$$

(19)

which defines a set of non-linear equations to be solved for $u_{n+1}$. Note that the equations are equally well defined for the rate-dependent as well as for the rate-independent cases. This solution procedure is due to Simo and Taylor,\textsuperscript{25} although an historical precedent may be found in an earlier article of Hughes and Taylor.\textsuperscript{5} The method overcomes some difficulties associated with earlier approaches in which the state was updated at every equilibrium iteration. Such difficulties include the occurrence of spurious loading–unloading cycles\textsuperscript{9} and the loss of quadratic convergence of the equilibrium iterations.\textsuperscript{14} Furthermore, the formulation is conceptually appealing in that it makes explicit unequivocally the discretized field equations which the solution is required to satisfy.

Note that the nature of equation (19) is strongly influenced by the choice of update algorithm. In particular, in a Newton–Raphson solution of (19), the consistent tangents obtained by linearization take the form

$$
K = \sum_e \int_{\Omega_e} B^T \tilde{D} B d\Omega_e
$$

(20)

where $\tilde{D}$ is the consistent tangent of Simo and Taylor,\textsuperscript{25} defined as

$$
\tilde{D} = \frac{\partial \delta(\varepsilon_{n+1}; \Lambda_n, \Delta t)}{\varepsilon_{n+1}}
$$

(21)

Note that the consistent tangent is dependent on the choice of update procedure.

In the rate-independent limit, the consistent moduli generally differ from the classical elastic–plastic tangents $D^{\text{ep}}$ expressed in (8). However, $\tilde{D}$ does reduce to $D^{\text{ep}}$ at the beginning of a time step, i.e. for $\varepsilon_{n+1} = \varepsilon_n$. That this is so, is a direct consequence of the consistency of the update algorithm. For a given strain history $\varepsilon(t)$ and exact initial conditions at, say, $t_n$, consistency requires that the incremental solution match the exact one to within terms of order $O(\Delta t^2)$, i.e.

$$
\delta(\varepsilon(t_n + \Delta t); \Lambda_n) = \sigma_n + \dot{\varepsilon}_n \Delta t + O(\Delta t^2)
$$

(22)

$$
\dot{\varepsilon}_n = D^{\text{ep}} \varepsilon_n
$$

Take a particular strain history of the form

$$
\varepsilon(t) = \varepsilon_n + \tau \Delta \varepsilon, \quad \tau = t - t_n
$$

(23)

for some given $\Delta \varepsilon$. Then, the first of (22) becomes

$$
\delta(\varepsilon_n + \tau \Delta \varepsilon; \Lambda_n) = \sigma_n + D^{\text{ep}} \Delta \varepsilon_n + O(\Delta t^2)
$$

(24)
Differentiate with respect to \( \varepsilon_{n+1} \) to obtain

\[
\frac{\partial \hat{\sigma}}{\varepsilon_{n+1}} (\varepsilon_{n} + \tau \Delta \varepsilon; \Lambda_n) = D_{n}^{ep} + O(\tau^2)
\]  

(25)

Finally, take the limit as \( \tau \to 0 \) to arrive at the sought result

\[
\frac{\partial \hat{\sigma}}{\varepsilon_{n+1}} (\varepsilon_{n}; \Lambda_n) = D_{n}^{ep}
\]  

(26)

3. SYMMETRY CONDITIONS

Next we seek to determine conditions under which the consistent tangents associated with an update algorithm preserve the symmetries of the constitutive equations. In particular, we wish to ascertain under what conditions the consistent tangents are assured to be symmetric whenever plastic flow is associative. We start by addressing the relatively simpler case of ideal plasticity, following which we endeavor to generalize the results to the hardening range.

3.1 Ideal plasticity

Start by considering constitutive relations appropriate to ideal plasticity,

\[
\sigma = D : (\varepsilon - \varepsilon^p) \\
\dot{\varepsilon}^p = \gamma r(\sigma) \\
\dot{\gamma} = \phi(\sigma) / \eta
\]  

(27)

Let these equations be integrated by means of the generalized midpoint rule (11). Then, a direct computation yields the following expression for the consistent tangent:

\[
\dot{D} = \dot{D}_{ep} + \frac{\gamma}{\eta} \frac{\partial \psi}{\partial \sigma} \dot{\varepsilon}^p
\]  

(28)

where the effective elastic moduli are given by

\[
\dot{D}_{ep} = \frac{\gamma}{\eta + \gamma} \frac{\partial \psi}{\partial \sigma} \dot{\varepsilon}^p
\]  

(29)

We turn now to the question of the symmetry of \( \dot{D} \). The usual concept of normality expressed in (10) proves too restrictive in the rate-dependent case. Many material models which do result in symmetric tangents do not obey normality in the sense of (10). A simple example is furnished by the model defined by equations (37). A less stringent concept of normality which is better suited to rate-dependent behaviour is to require that

\[
r = \frac{1}{f} \frac{\partial \phi}{\partial \sigma} = \frac{\nu}{f}
\]  

(30)

for some scalar function \( f(\sigma, q) \). For example, normality in this particular sense is complied with by models for which a flow potential \( \psi(\sigma, q) \) exists, i.e. for which \( r = \partial \psi / \partial \sigma \), such that \( \phi = F(\psi) \) for some function \( F \). Then it follows that \( f = F'(\psi) \).

If normality is to be expressed in the sense of (30), for some functions \( \sigma = \sigma(q) \) and for some function \( \psi \) defined on the hardening surface, then

\[
r = \frac{1}{f} \frac{\partial \phi}{\partial \sigma} = \frac{\nu}{f}
\]  

(32)

i.e. only the flow potential \( \psi \) enters.

3.2 Hardening

By virtue of our uniqueness assumption, for \( \alpha < 1 \) as candidate for the modulus of the hardening law, we can express the generalized hardening law by

\[
\frac{\partial \psi}{\partial \sigma} = \psi
\]  

(33)

within the given hardening law. For a more detailed investigation, we assume the only exact integral of \( \phi \) is

\[
\phi = \psi
\]  

(34)

for some function \( \psi \) defined on the hardening surface.

It is tempting to introduce constitutive relations a third time,

\[
\sigma = D : (\varepsilon - \varepsilon^p)
\]  

(35)

but this can be shown to be inconsistent with the rate-dependent plasticity and normality conditions.

There are then two possibilities: Either the hardening function \( \psi \) depends on the hardening variables, or it depends on the effective stress \( \sigma \). In such dependence, the normality condition (30) is assured.

...
If normality holds in the sense of (30), the consistent moduli take the form

\[ \bar{D} = \bar{D} + \bar{D} = \frac{\bar{D}^e_{n+1} \otimes \bar{D}^e_{n+1}}{\bar{f}_{n+1} + v_{n+1} \cdot \bar{D}^e_{n+1}} \]  

(31)

From this expression it is apparent that

\[ \bar{D}^T = \bar{D} \Rightarrow \xi = 1 \]  

(32)

i.e. only the fully implicit algorithm is symmetry preserving.

3.2. Hardening plasticity

By virtue of the results of the preceding section, we may now rule out from the outset the cases \( \alpha < 1 \) as candidates for symmetry-preserving algorithms. Thus, it suffices to confine our attention to the fully implicit algorithm \( \alpha = 1 \). For work-hardening materials, however, normality cannot be generally expected to result in symmetric consistent tangents unless additional restrictions are imposed on the constitutive equations. In this section, we discuss a particular constitutive framework wherein the consistent tangents associated with the fully implicit algorithm are indeed symmetry preserving. This constitutive framework is closely related to that considered by Carter and Martin \(^1\) in their discussion of work bounding functions in work-hardening materials and falls within the general class of models discussed by Rice \(^2\). These points of contact are explored in more detail in Section 4. It is not known to us whether the constitutive restrictions outlined next are the only ones which insure symmetry, although such a scenario appears quite plausible.

Assume that the direction of plastic flow \( r \), hardening moduli \( h \) and flow potential \( \phi \) can be expressed in the form

\[ r(\sigma, q) = \frac{\partial \psi}{\partial \sigma}(\sigma, Q(q)), \quad h(\sigma, q) = \frac{\partial \psi}{\partial Q}(\sigma, Q(q)), \quad \phi(\sigma, q) = F(\psi(\sigma, Q(q))) \]  

(33)

for some function \( F \) and scalar potential \( \psi(\sigma, Q) \). The quantities \( Q(q) \) themselves are assumed to derive from a potential, i.e.

\[ Q(q) = -\frac{\partial U^p(q)}{\partial q} \]  

(34)

for some function \( U^p(q) \). Thus, \( \psi \) acts as a join potential for both the direction of plastic flow and the hardening moduli, and the rate of effective plastic flow can be expressed as a direct function of \( \psi \).

It is tempting, although not critical to the present discussion, to attribute to the foregoing relations a thermodynamic meaning. Indeed, a function of the type

\[ U(\varepsilon - \varepsilon^p, q) = \frac{1}{2}(\varepsilon - \varepsilon^p)^T D^{e-1}(\varepsilon - \varepsilon^p) + U^p(q) \]  

(35)

can be shown to be the most general form of the internal energy of a solid exhibiting linear elasticity and decoupled elastic and plastic responses. The quantities

\[ \sigma = -\frac{\partial U}{\partial \varepsilon^p}, \quad Q = -\frac{\partial U}{\partial q} = -\frac{\partial U^p}{\partial q} \]  

(36)

are then the thermodynamic forces conjugate to the internal variables \((\varepsilon^p, q)\). From this perspective, the assumptions outlined above imply that the rate of variation of all internal variables depends solely on the current value of the conjugate thermodynamic forces, and that such dependence possesses a potential structure.
Example. Consider the case of $J_2$ flow theory with isotropic power hardening and power viscosity

\[ \dot{\gamma}_{ij} = \frac{3s_{ij}}{2\sigma} \]
\[ \dot{\gamma} = \dot{\gamma}_0 \left( \frac{\sigma}{\sigma_0} - 1 \right)^m \]
\[ \sigma_0 = \sigma_0 \left( 1 + \frac{\gamma}{\gamma_0} \right)^{1/m} \]

where

\[ \sigma = \left( \frac{3}{2} s_{ij} s_{ij} \right)^{1/2}, \quad s_{ij} = \sigma_{ij} - (\sigma_{kk}/3)\delta_{ij} \]

and $\dot{\gamma}_0$, $\sigma_0$, $\sigma_\eta$, $\gamma_0$, $m$ and $n$ are material constants. In this simple example, the only independent internal variable may be identified with $\gamma$. The hardening law for $\gamma$ is the trivial equation $\dot{\gamma} = \gamma$, which is of the type (3) with $h = 1$. The variable conjugate to $\gamma$ may be taken to be the flow stress

\[ Q(\gamma) = \sigma_0 = \sigma_\eta \left( 1 + \frac{\gamma}{\gamma_0} \right)^{1/m} \]

and the flow potential to be

\[ \psi(\sigma, Q) = \sigma - \sigma_0 = \sigma - Q \]

Since the only internal variable in this example is a scalar, the potential relation (34) is trivially obtained by integrating the last of (37) and letting

\[ U^\eta(\gamma) = -\frac{n\sigma_0 \gamma_0}{n+1} \left( 1 + \frac{\gamma}{\gamma_0} \right)^{(n+1)/n} \]

It is readily checked that

\[ \frac{\partial \psi}{\partial \sigma_{ij}} = \frac{3s_{ij}}{2\sigma} = r_{ij}, \quad \frac{\partial \psi}{\partial Q} = -1 = h \]

Finally, one verifies that the viscosity law may be recast as

\[ \dot{\gamma} = \frac{\phi}{\eta} \]

by choosing $\eta = \sigma_0/\gamma_0$ and letting

\[ \phi = \frac{\sigma}{\eta} \left( \frac{\psi}{\sigma_0} \right)^m \]

A direct computation of the consistent tangents yields

\[ \dot{D} = \langle 1, 0 \rangle \cdot \left( A - \frac{(A \cdot N) \otimes (A \cdot N)}{f\eta/\Delta t + N \cdot A \cdot N} \right) \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\} \]

where

\[ A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \]

and

In these expressions, the state $\Lambda_{n+1}$. The constitutive functions from the stress space (c) in Section 3.1. In the constitutive model for the stress vector $\sigma$, the constitutive relation is

These are for the second the flow stress depends on the

A mathematical description of the incremental potential

i.e. the stress

In some situations, the rate-independent constitutive model projects

\[ \mu \]

where $I$ signifies $0$ and $W$ is

It is intriguing to

and based on increase
where
\[
A^{-1} = \begin{pmatrix}
D^{-1} + \Delta \gamma \partial^2 \psi / \partial \sigma \partial \sigma & \Delta \gamma \partial^2 \psi / \partial \sigma \partial Q \\
\Delta \gamma \partial^2 \psi / \partial Q \partial \sigma & M^{-1} + \Delta \gamma \partial^2 \psi / \partial Q \partial Q
\end{pmatrix}, \quad N = \begin{pmatrix}
\partial \psi / \partial \sigma \\
\partial \psi / \partial Q
\end{pmatrix}
\]
(46)

and
\[
M = \frac{\partial \bar{Q}}{\partial \bar{q}} = -\frac{\partial^2 \bar{U}^p}{\partial \bar{q}^2} = M^T
\]
(47)

In these expressions it is tacitly understood that all functions are to be evaluated at the updated state $A_{n+1}$. The symbol $I$ in (45) is used to signify the identity matrix in stress space.

The consistent tangent (45) is obviously symmetric. That this should indeed be so is hardly surprising once one realizes that, by virtue of the restrictions formulated at the outset, the constitutive framework under consideration is identical to that of ideal plasticity in the generalized stress space $(\sigma, Q)$. Thus, the present results are but a straightforward extension of those given in Section 3.1. In order to bring out this correspondence more clearly, let $X = (\sigma, Q)$ be a generalized stress vector and $\chi = (\epsilon^p, q)$ an extended internal variable vector. In this notation, the assumed constitutive relations may be expressed as
\[
X = \frac{\partial U}{\partial \chi},
\]
\[
\dot{X} = \frac{\partial \psi}{\partial X},
\]
\[
\dot{\chi} = \frac{\partial \psi}{\partial X}
\]
(48)

These are formally identical to equation (28), with the first replacing the elastic response, the second the flow rule and the third the viscosity law.

4. RELATION TO HOLONOMIC METHODS BASED ON INCREMENTALLY EXTREMAL PATHS

A mathematical consequence of the symmetry of the consistent tangent $D$ is the existence of an incremental potential $\hat{W}(\epsilon_{n+1}; A_n, \Delta t)$ such that
\[
\sigma_{n+1} = \delta(\epsilon_{n+1}; A_n, \Delta t) = \frac{\partial \hat{W}}{\partial \epsilon_{n+1}}
\]
(49)
i.e. the stress update relations derive from the incremental potential $\hat{W}$.

In some simple cases, $\hat{W}$ can be written down explicitly. Consider, for example, the case of ideal rate-independent plasticity. Let $C$ denote a convex elastic domain and denote by $P_C$ the closest point projection onto $C$. Then one has
\[
\hat{W}(\epsilon_{n+1}; A_n, \Delta t) = \frac{1}{2} \Delta \epsilon : D^e : \Delta \epsilon + \frac{1}{2} [(I - P_C)(\sigma^e_{n+1})] : D^{-1} : [(I - P_C)(\sigma^e_{n+1})]
\]
\[
\sigma^e_{n+1} = \sigma_n + D^e : \Delta \epsilon, \quad \Delta \epsilon = \epsilon_{n+1} - \epsilon_n
\]
(50)

where $I$ signifies the identity mapping. Note that for $\sigma^e_{n+1}$ in the interior of $C$ one has $(I - P_C)(\sigma^e_{n+1}) = 0$ and $\hat{W}$ reduces to the elastic potential, as befits elastic unloading.

It is intriguing that a seemingly disconnected family of algorithms, namely holonomic methods based on incrementally extremal paths,\(^3\) results precisely in the same potential structure. The
essence of these methods is to integrate exactly the constitutive relations along strain paths which minimize the plastic dissipation. The results of this section demonstrate that these similarities are not fortuitous. Indeed, we show that the fully implicit algorithm and holonomic methods can be brought into correspondence.

A succinct review of basic concepts of extremal paths which bear on the present discussion is presented first. For further details, the interested reader is referred to the monograph of Martin.7 Throughout this section attention is confined to the rate-independent solid. We also assume throughout that the material is stable in the sense of Drucker.5 Consider all possible paths of deformation \( \varepsilon(\tau), \tau = t - t_n \), joining the initial and updated strain tensors \( \varepsilon_n \) and \( \varepsilon_{n+1} \), respectively. Thus, all paths under consideration are subject to the constraints

\[
\varepsilon(0) = \varepsilon_n, \quad \varepsilon(\Delta t) = \varepsilon_{n+1}
\]

The work done in deforming a material element in the period \( 0 \leq \tau \leq \Delta t \) is

\[
W[\varepsilon] = \int_0^{\Delta t} \sigma(\tau) : \varepsilon(\tau) \, d\tau
\]

where the notation adopted purports to indicate that \( W \) is a functional of the path \( \varepsilon(\tau) \).

We seek to obtain a bounding function \( \hat{W}(\varepsilon_{n+1}; \Lambda_n) \) such that

\[
\hat{W}(\varepsilon_{n+1}; \Lambda_n) \leq W[\varepsilon]
\]

for any choice of \( \varepsilon_{n+1} \) and \( \varepsilon(\tau) \). To be consistent with the notation of previous sections, we make explicit the parametric dependence of \( \hat{W} \) on the initial conditions \( \Lambda_n \) for the time step.

The concept of extremal path was introduced by Martin11,12 and led to the consistent development of work bounding principles by Ponter,16,19 Soechting and Lange,26 Martin13 and Ponter and Martin.20 A property of work bounding functions which is of particular significance here is the fact that they act as a potential for the terminal state of stress, i.e.

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

Martin et al.3 have proposed using relations of this type as stress update algorithms. In essence, the critical step in the method is to characterize the incremental extremal paths for particular material models. Bounding functions for specific constitutive models are available in the literature.4,8,11,12,20,26 It bears emphasis that equation (54) provides the exact solution for the updated stresses when the strain history does in fact coincide with the incremental extremal path. It should also be noted that, by virtue of the potential structure of the update (54), the consistent tangents are automatically symmetric.

4.1. Perfect plasticity

Here we seek to characterize the extremal strain paths in an ideally plastic solid. Making use of the elastic–plastic additive decomposition of the strain rates and the path independence of the elastic work of deformation, equation (52) may be recast as

\[
W[\varepsilon] = \int_0^{\Delta t} \sigma(\tau) : \varepsilon(\tau) \, d\tau = \int_0^{\Delta t} [\sigma(\tau) : \dot{\varepsilon}^e(\tau) + \sigma(\tau) : \dot{\varepsilon}^p(\tau)] \, d\tau
\]

where \( W^e \) is of no further interest.

Next, considering \( \varepsilon(0) = \varepsilon_n \) and \( \varepsilon(\Delta t) = \varepsilon_{n+1} \),

\[
\varepsilon(\Delta t) = \varepsilon_{n+1}
\]

In addition, the stress update

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

It follows immediately that

\[
\varepsilon(\Delta t) = \varepsilon_{n+1}
\]

and, hence, plastic strain.

Inserting this into (54) and

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

Since the yield criterion must be consistent with the update, the yield function

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

must in turn

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

Integration boundary

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

where use has been made of

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

which necessarily

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

during plasticity.

A path which satisfies

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

An elastic path which satisfies

\[
\sigma_{n+1} = \frac{\partial \hat{W}}{\partial \varepsilon_{n+1}}(\varepsilon_{n+1}; \Lambda_n)
\]

is allowed to be non-Drucker-stable. This relation is the incremental version of the plastic strain evolution law.
where $W^*$ is the elastic strain energy potential.

Next, consider arbitrary variations $\varepsilon^e(\tau) + \delta \varepsilon^e(\tau)$ in the strain path. Since the end values of the path $\varepsilon(0) = \varepsilon_0$ and $\varepsilon(\Delta t) = \varepsilon_{n+1}$ are given, the corresponding variations must vanish, i.e.,

$$\delta\varepsilon(0) = \delta\varepsilon_n = 0, \quad \delta\varepsilon(\Delta t) = \delta\varepsilon_{n+1} = 0$$  \hspace{1cm} (56)

In addition, the initial values of stress, elastic and plastic strain are given, and hence

$$\delta\sigma_n = 0, \quad \delta\varepsilon^e_n = 0, \quad \delta\varepsilon^p_n = 0$$  \hspace{1cm} (57)

The stationarity of (55) demands

$$0 = \delta W = \sigma_{n+1} : \delta \varepsilon^e_{n+1} + \int_0^{\Delta t} \left[ \delta\sigma(\tau) : \dot{\varepsilon}^p(\tau) + \sigma(\tau) : \delta \varepsilon^p(\tau) \right] d\tau$$  \hspace{1cm} (58)

It follows immediately from this expression that elastic paths, whenever admissible, are extremal. Assume on the contrary that

$$\phi(D^e:(\varepsilon(\tau) - \varepsilon_n^e)) > 0, \quad 0 < \tau \leq \Delta t$$  \hspace{1cm} (59)

and, hence, plastic flow must occur throughout the interval.

Inserting the flow rule into (58) we obtain

$$0 = \sigma_{n+1} : \delta \varepsilon^e_{n+1} + \int_0^{\Delta t} \left[ \frac{\partial \phi}{\partial \sigma}(\sigma(\tau)) : \dot{\sigma}(\tau) : \dot{\sigma}(\tau) + \sigma(\tau) : \delta \varepsilon^p(\tau) \right] d\tau$$  \hspace{1cm} (60)

Since the yield criterion $\phi(\sigma(\tau)) = 0$ must be identically satisfied during plastic loading, the plastic consistency condition

$$\frac{\partial \phi}{\partial \sigma}(\sigma(\tau)) : \dot{\sigma}(\tau) = 0$$  \hspace{1cm} (61)

must in turn hold and (60) simplifies to

$$0 = \sigma_{n+1} : \delta \varepsilon^e_{n+1} + \int_0^{\Delta t} \sigma(\tau) : \delta \varepsilon^p(\tau) d\tau$$  \hspace{1cm} (62)

Integration by parts yields

$$0 = \sigma_{n+1} : \delta \varepsilon^e_{n+1} - \int_0^{\Delta t} \dot{\phi}(\tau) : \delta \varepsilon^p(\tau) d\tau$$  \hspace{1cm} (63)

where use has been made of the identity $\delta \varepsilon_{n+1} = \delta \varepsilon^e_{n+1} + \delta \varepsilon^p_{n+1}$. The essential conditions (56) further reduce (63) to

$$0 = \int_0^{\Delta t} \dot{\phi}(\tau) : \delta \varepsilon^p(\tau) d\tau$$  \hspace{1cm} (64)

which necessitates

$$\dot{\sigma} = 0$$  \hspace{1cm} (65)

during plastic loading.

A path which complies with all the above stationarity conditions can be constructed as follows. An elastic path is followed from $\sigma_n$ to the final state of stress $\sigma_{n+1}$. Then, if necessary, the material is allowed to yield at constant stress $\sigma_{n+1}$. From the general properties of extremal paths in Drucker-stable materials, it follows that this is indeed the unique path which minimizes the incremental work of deformation.
For the extremal path just described, plastic flow takes place at the final stress and, hence, the flow rule reduces to

\[
\dot{\epsilon}^p = \dot{\gamma} \frac{\partial \phi}{\partial \sigma} (\sigma_{n+1})
\]

(66)

Since the direction of plastic flow is fixed, this equation can be integrated explicitly to yield

\[
\epsilon_{n+1}^p = \epsilon_n^p + \Delta \gamma \frac{\partial \phi}{\partial \sigma} (\sigma_{n+1})
\]

(67)

Hooke's law thus becomes

\[
\sigma_{n+1} = \sigma_n + D^p \left[ \epsilon_{n+1} - \Delta \gamma \frac{\partial \phi}{\partial \sigma} (\sigma_{n+1}) \right]
\]

(68)

To these equations one needs to append the requirement that \( \sigma_{n+1} \) satisfy the yield criterion

\[
\phi(\sigma_{n+1}) = 0
\]

(69)

Equations (68) and (69) constitute a statement of the fully implicit or closest point algorithm. Thus one concludes that for ideally plastic, rate-independent materials, holonomic methods based on incrementally external paths and the fully implicit algorithm coincide. In particular, it is seen that the closest point algorithm minimizes the work of deformation among all possible strain paths with prescribed terminal values. Furthermore, the fully implicit algorithm integrates the constitutive equations exactly along the extremal path. It also shares with holonomic methods the potential structure of the incremental stress–strain relations resulting in the symmetry of the consistent tangents, as was determined by direct computation in Section 3.1.

4.2. Hardening plasticity

Next we extend the analysis of the preceding section to hardening plasticity. Except for notational differences, our present discussion parallels that of Carter and Martin on work bounding functions.\(^1\) We assume that the material obeys the generalized normality conditions derived in Section 3.2. For the rate-independent solid, these amount to the expressibility of the constitutive equations in the form

\[
\dot{\epsilon}_p = \dot{\gamma} \frac{\partial \phi}{\partial \sigma}
\]

\[
\dot{q} = \dot{\gamma} \frac{\partial \phi}{\partial Q}
\]

(70)

where the function \( \mathbf{Q}(\mathbf{q}) \) derives from a potential as in (34).

As in the case of ideal plasticity, we seek to minimize the work of deformation (52) done by external agencies as the element of material is deformed along a strain path \( \epsilon(\tau) \) with prescribed terminal values \( \epsilon_n \) and \( \epsilon_{n+1} \). The deformation power at some time \( \tau \) takes the familiar form

\[
\dot{W} = \sigma : \dot{\epsilon}
\]

(71)

The rate of variation of the internal energy potential (35), on the other hand, is

\[
\dot{U} = \sigma : \dot{\epsilon} - \sigma : \dot{\epsilon} - \mathbf{Q} : \dot{\mathbf{q}}
\]

(72)
Combining these two expressions, the work functional (52) may be recast as

$$ W[e] = U_{n+1} - U_n + \int_0^\Delta \left[ \sigma(t) \cdot \dot{\varepsilon}^p(t) + Q(t) \cdot \dot{q}(t) \right] dt $$

(73)

Using the essential conditions (56), the condition for stationarity of $W$ may be expressed as

$$ 0 = \delta W = -\mathbf{\sigma}_{n+1} : \delta \varepsilon^p_{n+1} - \mathbf{Q}_{n+1} : \delta \mathbf{q}_{n+1} $$

$$ + \int_0^\Delta \left[ \mathbf{\delta \sigma}(t) \cdot \dot{\varepsilon}^p(t) + \mathbf{\delta Q}(t) \cdot \dot{q}(t) + \mathbf{\sigma}(t) \cdot \delta \varepsilon^p(t) + \mathbf{Q}(t) \cdot \delta \mathbf{q}(t) \right] dt $$

(74)

As in the ideally plastic case, it follows immediately from this condition that elastic paths are extremal. If plastic flow is assumed to occur throughout the strain increment, use may be made of the flow and hardening rules to rephrase (74) as

$$ 0 = -\mathbf{\sigma}_{n+1} : \delta \varepsilon^p_{n+1} - \mathbf{Q}_{n+1} : \delta \mathbf{q}_{n+1} $$

$$ + \int_0^\Delta \left[ \left( \frac{\partial \dot{\varepsilon}^p}{\partial \mathbf{\sigma}}(t) \right) \cdot \delta \mathbf{\sigma}(t) + \frac{\partial \dot{q}}{\partial \mathbf{Q}}(t) \cdot \delta \mathbf{Q}(t) \right] dt $$

(75)

But the satisfaction of the yield criterion requires the factor of $\dot{\varepsilon}^p$ to vanish identically, and (75) reduces to

$$ 0 = -\mathbf{\sigma}_{n+1} : \delta \varepsilon^p_{n+1} - \mathbf{Q}_{n+1} : \delta \mathbf{q}_{n+1} + \int_0^\Delta \left[ \left( \frac{\partial \dot{\varepsilon}^p}{\partial \mathbf{\sigma}}(t) \right) \cdot \delta \mathbf{\sigma}(t) + \frac{\partial \dot{q}}{\partial \mathbf{Q}}(t) \right] \delta \mathbf{Q}(t) dt $$

(76)

Integrating this expression by parts and using the essential conditions (56) one finds

$$ 0 = \int_0^\Delta \left[ \delta \dot{\varepsilon}^p(t) + \dot{Q}(t) \cdot \delta \mathbf{q}(t) \right] dt $$

(77)

Thus, the work functional is minimized by taking

$$ \dot{\varepsilon} = 0, \quad \dot{Q} = 0 $$

(78)

i.e. by holding both $\dot{\varepsilon}$ and $Q$ constant.

As noted by Carter and Martin, one cannot generally expect strain paths to exist resulting in the concurrent satisfaction of conditions (78). In fact, the variations $\delta \varepsilon^p$ and $\delta \mathbf{q}$ are not independent and, thus, the unconstrained minimality condition (78) furnishes a generally unrealizable lower bound to the work of deformation. In spite of its physical unattainability, the incremental potentials derived from the unconstrained extremum conditions (78) do possess all the properties which render them valuable in the present context. In particular, the bounding function $\tilde{W}$$\varepsilon_{n+1}; A_n)$ obtained by unconstrained minimization of the work functional is a potential for the terminal stresses in the sense of (54).

Conditions (78) imply that the unique unconstrained minimum is obtained by letting the inelastic processes take place at constant $\dot{\mathbf{\sigma}}$ and $Q$. Thus, the work potential $\tilde{W}$ follows by formulating the flow and hardening rules at the terminal values of $\dot{\mathbf{\sigma}}$ and $Q$, i.e.

$$ \dot{\varepsilon}^p = \dot{\varepsilon}^p (\sigma_{n+1}, Q_{n+1}) $$

$$ \dot{q} = \dot{q} (\sigma_{n+1}, Q_{n+1}) $$

(79)
where \( \varepsilon_{n+1} \) and \( Q_{n+1} \) must comply with the yield criterion

\[
\phi(\varepsilon_{n+1}, Q_{n+1}) = 0
\]

Equations (79) can be trivially integrated to yield

\[
\varepsilon_{n+1} = \varepsilon_n + \Delta \gamma \frac{\partial \phi}{\partial \varepsilon}(\varepsilon_{n+1}, Q_{n+1})
\]

\[
Q_{n+1} = Q_n + \Delta \gamma \frac{\partial \phi}{\partial Q}(\varepsilon_{n+1}, Q_{n+1})
\]

Equations (80) and (81), in conjunction with Hooke's law, constitute a statement of the fully implicit algorithm. Thus, as in the case of ideal plasticity, the fact that the generalized normality conditions (33) and (34) result in symmetric consistent tangents may again be traced to the extremal properties of the resulting fully implicit algorithm. In particular, the potential from which the stress update relations derive is the bounding work function of Carter and Martin. In contrast to the ideally plastic case, the incremental potential now generally provides a physically unattainable lower bound to the work of deformation. Furthermore, except for the exceptional materials for which the extremal paths are physically realizable, the closest point algorithm does not represent an exact solution of the constitutive equations for any strain path.

REFERENCES