INTERNATIONAL JOURNAL FOR NUMERICAL AND ANALYTICAL METHODS IN GEOMECHANICS
Int. J. Numer. Anal. Meth. Geomech., **23**, 467–492 (1999)

BIOT CONSOLIDATION ANALYSIS WITH AUTOMATIC TIME STEPPING AND ERROR CONTROL PART 1: THEORY AND IMPLEMENTATION

SCOTT W. SLOAN[1]*† AND ANDREW J. ABBO[2]‡

[1]*Department of Civil, Surveying and Environmental Engineering, University of Newcastle, NSW 2308, Australia*
[2]*Formation Design Systems, Fremantle, WA 6160, Australia*

SUMMARY

A new finite element algorithm for solving elastic and elastoplastic coupled consolidation problems is described. The procedure treats the governing consolidation relations as a system of first-order differential equations and is based on the backward Euler and Thomas and Gladwell schemes with automatic subincrementation of a prescribed series of time increments. The prescribed time increments, which are called coarse time steps, serve to start the procedure and are chosen by the user. The automatic consolidation algorithm attempts to select the time subincrements such that, for a given mesh, the time-stepping (or temporal discretisation) error in the displacements lies close to a specified tolerance.

Unlike existing solution techniques, the new algorithm computes not only the displacements and pore pressures, but also their derivatives with respect to time. These extra variables permit a family of unconditionally stable integration algorithms to be constructed which automatically provide an estimate of the local truncation error for each time step. This error estimate is inexpensive to compute and may be used to develop a simple and efficient automatic time stepping mechanism. For the elastic case, the displacements and pore pressures at the end of each subincrement may be solved directly without the need for iteration. For elastoplastic behaviour, however, the governing relationships are non-linear and a system of non-linear equations must be solved to compute the updates. Copyright © 1999 John Wiley & Sons, Ltd.

Key words: consolidation; automatic; time stepping; algorithm; finite element

1. INTRODUCTION

The finite element method can be used to model coupled consolidation using a mixed formulation which incorporates displacement and pore pressure variables. Even for elastic material behaviour, the resulting governing equations may be non-linear due to the dependence of the consolidation coefficient on the excess pore water pressures. The solution of these equations requires the discretization of the time domain into a number of time increments which can be difficult to

*Correspondence to: S. W. Sloan, Department of Civil, Surveying and Environmental Engineering, University of Newcastle NSW 2308, Australia, Email: scott.sloan@newcastle.edu.au
† Professor
‡ Software Engineer

CCC 0363–9061/99/060467–26$17.50

Received 24 September 1997
Revised 23 January 1998

attempt on a trial and error basis. The key problem is that an acceptable increment size may vary by several orders of magnitude throughout the analysis. During the early stages of consolidation, where excess pore water pressure gradients are usually high, relatively small time increments are necessary to obtain an accurate solution. As the analysis proceeds, however, much larger time increments can be used with good accuracy. Indeed, large time increments are often mandatory in order to obtain an efficient solution over a long period of consolidation.

The first authors to develop a finite element formulation of Biot[1] consolidation theory were Sandhu and Wilson.[2] Since then, numerous researchers, including Christian and Boehmer,[3] Hwang *et al.*,[4] Yokoo *et al.*,[5] Krause[6] and Borja[7] have formulated the governing finite element equations for elastic materials. For non-linear materials, various incremental solution strategies have been given by Lewis *et al.*,[8] Small *et al.*,[9] Prevost[10] and Borja[11]. Other solution methods have been presented by Carter *et al.*,[12] who incorporated the effects of finite deformations, and Ghaboussi and Wilson,[13] who accounted for pore fluid compressibility. In all of these linear and non-linear formulations, the governing finite element relations can be expressed as a system of coupled differential equations.

Techniques for the finite element solution of consolidation problems are often based on the well-known θ-method, the stability and accuracy of which has been investigated by Booker and Small[14] and Vermeer and Verruijt.[15] For elastic soils, the resulting time-stepping schemes are essentially the same as those used in the solution of first-order systems of differential equations. Since these types of equations arise in many areas of the physical sciences, they have been studied extensively and a vast amount of literature exists on their solution. An excellent summary of the stability and accuracy of various algorithms can be found in Reference 16. In order to solve elastic-coupled consolidation problems efficiently with the θ-method, it is generally necessary to use an implicit time integration scheme with $\theta \geqslant 0{\cdot}5$. With this choice of integration parameter, Booker and Small[14] proved that the solution process is unconditionally stable so that large time increments may be used with safety. Explicit integration methods, which employ $\theta = 0$, are only conditionally stable and may require the use of very small time steps.

In the analysis of elastoplastic soils, the application of implicit time integration schemes requires the solution of a system of non-linear equations for each time step. Small *et al.*[9] solved these non-linear equations with an initial stiffness iteration scheme which used averaged values of all time-varying quantities. This approach has been adopted by Siriwardane and Desai,[17] who also present an alternative tangent stiffness update with no iteration.

Somewhat surprisingly, very little work has been done on the development of automatic time stepping algorithms for finite element analysis of consolidation. A number of general integration methods, which were primarily developed for second-order systems of dynamics equations but are also applicable to first-order systems of consolidation equations, have been presented by Zienkiewicz *et al.*[18] and Thomas and Gladwell.[19] All of these schemes use an estimate of the local truncation error to control the time step size. In the methods of Zienkiewicz *et al.*,[18] the local error is found from a Taylor series expansion but, although the time steps may expand or contract as the analysis proceeds, no effort is made to control the error in the solution precisely. Thomas and Gladwell,[19] on the other hand, use the difference between solutions from pth- and $(p+1)$th-order schemes to estimate the local truncation error. This quantity can be used to adjust the size of each time step. A key advantage of these methods is that they operate in single-step mode and, hence, do not need to use values generated in previous time increments. Moreover, their local error estimators are embedded and can be computed cheaply with no extra matrix factorisations.

In a companion paper, Gladwell and Thomas[20] discuss issues associated with the implementation of their formulas and give an example code to illustrate the computational detail.

The scheme presented in this paper uses a first-order accurate one-stage method, together with a second-order accurate two-stage method, to estimate the local truncation error for the governing system of consolidation equations. As in the Thomas and Gladwell[19] strategy, the truncation error is measured with a single factorization and solution of the matrix equations for each time step.

2. FINITE ELEMENT FORMULATION OF BIOT CONSOLIDATION EQUATIONS

During consolidation, the distribution of the stresses in a saturated porous body is governed by the principles of effective stress, equilibrium, and continuity of flow. These relationships are all invoked to obtain the governing finite element equations.

The effective stress principle assumes that the total stresses, $\boldsymbol{\sigma}$, are equal to the sum of the effective stresses, $\boldsymbol{\sigma}'$, and the pore water pressure, p, according to

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}' + \mathbf{m}p$$

where $\boldsymbol{\sigma} = [\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{xz}, \tau_{yz}]^{\mathrm{T}}$, $\boldsymbol{\sigma}' = [\sigma'_x, \sigma'_y, \sigma'_z, \tau'_{xy}, \tau'_{xz}, \tau'_{yz}]^{\mathrm{T}}$, $\mathbf{m} = [1, 1, 1, 0, 0, 0]^{\mathrm{T}}$ and tensile stresses and pore pressures are taken as positive. Differentiating this equation with respect to time gives

$$\dot{\boldsymbol{\sigma}} = \dot{\boldsymbol{\sigma}}' + \mathbf{m}\dot{p} \tag{1}$$

In geotechnical analysis, it is conventional to decompose the total pore pressure into a steady-state component, p_{s}, and a time-varying excess component, p_{e}, according to

$$p = p_{\mathrm{s}} + p_{\mathrm{e}} \tag{2}$$

Noting that p_{s} is constant, differentiation of (2) with respect to time gives

$$\dot{p} = \dot{p}_{\mathrm{e}} \tag{3}$$

For applications which involve a horizontal phreatic surface, the steady-state pore pressure component corresponds to a hydrostatic stress distribution. To determine the governing relations for finite element analysis of consolidation, the effective stresses and pore pressures are treated separately and the primary nodal variables are the displacement rates (velocities) and the pore pressure rates. Because of the equivalence established in equation (3), there is little to choose between working with total pore pressure rates or excess pore pressure rates. In this paper, the formulation will be developed in terms of total pore pressure rates.

For an element with velocity variables at n nodes, the velocity field at any internal point is assumed to be of the form

$$\dot{\mathbf{d}} = \mathbf{N}_u \dot{\mathbf{u}}$$

where $\dot{\mathbf{d}} = [\dot{u}_x, \dot{u}_y, \dot{u}_z]^{\mathrm{T}}$ is a velocity vector with components $\dot{u}_x$, $\dot{u}_y$ and $\dot{u}_z$ in each co-ordinate direction, $\mathbf{N}_u$ is a matrix of shape functions given by

$$\mathbf{N}_u = \begin{bmatrix} N_{u1} & 0 & 0 & N_{u2} & 0 & 0 & \cdots & N_{un} & 0 & 0 \\ 0 & N_{u1} & 0 & 0 & N_{u2} & 0 & \cdots & 0 & N_{un} & 0 \\ 0 & 0 & N_{u1} & 0 & 0 & N_{u2} & \cdots & 0 & 0 & N_{un} \end{bmatrix}$$

and $\dot{\mathbf{u}} = [\dot{u}_{x1}, \dot{u}_{y1}, \dot{u}_{z1}, \dot{u}_{x2}, \dot{u}_{y2}, \dot{u}_{z2}, \ldots \dot{u}_{xn}, \dot{u}_{yn}, \dot{u}_{zn}]^{\mathrm{T}}$ is a vector of element nodal velocities. Similarly, the field of pore pressure rates for an element with pore pressure freedoms at m nodes is assumed to be of the form

$$\dot{p} = \mathbf{N}_p \dot{\mathbf{p}} \tag{4}$$

where

$$\mathbf{N}_p = [N_{p1}, N_{p2}, \ldots, N_{pm}] \tag{5}$$

and

$$\dot{\mathbf{p}} = [\dot{p}_1, \dot{p}_2, \ldots, \dot{p}_m]^{\mathrm{T}}$$

are, respectively, a matrix of shape functions and a vector of nodal pore pressure rates. Note that, for generality, different sets of shape functions may be used to describe the variation of the velocities and the pore pressure rates. This implies that the nodes in the finite element mesh may have varying degrees of freedom, with some being associated with velocities, some being associated with pore pressure rates, and some being associated with both. In order for the pore pressure rates to be consistent with the stress rates, it is usual to choose the polynomial describing the pore pressure rates to be one order lower than the polynomial describing the velocities. As discussed by a number of researchers, including Yokoo *et al.*[5] and Sandhu *et al.*,[21] this approach leads to less accurate estimates of the settlements but much smaller oscillations in the pore pressures.

The equations of equilibrium for a three-dimensional solid can be expressed in the compact form

$$\bar{\nabla}^{\mathrm{T}} \boldsymbol{\sigma} + \mathbf{b} = \mathbf{0} \tag{6}$$

where $\bar{\nabla}$ denotes the differential operator

$$\bar{\nabla}^{\mathrm{T}} = \begin{bmatrix} \dfrac{\partial}{\partial x} & 0 & 0 & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} & 0 \\ 0 & \dfrac{\partial}{\partial y} & 0 & \dfrac{\partial}{\partial x} & 0 & \dfrac{\partial}{\partial z} \\ 0 & 0 & \dfrac{\partial}{\partial z} & 0 & \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} \end{bmatrix}$$

and $\mathbf{b} = [b_x, b_y, b_z]^{\mathrm{T}}$ represents a body force vector whose components lie in the x-, y- and z-directions respectively. These conditions may be converted to a weak integral form using the method of weighted residuals. This ensures that equilibrium is satisfied only in some weighted average sense throughout the domain.

Applying the Green–Gauss theorem and Galerkin weighted residual method to the equilibrium equations for a single element leads to

$$\int_{V^e} \mathbf{B}_u^{\mathrm{T}} \dot{\boldsymbol{\sigma}}' \, \mathrm{d}V + \int_{V^e} \mathbf{B}_u^{\mathrm{T}} \mathbf{m} \, \dot{p} \, \mathrm{d}V - \int_{S^e} \mathbf{N}_u^{\mathrm{T}} \dot{\mathbf{t}} \, \mathrm{d}S - \int_{V^e} \mathbf{N}_u^{\mathrm{T}} \dot{\mathbf{b}} \, \mathrm{d}V = \mathbf{0} \tag{7}$$

where V^e is the element volume, S^e is the element surface area, $\mathbf{t} = [t_x, t_y, t_z]^{\mathrm{T}}$ is a vector of external surface tractions, and $\mathbf{B}_u$ is the strain rate–velocity matrix defined by

$$\mathbf{B}_u = \bar{\nabla} \mathbf{N}_u$$

In the theory of elastoplasticity, the effective stress rates are assumed to be related to the strain rates $\dot{\boldsymbol{\varepsilon}} = [\dot{\varepsilon}_x, \dot{\varepsilon}_y, \dot{\varepsilon}_z, \dot{\gamma}_{xy}, \dot{\gamma}_{xz}, \dot{\gamma}_{yz}]^{\mathrm{T}}$ via the constitutive law

$$\dot{\boldsymbol{\sigma}}' = (\mathbf{D}_{\mathrm{e}} - \mathbf{D}_{\mathrm{p}})\dot{\boldsymbol{\varepsilon}} = \mathbf{D}_{\mathrm{ep}}\dot{\boldsymbol{\varepsilon}}$$

where $\mathbf{D}_{\mathrm{e}}$ is the elastic constitutive matrix, $\mathbf{D}_{\mathrm{p}}$ is the plastic component of the standard elastoplastic constitutive matrix $\mathbf{D}_{\mathrm{ep}}$, and

$$\dot{\boldsymbol{\varepsilon}} = \bar{\nabla}\dot{\mathbf{d}} = \bar{\nabla}\mathbf{N}_u\dot{\mathbf{u}} = \mathbf{B}_u\dot{\mathbf{u}}$$

Combining the two previous equations permits the effective stress at any point inside an element to be expressed in terms of the element nodal velocities according to

$$\dot{\boldsymbol{\sigma}}' = \mathbf{D}_{\mathrm{ep}}\mathbf{B}_u\dot{\mathbf{u}} \tag{8}$$

Substituting (4) and (8) into (7) gives a weak statement of the conditions of equilibrium for a single element in rate form. These equations may be written as

$$\mathbf{k}_{\mathrm{ep}}\dot{\mathbf{u}} + \mathbf{l}\dot{\mathbf{p}} = \dot{\mathbf{f}}^{\mathrm{ext}} \tag{9}$$

where

$$\begin{aligned} \mathbf{k}_{\mathrm{ep}} &= \int_{V^e} \mathbf{B}_u^{\mathrm{T}}\mathbf{D}_{\mathrm{ep}}\mathbf{B}_u \,\mathrm{d}V \\ \mathbf{l} &= \int_{V^e} \mathbf{B}_u^{\mathrm{T}}\mathbf{m}\mathbf{N}_p \,\mathrm{d}V \end{aligned} \tag{10}$$

are the elemental elastoplastic stiffness and coupling matrices and

$$\dot{\mathbf{f}}^{\mathrm{ext}} = \int_{V^e} \mathbf{N}_u^{\mathrm{T}}\dot{\mathbf{b}} \,\mathrm{d}V + \int_{S^e} \mathbf{N}_u^{\mathrm{T}}\dot{\mathbf{t}} \,\mathrm{d}S$$

is the elemental vector of external force rates.

In order to complete the mathematical description of the consolidation process, we consider the continuity of flow for an element of soil. Assuming that the pore water and soil grains are incompressible, continuity of flow demands that the rate at which water is drained from the soil skeleton must be equal to the rate of volume decrease of the soil mass. This condition may be expressed mathematically as

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = -(\dot{\varepsilon}_x + \dot{\varepsilon}_y + \dot{\varepsilon}_z)$$

or

$$\operatorname{div}\mathbf{v} + \mathbf{m}^{\mathrm{T}}\dot{\boldsymbol{\varepsilon}} = 0 \tag{11}$$

where $\mathbf{v} = [v_x, v_y, v_z]^{\mathrm{T}}$ denotes a vector of superficial (or Darcy) fluid velocities. The continuity equation may be expressed in terms of the pore water pressure using Darcy's law which states that the fluid velocities are given by

$$v_x = \frac{k_x}{\gamma_{\mathrm{w}}}\frac{\partial p}{\partial x}; \qquad v_y = \frac{k_y}{\gamma_{\mathrm{w}}}\left(\frac{\partial p}{\partial y} - \gamma_{\mathrm{w}}\right); \qquad v_z = \frac{k_z}{\gamma_{\mathrm{w}}}\frac{\partial p}{\partial z}$$

where γ_w is the unit weight of water and k_x, k_y, k_z are soil permeabilities in the three coordinate directions. Note that, because compressive pore pressures are taken as negative, the sign convention in the above equations is different to that normally used in soil mechanics. Darcy's law may also be written in the more compact matrix form

$$\mathbf{v} = \frac{\mathbf{k}}{\gamma_w}(\nabla p - \mathbf{b}_w) \tag{12}$$

where $\mathbf{b}_w = [0, \gamma_w, 0]^T$ is a body force vector and $\mathbf{k}$ is a matrix of permeability coefficients of the form

$$\mathbf{k} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{bmatrix}$$

Inserting (12) into (11) gives the continuity equation in terms of the pore pressures according to

$$\operatorname{div}\left(\frac{\mathbf{k}}{\gamma_w}(\nabla p - \mathbf{b}_w)\right) + \mathbf{m}^T\dot{\boldsymbol{\varepsilon}} = 0$$

The weak form of this equation for a single element is obtained by again applying the Green–Gauss theorem and Galerkin method of weighted residuals to give

$$\int_{V^e} \mathbf{N}_p^T\mathbf{m}^T\mathbf{B}_u \, dV\dot{\mathbf{u}} - \int_{V^e} \mathbf{B}_p^T \frac{\mathbf{k}}{\gamma_w}\mathbf{B}_p \, dV\,\mathbf{p} + \int_{S^e} \mathbf{N}_p^T q \, dS + \int_{V^e} \mathbf{B}_p^T \frac{\mathbf{k}}{\gamma_w}\mathbf{b}_w \, dV = \mathbf{0} \tag{13}$$

where

$$q = (\nabla p - \mathbf{b}_w)^T \frac{\mathbf{k}}{\gamma_w}\mathbf{n} = \mathbf{v}^T\mathbf{n}$$

is a prescribed outward flow per unit area, $\mathbf{n} = [n_x, n_y, n_z]^T$ is a vector of direction cosines for the unit outward normal to S^e, and

$$\mathbf{B}_p = \nabla\mathbf{N}_p \tag{14}$$

Equation (13) is usually written in the more compact form

$$\mathbf{l}^T\dot{\mathbf{u}} + \mathbf{h}\mathbf{p} = \mathbf{q} \tag{15}$$

where the coupling matrix $\mathbf{l}$ is given by (10), the flow matrix $\mathbf{h}$ is defined as

$$\mathbf{h} = -\int_{V^e} \mathbf{B}_p^T \frac{\mathbf{k}}{\gamma_w}\mathbf{B}_p \, dV$$

and

$$\mathbf{q} = -\int_{S^e} \mathbf{N}_p^T q \, dS - \int_{V^e} \mathbf{B}_p^T \frac{\mathbf{k}}{\gamma_w}\mathbf{b}_w \, dV$$

is a fluid supply vector. Equations (9) and (15) define the governing relations for Biot consolidation at the element level. Assembling the element matrices in the usual way produces a global

system of equations of the form

$$\begin{bmatrix} \mathbf{K}_{ep} & \mathbf{L} \\ \mathbf{L}^{T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{U}} \\ \dot{\mathbf{P}} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{P} \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{F}}^{ext} \\ \mathbf{Q} \end{bmatrix} \tag{16}$$

where

$$\mathbf{K}_{ep} = \sum_{elem.} \mathbf{k}_{ep} = \sum_{elem.} \int_{V^e} \mathbf{B}_u^{T} \mathbf{D}_{ep} \mathbf{B}_u \, dV \tag{17}$$

$$\mathbf{L} = \sum_{elem.} \mathbf{l} = \sum_{elem.} \int_{V^e} \mathbf{B}_u^{T} \mathbf{m} \mathbf{N}_p \, dV$$

$$\mathbf{H} = \sum_{elem.} \mathbf{h} = - \sum_{elem.} \int_{V^e} \mathbf{B}_p^{T} \frac{\mathbf{k}}{\gamma_w} \mathbf{B}_p \, dV$$

are the global elastoplastic stiffness, coupling and flow matrices and

$$\dot{\mathbf{F}}^{ext} = \sum_{elem.} \dot{\mathbf{f}}^{ext} = \sum_{elem.} \int_{V^e} \mathbf{N}_u^{T} \dot{\mathbf{b}} \, dV + \sum_{elem.} \int_{S^e} \mathbf{N}_u^{T} \dot{\mathbf{t}} \, dS$$

$$\mathbf{Q} = \sum_{elem.} \mathbf{q} = - \sum_{elem.} \int_{S^e} \mathbf{N}_p^{T} q \, dV - \sum_{elem.} \int_{V^e} \mathbf{B}_p^{T} \frac{\mathbf{k}}{\gamma_w} \mathbf{b}_w \, dV$$

are the global external force rate and fluid supply vectors. For the case of an elastic soil, the elastoplastic stiffness matrix defined by (17) is replaced by the elastic stiffness matrix

$$\mathbf{K}_e = \sum_{elem.} \mathbf{k}_e = \sum_{elem.} \int_{V^e} \mathbf{B}_u^{T} \mathbf{D}_e \mathbf{B}_u \, dV$$

where $\mathbf{D}_e$ is the elastic stress–strain matrix.

3. SOLUTION OF ELASTIC CONSOLIDATION EQUATIONS

For the analysis of linear elastic solids with constant permeabilities, the relations (16) constitute a system of linear first-order differential equations of the form

$$\mathbf{C}_e \dot{\mathbf{X}} + \mathbf{K}\mathbf{X} = \mathbf{F}(t) \tag{18}$$

where

$$\mathbf{C}_e = \begin{bmatrix} \mathbf{K}_e & \mathbf{L} \\ \mathbf{L}^{T} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{bmatrix} \tag{19}$$

are matrices of constants, $\mathbf{F}(t)$ is a time-dependent forcing function defined by

$$\mathbf{F}(t) = \begin{bmatrix} \dot{\mathbf{F}}^{ext} \\ \mathbf{Q} \end{bmatrix} \tag{20}$$

and $\mathbf{X} = [\mathbf{U}, \mathbf{P}]^{T}$ with $\dot{\mathbf{X}} = [\dot{\mathbf{U}}, \dot{\mathbf{P}}]^{T}$. This type of system occurs in many areas of engineering science and has been widely studied. A very comprehensive summary of the stability and accuracy of various solution strategies for solving (18) may be found in Reference 16.

3.1. The θ method

The simplest strategy for solving (18) is commonly known as the θ-method. For the nth time step ranging from t_{n-1} to $t_n = t_{n-1} + h$, this algorithm may be expressed in the form

$$[\mathbf{C}_e + \theta h\mathbf{K}]\mathbf{X}_n = [\mathbf{C}_e - (1-\theta)h\mathbf{K}]\mathbf{X}_{n-1} + h[(1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n] \tag{21}$$

where θ is an integration parameter in the interval $0 \leqslant \theta \leqslant 1$, the subscripts n and $n-1$ denote, respectively, quantities evaluated at the start and end of the step, and all values except $\mathbf{X}_n$ are known. The process assumes that the initial condition, $\mathbf{X}_0$ at time t_0, is known. For the case of a linear elastic soil with constant permeabilities, the matrices $\mathbf{C}_e$ and $\mathbf{K}$ are independent of $\mathbf{X}$ and (21) defines a system of linear equations which can be solved for $\mathbf{X}_n$. Note that if h is kept constant, then the matrix $[\mathbf{C}_e + \theta h\mathbf{K}]$ needs to be factorised only once to obtain the solution for all time steps. Because θ may assume a range of values, equation (21) can be used to generate a number of single-stage single-step schemes that all march the solution $\mathbf{X}$ forward in time without the need for information from previous steps.

The θ-method is at least first order accurate and, provided $\theta \geqslant 0{\cdot}5$, is unconditionally stable.[14] Unconditional stability is an essential characteristic for an efficient consolidation scheme since it is often necessary to integrate over very long time periods using large time steps. For the special case of $\theta = 0{\cdot}5$, the θ-method is second order accurate and corresponds to the ubiquitous Crank–Nicolson scheme. Although appealing because of its high accuracy, the Crank–Nicolson method may generate spurious oscillations in the solution, especially if there are abrupt changes in the forcing function, and often requires special smoothing procedures such as those advocated by Wood and Lewis[22] and Wood.[23] Choosing a value of $\theta = 1$ gives the well-known backward Euler scheme which is first order accurate and unconditionally stable.[16] The fact that the backward Euler method also damps out unwanted oscillations has led to it being widely used in finite element consolidation studies, even though it is less accurate than the Crank–Nicolson scheme.

As an alternative to (21), the θ-method can be expressed in the more compact form

$$[\mathbf{C}_e + \theta h\mathbf{K}]\mathbf{V} = (1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n - \mathbf{K}\mathbf{X}_{n-1} \tag{22}$$

where

$$\mathbf{V} = (\mathbf{X}_n - \mathbf{X}_{n-1})/h$$

is an average estimate of $\dot{\mathbf{X}}$ over the time step h and $\mathbf{X}$ is updated according to

$$\mathbf{X}_n = \mathbf{X}_{n-1} + h\mathbf{V}$$

In the finite element literature, this form of the θ-method is known as the SS11 procedure, where the terminology SSpj stands for a Single-Step algorithm which uses an approximation of degree p to solve a differential equation of order j. These relations will be used to develop an automatic time stepping scheme in Section 4.1.

3.2. The Thomas and Gladwell method

A general class of multistage single-step methods for solving systems of first- and second-order differential equations, such as (18), has been proposed by Zienkiewicz *et al.*[18] Of particular

interest here is the two-stage single-step scheme defined by

$$[\theta_1 h\mathbf{C}_e + \tfrac{1}{2}\theta_2 h^2\mathbf{K}]\,\mathbf{A} = (1-\theta_1)\mathbf{F}_{n-1} + \theta_1\mathbf{F}_n - \mathbf{C}_e\dot{\mathbf{X}}_{n-1} - \mathbf{K}[\mathbf{X}_{n-1} + \theta_1 h\dot{\mathbf{X}}_{n-1}] \tag{23}$$

where

$$\mathbf{A} = (\dot{\mathbf{X}}_n - \dot{\mathbf{X}}_{n-1})/h$$

is an average estimate of $\ddot{\mathbf{X}}$ over the time step h and the updates are

$$\mathbf{X}_n = \mathbf{X}_{n-1} + h\dot{\mathbf{X}}_{n-1} + \tfrac{1}{2}h^2\mathbf{A} \tag{24}$$

$$\dot{\mathbf{X}}_n = \dot{\mathbf{X}}_{n-1} + h\mathbf{A} \tag{25}$$

This scheme, which is commonly known as the SS21 algorithm, is second order accurate and unconditionally stable provided $\theta_2 > \theta_1 \geqslant 0{\cdot}5$. For the special case of $\theta_1 - \theta_2 = 0{\cdot}5$ and $\theta_1 > 0{\cdot}5$, the SS21 scheme is third order accurate but only conditionally stable. During a given time step, $\mathbf{X}_n$ and $\dot{\mathbf{X}}_n$ are updated using (24) and (25) after first solving for $\mathbf{A}$ in the linear system (23). Because the SS21 procedure advances the solution for both $\mathbf{X}$ and $\dot{\mathbf{X}}$ and only uses values from the current time step, the algorithm is termed a two-stage single-step method. The chief advantage of this type of scheme is that the step size may be adjusted easily as the integration proceeds. The price of this flexibility is the need to compute and store $\dot{\mathbf{X}}$ for each time step.

More recently, Thomas and Gladwell[19] have proposed a generalized form of the SS21 algorithm. Their procedure uses three integration parameters instead of two and may be written as

$$[\varphi_2 h\mathbf{C}_e + \varphi_3 h^2\mathbf{K}]\,\mathbf{A} = \mathbf{F}(t_{n-1} + \varphi_1 h) - \mathbf{C}_e\dot{\mathbf{X}}_{n-1} - \mathbf{K}[\mathbf{X}_{n-1} + \varphi_1 h\dot{\mathbf{X}}_{n-1}] \tag{26}$$

where $\mathbf{A}$ has the same meaning as before and the updates are again given by (24) and (25). This scheme is second order accurate and unconditionally stable provided $2\varphi_3 > \varphi_1 \geqslant 0{\cdot}5$ and $\varphi_2 \geqslant 0{\cdot}5$. Using the fact that

$$\mathbf{F}(t_{n-1} + \varphi_1 h) = (1-\varphi_1)\mathbf{F}_{n-1} + \varphi_1\mathbf{F}_n + O(h^2) \tag{27}$$

a comparison of (23) with (26) reveals that the SS21 algorithm of Zienkiewicz *et al.*[18] is a special case of the Thomas and Gladwell[19] algorithm with $\theta_1 = \varphi_1 = \varphi_2$ and $\theta_2 = 2\varphi_3$. For cases where $\theta_1 = \varphi_1 = 1$, this equivalence holds without the need to approximate the forcing function by (27). Because of the additional freedom that is introduced by having three integration parameters, the Thomas and Gladwell[19] method is ideally suited to the design of automatic integration schemes with embedded error estimators. Indeed, it will be employed for this purpose in the next section.

4. AUTOMATIC TIME-STEPPING SCHEME FOR ELASTIC CONSOLIDATION

We now describe an adaptive integration scheme that automatically adjusts the time step according to a specified error criterion. For ease of use, the algorithm assumes that a number of (coarse) time increments are defined and automatically breaks these up into a number of smaller subincrements if necessary. The coarse time step is assumed to start at t_0 and end at $t_0 + \Delta t$, and is thus of size Δt. The nth time subincrement is assumed to be of size h, and ranges from t_{n-1} to $t_n = t_{n-1} + h$. This arrangement is shown schematically in Figure 1.

The adaptive procedure uses two integration methods, of different accuracy, to provide an estimate of the local truncation error in the displacements and pore pressures and is thus based on

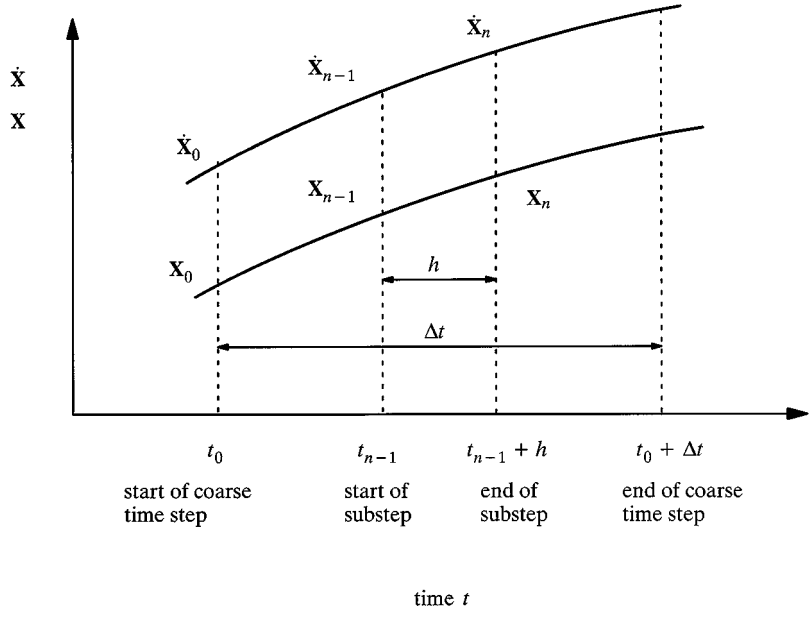

Figure 1. Coarse and subincremental time steps

the same idea that has been widely exploited in the design of solvers for ordinary differential equations (see, for example, Reference 24). In the new algorithm, the SS11 version of the θ-method and the Thomas and Gladwell[19] method are employed, respectively, to generate first and second-order accurate solutions for the error control.

4.1. *Theory for elastic algorithm*

To derive an efficient solution scheme with an in-built error estimator, the integration parameters for the SS11 and Thomas and Gladwell algorithms are selected so that only one matrix factorisation is needed for each time step. The constraints that this imposes on the integration parameters may be seen by rewriting equations (22) and (26) as

$$\mathbf{C}_e\mathbf{V} + \mathbf{K}[\mathbf{X}_{n-1} + \theta h\mathbf{V}] = (1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n \tag{28}$$

$$\mathbf{C}_e[\dot{\mathbf{X}}_{n-1} + \varphi_2 h\mathbf{A}] + \mathbf{K}[\mathbf{X}_{n-1} + \varphi_1 h\dot{\mathbf{X}}_{n-1} + \varphi_3 h^2\mathbf{A}] = \mathbf{F}(t_{n-1} + \varphi_1 h)$$

Assuming that both methods are used with the same starting value, $\mathbf{X}_{n-1}$, these equations indicate that the SS11 and Thomas and Gladwell schemes give rise to an identical system of equations if

$$\mathbf{V} = \dot{\mathbf{X}}_{n-1} + \varphi_2 h\mathbf{A} \tag{29}$$

$$\theta h\mathbf{V} = \varphi_1 h\dot{\mathbf{X}}_{n-1} + \varphi_3 h^2\mathbf{A} \tag{30}$$

and

$$(1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n = \mathbf{F}(t + \varphi_1 h) \tag{31}$$

Neglecting, for the moment, the forcing function terms, the required constraints on the integration parameters are obtained by substituting (29) into (30) to give

$$\theta = \varphi_1 = \varphi_3/\varphi_2$$

Because the error estimator assumes that the SS11 method is only first-order accurate, the Crank–Nicolson special case must be excluded so that

$$\theta \neq 0{\cdot}5$$

The previous two sets of constraints ensure that first- and second-order accurate solutions can be obtained, respectively, from the SS11 the Thomas and Gladwell methods with only a single matrix factorization for each time subincrement. Combining these with the unconditional stability requirements

$$\theta \geqslant 0{\cdot}5$$

$$2\varphi_3 > \varphi_1 \geqslant 0{\cdot}5$$

$$\varphi_2 \geqslant 0{\cdot}5$$

gives the final set of constraints as

$$2\varphi_3 > \theta = \varphi_1 = \varphi_3/\varphi_2 > 0{\cdot}5 \tag{32}$$

From (31), we see that the forcing functions for the two schemes are identical if $\theta = \varphi_1 = 1$. For other choices of θ and φ_1, it is necessary to invoke the second-order approximation for $\mathbf{F}(t_{n-1} + \varphi_1 h)$ given in equation (27). Using this approximation, equation (31) becomes

$$(1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n = (1-\varphi_1)\mathbf{F}_{n-1} + \varphi_1\mathbf{F}_n + O(h^2)$$

which is automatically satisfied, with error $O(h^2)$, by the constraint $\theta = \varphi_1$ embedded in (32).

During a typical time subincrement h, first- and second-order accurate estimates for $\mathbf{X}_n$ may be found by using (22). The first-order accurate solution, denoted by $\overset{*}{\mathbf{X}}_n$, is found by solving

$$\mathbf{V} = [\mathbf{C}_e + \theta h\mathbf{K}]^{-1}[(\mathbf{1}-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n - \mathbf{K}\mathbf{X}_{n-1}] \tag{33}$$

so that

$$\overset{*}{\mathbf{X}}_n = \mathbf{X}_{n-1} + h\mathbf{V} \tag{34}$$

Once $\mathbf{V}$ is known, the second-order accurate updates can then be found by using equations (24) and (25)

$$\mathbf{X}_n = \mathbf{X}_{n-1} + h\dot{\mathbf{X}}_{n-1} + \tfrac{1}{2}h^2\mathbf{A} \tag{35}$$

$$\dot{\mathbf{X}}_n = \dot{\mathbf{X}}_{n-1} + h\mathbf{A} \tag{36}$$

where $\mathbf{A}$ is found from (29) according to

$$\mathbf{A} = \frac{\mathbf{V} - \dot{\mathbf{X}}_{n-1}}{\varphi_2 h} \tag{37}$$

and $\dot{\mathbf{X}}_{n-1}$ is assumed known.

Since the local truncation errors in the updates (35) and (34) are, respectively, $O(h^3)$ and $O(h^2)$, the lower-order estimate may be subtracted from the higher-order estimate to give the local truncation error measure

$$\mathbf{E}_n = \mathbf{X}_n - \overset{*}{\mathbf{X}}_n = h\dot{\mathbf{X}}_{n-1} + \tfrac{1}{2}h^2\mathbf{A} - h\mathbf{V}$$

Substituting equations (36) and (37), this estimator may be expressed in the form

$$\mathbf{E}_n = h\left[(\varphi_2 - \tfrac{1}{2})\dot{\mathbf{X}}_{n-1} + (\tfrac{1}{2} - \varphi_2)\dot{\mathbf{X}}_n\right]$$

Note that the undesirable special case of $\varphi_2 = 0{\cdot}5$, which gives a zero estimate of the local error regardless of h, is automatically excluded by the constraints (32). For the purposes of error control, $\mathbf{E}_n$ may be replaced by the more useful dimensionless relative error measure

$$R_n = \max\left\{\frac{\|\mathbf{E}_n^u\|}{\|\mathbf{U}_n\|}, \frac{\|\mathbf{E}_n^p\|}{\|\mathbf{P}_n\|}\right\} \tag{38}$$

where

$$\begin{aligned}\mathbf{E}_n^u &= h\left[(\varphi_2 - \tfrac{1}{2})\dot{\mathbf{U}}_{n-1} + (\tfrac{1}{2} - \varphi_2)\dot{\mathbf{U}}_n\right]\\ \mathbf{E}_n^p &= h\left[(\varphi_2 - \tfrac{1}{2})\dot{\mathbf{P}}_{n-1} + (\tfrac{1}{2} - \varphi_2)\dot{\mathbf{P}}_n\right]\end{aligned} \tag{39}$$

and $(\mathbf{U}_n, \mathbf{P}_n)$ are the displacement and pore pressure components of $\mathbf{X}_n$ and $(\dot{\mathbf{U}}_n, \dot{\mathbf{P}}_n)$ are the velocity and pore pressure rate components of $\dot{\mathbf{X}}_n$. For cases where a weightless soil and fluid are assumed, the total pore pressures correspond to the excess pore pressures and $\|\mathbf{P}_n\|$ will approach zero in later stages of consolidation. To avoid ill-conditioning of the relative error estimator in this situation, it is possible to measure the error in the displacements only and replace (38) by

$$R_n^u = \frac{\|\mathbf{E}_n^u\|}{\|\mathbf{U}_n\|} \tag{40}$$

This simplification is sufficiently accurate for practical computations and removes the need to design a test which uses both absolute and relative error tolerances.

Once R_n^u has been computed, the current time subincrement is accepted if it is less than some specified tolerance on the local truncation error, $DTOL$, and rejected otherwise. In either case, the size of the next time step h_{n+1} is found from

$$h_{n+1} = qh_n \tag{41}$$

where q is a factor which is chosen to limit the predicted truncation error. Since the truncation error for the next time subincrement, R_{n+1}^u, is approximately related to R_n^u by

$$R_{n+1}^u \approx q^2 R_n^u$$

the required factor q is found by insisting that $R_{n+1}^u \leqslant DTOL$ to give

$$q \leqslant \sqrt{DTOL/R_n^u}$$

The above procedure for determining q is based on the dominant error term for the SS11 method and assumes that this first-order scheme is used for the update. Because this approximation may become inaccurate for strongly non-linear behaviour, it is wise to choose q conservatively so as to

minimise the number of rejected time subincrements. Numerical experiments suggest that a suitable strategy for computing q is to set

$$q = 0{\cdot}9\sqrt{DTOL/R_n^u} \tag{42}$$

with the additional constraint that

$$0{\cdot}1 \leqslant q \leqslant 2 \tag{43}$$

Note that the coefficient of 0·9 acts merely as a safety factor, since it usually prevents the step control mechanism from choosing a time subincrement whose local truncation error just fails to meet the local error tolerance. Constraining the growth in consecutive time subincrements to a factor of two also has this effect. The value of the safety factor in (42), and the limits in (43), were determined by numerical experiments on a wide variety of examples and ensure that most of the substeps are successful without making the step selection mechanism unduly conservative. As well as imposing the above conditions, it is prudent to prohibit the step size from growing immediately after a failed load subincrement. This ensures that there are at least two load subincrements of similar size following a failure, and is useful for cases where the load path has sharp changes in curvature.

At the start of a typical substep, $\mathbf{V}$ is found from (33) and the updates for $\mathbf{X}$ and $\dot{\mathbf{X}}$ are computed using (35) and (36). Note that for the first coarse time step h is typically initialised to Δt, but in subsequent coarse time steps h may be initialised to equal the last untruncated subincrement. Once $\mathbf{X}$ and $\dot{\mathbf{X}}$ have been updated, the relative error R_n^u is then determined using equations (39) and (40). If this error is less than or equal to the specified tolerance $DTOL$, then the current time subincrement is accepted and the step size for the next time subincrement is found using equations (41)–(43). If R_n^u exceeds $DTOL$, then the solution is rejected and the same equations are used to predict a smaller step size that will hopefully satisfy the constraint on the local error. In this case, the stage is repeated and, if necessary, the step size is reduced further until a successful time substep size is obtained. Note that we choose to propagate the second-order accurate solution throughout the analysis, even though the error estimate assumes that we would march forward the first order solution. This practice is used in most modern algorithms for solving systems of ordinary differential equations and is known as local extrapolation. Because it compensates for the fact that the error measure is only local, local extrapolation leads to better control of the global temporal discretization error which accumulates over many time steps. Although we could design an algorithm which marches forward the first-order solution with little difficulty, this option will not be pursued here.

4.2. Starting conditions

Assuming initial values for h and $\mathbf{X}_0$, with the latter typically zero, the integration scheme is started by solving (33) for $\mathbf{V}$. For the first coarse time increment, h is typically set to Δt, but in subsequent coarse time increments it may be initialised to the value of the last untruncated subincrement. In order to compute the second-order update for $\mathbf{X}$ using equation (35), a starting value for $\dot{\mathbf{X}}$ at $t = 0$ is needed. Assuming that the matrix $\mathbf{C}_e$ has an inverse, $\dot{\mathbf{X}}_0$ may be found by solving the governing differential equation (18) at $t = 0$ according to

$$\dot{\mathbf{X}}_0 = [\mathbf{C}_e]^{-1}[\mathbf{F}_0 - \mathbf{K}\mathbf{X}_0]$$

This type of procedure is valid for elements with a pore pressure expansion which is one order lower than the displacement expansion. For elements where the expansions are the same, $\mathbf{C}_e$ does not have an inverse and the above equation cannot be used. The simplest alternative in this case is to take a very small time step δh and apply equations (33), (36) and (37) with $\dot{\mathbf{X}}_0 = \mathbf{0}$ to give

$$\mathbf{V} = [\mathbf{C}_e + \theta\delta h\mathbf{K}]^{-1}[(1-\theta)\mathbf{F}_{0^+} + \theta\mathbf{F}_{0^+} - \mathbf{K}\mathbf{X}_0]$$

$$\dot{\mathbf{X}}_{0^+} = \mathbf{V}/\varphi_2$$

where the subscript 0^+ indicates quantities evaluated at time $t = \delta h$. For a sufficiently small value of δh, $\dot{\mathbf{X}}_{0^+}$ may be used as the initial value for $\dot{\mathbf{X}}$ at time $t = 0$.

4.3. Evaluation of the forcing function rate

The discussion has, so far, assumed that it is convenient to evaluate the external force rate, $\dot{\mathbf{F}}^{\text{ext}} = \mathrm{d}\mathbf{F}^{\text{ext}}/\mathrm{d}t$, analytically in the overall forcing function defined by equation (20). For cases where this is not so, this derivative can be approximated using discrete values of the external force vector. Examples of four useful approximations are

$$\dot{\mathbf{F}}^{\text{ext}}_n = (\mathbf{F}^{\text{ext}}_n - \mathbf{F}^{\text{ext}}_{n-1})/h + O(h) \tag{44}$$

$$\dot{\mathbf{F}}^{\text{ext}}_{n-1} = (\mathbf{F}^{\text{ext}}_n - \mathbf{F}^{\text{ext}}_{n-1})/h + O(h) \tag{45}$$

$$\dot{\mathbf{F}}^{\text{ext}}_n = (3\mathbf{F}^{\text{ext}}_n - 4\mathbf{F}^{\text{ext}}_{n-1/2} + \mathbf{F}^{\text{ext}}_{n-1})/h + O(h^2) \tag{46}$$

$$\dot{\mathbf{F}}^{\text{ext}}_{n-1} = (-3\mathbf{F}^{\text{ext}}_{n-1} + 4\mathbf{F}^{\text{ext}}_{n-1/2} - \mathbf{F}^{\text{ext}}_n)/h + O(h^2) \tag{47}$$

where h is the current time step and the subscripts $n-1$, $n-1/2$, and n denote values computed at the times t_{n-1}, $t_{n-1/2} = t_{n-1} + h/2$ and $t_n = t_{n-1} + h$. Since the adaptive integration scheme described here is second-order accurate, equations (46) and (47) should be employed when the variation of the external forcing function is non-linear with time. For problems where the external loading is piecewise linear with time, which covers most practical situations, the approximations (44) and (45) are exact and therefore appropriate. These approximations are used throughout this paper.

4.4. Scaling of linear equations

The automatic scheme described in Section 4.1 requires the solution of the linear system of equations (33). These may be written in the simple form

$$\begin{bmatrix} \mathbf{K}_e & \mathbf{L} \\ \mathbf{L}^{\mathrm{T}} & \theta h\mathbf{H} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{U}} \\ \dot{\mathbf{P}} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_u \\ \mathbf{R}_p \end{bmatrix} \tag{48}$$

where $\dot{\mathbf{U}}$ and $\dot{\mathbf{P}}$ are, respectively, average velocities and pore pressure rates over a time step h, and $\mathbf{R}_u$ and $\mathbf{R}_p$ are arbitrary vectors. For small time steps h, this system may become ill-conditioned as the diagonal terms in $\mathbf{K}_e$ can be many orders of magnitude greater than the terms in $\theta h\mathbf{H}$. The effects of ill-conditioning in a Biot formulation were first noted by Ghaboussin and Wilson,[13] who developed a criterion for selecting a minimum value of h. This criterion is valid for

consolidation of an elastic soil which is isotropic and homogeneous. Rather than limiting the size of the time step, which can introduce another source of error due to the time dependence of the governing equations, it is possible to scale the various terms in (48) so that ill-conditioning is avoided. This approach, suggested by Reed[25] and used in Lewis and Schrefler,[26] has proven successful and is adopted in this paper. The scaling preserves any symmetry of the original linear equations and takes the form

$$\begin{bmatrix} \mathbf{K}_e & s\mathbf{L} \\ s\mathbf{L}^T & s^2\theta h\mathbf{H} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{U}} \\ \frac{1}{s}\dot{\mathbf{P}} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_u \\ s\mathbf{R}_p \end{bmatrix}$$

where s is a scalar parameter which is chosen so as to roughly equate the size of the diagonal terms in $\mathbf{K}_e$ and $s^2\theta h\mathbf{H}$. Ignoring the effects of element geometry, the diagonal terms of $\mathbf{K}_e$ are proportional to the terms of the elastic stress strain matrix $\mathbf{D}_e$, and hence are of the same order as Young's modulus E. Similarly, the terms of the matrix $\mathbf{H}$ are approximately proportional to k/γ_w, where k is a representative value of the permeability. Equating these contributions and noting that θ is of order 1, a suitable value for the scaling parameter can be estimated as

$$s = \sqrt{\frac{E\gamma_w}{hk}}$$

For problems involving an homogeneous isotropic soil, the choice of E and k in the above equation is straightforward. In other situations, however, it is necessary to choose representative values of these material parameters. One such scaling scaling strategy can be found in the code of Lewis and Schrefler.[26]

4.5. Implementation of elastic algorithm

When implementing the automatic integration scheme, it is necessary to select specific values of the integration parameters θ, φ_1, φ_2 and φ_3 which satisfy the constraints (32). A series of numerical experiments covering a wide range of problems suggested that suitable choices for these parameters are

$$\theta = \varphi_1 = \varphi_2 = \varphi_3 = 1 \tag{49}$$

The advantages of these values are as follows:

(i) Setting $\theta = 1$ implies that the first order method corresponds to the backward Euler scheme. This procedure is known to damp out unwanted oscillations quickly and thus provides a reliable first-order solution for estimating the local truncation error.
(ii) Selecting $\varphi_1 = \theta = 1$ means that the forcing functions for the first-order and second-order schemes are identical without any need to make the approximation shown in equation (27). Moreover, there is no need to evaluate the forcing function outside the current time step.
(iii) Setting $\varphi_2 = 1$ means that the vector $\mathbf{V}$, computed from equation (33), corresponds automatically to $\dot{\mathbf{X}}_n$, the value of $\dot{\mathbf{X}}$ at the end of the current step for the second order scheme. This feature, which can be seen by comparing equations (29) and (36), results in a simple, compact algorithm with low storage requirements.

Other settings for these parameters are of course possible, and may lead to a scheme with improved performance for certain cases. The values in (49), however, give excellent results for a broad range of practical problems.

As mentioned previously, the implementation of the adaptive integration algorithm assumes that a series of coarse time increments have been specified. These coarse increments are, if necessary, subincremented automatically to satisfy a tolerance on the local truncation error.

Algorithm for Elastic Consolidation. The automatic time stepping algorithm for elastic consolidation may be summarised as follows.

Step 1. Initialisation

1.1. Enter with the time at the start of the coarse increment t_0, the current displacements and pore pressures $\mathbf{X}_{t_0}$, their corresponding derivatives $\dot{\mathbf{X}}_{t_0}$, the coarse time increment Δt, the last untruncated time substep h_{last}, the current effective stresses at each integration point $\boldsymbol{\sigma}'_{t_0}$, and the specified displacement error tolerance $DTOL$. For the first coarse time step, set $h_{\text{last}} = \Delta t$.

1.2. Set $t = t_0$ and $h = \min\{h_{\text{last}}, \Delta t\}$

Step 2. Main subincrement loop over time step Δt

2.1. Do steps 2.2–2.8

2.2. Compute $\dot{\mathbf{X}}_{t+h}$ using

$$\dot{\mathbf{X}}_{t+h} = [\mathbf{C}_e + h\mathbf{K}]^{-1}[\mathbf{F}_{t+h} - \mathbf{K}\mathbf{X}_t]$$

where

$$\mathbf{F}_{t+h} = \begin{bmatrix} \dot{\mathbf{F}}^{\text{ext}}_{t+h} \\ \mathbf{Q}_{t+h} \end{bmatrix} = \begin{bmatrix} (\mathbf{F}^{\text{ext}}_{t+h} - \mathbf{F}^{\text{ext}}_{t})/h \\ \mathbf{Q}_{t+h} \end{bmatrix}$$

2.3. Estimate the local truncation error in the displacements for the current subincrement using

$$E^u_{t+h} = \tfrac{1}{2} h \, \| \dot{\mathbf{U}}_t - \dot{\mathbf{U}}_{t+h} \|$$

where $\dot{\mathbf{U}}$ denotes the velocity component of $\dot{\mathbf{X}}$.

2.4. Update the displacements and pore water pressures and hold them in temporary storage according to

$$\bar{\mathbf{X}}_{t+h} = \mathbf{X}_t + \frac{h}{2}(\dot{\mathbf{X}}_t + \dot{\mathbf{X}}_{t+h})$$

2.5. Estimate the relative error for current subincrement using

$$R^u_{t+h} = \max\{EPS, E^u_{t+h} / \| \bar{\mathbf{U}}_{t+h} \|\}$$

where $\bar{\mathbf{U}}_{t+h}$ is the displacement component of $\bar{\mathbf{X}}_{t+h}$ and EPS is a machine constant.

2.6. If $R^u_{t+h} > DTOL$ then go to step 2·8. Else this step is successful so update displacements, pore pressures and integration point effective stresses according to

$$\mathbf{X}_{t+h} = \bar{\mathbf{X}}_{t+h}$$

$$\boldsymbol{\sigma}'_{t+h} = \mathbf{D}_e \mathbf{B} \mathbf{u}_{t+h}$$

If $t + h = t_0 + \Delta t$ then integration is complete, so set $t \leftarrow t + h$ and go to step 3.1

2.7. Estimate a new subincrement size factor by computing

$$q = \min\{0{\cdot}9\sqrt{DTOL/R^u_{t+h}}, 2\}$$

If the previous subincrement was unsuccessful, then prevent substeps from growing by setting

$$q = \min\{q, 1\}$$

Then update time and compute and store next substep size according to

$$t \leftarrow t + h$$

$$h \leftarrow qh$$

$$h_{\text{last}} = \min\{h, \Delta t\}$$

If necessary, truncate substep size so that integration does not proceed beyond $t_0 + \Delta t$ by setting

$$h \leftarrow \min\{h, t_0 + \Delta t - t\}$$

before returning to step 2.1.

2.8. This subincrement has failed, so estimate smaller time substep by computing

$$q = \max\{0{\cdot}9\sqrt{DTOL/R^u_{t+h}}, 0{\cdot}1\}$$

and then setting

$$h \leftarrow qh$$

before returning to step 2.1.

Step 3. Exit

3.1. Exit with displacements and pore pressures, $\mathbf{X}_{t_0+\Delta t}$, their corresponding rates, $\dot{\mathbf{X}}_{t_0+\Delta t}$, and integration point effective stresses, $\boldsymbol{\sigma}'_{t_0+\Delta t}$, at end of coarse time increment.

Upon exiting the above algorithm, the variable h_{last} stores the size of the last untruncated subincrement for the current coarse time step. This can be used as the starting value for h in the next coarse time step in order to minimize the number of rejected subincrements. Note that it is necessary to store the value of the last untruncated substep, rather than the last actual substep, since the latter is usually trimmed to avoid overshooting the end of the integration interval.

In steps 2.3 and 2.5, we use the max norm to measure the size of the error vectors. Other norms may, of course, be used but the max norm is generally the cheapest to compute. In step 2.5, *EPS* represents the smallest relative error that can be computed on the host machine, and is typically set to around 10^{-16} for double precision arithmetic on a 32-bit architecture. Typical values for the tolerance on the truncation error in the displacements, $DTOL$, are in the range 10^{-2} to 10^{-4}, with a value of 10^{-3} being adequate for most practical computations.

Note that, in step 2.2 of the above algorithm, it is generally necessary to form and factorise the matrix $[\mathbf{C}_e + h\mathbf{K}]$ afresh for each subincrement in order to find $\dot{\mathbf{X}}_{t+h}$. This is because h may vary

throughout the integration. Although the factorisation cannot be avoided, the cost of the formation step can be reduced significantly by computing

$$\mathbf{C}_e = \begin{bmatrix} \mathbf{K}_e & \mathbf{L} \\ \mathbf{L}^T & \mathbf{0} \end{bmatrix}$$

only once, at the start of the analysis, and storing it on disk. After loading $\mathbf{C}_e$ into memory at the start of each time subincrement, the contributions from

$$h\mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & h\mathbf{H} \end{bmatrix}$$

may then be added, element by element, to form $[\mathbf{C}_e + h\mathbf{K}]$. Depending on the type of element used, further economies may be realized by storing the element flow matrices, $\mathbf{h}$, on disk to minimize the cost of recomputing $\mathbf{H}$.

5. AUTOMATIC TIME STEPPING SCHEME FOR ELASTOPLASTIC CONSOLIDATION

For the analysis of elastoplastic soils, the governing relations given by (16) can be represented as a system of nonlinear equations of the form

$$\mathbf{R}(\mathbf{X}, \dot{\mathbf{X}}) = \mathbf{F}(t) - \mathbf{C}_{ep}(\mathbf{X})\dot{\mathbf{X}} - \mathbf{K}\mathbf{X} = \mathbf{0} \tag{50}$$

where

$$\mathbf{C}_{ep}(\mathbf{X}) = \begin{bmatrix} \mathbf{K}_{ep}(\mathbf{X}) & \mathbf{L} \\ \mathbf{L}^T & \mathbf{0} \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{bmatrix}, \quad \mathbf{F}(t) = \begin{bmatrix} \dot{\mathbf{F}}^{ext} \\ \mathbf{Q} \end{bmatrix} \tag{51}$$

and $\mathbf{X} = [\mathbf{U}, \mathbf{P}]^T$ with $\dot{\mathbf{X}} = [\dot{\mathbf{U}}, \dot{\mathbf{P}}]^T$. We again assume that the permeabilities are independent of time, so that the matrix $\mathbf{K}$ is constant. Due to the difficulty in measuring the permeability of soil accurately, this is the assumption that is usually made in practice. The algorithm described in the following sections can be extended to deal with cases where $\mathbf{K}$ is time dependent, but this generalisation will not be covered here. The major complication introduced by elastoplasticity is that the matrix $\mathbf{C}_{ep}$ is dependent on the current stress state (and hence the displacements).

5.1. Theory for elastoplastic algorithm

The theory for developing an automatic time-stepping scheme for elastoplastic soils is essentially the same as that for the elastic case described in Section 4.1. For a given time step, the local truncation error is again measured by taking the difference between a pair of first- and second-order solutions which, as before, are provided by the SS11 and Thomas and Gladwell methods. The key change from the elastic scheme is that it is now necessary to solve a system of nonlinear equations in order to update the displacements and pore pressures. It is once again assumed that the user defines a series of coarse time steps which, if required, are subdivided into substeps to keep the local truncation error below a specified tolerance.

Applying the SS11 and Thomas and Gladwell algorithms to (50) yields the systems of non-linear equations

$$\mathbf{R}_1(\mathbf{V}) = (1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n - \mathbf{C}_{\mathrm{ep}}(\mathbf{X}_{n-1} + \theta h\mathbf{V})\mathbf{V} - \mathbf{K}[\mathbf{X}_{n-1} + \theta h\mathbf{V}] = \mathbf{0} \tag{52}$$

$$\mathbf{R}_2(\mathbf{A}) = \mathbf{F}(t_{n-1} + \varphi_1 h) - \mathbf{C}_{\mathrm{ep}}(\mathbf{X}_{n-1} + \theta h\mathbf{V})[\dot{\mathbf{X}}_{n-1} + \varphi_2 h\mathbf{A}]$$
$$- \mathbf{K}[\mathbf{X}_{n-1} + \varphi_1 h\dot{\mathbf{X}}_{n-1} + \varphi_3 h^2\mathbf{A}] = \mathbf{0}$$

Following the procedure outlined in Section 4.1, these equations are identical if the constraints (32) are satisfied and the same starting values, $\mathbf{X}_{n-1}$, are used for both methods. Under these conditions, it is necessary to solve only (52) for $\mathbf{V}$, with the known initial value $\mathbf{X}_{n-1}$, in order to march the solution forward for each time substep. As in the elastic case, the updates for the displacements and pore pressures, $\mathbf{X}_n$, and their corresponding rates, $\dot{\mathbf{X}}_n$, are found from equations (35) and (36) and the local error estimator is given by (40). Thus, dropping the subscript on $\mathbf{R}$, the system of nonlinear equations to be solved for each time substep may be written as

$$\mathbf{R}(\mathbf{V}) = (1-\theta)\mathbf{F}_{n-1} + \theta\mathbf{F}_n - \mathbf{C}_{\mathrm{ep}}(\tilde{\mathbf{X}})\mathbf{V} - \mathbf{K}\tilde{\mathbf{X}} = \mathbf{0} \tag{53}$$

where

$$\tilde{\mathbf{X}} = \mathbf{X}_{n-1} + \theta h\mathbf{V}$$

The solution to the system (53) may be found using the Newton–Raphson algorithm. Letting the superscript i denote iteration number, this scheme takes the form

$$\mathbf{V}^i = \mathbf{V}^{i-1} + \delta\mathbf{V}^i$$

$$\tilde{\mathbf{X}}^i = \mathbf{X}_{n-1} + \theta h\mathbf{V}^i$$

where the iterative update for $\mathbf{V}^i$ is

$$\delta\mathbf{V}^i = -\left[\frac{\partial\mathbf{R}}{\partial\mathbf{V}}(\tilde{\mathbf{X}}^{i-1})\right]^{-1}\mathbf{R}(\mathbf{V}^{i-1}) \tag{54}$$

and

$$\frac{\partial\mathbf{R}}{\partial\mathbf{V}}(\tilde{\mathbf{X}}^{i-1}) \tag{55}$$

is the Jacobian matrix $\partial\mathbf{R}/\partial\mathbf{V}$ evaluated at $\tilde{\mathbf{X}}^{i-1}$. Suitable values for starting the iterations may be obtained by setting

$$\mathbf{V}^0 = \dot{\mathbf{X}}_{n-1}$$

$$\tilde{\mathbf{X}}^0 = \mathbf{X}_{n-1} + \theta h\dot{\mathbf{X}}_{n-1}$$

To complete the description of the Newton–Raphson algorithm, it is necessary to evaluate $\partial\mathbf{R}/\partial\mathbf{V}$. Differentiating (53) and neglecting second-order terms and above gives the required Jacobian matrix as

$$\left[\frac{\partial\mathbf{R}}{\partial\mathbf{V}}(\tilde{\mathbf{X}}^{i-1})\right] = -[\mathbf{C}_{\mathrm{ep}}(\tilde{\mathbf{X}}^{i-1}) + \theta h\mathbf{K}] \tag{56}$$

Note that if the $\mathbf{K}_{ep}$ component of $\mathbf{C}_{ep}$ is formed using the standard elastoplastic constitutive matrix, $\mathbf{D}_{ep}$, then the rate of convergence of the iteration scheme will be linear rather than quadratic. With the proposed integration method, however, the number of iterations required for a typical time substep is usually low due to the fact that the error control mechanism automatically chooses small time steps in the vicinity of highly non-linear behaviour. At the cost of additional complexity, quadratic convergence may be obtained by using the so-called 'consistent' form of $\mathbf{D}_{ep}$ introduced by Simo and Taylor.[27] For cases which are only mildly nonlinear, it is possible to replace (56) by the initial stiffness approximation

$$\left[\frac{\partial \mathbf{R}}{\partial \mathbf{V}}\right] \approx -[\mathbf{C}_e + \theta h \mathbf{K}]$$

where $\mathbf{C}_e$ is given by (19). Although much slower to converge, this approach does not require a fresh factorisation for each iteration and always has a well conditioned Jacobian matrix.

A convenient check for terminating the iteration procedure is to test whether the relative change in the displacement component of $\tilde{\mathbf{X}}$ is less than or equal to a specified tolerance, $ITOL$. This can be expressed as

$$\theta h \parallel \delta \dot{\tilde{\mathbf{U}}}^i \parallel / \parallel \tilde{\mathbf{U}}^i \parallel \leqslant ITOL$$

where $\tilde{\mathbf{U}}$ corresponds to the displacement entries in $\tilde{\mathbf{X}}$ and $\dot{\tilde{\mathbf{U}}}$ corresponds to the velocity entries in $\mathbf{V}$.

A number of alternative strategies can be used to evaluate the residual $\mathbf{R}(\mathbf{V}^{i-1})$ in equation (54). The various options available become evident upon substituting equations (51) in (53) to give the expanded form

$$\mathbf{R}(\mathbf{V}^{i-1}) = \begin{bmatrix} (1-\theta)\dot{\mathbf{F}}_{n-1}^{ext} + \theta \dot{\mathbf{F}}_n^{ext} \\ (1-\theta)\mathbf{Q}_{n-1} + \theta \mathbf{Q}_n \end{bmatrix} - \begin{bmatrix} \mathbf{K}_{ep}(\tilde{\mathbf{X}}^{i-1}) & \mathbf{L} \\ \mathbf{L}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\tilde{\mathbf{U}}}^{i-1} \\ \dot{\tilde{\mathbf{P}}}^{i-1} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{H}\tilde{\mathbf{P}}^{i-1} \end{bmatrix}$$

Assuming that an analytic form for $\dot{\mathbf{F}}^{ext}$ is available, the first term on the right-hand side is straightforward to evaluate and the remaining matrix–vector products can be performed element by element. This strategy is appropriate for a general value of θ and is simple to implement. For the backward Euler case with $\theta = 1$, $\mathbf{V}$ and $\mathbf{X}$ contain values at the end of the current time step and the residual becomes

$$\mathbf{R}(\mathbf{V}_n^{i-1}) = \begin{bmatrix} \dot{\mathbf{F}}_n^{ext} \\ \mathbf{Q}_n \end{bmatrix} - \begin{bmatrix} \mathbf{K}_{ep}(\tilde{\mathbf{X}}_n^{i-1}) & \mathbf{L} \\ \mathbf{L}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\tilde{\mathbf{U}}}_n^{i-1} \\ \dot{\tilde{\mathbf{P}}}_n^{i-1} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{H}\tilde{\mathbf{P}}_n^{i-1} \end{bmatrix}$$

or

$$\mathbf{R}(\mathbf{V}_n^{i-1}) = \begin{bmatrix} \dot{\mathbf{F}}_n^{ext} - (\dot{\mathbf{F}}_n^{int})^{i-1} \\ \mathbf{Q}_n - \mathbf{L}^T \dot{\tilde{\mathbf{U}}}_n^{i-1} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{H}\tilde{\mathbf{P}}_n^{i-1} \end{bmatrix} \tag{57}$$

where

$$(\dot{\mathbf{F}}_n^{int})^{i-1} = \int_V \mathbf{B}_u^T \dot{\boldsymbol{\sigma}}_n^{i-1} \, dV = \mathbf{K}_{ep}(\tilde{\mathbf{X}}_n^{i-1}) \dot{\tilde{\mathbf{U}}}_n^{i-1} + \mathbf{L}\dot{\tilde{\mathbf{P}}}_n^{i-1}$$

is the internal force rate vector at the start of the current iteration. Using the first-order approximations

$$(\dot{\mathbf{F}}_n^{\text{int}})^{i-1} = ((\mathbf{F}_n^{\text{int}})^{i-1} - \mathbf{F}_{n-1}^{\text{int}})/h + O(h)$$

$$\dot{\mathbf{F}}_n^{\text{ext}} = (\mathbf{F}_n^{\text{ext}} - \mathbf{F}_{n-1}^{\text{ext}})/h + O(h)$$

the residual (57) may written in the alternative form

$$\mathbf{R}(\mathbf{V}_n^{i-1}) = \begin{bmatrix} \dfrac{\mathbf{F}_n^{\text{ext}} - (\mathbf{F}_n^{\text{int}})^{i-1}}{h} \\ \mathbf{Q}_n - \mathbf{L}^{\mathrm{T}}\dot{\tilde{\mathbf{U}}}_n^{i-1} \end{bmatrix} - \begin{bmatrix} \dfrac{\mathbf{F}_{n-1}^{\text{ext}} - \mathbf{F}_{n-1}^{\text{int}}}{h} \\ \mathbf{H}\tilde{\mathbf{P}}_n^{i-1} \end{bmatrix} \tag{58}$$

The top rows of the above equations indicate how well the rate form of the equilibrium equations are satisfied for the current iteration. Physically, the terms $\mathbf{F}_{n-1}^{\text{ext}} - \mathbf{F}_{n-1}^{\text{int}}$ and $\mathbf{F}_n^{\text{ext}} - (\mathbf{F}_n^{\text{int}})^{i-1}$ correspond, respectively, to the unbalanced forces at the start and end of the current time substep. Note that the latter forces change from iteration to iteration, whereas the former are constant over the time step. Provided the iteration tolerance $ITOL$ is sufficiently small, both the rate and absolute forms of the equilibrium equations are satisfied approximately throughout the time stepping process since the Newton–Raphson iterations ensure that the unbalanced forces are close to zero. If $ITOL$ is less stringent, however, the unbalanced forces at the start of each time substep may be significant and the solution will tend to drift from absolute equilibrium as it is marched forward. This drift, which is cumulative, may be minimised by adding the unbalanced forces at the start of each time step to the externally applied forces. Assuming that the unbalanced forces are imposed at a constant rate over the time interval h, the right-hand side of (58) is thus modified to give

$$\mathbf{R}(\mathbf{V}_n^{i-1}) = \begin{bmatrix} \dfrac{\mathbf{F}_n^{\text{ext}} - (\mathbf{F}_n^{\text{int}})^{i-1}}{h} \\ \mathbf{Q}_n - \mathbf{L}^{\mathrm{T}}\dot{\tilde{\mathbf{U}}}_n^{i-1} \end{bmatrix} - \begin{bmatrix} \dfrac{\mathbf{F}_{n-1}^{\text{ext}} - \mathbf{F}_{n-1}^{\text{int}}}{h} \\ \mathbf{H}\tilde{\mathbf{P}}_n^{i-1} \end{bmatrix} + \begin{bmatrix} \dfrac{\mathbf{F}_{n-1}^{\text{ext}} - \mathbf{F}_{n-1}^{\text{int}}}{h} \\ \mathbf{0} \end{bmatrix}$$

or

$$\mathbf{R}(\mathbf{V}_n^{i-1}) = \begin{bmatrix} \dfrac{\mathbf{F}_n^{\text{ext}} - (\mathbf{F}_n^{\text{int}})^{i-1}}{h} \\ \mathbf{Q}_n - \mathbf{L}^{\mathrm{T}}\dot{\tilde{\mathbf{U}}}_n^{i-1} - \mathbf{H}\tilde{\mathbf{P}}_n^{i-1} \end{bmatrix}$$

This expression clearly requires the unbalanced forces to be small before convergence can occur, and thus guarantees that equilibrium is satisfied, in an absolute sense, for each time step. To complete the residual evaluation, the matrix–vector products

$$\mathbf{L}^{\mathrm{T}}\dot{\tilde{\mathbf{U}}}_n^{i-1}$$

and

$$\mathbf{H}\tilde{\mathbf{P}}_n^{i-1}$$

may be performed element by element.

Once the solution for $\mathbf{V}$ has been obtained for each time substep, the error control and step adjustment mechanism is almost identical to that for the elastic case discussed in Section 4.1. The only minor change in that the safety factor in (42) is lowered from 0·9 to 0·8 so that

$$q = 0{\cdot}8\sqrt{DTOL/R_n^u}$$

This means the step control mechanism is slightly more conservative than that for the elastic case, and allows for the material non-linearity introduced by elastoplasticity.

5.2. Implementation of elastoplastic algorithm

The non-linear automatic time-stepping scheme described here again assumes that a number of coarse time increments are defined which, if necessary, are automatically subincremented so that the local truncation error for each substep does not exceed a prescribed tolerance, $DTOL$. Because of the advantages discussed in Section 4.5, the integration parameters are again set to the values given in (49).

Algorithm for Elastoplastic Consolidation. For the sake of brevity, The automatic time-stepping algorithm described below assumes an elastoplastic soil model with isotropic strain or work hardening. The scheme may be extended easily to deal with more complex soil models.

Step 1. Initialization

1.1. Enter with the time at the start of the coarse increment t_0, the current displacements and pore pressures $\mathbf{X}_{t_0}$, their corresponding derivatives $\dot{\mathbf{X}}_{t_0}$, the coarse time increment Δt, the last untruncated time substep h_{last}, the current effective stresses at each integration point $(\boldsymbol{\sigma}'_{t_0}, \kappa_{t_0})$, the iteration tolerance $ITOL$, and the specified displacement error tolerance $DTOL$. For the first coarse time step, set $h_{\text{last}} = \Delta t$.

1.2. Set $t = t_0$ and $h = \min\{h_{\text{last}}, \Delta t\}$

Step 2. Main subincrement loop over time step Δt

2.1. Do steps 2.2 to 2.7

2.2. Compute $\dot{\mathbf{X}}_{t+h}$ and $\tilde{\mathbf{X}}_{t+h}$ using the Newton–Raphson or initial stiffness algorithm. If the solution fails to converge, set

$$h \leftarrow 0{\cdot}25h$$

and try again.

2.3. Estimate the local truncation error in the displacements for the current subincrement using

$$E_{t+h}^u = \tfrac{1}{2}h \,\|\, \dot{\mathbf{U}}_t - \dot{\mathbf{U}}_{t+h} \,\|$$

where $\dot{\mathbf{U}}$ denotes the velocity component of $\dot{\mathbf{X}}$.

2.4. Estimate the relative error for current subincrement using

$$R_{t+h}^u = \max\{EPS, E_{t+h}^u / \,\|\, \tilde{\mathbf{U}}_{t+h} \,\|\}$$

where $\tilde{\mathbf{U}}_{t+h}$ is the displacement component of $\tilde{\mathbf{X}}_{t+h}$ and EPS is a machine constant.

2.5. If $R^u_{t+h} > DTOL$ then go to step 2.7. Else this step is successful so compute new displacements and pore pressures using the second-order update

$$\mathbf{X}_{t+h} = \mathbf{X}_t + \frac{h}{2}(\dot{\mathbf{X}}_t + \dot{\mathbf{X}}_{t+h})$$

For each integration point, compute the strains

$$\Delta\boldsymbol{\varepsilon} = \mathbf{B}\left\{\frac{h}{2}(\dot{\mathbf{u}}_t + \dot{\mathbf{u}}_{t+h})\right\}$$

and integrate constitutive law to find corresponding increments in stresses, $\Delta\boldsymbol{\sigma}'$, and hardening parameter $\Delta\kappa$. Then update stress state according to

$$\boldsymbol{\sigma}'_{t+h} = \boldsymbol{\sigma}'_t + \Delta\boldsymbol{\sigma}'$$

$$\kappa_{t+h} = \kappa_t + \Delta\kappa$$

If $t + h = t_0 + \Delta t$ then integration is complete, so set $t \leftarrow t + h$ and go to step 3.1

2.6. Estimate a new subincrement size factor by computing

$$q = \min\{0{\cdot}8\sqrt{DTOL/R^u_{t+h}}, 2\}$$

If the previous subincrement was unsuccessful, then prevent substeps from growing by setting

$$q = \min\{q, 1\}$$

Then update time and compute and store next substep size according to

$$t \leftarrow t + h$$

$$h \leftarrow qh$$

$$h_{\text{last}} = \min\{h, \Delta t\}$$

If necessary, truncate substep size so that integration does not proceed beyond $t_0 + \Delta t$ by setting

$$h \leftarrow \min\{h, t_0 + \Delta t - t\}$$

before returning to step 2.1.

2.7. This subincrement has failed, so estimate smaller time substep by computing

$$q = \max\{0{\cdot}8\sqrt{DTOL/R^u_{t+h}}, 0{\cdot}1\}$$

and then setting

$$h \leftarrow qh$$

before returning to step 2.1.

Step 3. Exit

3.1. Exit with displacements and pore pressures, $\mathbf{X}_{t_0+\Delta t}$, their corresponding rates, $\dot{\mathbf{X}}_{t_0+\Delta t}$, and integration point effective stress state, $(\boldsymbol{\sigma}'_{t_0+\Delta t}, \kappa_{t_0+\Delta t})$, at end of coarse time increment.

Newton–Raphson Algorithm for Elastoplastic Consolidation. The Newton–Raphson procedure for solving the non-linear equations (53) may be summarized as

Step 1. Initialization

1.1. Enter with the current displacements and pore pressures $\mathbf{X}_t$, their corresponding derivatives $\dot{\mathbf{X}}_t$, the current step size h, the iteration tolerance $ITOL$, and the maximum number of iterations $MAXITS$.

1.2. Compute an estimate of new displacements/pore pressures and their corresponding rates using

$$\tilde{\mathbf{X}}^0_{t+h} = \mathbf{X}_t + h\dot{\mathbf{X}}_t$$

$$\dot{\mathbf{X}}^0_{t+h} = \dot{\mathbf{X}}_t$$

1.3. For tangent stiffness scheme only set

$$\alpha^0 = h \,\|\, \dot{\mathbf{U}}_t \,\| \,/\, \|\, \tilde{\mathbf{U}}^0_{t+h} \,\|$$

where $\dot{\mathbf{U}}_t$ is the velocity component of $\dot{\mathbf{X}}_t$ and $\tilde{\mathbf{U}}^0_{t+h}$ is the displacement component of $\tilde{\mathbf{X}}^0_{t+h}$.

Step 2. Iteration loop

2.1. Repeat steps 2.2–2.5 for $i = 1$ to $MAXITS$

2.2. Compute residual vector

$$\mathbf{R}^i = \begin{bmatrix} \dfrac{\mathbf{F}^{\text{ext}}_{t+h} - \mathbf{F}^{\text{int}}_{t+h}}{h} \\ \mathbf{Q}_{t+h} - \mathbf{L}^{\mathrm{T}}\dot{\mathbf{U}}^{i-1}_{t+h} - \mathbf{H}\tilde{\mathbf{P}}^{i-1}_{t+h} \end{bmatrix}$$

and solve for $\delta\dot{\mathbf{X}}^i$ using

$$\delta\dot{\mathbf{X}}^i = [\mathbf{C}_{\text{ep}} + h\mathbf{K}]^{-1}\mathbf{R}^i \quad \text{(tangent stiffness)}$$

or

$$\delta\dot{\mathbf{X}}^i = [\mathbf{C}_{\text{e}} + h\mathbf{K}]^{-1}\mathbf{R}^i \quad \text{(initial stiffness)}$$

where $\dot{\mathbf{F}}^{\text{int}}_{t+h}$ and $\mathbf{C}_{\text{ep}}$ are evaluated at $\tilde{\mathbf{X}}^{i-1}_{t+h}$.

2.3. Update the displacements and pore pressures and their rates according to

$$\dot{\mathbf{X}}^i_{t+h} = \dot{\mathbf{X}}^{i-1}_{t+h} + \delta\dot{\mathbf{X}}^i$$

$$\tilde{\mathbf{X}}^i_{t+h} = \mathbf{X}_t + h\dot{\mathbf{X}}^i_{t+h}$$

2.4. Compute convergence criterion

$$\alpha^i = h \,\|\, \delta\dot{\mathbf{U}}^i \,\| \,/\, \|\, \tilde{\mathbf{U}}^i_{t+h} \,\|$$

where $\delta\dot{\mathbf{U}}^i$ is the velocity component of $\delta\dot{\mathbf{X}}^i$ and $\tilde{\mathbf{U}}^i_{t+h}$ is the displacement component of $\tilde{\mathbf{X}}^i_{t+h}$. If $\alpha^i \leqslant ITOL$ then go to step 4.1.

2.5. For tangent stiffness only, check rate of convergence. If $\alpha^i/\alpha^{i-1} > 0{\cdot}5$ for more than two consecutive iterations then exit with 'no convergence' warning.

Step 3. Check for too many iterations

3.1. Maximum number of iterations exceeded. Exit with 'no convergence' warning.

Step 4. Exit

4.1. Exit with displacements/pore pressures, $\tilde{\mathbf{X}}_{t+h} = \tilde{\mathbf{X}}^i_{t+h}$, and their rates, $\dot{\mathbf{X}}_{t+h} = \dot{\mathbf{X}}^i_{t+h}$, at time $t + h$.

Typical values for the iteration tolerance, $ITOL$, are in the range 10^{-3} to 10^{-6}, with the lower limit ensuring that the drift from equilibrium is very small. For the tangent stiffness scheme, the maximum number of iterations permitted for each subincrement, $MAXITS$, is typically set to around 15. To minimize the high cost associated with failed substeps, and to allow for possible divergence of the iteration scheme, h is automatically reduced by a factor of four if no convergence is obtained within this limit. This feature may result in an excessive number of substeps if $MAXITS$ is set to a very low value. The optimum setting for $MAXITS$ is dependent on the specified value of $DTOL$, since loose values of this tolerance may give large time steps and hence large numbers of iterations. It is also a function of the solution scheme used to solve the governing nonlinear equations. The tangent stiffness algorithm generally requires far fewer iterations than the initial stiffness algorithm, especially if the behaviour is highly non-linear.

For the initial stiffness method, the matrix $[\mathbf{C}_e + h\mathbf{K}]$ needs to be formed and factorised afresh only once per subincrement to obtain the iterates $\delta\dot{\mathbf{X}}^i$. The cost associated with each formation phase may be reduced significantly using the elastic procedure outlined in Section 4.5. With the tangent stiffness Newton–Raphson scheme it is necessary to reform and refactorise the Jacobian matrix $[\mathbf{C}_{ep} + h\mathbf{K}]$ once per iteration, since $\mathbf{C}_{ep}$ is dependent on the displacements held in $\tilde{\mathbf{X}}^{i-1}_{t+h}$. In this case, it is possible to exploit the decomposition

$$\mathbf{C}_{ep} = \begin{bmatrix} \mathbf{K}_{ep} & \mathbf{L} \\ \mathbf{L}^T & \mathbf{0} \end{bmatrix} = \mathbf{C}_e - \mathbf{C}_p = \begin{bmatrix} \mathbf{K}_e & \mathbf{L} \\ \mathbf{L}^T & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{K}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where the matrix $\mathbf{C}_e$ need only be formed once and stored. In order to generate $\mathbf{C}_{ep}$ efficiently for each iteration, $\mathbf{C}_e$ is located into memory and the plastic element stiffness matrices, defined by

$$\mathbf{k}_p = \int_{V^e} \mathbf{B}_u^T \mathbf{D}_p \mathbf{B}_u \, dV$$

are subtracted element by element. Since only plastic Gauss points contribute to $\mathbf{K}_p$, the effort required to form $\mathbf{K}_{ep}$, and hence $\mathbf{C}_{ep}$, is usually small for much of the loading range.

Finally, it is also possible to hold the Jacobian matrix constant for several subincrements and only refactorise it when convergence becomes slow or when the step size h changes significantly. These options have not been pursued in this paper, but may lead to reductions in the overall CPU time.

6. CONCLUSIONS

This paper describes the theory and implementation of an efficient automatic time stepping strategy for solving elastic and elastoplastic consolidation problems. The method controls the step size by using an estimate of the local truncation error for each time step. The error estimator is computed by taking the difference between a first-order accurate solution and a second-order accurate solution, and is very cheap to compute. A detailed demonstration of the numerical performance of the proposed automatic time stepping scheme is given in the companion paper of Sloan and Abbo[28].

REFERENCES

1. M. A. Biot, 'General theory of three dimensional consolidation', *J. Appl. Phys.* **12,** 155–164 (1941).
2. R. S. Sandhu and E. L. Wilson, 'Finite element analysis of seepage in elastic media', *J. Engng. Mech. ASCE*, **95,** 641–652 (1969).
3. J. T. Christian and J. W. Boehmer, 'Plain strain consolidation by finite elements', *J. Soil Mech. and Foundation Div. ASCE*, **SM4,** 1435–1457 (1970).
4. C. T. Hwang, N. R. Morgenstern and D. W. Murray, 'On solutions of plane strain consolidation problems by finite elements methods', *Can. Geotech. J.* **8,** 109–118 (1971).
5. Y. Yokoo, K. Yamagata and H. Nagaoka, 'Finite element method applied to Biot's consolidation theory', *Soils Found.* (Japanese Society for Soil Mechanics and Foundation Engineering), **11**(1)**,** 29–46 (1971).
6. G. Krause, 'Finite element schemes for porous media', *J. Eng. Mech. Div. ASCE*, **104**(EM3)**,** 605–620 (1978).
7. R. I. Borja, 'Finite element formulation for transient pore pressure dissipation: A variational approach', *Int. J. Solids Struct.* **22,** 1201–1211 (1986).
8. R. W. Lewis, G. K. Roberts and O. C. Zienkiewicz, 'A nonlinear flow and deformation analysis of consolidated problems', in *Proc. 2nd Int. Conf. on Numerical Methods in Geomechanics, ASCE*, New York, 1976.
9. J. C. Small, J. R. Booker and E. H. Davis, 'Elasto-plastic consolidation of soil', *Int. J. Solids Struct.* **12**, 431–448 (1976).
10. J. H. Prevost, 'Nonlinear transient phenomena in saturated porous media', *Comput. Meth. Appl. Mech. Engng.*, **20,** 3–18 (1982).
11. R. I. Borja, 'Linearisation of elasto-plastic consolidation equations', *Engng. Comput.* **6**, 163–168 (1989).
12. J. P. Carter, J. C. Small and J. R. Booker, 'The analysis of finite elasto-plastic consolidation', *Int. J. Numer. Anal. Meth. Geomech.*, **3,** 107–129 (1979).
13. J. Ghaboussi and E. L. Wilson, 'Flow of compressible fluid in porous elastic media', *Int. J. Numer. Meth. Engng.*, **5,** 419–442 (1973).
14. J. R. Booker and J. C. Small, 'An investigation of the stability of numerical solutions of Biot's equations of consolidation', *Int. J. Solids Struct.*, **11,** 907–917 (1975).
15. P. A. Vermeer and A. Verruijt 'An accuracy condition for consolidation by finite elements', *Int. J. Numer. Anal. Meth. Geomech.*, **5,** 1–14 (1981).
16. W. L. Wood, *Practical Time Steping Schemes*, Clarendon Press, Oxford (1990).
17. H. J. Siriwardane and C. S. Desai, 'Two numerical schemes for nonlinear consolidation', *Int. J. Numer. Meth. Engng.*, **17,** 405–426 (1981).
18. O. C. Zienkiewicz, W. L. Wood, N. W. Hine and R. L. Taylor, 'A unified set of single step algorithms. Part 1: General formulation and applications', *Int. J. Numer. Meth. Engn.*, **20,** 1529–1552 (1984).
19. R. M. Thomas and I. Gladwell, 'Variable-order variable step algorithms for second-order systems. Part 1: The methods', *Int. J. Numer. Meth. Engng.*, **26,** 39–53 (1988).
20. I. Gladwell and R. M. Thomas, 'Variable-order variable step algorithms for second-order systems. Part 2: The codes', *Int. J. Numer. Meth. Engng.*, **26,** 55–80 (1988).
21. R. S. Sandhu, H. Lui and K. J. Singh, 'Numerical performance of some finite element schemes for analysis of seepage in porous elastic media', *Int. J. Numer. Anal. Meth. Geomech.* **1,** 177–194 (1977).
22. W. L. Wood and R. W. Lewis, 'A comparison of time-marching schemes for the transient heat conduction equation', *Int. J. Numer. Meth. Engng.*, **9,** 679–689 (1975).
23. W. L. Wood, 'Control of Crank–Nicholson noise in the numerical solution of the heat conduction equation', *Int. J. Numer. Meth. Engng.*, **11,** 1059–1065 (1977).
24. L. F. Shampine *Numerical Solution of Ordinary Differential Equations*, Chapman & Hall, London, 1994.
25. M. B. Reed, 'An investigation of numerical errors in the analysis of consolidation by finite element analysis', *Int. J. Numer. Anal. Meth. Geomech.*, **8,** 243–257 (1984).
26. R. W. Lewis and B. A. Schrefler, *The Finite-Element Method in the Deformation and Consolidation of Porous Media*, Wiley, Chichester, (1987).
27. J. C. Simo and R. L. Taylor, 'Consistent tangent operators for rate-independent elastoplasticity', *Comput. Meth. Appl. Mech. Engng.*, **48,** 101–118 (1985).
28. S. W. Sloan and A. J. Abbo 'An automatic time stepping scheme for elastic and elastoplastic consolidation. Part 2: Applications', *Int. J. Numer. Anal. Meth. Geomech.*, (1997).

