The Alliance of Laboratories in Europe for
Education, Research and Technology

# ALERT Doctoral School 2024

*Numerical Methods in Geomechanics*

Editors:

Claudio Tamagnini

Lorenzo Sanavia

Manuel Pastor

# Preface

*From 3rd October to 5th October 2024, the ALERT Doctoral School 2024 will take place in Aussois and will be dedicated to "Numerical Methods in Geomechanics". The School has been organized by Claudio Tamagnini (Università degli Studi di Perugia), Lorenzo Sanavia (Università degli Studi di Padova) and Manolo Pastor (Universidad Politécnica de Madrid). I sincerely thank the organizers and all the contributors to this book for their effort!*

*Numerical methods in geomechanics involve different approaches, but also common ingredients to deal with: non-linear behavior, interactions with fluids included in the pores of the geomaterial, non-uniqueness of the material response... These methods are now part of the life of researchers but also of practitioners. This year's doctoral school focuses on the finite element method, which has been developed over many years. The school aims to provide participants with the tools to make informed use of this method, covering a wide range of aspects such as constitutive laws, their development and numerical integration, as well as modeling under partially saturated conditions or for thermo-hydro-mechanical problems.*

*The school will take place over three days, and will aim to explain the basics of finite element analysis, the potential problems associated with different types of problem, and possible solutions. Numerical aspects of dealing with non-linear problems will also be covered, as these non-linearities arise from the behavior of geomaterials and the multiphysical couplings taking place within them. Practical sessions will be organized in order to better grasp the theoretical concepts and put them into practice.*

*As usual, the pdf file of the book can be downloaded for free from the website of ALERT Geomaterials (`https://alertgeomaterials.eu/publications/`) after the school. On behalf of the ALERT Board of Directors I wish all participants a successful ALERT Doctoral School 2024!*

Frédéric Collin
Director of ALERT Geomaterials
University of Liege

# Contents

# Foreword

The contributions assembled in the present volume proceed from the lectures of the 2024 ALERT Geomaterials Doctoral School devoted to Numerical methods in geomechanics. The school has been organized and coordinated by Claudio Tamagnini (Università degli Studi di Perugia), Lorenzo Sanavia (Università degli Studi di Padova) and Manuel Pastor (Universidad Politécnica de Madrid). It follows the 1st and the 2nd ALERT Olek Zienkiewicz Course organized at the Universidad Politécnica de Madrid by Manuel Pastor and Claudio Tamagnini in 2009 (Numerical Methods in Geomechanics) and in 2014 (Advanced Numerical Modelling in Geomechanics), respectively.

The study and application of rock and soil mechanics require solving mainly nonlinear initial boundary value problems on complex domains, for which the analytical solution is usually unavailable. It also needs to consider the material as a multiphase porous system characterized by coupled multiphysics phenomena. Consequently, only numerical methods can be applied successfully to solve real problems and, because they are approximate methods, need to be thoroughly understood and used carefully and critically.

This volume contains eleven chapters presenting the fundamentals that help in understanding numerical methods applied to multiphase porous systems. The volume is divided into five main parts: (i) an introduction to the finite element method for ellipt, parabolic and hyperbolic equations, (ii) the constitutive modelling of geomaterials within the Theory of plasticity and Generalized plasticity, also for rate dependent materials and unsaturated soils, respectively, (iii) the formulation of a mathematical model for non-isothermal multiphase porous materials based on the Hybrid Mixture Theory, which includes, as a particular case, the well-known Biot poromechanical model, (iv) the numerical approaches for the solution of nonlinear problems, the computational plasticity, and the space and time discretization of a multiphase porous media model at large elasto-plastic strains as an example of the application of the previous sections, and (v) the finite element modelling of non-isothermal variably saturated soils under quasi-statics or dynamics conditions.

The 2024 ALERT Geomaterials Doctoral School also includes some practical sessions to practice with the numerical solution of some geomechanical problems with the finite element code GeHoMadrid.

We believe that this volume may provide to postgraduate students, researchers and practitioners, a valuable introduction and a sound basis for further progress in the challenging field of virtual modelling of coupled and multiphysics phenomena in multiphase porous systems, which extends not only to geomechanics but far beyond.

Claudio Tamagnini (Università degli Studi di Perugia)
Lorenzo Sanavia (Università degli Studi di Padova)
Manuel Pastor (Universidad Politécnica de Madrid)

# Introduction to finite elements (I): steady state problems of elliptic type

**Manuel Pastor, Saeid M. Tayyebi, Pablo Mira, Miguel M. Stickle, Diego Manzanal, J.A. Fernandez Merodo**

*ETS de Ingenieros de Caminos, Universidad Politecnica de Madrid,
Ciudad Universitaria s/n, 28040, Madrid, Spain*

*The Finite Element metod has reached its maturity after some 75 years since their development. During these years it has become a tool for engineers and scientists in many fields. This introductory pair of lectures aim to present the fundamentals of the method in both steady state and transient problems- We have chosen to present the material at a MSc level, as in the reference books by Zienkiewicz and his coworkers. This first Chapter is devoted to steady state problems, trying to show the common equations of various problems.*

## 1 1. Introduction

Late Professor O.C.Zienkiewicz, one of the founders and pioneers of the Finte Element Method, used to include in some talks a quote from Von Karman, ".engineers build things that did not exist before.." and explained that engineers must therefore predict the behaviour of the structures they projected and their interaction with the environment - the latter, is sort of a live-hate relation. as engineers structures affect the environment and suffer actions from it which may result on disaster,

For such predictions, there exist three alternative ways:

(i) Building of prototypes and studying their behaviour

(ii) small scale laboratory models (such in harbour or hydraulics engineering)

and (iii) create mathematical models describing the fundamental phenomena involved.

This course, and this talk, deal precisely with this latter way. It is a course on modelling for which in most of occasions three ingredients are needed:

(a) A mathematical model

(b) A constitutive -sometimes called rheological- model describing the behaviour of the materials, and

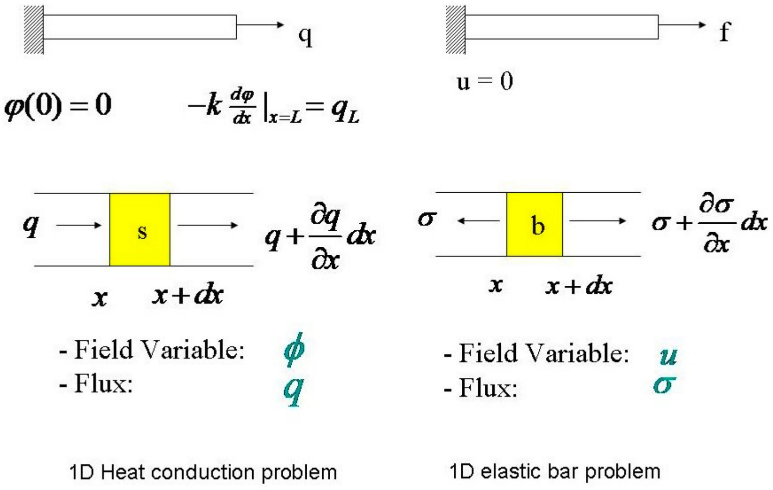
(c) A numerical model to discretize the two previous ingredients.

This is why these two Chapters will be devoted to present elliptic, parabolic and hyperbolic models, rather than to explain a series of different phenomena and the models that describe them.

The reader is encouraged to find applications in different areas of engineering where models are similar. For instance, landslide propagation is described by models which were derived first for hydraulics engineering.

## 2 Strong formulation of steady state problems of elliptic time

We will consider first two simple 1D problems, the transport of heat in a 1D bar, and the behaviour of a simple component of a structure, 1 1D bar (figure 1).


q
$\varphi(0)=0$
$-k\frac{d\varphi}{dx}|_{x=L}=q_L$
f
u = 0
q
s
$q+\frac{\partial q}{\partial x}dx$
x
x+dx
$\sigma$
b
$\sigma+\frac{\partial \sigma}{\partial x}dx$
x
x+dx
- Field Variable: $\phi$
- Flux: $q$
- Field Variable: $u$
- Flux: $\sigma$

1D Heat conduction problem

1D elastic bar problem

Figure 1: Two simple 1D problems

In both cases, formulations are very similar:

Fluxes are related to field variables as:

$$q=-k\frac{\partial \phi}{\partial x} \quad \text{and} \quad \sigma=-E\frac{\partial u}{\partial x} \tag{1}$$

(a) In the heat conduction problem, the balance equation reads:

$$-\frac{\partial}{\partial}q + s = 0 \qquad x \in (o, L) \tag{2}$$

from where, substituting the flux q, we have:

$$\frac{\partial}{\partial}\left(k\frac{\partial \phi}{\partial x}\right) + s = 0 \qquad x \in (o, L) \tag{3}$$

We will assume that on the left boundary temperature is prescribed as zero, and at the right boundary, flux is prescribed equal to $q_L$

$$\phi\left(0\right) = 0 \qquad\quad -k\left.\frac{\partial \phi}{\partial x}\right|_{x=L} = q_L \tag{4}$$

where $k$ is the thermal conductivity, $s$ the heat sources and $q_L$ the prescribed flux at $x = L$

(b) For the elastic bar, the equations are:

$$\frac{\partial}{\partial}\left(EA\frac{\partial u}{\partial x}\right) - b = 0 \qquad x \in (o, L) \tag{5}$$

$$u\left(0\right) = 0 \qquad\quad -EA\left.\frac{\partial u}{\partial x}\right|_{x=L} = F_L \tag{6}$$

where $E$ is the elastic modulus, $A$ the cross sectional area, $b$ the bodyf orces and $F_L$ the force acting at at $x = L$

In 3D situations, the equations are obtained in a similar way figure 2:

In figure 2 we have sketched the fluxes q along x,y and z. The balance equations are written in term of fluxes as:

$$-\frac{\partial q_x}{\partial x}dxdtdz - \frac{\partial q_x}{\partial x}dxdtdz - \frac{\partial q_x}{\partial x}dxdtdz + s = 0 \tag{7}$$

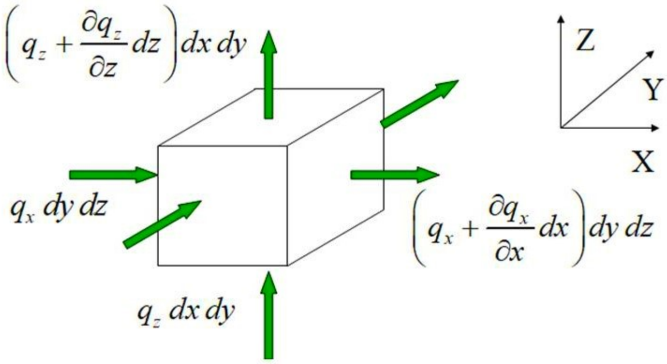

$\left(q_z + \frac{\partial q_z}{\partial z} dz\right) dx\, dy$
Z
Y
X
$q_x\, dy\, dz$
$\left(q_x + \frac{\partial q_x}{\partial x} dx\right) dy\, dz$
$q_z\, dx\, dy$

Figure 2: Heat conduction problem in 3D.

from where we obtain

$$-\frac{\partial q_x}{\partial x} - \frac{\partial q_y}{\partial y} - \frac{\partial q_z}{\partial z} + s = 0 \tag{8}$$

which can be written in a more compact form as

$$-\operatorname{div}\mathbf{q} + s = 0 \tag{9}$$

with

$$\mathbf{q} = (q_x, q_y, q_z)^T \tag{10}$$

from where the balance equation is

$$\operatorname{div}(k \operatorname{grad}\phi) + s = 0 \tag{11}$$

In cases where the conductivity k is different along x, y and z, the tensor of conductivity $\mathbf{k}$ is introduced as:

$$\mathbf{k} = \begin{pmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{pmatrix} \tag{12}$$

and the balance equation is now

$$\operatorname{div}(\mathbf{k}\,\operatorname{grad}\phi) + s = 0 \tag{13}$$

Boundary conditions on the boundary $\partial\Omega$ (figure 3) can be of two main types:

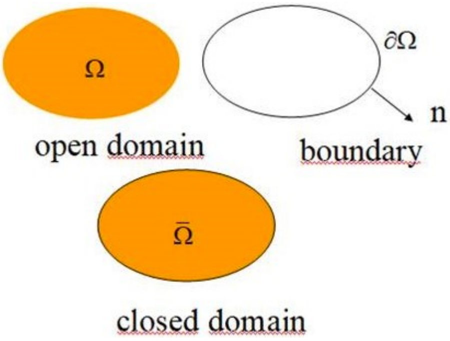


Figure 3: Boundary conditions on the boundary $\partial\Omega$ .

(i) Dirichlet, in a part of the boundary $\partial_\phi\Omega \subset \partial\Omega \quad \phi - \phi = 0$

(ii) Neumann, at $\partial_q\Omega \subset \partial\Omega \quad D\frac{\partial\phi}{\partial\mathbf{n}} + \overline{q}$ were $\mathbf{n}$ is the outer normal to the boundary

Other boundary conditions for the heat conduction problem are the convection and the radiation boundary conditions.

The formulation presented is found, in addition to the Fourier problem described above in other transport problems such as Darcy, Fick and Navier.

The equations can be written more formally in what is called strong formulation as

$$\operatorname{div}(D\,\operatorname{grad}\phi) + s = 0 \quad \text{in} \quad \Omega \qquad \phi \in C^2(\Omega) \tag{14}$$

$$\phi - \phi = 0 \quad \text{on} \quad \partial_\phi\Omega \qquad D : \overline{\Omega} \to \mathbb{R} \tag{15}$$

$$D\frac{\partial\phi}{\partial\mathbf{n}} + \overline{q} = 0 \quad \text{on} \quad \partial_q\Omega \qquad Ds : \overline{\Omega} \to \mathbb{R} \tag{16}$$

## 3 Finite element approximation of functions: Nodes, elements and shape functions. Isoparametric elements

If we had to choose the two main ingredients of finite elements, we would undoubtely mention the concept of finite element approximation of functions and the weak formulations of the field equations i.e, Galerkin and virtual work methodd of approximation.

We will address here the former, which we could characterize as a ”divide and conquer” based method.

## 3.1 1D problems: linear 2 node elements

Functions can be approximated in many alternative ways, being interpolation the simplest. Let us consider the simple prismatic bar depicted in figure 4, and a field magnitude $\phi$ of which we know its values at a set of points 1...n+1 which we will call nodes. The spacing can be either fixed or variable, using closer nodes where the function varies more (larger gradients). We see that the bar has been divided into n segments, which we will call elements, where we will approximate using simpler functions, for instance, linear functions.

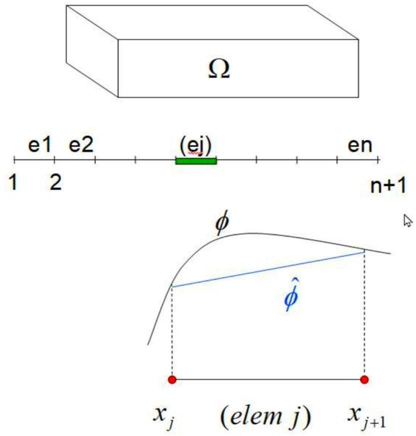


Figure 4: Concepts of nodes, elements and linear 1D elements.

A simple way to describe the approximation done in each element (ej), is to define the two linear functions **-which will be called shape function**s -shown in figure 5.

$$N_j = \frac{x_{j+1} - x}{x_{j+1} - x_{jc}} \tag{17}$$

$$N_{j+1} = \frac{x - x_j}{x_{j+1} - x_j} \tag{18}$$

It is easily seen that $N_j$ is 1 at node j and 1 at node j+1, while $N_{j+1}$ is 0 at node j and 1 at node j+1

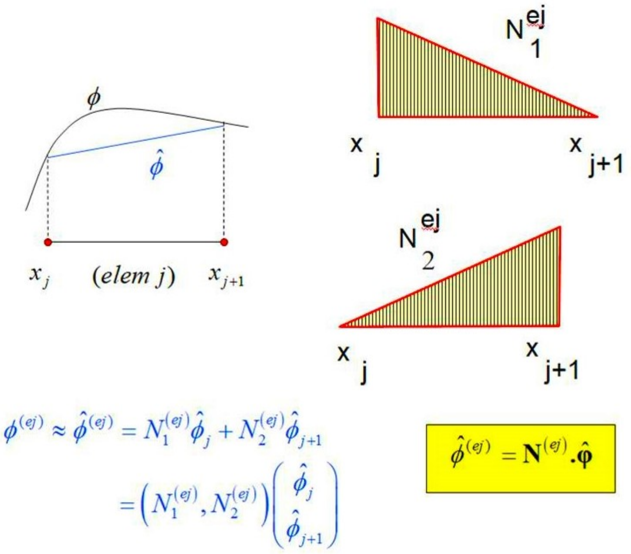

$\phi$
$\hat{\phi}$
$x_j$
$(elem\ j)$
$x_{j+1}$
$N_1^{ej}$
$x_j$
$x_{j+1}$
$N_2^{ej}$
$x_j$
$x_{j+1}$
$\phi^{(ej)} \approx \hat{\phi}^{(ej)} = N_1^{(ej)}\hat{\phi}_j + N_2^{(ej)}\hat{\phi}_{j+1}$
$= \left(N_1^{(ej)}, N_2^{(ej)}\right)\begin{pmatrix} \hat{\phi}_j \\ \hat{\phi}_{j+1} \end{pmatrix}$
$\hat{\phi}^{(ej)} = \mathbf{N}^{(ej)}.\hat{\boldsymbol{\varphi}}$

Figure 5: Nodes, elements and shape functions (linear 1D element).

The interpolation, which will be denoted as $\widehat{\phi}\,(x)$ can be written as:

$$\widehat{\phi}\,(x) = N_j\widehat{\phi}_j + N_{j+1}\widehat{\phi}_{j+1} \tag{19}$$

where $\widehat{\phi}_j$ and $\widehat{\phi}_{j+1}$are the nodal values at the element nodes left and right.. Above expression can be written in a more compact manner as

$$\widehat{\phi}\,(x) = \mathbf{N}^T\widehat{\phi} \tag{20}$$

In above, we have introduced two vectors of nodal variables and shape functions, respectively.

$$\mathbf{N}^{(ej)} = \begin{pmatrix} N_j^{(ej)} & N_{j+1}^{(ej)} \end{pmatrix} \tag{21}$$

and

$$\hat{\mathbf{\Phi}} = \begin{pmatrix} \hat{\Phi}_j \\ \hat{\Phi}_{j+1} \end{pmatrix} \tag{22}$$

The superindex $(ej)$ refers to element j

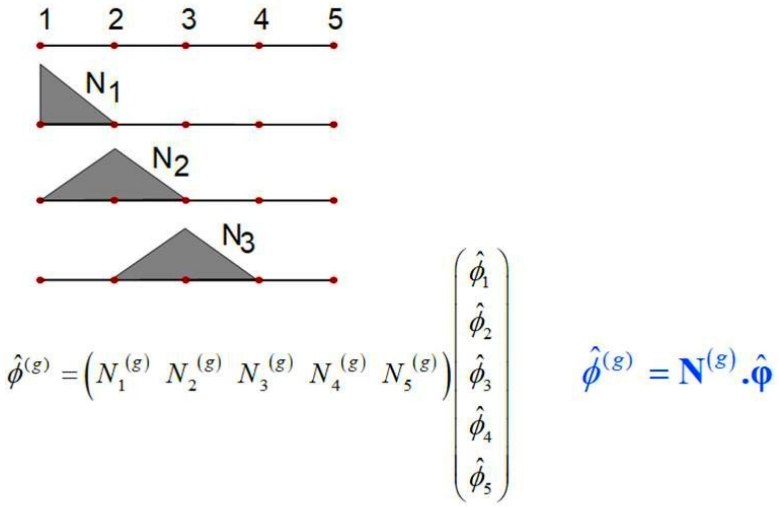

1
2
3
4
5
N1
N2
N3
$\hat{\phi}^{(g)} = \left( N_1^{(g)} \; N_2^{(g)} \; N_3^{(g)} \; N_4^{(g)} \; N_5^{(g)} \right) \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \\ \hat{\phi}_4 \\ \hat{\phi}_5 \end{pmatrix}$
$\hat{\phi}^{(g)} = \mathbf{N}^{(g)}.\hat{\boldsymbol{\varphi}}$

Figure 6: Global shape functions.

A global approximation can be built by combining the approximations in all elements, as sketched in figure 6.

Function $\Phi$ is approximated as

$$\hat{\Phi} = \mathbf{N}.\hat{\boldsymbol{\Phi}} \tag{23}$$

where

$$\mathbf{N} = \begin{pmatrix} N_0 & N_1 & ... & N_N \end{pmatrix} \tag{24}$$

and

$$\hat{\boldsymbol{\Phi}} = \begin{pmatrix} \hat{\Phi}_0 \\ \hat{\Phi}_1 \\ ... \\ \\ \hat{\Phi}_N \end{pmatrix} \tag{25}$$

Let us remember at this point that:

(i) The global approximation is continuous

(ii) First order derivatives are discontinuous at nodes

(iii) Second order derivatives are infinity at nodes

### 3.1.1 Mapping

Shape functions of element $(ej)$ are:

$$N_{j-1} = \frac{x_j - x}{x_j - x_{j-1}} \tag{26}$$

$$N_j = \frac{x - x_{j-1}}{x_j - x_{j-1}} \tag{27}$$

and vary from element to element. It would be convenient to define shape functions in a such way that they would be always the same. This is achieved by geometrical mappings which transform the elements into a standard one. In the case of the linear 1D elements we are considering, If we introduce the abscissa

$$\xi = \frac{x - x_j}{x_{j+1} - x_j} \tag{28}$$

the interval $[0, L]$ is normalized, being all elements transformed into the element $[0, 1]$ .This results in allelements having the same shape functions

$$N_A = 1 - \xi \tag{29}$$

$$N_B = \xi \tag{30}$$

where$A$ and $B$ are the nodes of the element considered..

Inside each element it is possible to obtain $x$ empleando la definición deusig the definition of $\xi$ as

$$x = x_{j-1} + \xi\left(x_j - x_{j-1}\right) \tag{31}$$

$$x = (1 - \xi)\, x_{j-1} + \xi x_j \tag{32}$$

$$x = N_A x_{j-1} + N_B x_j \tag{33}$$

Therefore, the mapping can be defined by the shape functions which were used for building the approximation, and the element is referred to as isoparametric

The main advantage of the proposed mapping isto express both the approximation and its derivatives in acommon form for all elements.

$$\hat{\Phi}^{(ej)} = N_A.\hat{\Phi}_A + N_B.\hat{\Phi}_B \tag{34}$$

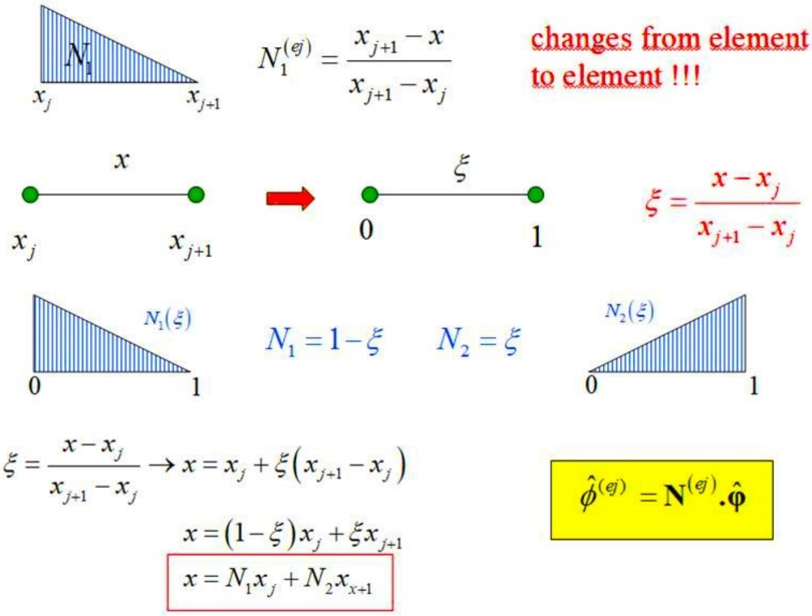

$N_1$
$x_j$
$x_{j+1}$
$N_1^{(ej)} = \frac{x_{j+1}-x}{x_{j+1}-x_j}$
changes from element to element !!!
$x$
$x_j$
$x_{j+1}$
$\xi$
0
1
$\xi = \frac{x-x_j}{x_{j+1}-x_j}$
$N_1(\xi)$
0
1
$N_1 = 1-\xi$
$N_2 = \xi$
$N_2(\xi)$
0
1
$\xi = \frac{x-x_j}{x_{j+1}-x_j} \rightarrow x = x_j + \xi(x_{j+1}-x_j)$
$x = (1-\xi)x_j + \xi x_{j+1}$
$x = N_1 x_j + N_2 x_{x+1}$
$\hat{\phi}^{(ej)} = \mathbf{N}^{(ej)}.\hat{\boldsymbol{\varphi}}$

Figure 7: The concept of isoparametric element.

or, in a more compact manner

$$\hat{\Phi}^{(ej)} = \mathbf{N}^{(e)}.\hat{\mathbf{\Phi}}^{(e)} \tag{35}$$

$$\hat{\Phi}^{(ej)} = \begin{bmatrix} N_A & N_B \end{bmatrix} . \begin{bmatrix} \hat{\Phi}_A \\ \hat{\Phi}_B \end{bmatrix} \tag{36}$$

Derivatives are obtained as follows:

$$\frac{\partial \hat{\Phi}^{(ej)}}{\partial x} = \frac{\partial \hat{\Phi}^{(ej)}}{\partial \xi} . \frac{\partial \xi}{\partial x} = \frac{\partial \hat{\Phi}^{(ej)}}{\partial \xi} . \frac{1}{L^{(e)}} \tag{37}$$

where $L^{(e)}$ is the length of the considered element.

From here, and expresing $\hat{\Phi}^{(ej)}$ using the shape functions,

$$\frac{\partial \hat{\Phi}^{(ej)}}{\partial x} = \frac{1}{L^{(e)}} \begin{bmatrix} \frac{\partial N_A}{\partial \xi} & \frac{\partial N_A}{\partial \xi} \end{bmatrix} . \begin{bmatrix} \hat{\Phi}_A \\ \hat{\Phi}_B \end{bmatrix} \tag{38}$$

$$= \frac{1}{L^{(e)}} \begin{bmatrix} \text{-1} & 1 \end{bmatrix} . \begin{bmatrix} \hat{\Phi}_A \\ \hat{\Phi}_B \end{bmatrix}$$

which is usually written as

$$\frac{\partial \hat{\Phi}^{(ej)}}{\partial x} = \mathbf{B}.\hat{\mathbf{\Phi}}^{(e)} \tag{39}$$

where matrix $\mathbf{B}$ is the discrte form of the derivative operator. If we introduce $\mathbf{S}$

$$\mathbf{S} = \frac{\partial}{\partial x} \tag{40}$$

it is easy to see that

$$\mathbf{B} = \mathbf{S}.\mathbf{N} \tag{41}$$

### 3.1.2 Finite elements in two and three dimensions

There exist today a wide variety of two dimensional elements, from the simplest linear triangles to much more complex elements such as, for instance, the 15 noded triangles which have become popular in geotechnical finite element codes because of their robustness. The choice of element depends, of course on the element availability in the computer code we are using. First finite element programmers favoured simple triangle and quadrilaterals for the simplicity and speed of computations. However, it soon became apparent that such simple elements presented important inconvenients (poor convergence rate, por behaviour in bending, locking in quasi-incompressible materials...). Because of this reason, and also because availability of frontal solvers, during a second period, more complex elements were used. Today, further theoretical developments have allowed to improve the simplest elements. This is the case of the ”enhanced strain” quadrilaterals, with an excellent behaviour in bending dominated situations and when the materials are close to incompressibility, or some pressure-displacement mixed triangles.

In the domain of Computational Fluid Dynamics, the situation is different, as triangles and tetrahedra are the favourite choices in large scale problems with more than $10^6$ degrees of freedom.

The simplest element in 2D is the linear triangle. Fig.8 shows the domain $\Omega$ and its approximation $\Omega^h$ by a mesh of triangular finite elements. It can be seen how the boundary $\partial\Omega$ is also approximated by $\partial\Omega^h$. All the elements can be mapped into the normalized triangle which can be seen in the figure, using the same shape functions of the triangle:

$$\begin{aligned} N_1^{(e)} &= 1 - \xi - \eta \\ N_2^{(e)} &= \xi \\ N_3^{(e)} &= \eta \end{aligned} \tag{42}$$

$$\mathbf{x} = \begin{pmatrix} N_1^{(e)}(\xi,\eta) & N_2^{(e)}(\xi,\eta) & N_3^{(e)}(\xi,\eta) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}_1^{(e)} \\ \mathbf{x}_2^{(e)} \\ \mathbf{x}_3^{(e)} \end{pmatrix} \tag{43}$$

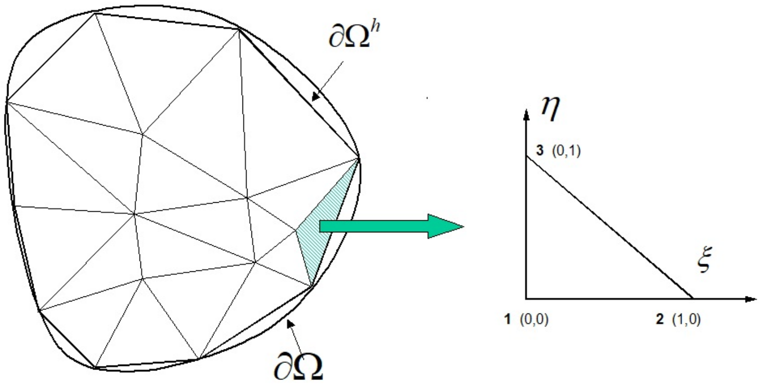


Figure 8: Mesh of linear triangles.

Among triangles of higher order, we can mention the 6 noded quadratic triangle, which is sketched in Fig.9. This element provides a better approximation to curved boundaries, as it approximate them

The bilinear quadrilateral is sketched in Fig.10, and the ”serendiptic” 8 noded quadrilateral can be seen in Fig.11.

Concerning 3D problems, we could mention the linear and quadratic tetrahedra, with 4 and 10 nodes, and the 8 and 20 nodes hexahedral or ”bricks”. All of them are isoparametric elements.

An important point is the evaluation of derivatives inside elements, which are needed to obtain approximations of the gradient operator, fluxes, strain and stress, for instance.

The gradient of the approximation, $\nabla u^{h(e)}$can be obtained substituing $u^{h(e)} = \mathbf{N}^{(e)}.\hat{\mathbf{u}}^{(e)}$, which results on

$$\nabla u^{h(e)} = \nabla \mathbf{N}^{(e)}.\hat{\mathbf{u}}^{(e)} = \mathbf{G}^{(e)}.\hat{\mathbf{u}}^{(e)} \tag{44}$$

where we have introduced the discrete gradient operator $\mathbf{G}^{(e)} = \nabla \mathbf{N}^{(e)}$. In the case of 3D, it is given by:

$$\mathbf{G}^{(e)} = \begin{pmatrix} \partial_x N_1^{(e)} & ... & \partial_x N_{nnode}^{(e)} \\ \partial_y N_1^{(e)} & & \partial_y N_{nnode}^{(e)} \\ \partial_z N_1^{(e)} & & \partial_z N_{nnode}^{(e)} \end{pmatrix} \tag{45}$$

We can follow a similar method to obtain the strain within a particular element. The strain is related to the displacements $\mathbf{u}^{h(e)}$ by $\varepsilon^{h(e)}$= $\mathbf{S}.\mathbf{u}^{h(e)}$ where $\mathbf{S}$ is the strain operator. If we write the displacement field in terms of the shape functions, we arrive at:

$$\varepsilon^{h(e)} = \mathbf{B}^{(e)}.\hat{\mathbf{u}}^{(e)} \tag{46}$$

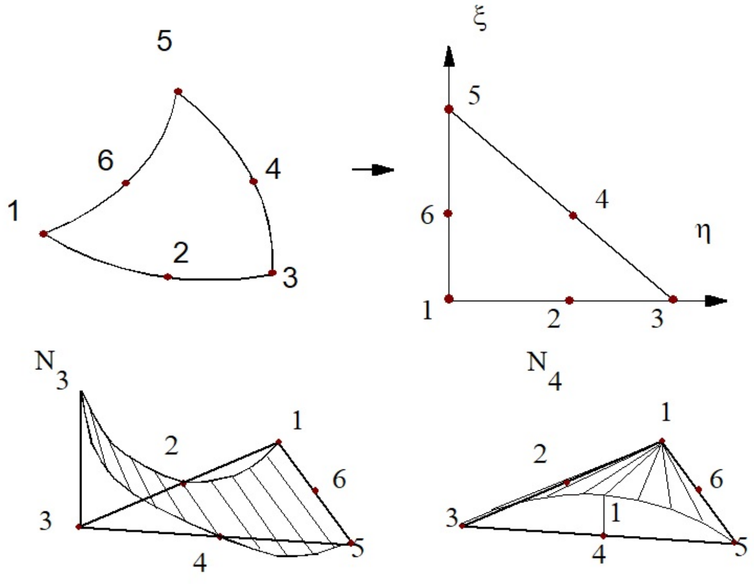


Figure 9: Quadratic triangle.

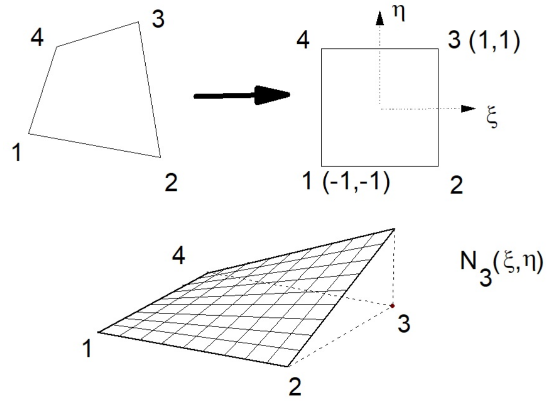


Figure 10: Bilinear Quadrilateral.

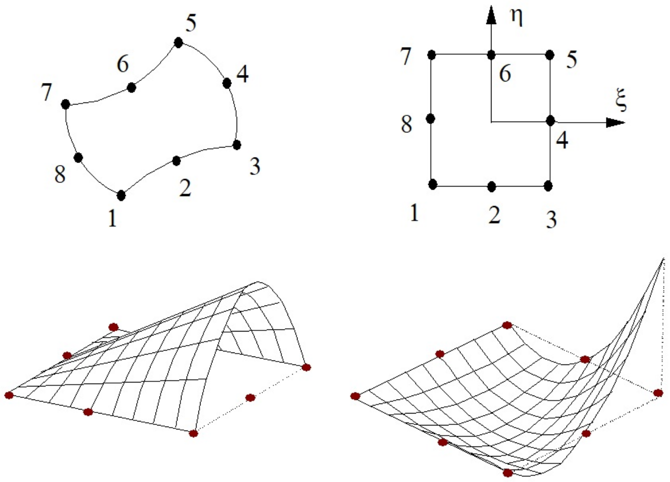


Figure 11: 8 noded quadrilateral.

where $\mathbf{B}^{(e)}$ is the discrete strain operator. In the case of a plane strain problem, it is given by

$$\mathbf{B}^{(e)} = \begin{pmatrix} \partial_x N_1^{(e)} & 0 & \cdots & \partial_x N_{nnode}^{(e)} & 0 \\ 0 & \partial_y N_1^{(e)} & \cdots & 0 & \partial_y N_{nnode}^{(e)} \\ \partial_y N_1^{(e)} & \partial_x N_1^{(e)} & \cdots & \partial_y N_{nnode}^{(e)} & \partial_x N_{nnode}^{(e)} \end{pmatrix} \tag{47}$$

The stress follows immediately as $\sigma^{(e)} = \mathbf{D}^e.\mathbf{B}^{(e)}.\hat{\mathbf{u}}^{(e)}$. It is important to notice that all entities related to derivatives will be discontinuous between elements.

# 4 A note on numerical integration techniques on finite element spaces

Quite often we will have to evaluate integrals over the domain $\Omega^h$ of functions which will involve products of shape functions and their derivatives. Of course, the integral will be decomposed into integral over the elements. In some rare cases the integrals can be easily obtained as it actually happens with linear triangles and tetrahedra. However, in the more general case, we will have to evaluate them using numerical integration techniques.

A numerical integration rule can be expressed as:

$$\int_{\Omega^{(e)}} f(\mathbf{x}) d\Omega = \sum_{k=1}^{ngauss} W_k \, f(\mathbf{x}_k) = \sum_{k=1}^{ngauss} W_k \, f(\xi_k) \det \mathbf{J} \tag{48}$$

where $W_k$ and $\mathbf{x}_k$ (with $k = 1..ngauss$) are the weights and the position of the integration points. If the integral is performed over the normalized element, we need to multiply by the determinant of the jacobian matrix of the mapping $\mathbf{J}$. The number of points in the integration rule $n_{gauss}$ depends on the particular rule which has been chosen. A key point is the order of precision of the formula, which indicates the higher order of the polynomial which can be exactly integrated. Sometimes, the analysts use integration rules of lower degree of precission than required. This is done for several reasons, among which we can mention (i) speed up the computations, (ii) avoid volumetric locking, etc. However, the use of these so-called ”reduced integration” formulae can cause spurious oscillations.

# 5 Method of Galerkin (Boubnov)

## 5.1 General description

When we substitute the solution of the PDE $\Phi(\mathbf{x}), \mathbf{x} \in \Omega \subset \mathbf{R}^2$ by its approximation $\hat{\Phi}(\mathbf{x})$, both the PDE and its boundary conditions will not be satisfied. We will denote residual to the error :

$$R_\Omega = \nabla^T \left(\mathbf{k}\nabla\hat{\Phi}\right) + s \neq 0 \tag{49}$$

in $\Omega$ the residuals in the boundary conditions being

$$R_{\Gamma\Phi} = \hat{\Phi} - \bar{\Phi} \neq 0 \tag{50}$$

$$R_{\Gamma q} = \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \bar{q} \neq 0 \tag{51}$$

where $\mathbf{n}$ is the unit vector normal to the contour. In the case that the material is isotropic $(k_x = k_y = k)$ the previous equation reduces to

$$R_{\Gamma q} = k\frac{\partial\hat{\Phi}}{\partial n} + \bar{q} \neq 0 \tag{52}$$

In the case where there is heat loss by convection in the contour $\Gamma_{h,}$

$$R_{\Gamma h} = \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + h_c\left(\hat{\Phi} - \Phi_\infty\right) \neq 0 \tag{53}$$

or

$$R_{\Gamma h} = k\frac{\partial\hat{\Phi}}{\partial n} + h_c\left(\hat{\Phi} - \Phi_\infty\right) \neq 0 \tag{54}$$

if the material is isotropic. Finally, the residual in the contour where there is heat loss by radiation is:

$$R_{\Gamma r} = \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \sigma\varepsilon\left(\hat{\Phi}^4 - \Phi_\infty^4\right) \neq 0 \tag{55}$$

where $\hat{\Phi} = \mathbf{N}.\hat{\mathbf{\Phi}}$, or $\hat{\Phi} = \sum_{j=0}^{N} N_j.\hat{\Phi}_j$.

Galerkin's method basically consists of determining the $N + 1$ unknowns$\hat{\Phi}_j$, $j = 1..N$, imposing the $N + 1$ conditions:

$$\int_\Omega N_i R_\Omega d\Omega + \int_{\Gamma_\Phi} \alpha_\phi N_i R_{\Gamma_\Phi} d\Gamma + \int_{\Gamma_q} \alpha_q N_i R_{\Gamma_q} d\Gamma + ... = 0 \tag{56}$$

$$\int_\Omega N_i \left[\nabla^T \left(\mathbf{k}\nabla\hat{\Phi}\right) + s\right] d\Omega + \int_{\Gamma_\Phi} \alpha_\phi N_i \left(\hat{\Phi} - \bar{\Phi}\right) d\Gamma \tag{57}$$

$$+ \int_{\Gamma_q} \alpha_q N_i \left(\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \bar{q}\right) d\Gamma + ... = 0$$

where $\alpha_q$, etc., are parameters whose value will be obtained later so that the expressions obtained are as simple as possible.

Applying Green's Theorem, we obtain

$$-\int_\Omega (\nabla N_i)^T \mathbf{k}\nabla\hat{\Phi} d\Omega + \int_{\Gamma_\Phi+\Gamma_q} N_i.\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} d\Gamma + \int_\Omega N_i s d\Omega + ... \tag{58}$$

$$+ \int_{\Gamma_\Phi} \alpha_\phi N_i \left(\hat{\Phi} - \bar{\Phi}\right) d\Gamma + \int_{\Gamma_q} \alpha_q N_i \left(\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \bar{q}\right) d\Gamma = 0 \tag{59}$$

where the first term is

$$-\int_\Omega \begin{pmatrix} \frac{\partial N_i}{\partial x} & \frac{\partial N_i}{\partial y} \end{pmatrix} \mathbf{k} \begin{pmatrix} \frac{\partial \hat{\Phi}}{\partial x} \\ \frac{\partial \hat{\Phi}}{\partial y} \end{pmatrix} d\Omega \tag{60}$$

Now considering only the terms corresponding to the integrals on the contour, we obtain

$$\int_{\Gamma_\Phi+\Gamma_q} \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} d\Gamma + \int_{\Gamma_\Phi} N_i \left(\hat{\Phi} - \bar{\Phi}\right) d\Gamma + \int_{\Gamma_q} \alpha_q N_i \left(\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \bar{q}\right) d\Gamma + .... = 0 \tag{61}$$

$$\int_{\Gamma_\Phi} N_i \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} d\Gamma + \int_{\Gamma_\Phi} N_i \left(\hat{\Phi} - \bar{\Phi}\right) d\Gamma + \int_{\Gamma_q} \left[N_i \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \alpha_q N_i \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} + \alpha_q \bar{q}\right] d\Gamma \tag{62}$$

where it can be seen that if we choose $\alpha_q = -1$, it results

$$\int_{\Gamma_\Phi} N_i \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} d\Gamma - \int_{\Gamma_q} N_i \bar{q} d\Gamma \tag{63}$$

having assumed that in the contour $\Gamma_\Phi$ $\hat{\Phi} - \bar{\Phi} = 0$ is approximately fulfilled. This implies that the number of unknowns is reduced by $N_\Phi\}$, since the values of $\Phi$ are then known. It should be noted that the terms $\int_{\Gamma_\Phi} N_i\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi}d\Gamma$ or $\int_{\Gamma_\Phi} N_i k\frac{\partial\hat{\Phi}}{\partial n}d\Gamma$ are not known in $\Gamma_\Phi$, , being, therefore, additional unknowns, so that the system that is finally solved has as unknowns to be determined the $N - N_\Phi$ values of $\hat{\mathbf{\Phi}}$, as well as the $N_\Phi$ values of the flows -$k\frac{\partial\hat{\Phi}}{\partial n}$at $\Gamma_\Phi$

These are, therefore, additional unknowns, so that the system that is finally solved has as unknowns to be determined the $N - N_\Phi$values of $\hat{\mathbf{\Phi}}$, as well as the $N_q$ values of the flows -$k\frac{\partial\hat{\Phi}}{\partial n}$

The resulting system is, then,

$$-\int_\Omega (\nabla N_i)^T\mathbf{k}\nabla\hat{\Phi}d\Omega + \int_\Omega N_i s d\Omega + \int_{\Gamma_\Phi} N_i\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi}d\Gamma - \int_{\Gamma_q} N_i\bar{q}d\Gamma = 0 \quad (64)$$

or,

$$\int_\Omega (\nabla N_i)^T\mathbf{k}\nabla\hat{\Phi}d\Omega = \int_\Omega N_i s d\Omega + \int_{\Gamma_\Phi} N_i k\frac{\partial\hat{\Phi}}{\partial n}d\Gamma - \int_{\Gamma_q} N_i\bar{q}d\Gamma \quad (65)$$

Substituting below the value of $\hat{\Phi}$ by $\hat{\Phi} = \sum_{j=0}^N N_j.\hat{\Phi}_j$, we obtain

$$\left(\int_\Omega\int_\Omega (\nabla N_i)^T\mathbf{k}\nabla N_j d\Omega\right)\hat{\Phi}_j = -\int_\Omega N_i s d\Omega + \int_{\Gamma_\Phi} N_i k\frac{\partial\hat{\Phi}}{\partial n}d\Gamma - \int_{\Gamma_q} N_i\bar{q}d\Gamma \quad (66)$$

or, in a more compact manner,

$$K_{ij}\hat{\Phi}_j = f_i \quad (67)$$

$$\mathbf{K}\hat{\mathbf{\Phi}} = \mathbf{f} \quad (68)$$

where $K_{ij}$ are the terms $(i,j)$of the coefficient matrix of a system of linear equations, and $f_i$ are the independent terms, given, respectively, by

$$K_{ij} = \int_\Omega\int_\Omega (\nabla N_i)^T\mathbf{k}(\nabla N_j)d\Omega \quad (69)$$

$$f_i = \int_\Omega N_i s d\Omega + \int_{\Gamma_\Phi} N_i\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi}d\Gamma - \int_{\Gamma_q} N_i\bar{q}d\Gamma \quad (70)$$

Taking into account that the terms $\mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi}$ are flows with a changed sign, the above equation can be written as

$$f_i = \int_\Omega N_i s d\Omega - \int_{\Gamma_\Phi} N_i\hat{q}d\Gamma - \int_{\Gamma_q} N_i\bar{q}d\Gamma \quad (71)$$

where $\hat{q}$ is the value of the normal flow (positive according to the normal) in the contour $\Gamma_\Phi$. The terms coming from the unknown flow in the boundary are usually expressed as

$$R_i = \int_{\Gamma_\Phi} N_i \hat{q} d\Gamma \tag{72}$$

The matrix **K** is called in some texts ”Stiffness Matrix”, and the vector of independent terms, ”Force Vector”, terms taken from applications to the structural analysis.

## 5.2 One dimensional problems

In the case of one-dimensional problems, the equations obtained are considerably simplified. The residues in the domain and in the boundary are given, respectively, by:

$$R_\Omega = \frac{d}{dx}\left(k\frac{d\hat{\Phi}}{dx}\right) + s \neq 0 \tag{73}$$

$$R_{\Gamma\Phi} = \hat{\Phi} - \bar{\Phi} \neq 0 \tag{74}$$

$$R_{\Gamma q} = k\frac{d\hat{\Phi}}{dn} + \bar{q} \neq 0 \tag{75}$$

$$R_{\Gamma h} = k\frac{d\hat{\Phi}}{dn} + h_c\left(\hat{\Phi} - \Phi_\infty\right) \neq 0 \tag{76}$$

$$R_{\Gamma r} = k\frac{d\hat{\Phi}}{dn} + \sigma\varepsilon\left(\hat{\Phi}^4 - \Phi_\infty^4\right) \neq 0 \tag{77}$$

where, $\hat{\Phi} = \mathbf{N}.\hat{\mathbf{\Phi}}$, or $\hat{\Phi} = \sum_{j=0}^{N} N_j.\hat{\Phi}_j$.

The $N+1$ equations obtained by the Galerkin method allow us to obtain the corresponding unknowns $\hat{\Phi}_j$, $j = 1..N$, by imposing the $N+1$ conditions:

$$\int_\Omega N_i R_\Omega d\Omega + \int_{\Gamma_\Phi} \alpha_\phi N_i R_{\Gamma_\Phi} d\Gamma + \int_{\Gamma_q} \alpha_q N_i R_{\Gamma_q} d\Gamma + ... = 0 \tag{78}$$

$$\int_\Omega \left[N_i \frac{d}{dx}\left(k\frac{\partial\hat{\Phi}}{\partial x}\right) + s\right] d\Omega + \int_{\Gamma_\Phi} \alpha_\phi N_i \left(\hat{\Phi} - \bar{\Phi}\right) d\Gamma \tag{79}$$

$$+\int_{\Gamma_q}\alpha_q N_i\left(D\frac{\partial\hat{\Phi}}{\partial n}+\bar{q}\right)d\Gamma+...=0$$

being in this case the contour integrals

$$\int_{\Gamma_\Phi}\alpha_\phi N_i\left(\hat{\Phi}-\bar{\Phi}\right)d\Gamma=\alpha_\phi N_i\left(\hat{\Phi}-\bar{\Phi}\right)A\mid_{\Gamma_\Phi} \tag{80}$$

$$\int_{\Gamma_q}\alpha_q N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)d\Gamma=\alpha_q N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A\mid_{\Gamma_q} \tag{81}$$

where $A$ is the section of the one-dimensional region considered, $\alpha_q$, etc., parameters whose value will be obtained later so that the expressions obtained are as simple as possible.

Integrating by parts, we obtain

$$-\int_\Omega\frac{dN_i}{dx}k\frac{d\hat{\Phi}}{dx}d\Omega+\left[N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A\right]_{\Gamma_\Phi+\Gamma_q}+\int_\Omega N_i s d\Omega+... \tag{82}$$

$$+\left[\alpha_\phi N_i\left(\hat{\Phi}-\bar{\Phi}\right)A\right]_{\Gamma_\Phi}+\left[\alpha_q N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A\right]_{\Gamma_q}=0$$

Now considering only the terms corresponding to the integrals in the contour, we obtain

$$\left[N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A\right]_{\Gamma_\Phi+\Gamma_q}+\left[\alpha_\phi N_i\left(\hat{\Phi}-\bar{\Phi}\right)A\right]_{\Gamma_\Phi} \tag{83}$$

$$+\left[\alpha_q N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A\right]_{\Gamma_q}$$

$$\left[\alpha_\phi N_i\left(\hat{\Phi}-\bar{\Phi}\right)A\right]_{\Gamma_\Phi}+\left[N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A+\alpha_q N_i\left(k\frac{d\hat{\Phi}}{dn}+\bar{q}\right)A\right]_{\Gamma_q} \tag{84}$$

where it can be seen that if we choose $\alpha_q = -1$ it results

$$\left[N_i k \frac{d\hat{\Phi}}{dn} A\right]_{\Gamma_\Phi} - \left[N_i \bar{q} A\right]_{\Gamma_q} \tag{85}$$

having assumed that in the contour $\Gamma_\Phi$ the condition $\hat{\Phi} - \bar{\Phi} = 0.$ is approximately satisfied.

This implies that the number of unknowns is reduced by $N_\Phi$, since the values of $\Phi$ are then known.

It should be noted that the terms $\left[N_i k \frac{d\hat{\Phi}}{dn} A\right]_{\Gamma_\Phi}$ are not known $\Gamma_\Phi$, thus dealing with $N_\Phi$ additional unknowns, so that the system that is finally solved has $N+1$ unknowns. In summary, the unknowns to be determined are the $N+1-N_\Phi$ values of $\hat{\mathbf{\Phi}}$, as well as the $N_\Phi$ values of the flows -$D\frac{\partial\hat{\Phi}}{\partial n}$

The resulting system is, then,

$$-\int_\Omega \frac{dN_i}{dx} k \frac{d\hat{\Phi}}{dx} d\Omega + \int_\Omega N_i s d\Omega + \left[N_i k \frac{d\hat{\Phi}}{dn} A\right]_{\Gamma_\Phi} - \left[N_i \bar{q} A\right]_{\Gamma_q} = 0 \tag{86}$$

or,

$$\int_\Omega \frac{dN_i}{dx} k \frac{d\hat{\Phi}}{dx} d\Omega = -\int_\Omega N_i s d\Omega + \left[N_i k \frac{d\hat{\Phi}}{dn} A\right]_{\Gamma_\Phi} - \left[N_i \bar{q} A\right]_{\Gamma_q} = 0 \tag{87}$$

If we substitute next the value of $\hat{\Phi}$ by $\hat{\Phi} = \sum_{j=0}^{N} N_j.\hat{\Phi}_j$, we arrive to

$$\left(\int_\Omega \frac{dN_i}{dx} k \frac{dN_j}{dx} d\Omega\right)\hat{\Phi}_j = -\int_\Omega N_i s d\Omega + \left[N_i k \frac{d\hat{\Phi}}{dn} A\right]_{\Gamma_\Phi} - \left[N_i \bar{q} A\right]_{\Gamma_q} \tag{88}$$

or, in a more compact form

$$K_{ij}\hat{\Phi}_j = f_i \tag{89}$$

$$\mathbf{K}\hat{\mathbf{\Phi}} = \mathbf{f} \tag{90}$$

where $K_{ij}$are the terms$(i,j)$ of the coefficient matrix of a system of linear equations, and $f_i$ the independent terms, given, respectively, by

$$K_{ij} = \int_\Omega \frac{dN_i}{dx} k \frac{dN_j}{dx} d\Omega = \int_0^L \frac{dN_i}{dx} k \frac{dN_j}{dx} A dx \tag{91}$$

$$f_i = -\int_0^L N_i s A dx + \left[ N_i k \frac{d\hat{\Phi}}{dn} A \right]_{\Gamma_\Phi} - [N_i \bar{q} A]_{\Gamma_q} \tag{92}$$

Taking into account that the terms are flows with changed signs, the previous equation can be written as

$$f_i = -\int_0^L N_i s A dx - [N_i \hat{q} A]_{\Gamma_\Phi} - [N_i \bar{q} A]_{\Gamma_q} \tag{93}$$

where $\hat{q}$ is the value of the normal flow (positive according to the normal) in the contour $\Gamma_\Phi$.

## 5.3 Examples

We will consider next some simple examples in order to fix the ideas.

**Example 1**

Solve the problem of heat transfer in a bar of length $L$ and section $A$ with the boundary conditions $\hat{\Phi} = 0$ at $x = 0$, and $q = \widehat{q}$ at $x = L$.

**Solution**

The coefficients $K_{ij}$ are obtained as:

$$K_{11} = \int_0^L \frac{dN_1}{dx} k \frac{dN_1}{dx} A dx \tag{94}$$

$$K_{12} = K_{21} = \int_0^L \frac{dN_1}{dx} k \frac{dN_2}{dx} A.dx \tag{95}$$

$$K_{22} = \int_0^L \frac{dN_2}{dx} k \frac{dN_2}{dx} A dx \tag{96}$$

from where

$$\mathbf{K} = \int_0^L \left\{ \begin{array}{cc} \frac{dN_1}{dx} k \frac{dN_1}{dx} & \frac{dN_1}{dx} k \frac{dN_2}{dx} \\ \frac{\partial N_1}{\partial x} D \frac{\partial N_2}{\partial x} & \frac{\partial N_2}{\partial x} k \frac{dN_2}{dx} \end{array} \right\} A dx \tag{97}$$

$$= \int_0^L \left\{ \begin{array}{c} \frac{dN_1}{dx} \\ \frac{dN_2}{dx} \end{array} \right\} k \left\{ \begin{array}{cc} \frac{dN_1}{dx} & \frac{dN_2}{dx} \end{array} \right\} A dx \tag{98}$$

$$= \int_0^L \mathbf{B}^T . D . \mathbf{B} A dx \tag{99}$$

Taking into account that

$$\mathbf{B} = \left\{ \begin{array}{cc} \text{-}\frac{1}{L} & \frac{1}{L} \end{array} \right\} \tag{100}$$

we obtain

$$\mathbf{K} = \int_0^L \left\{ \begin{array}{cc} \frac{k}{L^2} & \text{-}\frac{k}{L^2} \\ \text{-}\frac{k}{L^2} & \frac{k}{L^2} \end{array} \right\} A dx = \frac{kA}{L} \left[ \begin{array}{cc} 1 & \text{-1} \\ \text{-1} & 1 \end{array} \right] \tag{101}$$

On the other hand, the vector of independent terms is

$$f_i = -\int_0^L N_i s A dx - [N_i \hat{q} A]_{\Gamma_\Phi} - [N_i \bar{q} A]_{\Gamma_q} \tag{102}$$

being in this particular case

$$\mathbf{f} = \left( \begin{array}{c} -\hat{q}_1 A \\ \bar{q} A \end{array} \right) \tag{103}$$

The system is, therefore,

$$\frac{kA}{L} \left[ \begin{array}{cc} 1 & \text{-1} \\ \text{-1} & 1 \end{array} \right] \cdot \left[ \begin{array}{c} \hat{\Phi}_1 = 0 \\ \hat{\Phi}_2 \end{array} \right] = \left( \begin{array}{c} \text{-}\hat{q}_1 A \\ \text{-}\bar{q} A \end{array} \right) \tag{104}$$

the unknowns being $\hat{\Phi}_2$ and $\hat{q}_1$.

Starting with the second equation, we obtain

$$\frac{kA}{L} \hat{\Phi}_2 = -\bar{q} A \tag{105}$$

from where,

$$\hat{\Phi}_2 = -\frac{\bar{q} L}{k} \tag{106}$$

The flux at $x = 0$ is obtained using the first equation:

$$\frac{kA}{L} \left( \frac{\bar{q} L}{k} \right) = -\hat{q}_1 A \tag{107}$$

and

$$\hat{q}_1 = -\bar{q} \tag{108}$$

It is important to remember that the flow obtained at node 0 is

$$\hat{q}_1 = -k \frac{\partial \Phi}{\partial n} = k \frac{\partial \Phi}{\partial x} \tag{109}$$

so it is correct that it is equal to $-\bar{q}$.

**Example 2**

Solve the problem of heat transfer in a bar of length $L$ and section $A$ with the boundary conditions $\hat{\Phi} = 0$ at $x = 0$, and $\widehat{q} = 0$ at $x = L$ , existing source terms of constant value $s$.

**Solution**

The equations are analogous to those obtained in the previous example, with the difference that the contribution of the sources will now have to be added to the vector of independent terms. $s$ :

$$\int_{\Omega} N_i s d\Omega = \int_0^L N_i s A dx \tag{110}$$

which, taking into account that $s$ is constant results on

$$f_i = \frac{1}{2} sAL \tag{111}$$

From here, we arrive to

$$\frac{kA}{L} \begin{bmatrix} 1 & \text{-}1 \\ \text{-}1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \hat{\Phi}_1 = 0 \\ \hat{\Phi}_2 \end{bmatrix} = \begin{pmatrix} \text{-}\hat{q}_1 A + \frac{1}{2} sAL \\ 0 + \frac{1}{2} sAL \end{pmatrix} \tag{112}$$

which solution is

$$\hat{\Phi}_2 = \frac{sL^2}{2k} \tag{113}$$

being the heat flow in the bar

$$\mathbf{q} = \mathbf{B}.\hat{\mathbf{\Phi}} = \left(-\frac{1}{L}, \frac{1}{L}\right) \begin{pmatrix} \hat{\Phi}_1 = 0 \\ \hat{\Phi}_2 \end{pmatrix} = \frac{1}{2} s.L \tag{114}$$

Comparing the solution obtained with the analytical

$$\phi(x) = \frac{sL}{2} x - \frac{s}{2k} x^2 \tag{115}$$

where the flux is

$$q(x) = -k \frac{d\phi}{dx} = s(L - x) \tag{116}$$

It is observed that the values obtained at the nodes are exact, as well as the value of the heat flow at the midpoint of the element.

**Example 3**

Obtain the coefficient matrix for a linear triangular element with vertices $A(0,0)$, $B(1,0)$and $C(0,1)$ in the case that the conductivities according to x and y are $k$.

**Solution**

The coefficients $K_{ij}$ are:

$$\int_\Omega \begin{pmatrix} \frac{\partial N_i}{\partial x} & \frac{\partial N_i}{\partial y} \end{pmatrix} \mathbf{k} \begin{pmatrix} \frac{\partial N_j}{\partial x} \\ \frac{\partial N_j}{\partial y} \end{pmatrix} d\Omega \tag{117}$$

where $\mathbf{k}$ $=k\mathbf{I}$, being $\mathbf{I}$ the identity matrix:

$$K_{ij} = \int_\Omega k \left( \frac{\partial N_i}{\partial x} \frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y} \frac{\partial N_j}{\partial y} \right) d\Omega \tag{118}$$

If we introduce now the matrix $\mathbf{B}$

$$\mathbf{B} = \begin{pmatrix} \frac{\partial N_1}{\partial x} & \frac{\partial N_2}{\partial x} & \frac{\partial N_3}{\partial x} \\ \frac{\partial N_1}{\partial y} & \frac{\partial N_2}{\partial y} & \frac{\partial N_3}{\partial y} \end{pmatrix} \tag{119}$$

the coefficient matrix $\mathbf{K}$ can be expressed as

$$\mathbf{K} = \int \mathbf{B}^T \mathbf{k} \mathbf{B} d\Omega \tag{120}$$

or, defining $B_i$ as:

$$\mathbf{B}_i = \begin{pmatrix} \frac{\partial N_i}{\partial x} \\ \frac{\partial N_i}{\partial y} \end{pmatrix} \tag{121}$$

it results on

$$K_{ij} = \int_\Omega \mathbf{B}_i^T \mathbf{k} \mathbf{B}_j d\Omega \tag{122}$$

In the case of the element considered, which coincides with the normalized element, we have:

$$\mathbf{B}_1 \tag{123}$$

$$\mathbf{B}_2 \tag{124}$$

$$\mathbf{B}_3 \tag{125}$$

from where:

$$\mathbf{B} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \tag{126}$$

and

$$\mathbf{K} = \int_{\Omega} \mathbf{B}^T k \mathbf{B} d\Omega = \int_{\Omega} \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} k \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} d\Omega \tag{127}$$

$$= kA \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \tag{128}$$

with $A = \frac{1}{2}$

**Example 4**

Obtain the vector of independent terms of a linear triangle with nodes $A(0,0)$, $B(1,0)$ and $C(0,1)$ in the case that there is a constant source term $s$ in the element:

**Solution**

The contribution of the source term to the vector of element-independent terms is:

$$f_i = \int_{\Omega} N_i s d\Omega \tag{129}$$

or

$$\mathbf{f} = \int_{\Omega} \mathbf{N}^T s d\Omega \tag{130}$$

with

$$\mathbf{N}^T = \begin{pmatrix} N_A \\ N_B \\ N_C \end{pmatrix} \tag{131}$$

This integral is calculated taking into account that the volume of a tetrahedron of area $A$ and unit height is $(1/3)A$, so it results:

$$\mathbf{f} = \frac{1}{3} As \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{132}$$

**Example 5**

Given a triangular-shaped plate whose vertices are the points $1(0,0)$, $2(1,0)$ and $3(0,1)$ whose edge $12$ is maintained at a temperature $\Phi_0$, the other two edges being isolated, obtain the temperature at node 3, when there is a heat source of intensity $s$ on the plate and the thermal conductivity is $k$.

**Solution**

Taking into account the coefficient matrix and the independent term obtained in the previous examples, the system to be solved is:

$$kA\begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} \hat{\Phi}_1 = \Phi_0 \\ \hat{\Phi}_2 = \Phi_0 \\ \hat{\Phi}_3 \end{pmatrix} = \frac{1}{3}As\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} R_1 \\ R_2 \\ 0 \end{pmatrix} \tag{133}$$

where the terms $R_1$ and $R_2$ originated by the integral on the contour $\Gamma_\Phi$ have been included

$$R_1 = \int_\Omega N_1 k \frac{\partial \hat{\Phi}}{\partial n} d\Gamma \tag{134}$$

The system solution is immediate, obtaining:

$$\hat{\Phi}_3 = \Phi_0 + \frac{s}{3k} \tag{135}$$

# 6 Assembling of elements

In the previous examples, problems that have been discretized using a single element have been solved. In general, more elements will always be used, obtaining the coefficient matrix as the sum of those obtained by integration in each of the elements.

The contribution of an element to the global coefficient matrix will only be non-zero for those pairs of values $(i, j)$ in which both nodes belong to the element. Therefore, instead of adding $ne$ matrices of dimensions $(N+1)$x$(N+1)$, a simpler procedure called ”assembly” is used, which is described below.

## 6.1 One dimensional problems

The coefficients of the matrix $\mathbf{K}$ are given by

$$K_{ij} = \int_0^L \frac{dN_i}{dx} k \frac{dN_j}{dx} A dx \tag{136}$$

The integral in the domain $\Omega$ can be obtained as the sum of the integrals in the elements,

$$K_{ij} = \int_\Omega \frac{dN_i}{dx} k \frac{dN_j}{dx} A dx = \sum_{e=1}^{Nelem} \int_{\Omega^e} \frac{dN_i}{dx} k \frac{dN_j}{dx} A dx = \sum_{e=1}^{Nelem} K_{ij}^e \tag{137}$$

where the superindex $(e)$ refers to the element considered. On the other hand, the functions $N_i$ are the global shape functions defined above.

Considering the element $(j)$ of nodes $j-1$ and $j$, it can be easily verified that only the shape functions corresponding to the nodes of the element will be different from zero in $\Omega^e$, and therefore, they will only be different from zero the terms $(j-1, j-1), (j-1, j), (j, j-1)$ and $(j, j)$. The contribution of this element to the stiffness matrix will be therefore:

$$K_{j-1,j-1}^{(j)} = \int_{\Omega} \frac{dN_{j-1}}{dx} k \frac{dN_{j-1}}{dx} A dx = \int_{\Omega^e} \frac{dN_{j-1}}{dx} k \frac{dN_{j-1}}{dx} A dx \tag{138}$$

which, taking into account the nomenclature introduced for the functions used in the elements, can be written as

$$K_{j-1,j-1}^{(j)} = \int_{\Omega^e} \frac{dN_A}{dx} k \frac{dN_A}{dx} A dx \tag{139}$$

In a similar way, we will obtain

$$K_{j-1,j}^{(j)} = K_{j,j-1} = \int_{\Omega^e} \frac{dN_A}{dx} k \frac{dN_B}{dx} A dx \tag{140}$$

and

$$K_{j,j}^{(j)} = \int_{\Omega^e} \frac{dN_B}{dx} k \frac{dN_B}{dx} A dx \tag{141}$$

The contribution to the global matrix of the element considered would therefore be:

| nodes | 0 1 ... ... N |
|---|---|
| 0<br>1<br>...<br>...<br>N | $\begin{array}{\|cc\|} \hline K_{j-1,j-1}^{(j)} & K_{j-1,j}^{(j)} \\ K_{j,j-1}^{(j)} & K_{j,j}^{(j)} \\ \hline \end{array}$ |

or

| nodes | 0 1 ... ... N |
|---|---|
| 0<br>1<br>...<br>...<br>N | $\begin{array}{\|cc\|} \hline K_{A,A}^{(j)} & K_{A,B}^{(j)} \\ K_{B,A}^{(e)} & K_{B,B}^{(j)} \\ \hline \end{array}$ |

Under these conditions, when obtaining the contribution of each element, it is simpler to use a "matrix of element coefficients" $\mathbf{K}^{(e)}$

$$\mathbf{K}^{(e)} = \left[ \begin{array}{cc} K^{(e)}_{A,A} & K^{(e)}_{A,B} \\ K^{(e)}_{B,A} & K^{(e)}_{B,B} \end{array} \right] = \int_0^L \left\{ \begin{array}{cc} \frac{\partial N_1}{\partial x} D \frac{\partial N_1}{\partial x} & \frac{\partial N_1}{\partial x} D \frac{\partial N_2}{\partial x} \\ \frac{\partial N_1}{\partial x} D \frac{\partial N_2}{\partial x} & \frac{\partial N_2}{\partial x} D \frac{\partial N_2}{\partial x} \end{array} \right\} A dx$$

$$= \int_{\Omega^{(e)}} \left\{ \begin{array}{c} \frac{dN_A}{dx} \\ \frac{dN_B}{dx} \end{array} \right\} k \left\{ \begin{array}{cc} \frac{dN_A}{dx} & \frac{dN_B}{dx} \end{array} \right\} A dx \tag{142}$$

$$= \int \mathbf{B}^{(e)T}.k.\mathbf{B}^{(e)} A dx$$

where

$$\mathbf{B}^{(e)} = \left\{ \begin{array}{cc} -\frac{1}{L^{(e)}} & \frac{1}{L^{(e)}} \end{array} \right\} \tag{143}$$

Once obtained, their coefficients are placed in the corresponding positions of the global matrix. For this, the information of the nodes that each element has is used,

| Local | Global |
|---|---|
| A | j-1 |
| B | j |

(144)

the coefficient $(A, B)$ of the element matrix corresponds to the global $(j-1, j)$. This operation of distributing the terms is called "Assembly", and is usually represented in an abbreviated form as

$$\mathbf{K} = \bigcup_{e=1}^{N} \mathbf{K}^{(e)} \tag{145}$$

**Example 7**

Solve the problem of heat transfer in a bar of length $L$ whose left end is maintained at a temperature of zero degrees, with the flow being prescribed at the right end, using two elements of length $L/2$

**Solution**

The coefficient matrix of the first element is

$$\mathbf{K}^{(e)} = \left[ \begin{array}{cc} K^{(e)}_{A,A} & K^{(e)}_{A,B} \\ K^{(e)}_{B,A} & K^{(e)}_{B,B} \end{array} \right] = \int_{\Omega_1} \mathbf{B}^{(1)T}.k.\mathbf{B}^{(1)} A dx \tag{146}$$

where

$$\mathbf{B}^{(1)} = \left\{ \begin{array}{cc} -\frac{1}{L^{(1)}} & \frac{1}{L^{(1)}} \end{array} \right\} \tag{147}$$

and

$$L^{(1)} = L/2 \tag{148}$$

From here, we obtain

$$\mathbf{K}^{(1)}{=}\int \left\{ \begin{array}{cc} \frac{D}{L^{(1)2}} & \text{-}\frac{D}{L^{(1)2}} \\ \text{-}\frac{D}{L^{(1)2}} & \frac{D}{L^{(1)2}} \end{array} \right\} Adx = \frac{DA}{L/2} \left[ \begin{array}{cc} 1 & \text{-1} \\ \text{-1} & 1 \end{array} \right] \tag{149}$$

The coefficient matrix of the second element is obtained in a similar way, reaching the same result, since the lengths of the elements are equal.

The global array is formed by assembling the arrays of each element,

$$\mathbf{K} = \begin{pmatrix} K_{A,A}^{(1)} & K_{A,B}^{(1)} & 0 \\ K_{B,A}^{(1)} & K_{B,B}^{(1)} + K_{A,A}^{(2)} & K_{A,B}^{(2)} \\ 0 & K_{B,A}^{(2)} & K_{B,B}^{(2)} \end{pmatrix} \tag{150}$$

$$= \frac{kA}{L/2} \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \tag{151}$$

The system to be solved is, therefore,

$$\frac{kA}{L/2} \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\Phi}_1 = 0 \\ \hat{\Phi}_2 \\ \hat{\Phi}_3 \end{pmatrix} = \begin{pmatrix} -\hat{q}_1 A \\ 0 \\ -\bar{q} A \end{pmatrix} \tag{152}$$

Using the second and third equations we obtain the system

$$2\hat{\Phi}_2 - \hat{\Phi}_3 = 0 \tag{153}$$

$$-\hat{\Phi}_2 + \hat{\Phi}_3 = -\frac{qL}{2k} \tag{154}$$

which solution is

$$\hat{\Phi}_2 = -\frac{qL}{2k} \qquad \hat{\Phi}_3 = -\frac{qL}{k} \tag{155}$$

obtaining, from the first equation, $\hat{q}_1 = -\bar{q}$.

## 6.2 General case

In the general case, the coefficients of the matrix $\mathbf{K}$ are given by

$$K_{ij} = \int_\Omega \mathbf{B}_i^T \mathbf{k} \mathbf{B}_j d\Omega \tag{156}$$

integrals in the elements,

$$K_{ij} = \int_\Omega \mathbf{B}_i^T \mathbf{k} \mathbf{B}_j d\Omega = \sum_{e=1}^{Nelem} \int_{\Omega^e} \mathbf{B}_i^T \mathbf{k} \mathbf{B}_j d\Omega = \sum_{e=1}^{Nelem} K_{ij}^e \tag{157}$$

where the superindex $(e)$ refers to the element considered. On the other hand, the functions $N_i$ are the global shape functions defined above.

Considering the element $(e)$ of nodes $i, j$ and $k$, it can be easily verified that only the shape functions corresponding to the nodes of the element will be different from zero in $\Omega^e$ , and therefore, only they will be different from zero the terms $(i, i)$, $(i, j)$, $(i, k)$, $(j, j)$,$(j, k)$, $(k, k)$ and their symmetrical terms. The contribution of this element to the stiffness matrix will therefore be:

$$K_{\alpha\beta}^{(e)} = \textstyle\int_\Omega \mathbf{B}_\alpha^T \mathbf{k} \mathbf{B}_\beta d\Omega \qquad (\alpha, \beta) \in \{i, j, k\} \tag{158}$$

The contribution to the global matrix of the element considered would therefore be:

| nodes | 1 | ... | $i$ | ... | $j$ | ... | $k$ | ... | $N$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | |
| ... | | | | | | | | | |
| $i$ | | | $K_{ii}^{(e)}$ | | $K_{ij}^{(e)}$ | | $K_{ik}^{(e)}$ | | |
| ... | | | | | | | | | |
| $j$ | | | $K_{ji}^{(e)}$ | | $K_{jj}^{(e)}$ | | $K_{jk}^{(e)}$ | | |
| ... | | | | | | | | | |
| $k$ | | | $K_{ki}^{(e)}$ | | $K_{kj}^{(e)}$ | | $K_{kk}^{(e)}$ | | |
| ... | | | | | | | | | |
| $N$ | | | | | | | | | |

all other coefficients being null

Under these conditions, when obtaining the contribution of each element, it is simpler to use an ”element coefficient matrix” $\mathbf{K}^{(e)}$

$$\mathbf{K}^{(e)} = \int_{\Omega^{(e)}} \mathbf{B}^T \mathbf{k} \mathbf{B} d\Omega \tag{159}$$

Once obtained, their coefficients are placed in the corresponding positions of the global matrix. For this, the information of the nodes that each element has is used,

| Local | Global |
|---|---|
| 1 | i |
| 2 | j |
| 3 | k |

(160)

corresponding, for example, the coefficient $(2, 3)$ of the element matrix to the global $(j, k)$. As explained before, this operation of distributing the terms is called ”Assembly”, and is usually represented in an abbreviated form as

$$\mathbf{K} = \overset{N}{\underset{e=1}{\cup}} \mathbf{K}^{(e)} \tag{161}$$

- **Example 9**

A square plate with a unit side has its left edge at a temperature of 0°, the two vertical edges being insulated, and there being a unit flow of heat on the right edge, directed towards the interior of the plate. Obtain the temperature distribution using two three-node triangular elements, assuming that the thermal conductivity is $k$.

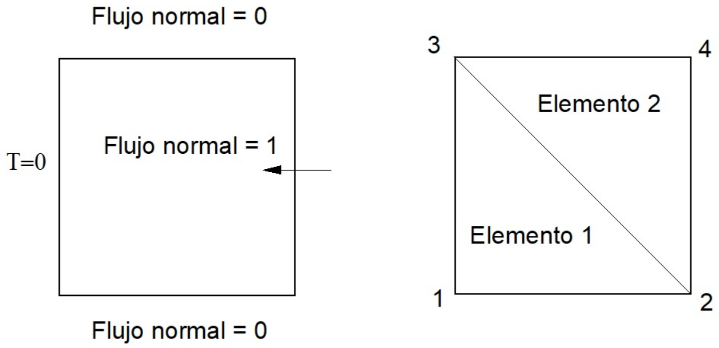


Figure 12: A square plate with a unit side.

**Solution**

- Contribution of the first element to the coefficient matrix

According to the results obtained in the preceding examples, and arranging the nodes in the order $(1, 2, 3)$ we obtain:

-

| Local | Global |
|---|---|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |

$$K^{(1)} = kA \begin{pmatrix} 1 & \text{-1} & 0 \\ \text{-1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{162}$$

- Contribution of the second element to the coefficient matrix:

The nodes of element 2 will be ordered as follows:

| Local | Global |
|---|---|
| 1 $(A)$ | 3 |
| 2 $(B)$ | 2 |
| 3 $(C)$ | 4 |

The coefficient matrix will be:

$$\mathbf{K}^{(e)} = \int_{\Omega^{(e)}} \mathbf{B}^T \mathbf{k} \mathbf{B} d\Omega \tag{163}$$

where $\mathbf{B}$ will be:

$$\begin{aligned} \mathbf{B} = & \begin{pmatrix} \frac{\partial N_1}{\partial x} & \frac{\partial N_2}{\partial x} & \frac{\partial N_3}{\partial x} \\ \frac{\partial N_1}{\partial y} & \frac{\partial N_2}{\partial y} & \frac{\partial N_3}{\partial y} \end{pmatrix} \\ = & \frac{1}{(x_{BA} . y_{CA} - y_{BA} . x_{CA})} \begin{pmatrix} y_{CA} & -y_{BA} \\ -x_{CA} & x_{BA} \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \end{aligned} \tag{164}$$

and

$$\begin{aligned} x_{BA} = & \quad x_2 - x_3 = & 1 \\ x_{CA} = & \quad x_4 - x_3 = & 1 \\ y_{BA} = & \quad y_2 - y_3 = & -1 \\ y_{CA} = & \quad y_4 - y_3 = & 0 \end{aligned} \tag{165}$$

Once these values have been substituted, we obtain

$$\begin{aligned} \mathbf{B} = & \frac{1}{1 * 0 - (-1) * 1} \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \\ = & \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \end{aligned} \tag{166}$$

This matrix could have been obtained directly, since the partial derivatives of the shape functions in the case considered are easy to calculate. The element stiffness matrix is, therefore,

$$\begin{aligned} \mathbf{K}^{(2)} = & \int_{\Omega^{(2)}} \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} k \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} d\Omega \\ = & kA \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix} \end{aligned} \tag{167}$$

- Global matrix of coefficients

The global coefficient matrix is obtained by assembling the element matrices. Once the first element has been assembled, and without having yet assembled the second, the global array is:

$$kA.\left(\begin{array}{ccccc} \text{nodes} & 1 & 2 & 3 & 4 \\ 1 & 1 & \text{-1} & 0 & \\ 2 & \text{-1} & 1 & 0 & \\ 3 & 0 & 0 & 1 & \\ 4 & & & & \end{array}\right) \tag{168}$$

The second element is then assembled by adding the terms of its coefficient matrix to the corresponding positions in the global matrix. In this way we obtain:

$$kA.\left(\begin{array}{ccccc} \text{nodes} & 1 & 2 & 3 & 4 \\ 1 & 1 & -1 & 0 & \\ 2 & -1 & 1{+}\mathbf{1} & 0{+}\mathbf{0} & -\mathbf{1} \\ 3 & 0 & 0{+}\mathbf{0} & 1{+}\mathbf{1} & -\mathbf{1} \\ 4 & & -\mathbf{1} & -\mathbf{1} & +\mathbf{2} \end{array}\right) \tag{169}$$

$$= kA\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}$$

- Independent terms

The contribution of each element to the vector of independent terms is given by:

$$f_i^{(e)} = \int_{\Gamma_\Phi^{(e)}} N_i \mathbf{n}^T.\mathbf{k}\nabla\hat{\Phi} d\Gamma - \int_{\Gamma_q} N_i \bar{q} d\Gamma \tag{170}$$

In the case studied, only (i) Flow terms should be included in the nodes where the temperature is prescribed (ii) integrals where $\bar{q}$ is different from zero. Therefore, it is only necessary to consider the contribution of element (2), in its contour 2-4, where a heat flow per unit length of unit value directed towards the inside of the plate has been imposed, and which will give rise to :

$$f_2^{(2)} = -\int_{\Gamma_q} N_2 \bar{q} d\Gamma = -\int_{\Gamma_{2-4}} N_2(-1) d\Gamma = \frac{1}{2} \tag{171}$$

$$f_4^{(2)} = -\int_{\Gamma_q} N_4 \bar{q} d\Gamma = -\int_{\Gamma_{2-4}} N_4(-1) d\Gamma = \frac{1}{2} \tag{172}$$

- The system of equations to be solved is, therefore,

$$kA\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}\begin{pmatrix} \hat{\Phi}_1 = 0 \\ \hat{\Phi}_2 \\ \hat{\Phi}_3 = 0 \\ \hat{\Phi}_4 \end{pmatrix} = \begin{pmatrix} R_1 \\ 1/2 \\ R_3 \\ 1/2 \end{pmatrix} \tag{173}$$

Taking equations (2) and (4), the reduced system is obtained:

$$2\hat{\Phi}_2 - \hat{\Phi}_4 = \frac{1}{2kA} = 1 \tag{174}$$

$$-\hat{\Phi}_2 + 2\hat{\Phi}_4 = \frac{1}{2kA} = 1 \tag{175}$$

where it has been taken into account that $A = 1/2$ and $k = 1$ . The solution is

$$\hat{\Phi}_2 = \hat{\Phi}_4 = 1 \tag{176}$$

Once the unknowns are calculated, the total flows in the nodes where the temperature has been prescribed are obtained.

$$R_1 = R_3 = -1 \tag{177}$$

# References

[Bre91] S.C.Brenner and L.R.Scott. *The Mathematical Theory of Finite Element Methods,* Springer Verlag, New York, 1991.

[Car83] G.F. Carey and J.T. Oden. *Finite Elements: A second Course,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1983.

[Hug87] T.J.R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[Zie83] O.Z. Zienkiewicz and K. Morgan. *Finite Element and Approximation,* John Wiley and Sons, New York, 1983.

[Zie00] O.Z. Zienkiewicz and R.L. Taylor. *The Finite Element Method 5th edition* (3 Vols.), Butterword-Heinemann, Oxford, 2000.

# Introduction to finite elements (II): transient problems of parabolic and hyperbolic types

**Manuel Pastor, Saeid M. Tayyebi, Pablo Mira, Miguel M. Stickle, Diego Manzanal, J.A. Fernandez Merodo**

*ETS de Ingenieros de Caminos, Universidad Politecnica de Madrid, Ciudad Universitaria s/n, 28040, Madrid, Spain*

*This Chapter is the second part of the material devoted to present the finite element method. As in the previous Chapter, we have chosen a level of mathematics accesible to MSc students of engineering.*

## 1 Introduction

The elliptic problems presented in the previous Chapter describe the steady state conditions which can be attained in time dependent problems. Indeed, the diffusion problems presented there correspond to a particular type of PDE referres to as parabolic In addition to them there is another class of PDE describing a large group of problems. Hyperbolic PDEs describe problems such as pollutant transport by a current, hydaulics of rivers and coasts, and dynamics of soils and structures.

This Chapter will be devoted to present both types of PDEs, hyperbolic and hyperbolic. We will present first the mathematical models describing transient diffusion and convective problems, showing the main differences between thei funsdamental solutions. While in diffusive problems we find physical damping of the solution, pure convective problems do not exhibit it. Moreover, discontnuities are propagated by the later bud diffused by the former.

Regarding discretization, we have chosen to use finite differences as a pedagogical vehicle to show why classical finite elements (Galerkin) will nor work for convection dominated problems. Then, we will show how one of the methods that work for convective problems, the Lax-Wendroff scheme, can be extended to finite elements, leading to the Taylor Galerkin method.

# 2 Model equations for time dependent problems

## 2.1 Parabolic problems

One of the simplest time-dependent problems is diffusion. Examples of this type of problems are the following:

(a) Heat transport by convection (Fourier's Law)

(b) Transport of matter by diffusion (Fick's Law)

(c) Linear momentum transport in viscous fluids (Navier's Law)

(d) Flow in porous medium (Darcy's Law)

All these problems can be described by Parabolic Partial Differential Equations. In the case of heat conduction, the equations in one-dimensional problems are:

$$\rho c \frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2} + s \tag{1}$$

where $D$ is the thermal conductivity coefficient, $\rho$ is the density and $c$ is the specific heat, with the term $s$ representing the heat sources (power generated per unit volume).

This equation has been obtained in a similar way to what was done in the case of the stationary problem, first considering the heat flow, given by Fourier's Law:

$$q = -D \frac{\partial \phi}{\partial x} \tag{2}$$

as well as the balance or conservation equation:

$$\rho c \frac{\partial \phi}{\partial t} = s - \frac{\partial q}{\partial x} \tag{3}$$

Both equations are combined, giving:

$$\rho c \frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x} \left( D \frac{\partial \phi}{\partial x} \right) + s \tag{4}$$

which is the EDP that describes the problem.

For the problem to be well posed, initial conditions must be defined:

$$\phi(x, t_0) = g(x) \tag{5}$$

together with suitable boundary conditions

$$\phi(x, t) - \bar{\phi}(t) \text{ in } \Gamma_\phi \text{ (Dirichlet)} \tag{6}$$

and

$$D\frac{\partial \phi}{\partial n} + \bar{q} = 0 \text{ en } \Gamma_q \text{ (Neumann)} \tag{7}$$

If we try to find elementary solutions of the type

$$\phi(x,t) = A\exp(i\kappa x - i\omega t) \tag{8}$$

where $\kappa$ is the number of waves and $\omega$ the angular frequency, in the problem without source type terms:

$$\rho c\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial \phi}{\partial x}\right) \tag{9}$$

we obtain:

$$\rho c(-i\omega)A\exp(i\kappa x - i\omega t) = -D\kappa^2 A\exp(i\kappa x - i\omega t) \tag{10}$$

from where

$$\rho c(-i\omega)A = -D\kappa^2 A \tag{11}$$

$$-i\omega t = -\frac{D}{\rho c}\kappa^2 t \tag{12}$$

being, therefore, the elementary solution to the problem 1:

$$\phi = A.\exp(-\frac{D}{\rho c}\kappa^2 t)\exp(i\kappa x) \tag{13}$$

In view of the solution obtained, the following fundamental aspects of its behavior can be deduced:

- The factor $\exp(-\frac{D}{\rho c}\kappa^2 t)$ will cause the solution to soften over time, as outlined in figure 1

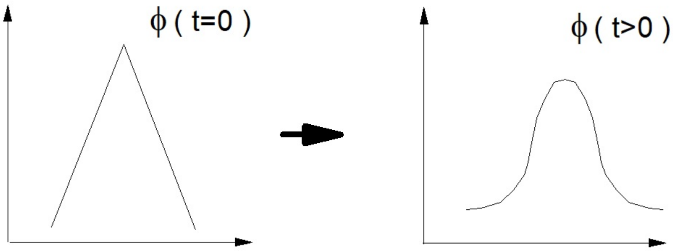


Figure 1: Softening over time.

- Damping increases with the wave number $\kappa$, being therefore greater for the higher modes having smaller wavelengths.

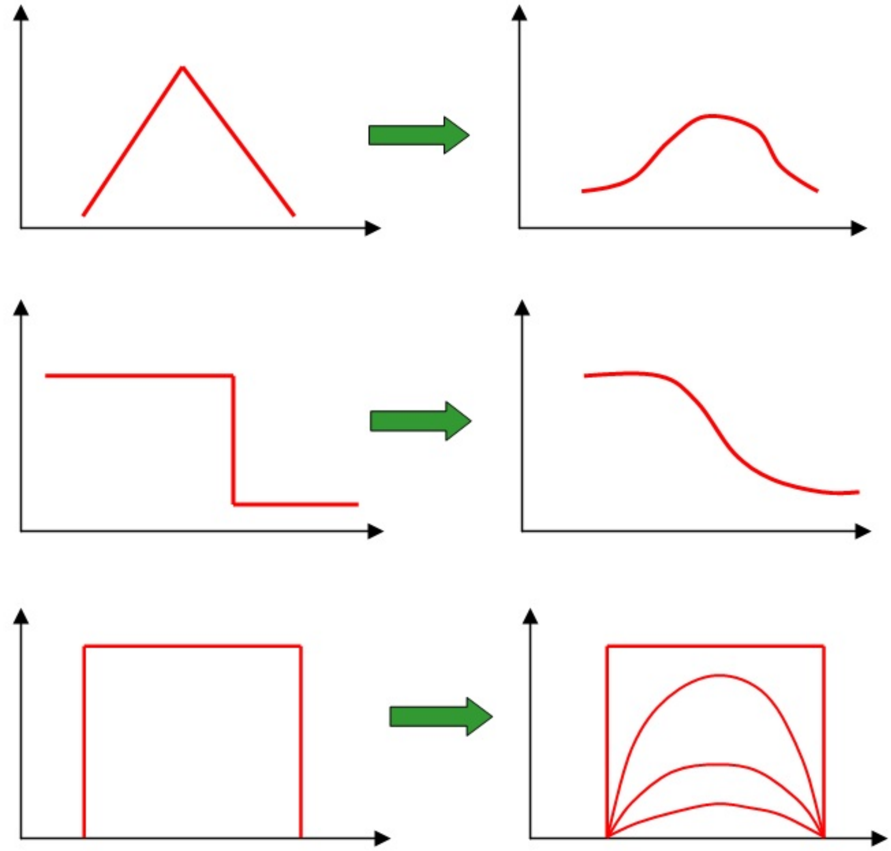


Figure 2: Smoothing of solutions.

- Discontinuities that may exist in the initial conditions are smoothed( figure 2).
- No discontinuities of any kind can spontaneously appear in the problem described.
- In two and three dimensions, the equations are similar. In the case of heat transport, the equations are:

  $$\rho c \frac{\partial \Phi}{\partial t} = \operatorname{div}\left(D.\operatorname{grad}\Phi\right) + s \tag{14}$$

  where div is the divergence operator in 2D, equal to the transposed of the gradient

  $$\operatorname{div} = \nabla^T = \operatorname{grad}^T = \left(\partial_x, \partial_y\right) \tag{15}$$

  $$orin3D\, \operatorname{div} = \nabla^T = \operatorname{grad}^{T=} \left(\partial_x, \partial_y, \partial_z\right) \tag{16}$$

## 2.2 1st order hyperbolic PDEs: convective transport problems

1st order PDEs are one of the simplest that can be found in Physics and Engineering, as they involve 1st order derivatives with respect to time and space. In addition to the pure convection of a scalar magnitude by a flow, they are found in Navier Stokes equations, when formulated in an eulerian framework; in the shallow water equations which describe coastal hydraulics problems, in flood waves caused by breaking of dams, and in fluidized soil avalanches, just to mention a few examples.

However their apparent simplicity, they present difficulties such as:

- The convective terms require special discretization techniques from the classical Boubnov-Galerkin Finite Elements -which are not stable.
- They present numerical diffusion and damping.
- Numerical dispersion, making shorter wavelenghts to travel with smaller speeds than the theoretical, appear.

The simplest model we will consider is the 1D scalar convection equation. If we introduce $\phi(x,t)$ as the concentration of a magnitude which is being convected by a fluid of constant velocity $u$, the balance equation is obtained by considering a control volume of length $dx$ and cross section $A$ as (figure 3):

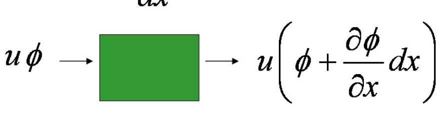

$\leftarrow dx \rightarrow$
$u\phi$
$u\left(\phi + \frac{\partial \phi}{\partial x} dx\right)$

Figure 3: 1D convective transport of a magnitude $\phi$.

$$Adx\frac{\partial \phi}{\partial t} = Au\phi - Au\left(\phi + \frac{\partial \phi}{\partial x}\right) \tag{17}$$

from where we arrive to:

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} = 0 \tag{18}$$

This equation is a particular case of the more general

$$a\frac{\partial \phi}{\partial t} + b\frac{\partial \phi}{\partial x} = c \tag{19}$$

or,

$$a\phi_t + b\phi_x = c \tag{20}$$

The equation is called linear when $a, b$ and $c$ depend on $(x, t)$ but not on $\phi$, and quasi-linear when they depend on $(x, t, \phi)$.

There exists an alternative formulation referred to as ”conservative”, written as:

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x}(u\,\phi) = 0 \tag{21}$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial F}{\partial x} = 0 \tag{22}$$

where the flux $F$ is

$$F = u\,\phi \tag{23}$$

The PDE can be derived by considering an arbitrary segment $[a, b]$ and expressing the rate of change of $\phi$ as the difference between the incoming and the outcoming fluxes:

$$\frac{d}{dt}\int_a^b \phi\,(x,t)\,dx = F\,(a) - F\,(b) \tag{24}$$

Taking into account that

$$\frac{d}{dt}\int_a^b \phi\,(x,t)\,dx = \int_a^b \frac{\partial \phi}{\partial t}\,(x,t)\;dx \tag{25}$$

and

$$F\,(a) - F\,(b) = -\int_a^b \frac{\partial F}{\partial x}\,(x,t)\;dx \tag{26}$$

we arrive to:

$$\int_a^b \frac{\partial \Phi}{\partial t}\,(x,t)\;dx = -\int_a^b \frac{\partial F}{\partial x}\,(x,t)\;dx \tag{27}$$

from where we obtain:

$$\int_a^b \left(\frac{\partial \Phi}{\partial t} + \frac{\partial F}{\partial x}\right)\;dx = 0 \tag{28}$$

Finally, as the chosen segment is arbitrary, we obtain the conservative formulation (22). The integral formulation is to be used when discontinuities appear in $\phi$,as 24 does not involve partial derivatives

Regarding initial and boundary conditions, it is important to notice that we will need:

(i) an initial condition $\phi\left(x, t=0\right) = h_0(x) \ \ x_0 \leq x \leq x_L$

(ii) one boundary condition at the end where the velocity enters the domain $\phi\left(x_0, t\right) = g\left(t\right) \qquad 0 < t < T_{end}$

where the domain is $\left[x_0, x_L\right] x \left[0, T_{end}\right]$

- If we try to find elementary solutions of the type

$$\phi(x,t) = A \exp(i\kappa x - i\omega t) \tag{29}$$

where $\kappa$ is the number of waves and $\omega$ the angular frequency, in the problem without source type terms:

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} = 0 \tag{30}$$

we obtain:

$$-i\omega A \exp(i\kappa x - i\omega t) + i\kappa u \, A \exp(i\kappa x - i\omega t) \tag{31}$$

i.e, $\quad -i\omega\phi + i\kappa u \, \phi = 0$ from where

$$\omega = u\kappa \tag{32}$$

$$-i\omega t = -\frac{D}{\rho c}\kappa^2 t \tag{33}$$

the elementary solution to the problem being:

$$\phi\left(x, t\right) = A. \exp(i\kappa\left(x - ut\right)) \tag{34}$$

which is a wave moving without changig of shape along $X$ axis with a constant velocity $u$. In view of the solution obtained, the following fundamental aspects of its behavior can be deduced:

- The solution is not damped
- The velocity does not depend on the wave length
- Discontinuities that may exist in the initial conditions are NOT smoothed
- In two and three dimensions, the equations are similar. In the case of heat transport, the equations are:

$$\frac{\partial \phi}{\partial t} + u\text{div}\left(\phi\right) = 0 \tag{35}$$

- In many practical cases, source terms and diffusion may exist. The PDE is then

$$\frac{\partial \phi}{\partial t} + \mathrm{div}\,(\mathbf{u}\phi) = s + \frac{\partial}{\partial x}\left(D\frac{\partial \phi}{\partial x}\right) \tag{36}$$

  where we have assumed that u depends on $x$

# 3 Finite difference approximations

The objective of this section is to introduce some basic finite difference (FD) schemes for both the 1D heat conduction and the 1D scalar convective transport equation. We have selected 3 schemes, (i) Forwards in time and centered in space (FTCS), (ii) Forwards in time and backwards in space (FTBS) and (iii) Lax-Wendroff. These schemes are interesting as they illustrate one fundamental difficulty presented for both Finite Elements and Differences and the ways to circumvent it.

These methods can be applied to more complex problems, such as system of 1st order hyperbolic PDEs and the convective part of Navier Stokes equations.

Finite differences schemes present the advantage of being simpler to present and understand, hence more pedagogic.

There exist many texts describing Finite Difference schemes, among which we can mention those of [Far82], [Fle88], [Hir88], [Lev92], [Roa98], [Smi78] and [Str89].

## 3.1 Finite difference schemes for the 1D heat conduction problem

The Finite Difference Method is based on constructing a grid in the domain $\Omega \times I$. For instance, if we consider the heat conduction in a one dimensional bar of length $L$ at times $t_0 \le t \le t_f$,

$$\rho c \frac{\partial \phi}{\partial t} = D\frac{\partial^2 \phi}{\partial x^2} \tag{37}$$

the mesh would be the one depicted in Figure 4. Any node $x = x_0 + j.\triangle x, t = t_0 + n\triangle t$, can be identified by $(j, n)$.

Partial derivatives with respect to time and space can then be approximated as combinations of the values at a set of nodes. For instance, the partial derivative with respect to time

$$\left.\frac{\partial \phi}{\partial t}\right|_j^n := \frac{\partial \phi}{\partial t}(x = x_j, t = t_n) \tag{38}$$

can be approximated as:

$$\left.\frac{\partial \phi}{\partial t}\right|_j^n = \frac{\phi_j^{n+1} - \phi_j^n}{\Delta t} + O(\Delta t) \tag{39}$$

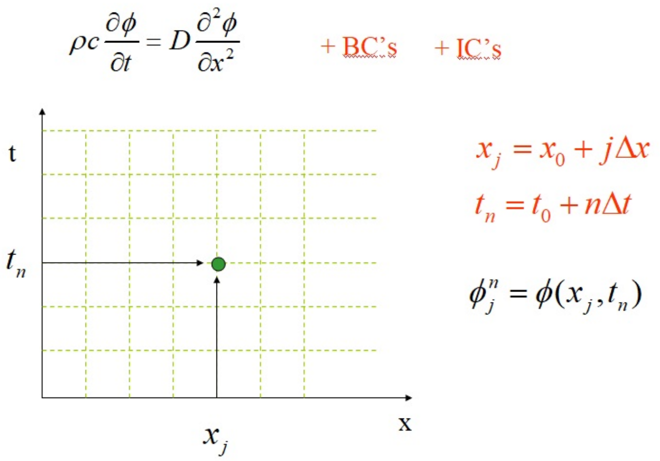

$\rho c \frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2}$
+ BC's
+ IC's
t
$t_n$
$x_j$
x
$x_j = x_0 + j\Delta x$
$t_n = t_0 + n\Delta t$
$\phi_j^n = \phi(x_j, t_n)$

Figure 4: Finite Difference Grid for the 1D heat conduction problem in a bar.

Another alternative is:

$$\left.\frac{\partial \phi}{\partial t}\right|_j^n = \frac{\phi_j^{n+1} - \phi_j^{n-1}}{2\Delta t} + O(\Delta t^2) \tag{40}$$

Second derivatives with respect to space can be obtained in the same way:

$$\left.\frac{\partial^2 \phi}{\partial x^2}\right|_j^n = \frac{\phi_{j+1}^n - 2\phi_j^n + \phi_{j-1}^n}{\Delta x^2} + O(\Delta x^2) \tag{41}$$

If we now substitute (39) and (41) into (37), we obtain:

$$\frac{\phi_j^{n+1} - \phi_j^n}{\Delta t} = \frac{D}{\rho c} \frac{\phi_{j+1}^n - 2\phi_j^n + \phi_{j-1}^n}{\Delta x^2} \tag{42}$$

from where we get:

$$\phi_j^{n+1} = \phi_j^n + D^* (\phi_{j+1}^n - 2\phi_j^n + \phi_{j-1}^n) \tag{43}$$

where we have introduced $D^* = D\Delta t / \rho c \Delta x^2$.

The finite difference scheme (43) is explicit, as the solution at $t^{n+1}$can be obtained directly without having to solve any system of equations, and it is conditionally stable as it will be shown later on, as the timestep $\Delta t$ has to be smaller than a critical value to avoid oscillations growing with time (see figure 5).

Of course, both the problem and the finite difference solution are complemented with suitable boundary and initial conditions.

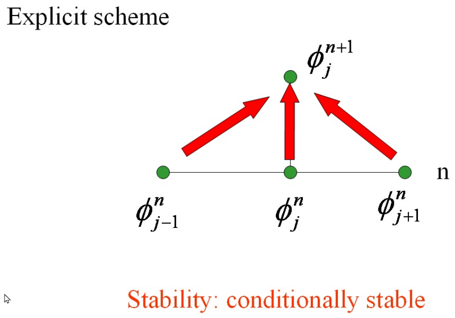


Figure 5: Stencil for the explicit FD FTCS scheme.

- **Example**

Given a one dimensional bar of constant cross section and length $L = 10$ with boundary conditions $\phi(0,t) = 0$ and $\phi(L,t) = 0$ and the initial distribution of temperature

$$\begin{array}{ll} \phi(x,0) = x/10 & 0 \leq x \leq 5 \\ \phi(x,0) = 1 - x/10 & 5 \leq x \leq 10 \end{array} \tag{44}$$

obtain the evolution with time of the bar temperature using the finite difference scheme given in this section using a grid with $\Delta x = 1.0$ in the two cases $\Delta t = 0.49$ and $\Delta t = 0.60$.Specific heat, density and thermal conductivity will be taken as unity.

**Solution**

We will use the FD scheme $\phi_j^{n+1} = \phi_j^n + D^*(\phi_{j+1}^n - 2\phi_j^n + \phi_{j-1}^n)$ with $j = 1..9$ (which can be easily done using a spreadsheet).

Figures 6 and 7 depict the results obtained in both cases. As expected, the scheme with $\Delta t = 0.60$ is unstable as $D^* = 0.60$.

## 3.2 Finite difference schemes for the 1D linear convective transport in 1D

We will consider two alternative explicit schemes, a FTCS (forward in time, centered in space) scheme and the Lax-Wendroff scheme. The former corresponds to the classical Galerkin formulation in finite elements, and it is unconditionally instable (i.e. never works), while the Lax Wendroff provides a different framework, and is the basis of the extension to finite elements (Taylor-Galerkin methods)

**FTCS scheme** Regarding the FTCS scheme, it approximates the time derivative in

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} = 0 \tag{45}$$

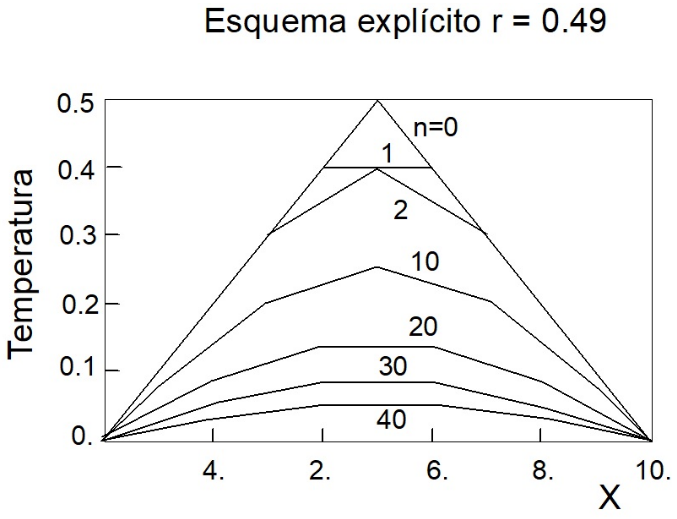


Figure 6: Explicit scheme with $D^* = 0.49$.

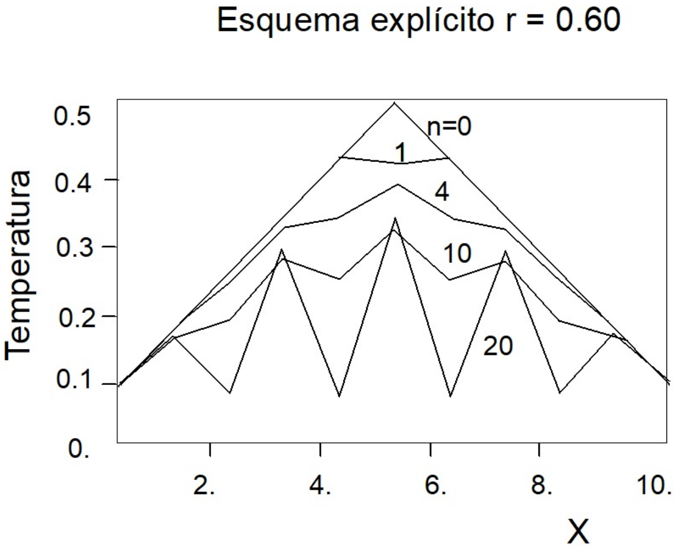


Figure 7: Explicit scheme with $D^* = 0.6$.

$$0 \leq x \leq L \quad 0 \leq t \leq T$$

by:

$$\left.\frac{\partial \phi}{\partial t}\right|_j^n = \frac{\phi_j^{n+1} - \phi_j^n}{\triangle t} + O(\triangle t) \tag{46}$$

and the spacial derivative by

$$\left.\frac{\partial \phi}{\partial x}\right|_j^n = \frac{\phi_{j+1}^n - \phi_{j-1}^n}{2 \triangle x} + O(\triangle x^2) \tag{47}$$

from where we obtain the FD scheme:

$$\phi_j^{n+1} = \phi_j^n - \frac{C}{2} \cdot \left(\phi_{j+1}^n - \phi_{j-1}^n\right) \tag{48}$$

where we have introduced the non dimensional number $C$ , referred to as *Courant number,*

$$C = \frac{u}{\triangle x / \triangle t} \tag{49}$$

which is the ratio between the physical convective velocity $u$ and $u_{num}$,the numerical velocity with which a signal can propagate in the FD grid,

$$u_{num} = \triangle x / \triangle t \tag{50}$$

The scheme is of explicit type, as once the value of $\phi$ is known at time $t_n$ , we can obtain the values of $\phi$ at $t_{n+1}$ directly, without having to solve a linear system of equations. The reader should notice that the space derivative used in the scheme is of second order, with an error which is of the order $\triangle x^2$ he derivatives at the ends of the domain, can be approximated by alternative expressions not involving the outer part of the domain, as for instance:

$$\left.\frac{\partial \phi}{\partial t}\right|_M^n = \frac{\phi_M^n - \phi_{M-1}^n}{\triangle x} + O(\triangle x) \tag{51}$$

- **Example**

Apply the FTCS to solve the problem described in the preceding example, taking again $\triangle x = 0.1$, and studying the three cases $C = \{0.5, 1.0, 1.5\}$.

**Solution**

The results obtained in the three cases can be seen in figure 8, where we have drawn the values of $\phi$ after 5 and 10 time increments in the first two cases, and 5 $\Delta t_s$ in the case $C = 1.5$ . In all the three cases we find growing oscillations, which correspond to unstable schemes.

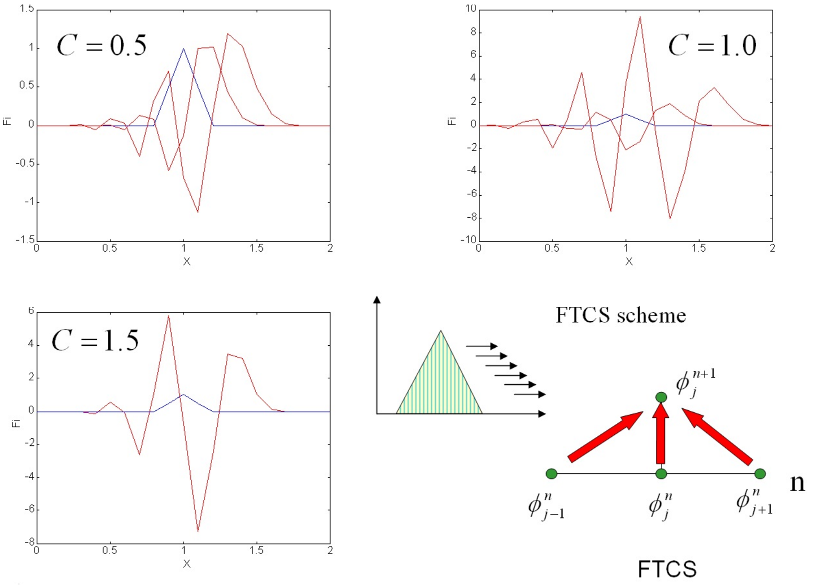


Figure 8: FTCS scheme. Solution for C=0.5,1 and 1.5.

**Lax-Wendroff scheme** Lax Wendroff scheme consists of two parts. First, we make a Taylor series expansion with respect to time of $\phi$ in $(j, n)$

$$\phi_j^{n+1} = \phi_j^n + \triangle t.\frac{\partial \phi}{\partial t} \mid_j^n + \frac{1}{2} \triangle t^2.\frac{\partial^2 \phi}{\partial t^2} \mid_j^n \tag{52}$$

Then, we will use the partial derivatives with respect to time provided by the PDE,

$$\frac{\partial \phi}{\partial t} = -u\frac{\partial \phi}{\partial x} \tag{53}$$

and

$$\begin{aligned} \frac{\partial^2 \phi}{\partial t^2} &= \frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial t}\right) = \frac{\partial}{\partial t}\left(-u\frac{\partial \phi}{\partial x}\right) \\ &= -u\frac{\partial}{\partial x}\left(\frac{\partial \phi}{\partial t}\right) = u^2\frac{\partial^2 \phi}{\partial x^2} \end{aligned} \tag{54}$$

We obtain, after substitution in the PDE

$$\phi_j^{n+1} = \phi_j^n - \triangle t.u\left.\frac{\partial \phi}{\partial t}\right|_j^n + \frac{1}{2} \triangle t^2.u^2\frac{\partial^2 \phi}{\partial x^2} \mid_j^n \tag{55}$$

where we can observe that all derivatives with respect to time have been replaced by space derivatives.

And now, it is possible to discretize the equation in space, using either Finite Differences or Elements. In the former case, we can use a 2nd order, centered scheme, approximating the derivatives as:

$$\left.\frac{\partial \phi}{\partial x}\right|_j^n = \frac{\phi_{j+1}^n - \phi_{j-1}^n}{2 \bigtriangleup x} \tag{56}$$

$$\left.\frac{\partial^2 \phi}{\partial x^2}\right|_j^n = \frac{\phi_{j+1}^n - 2\phi_j^n - \phi_{j-1}^n}{\bigtriangleup x^2} \tag{57}$$

After substitution in (55), we obtain:

$$\phi_j^{n+1} = \phi_j^n - \bigtriangleup t.u\left(\frac{\phi_{j+1}^n - \phi_{j-1}^n}{2 \bigtriangleup x}\right) + \frac{1}{2} \bigtriangleup t^2.u^2\left(\frac{\phi_{j+1}^n - 2\phi_j^n - \phi_{j-1}^n}{\bigtriangleup x^2}\right) \tag{58}$$

or,

$$\phi_j^{n+1} = \frac{1}{2}C(1+C)\phi_{j-1}^n + (1-C^2)\phi_j^n - \frac{1}{2}C(1-C)\phi_{j+1}^n \tag{59}$$

which is the Lax-Wendroff scheme.

- **Example**

Apply the Lax Wendroff scheme to solve the problem of examples 7 and 8, taking $\bigtriangleup x = 0.1$, and studying the cases $C = \{0.5, 1.0, 1.5\}$.

**Solution**

We obtain, for the three cases considered, the results shown in figure 9, where we have drawn the values of $\phi$ after 5, 10 and 15 increments of time in the first case, 5 and 10in the second, and 10 in the third case. We can observe that the exact solution is obtained for the case $C = 1$ while when $C = 0.5$ we observe numerical diffusion damping and smoothing the solution. In the case $C = 1.5$, the

# 4 Finite element approximations

## 4.1 1D finite elements for the heat conduction problem

The procedure followed usually consists of first discretizing the PDE4 in space, using the Galerkin method described above. The only difference with the elliptic type partial differential equation corresponding to the stationary problem is the term

$$\rho c \frac{\partial \Phi}{\partial t} \tag{60}$$

which, when using the Galerkin Method, gives rise to:

$$\int_\Omega N_i \rho c \frac{\partial \hat{\Phi}(x,t)}{\partial t} d\Omega \tag{61}$$

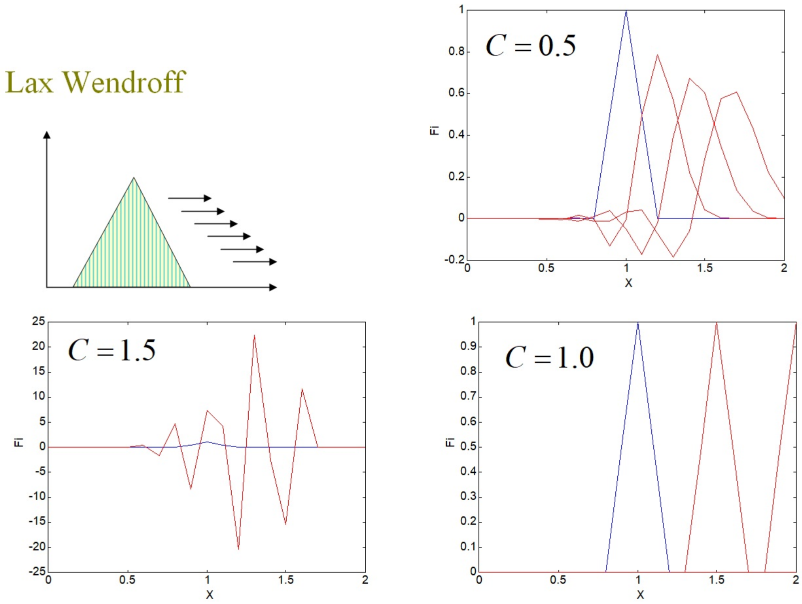


Figure 9: Lax Wendroff scheme. C=0.5,1.0 and 1.5.

where

$$\frac{\partial \hat{\Phi}(x,t)}{\partial t} = \sum_j \frac{\partial}{\partial t} N_j(x).\hat{\Phi}_j(t) \tag{62}$$

arriving, in this way, at:

$$\int_\Omega N_i \rho c \frac{\partial \hat{\Phi}(x,t)}{\partial t} d\Omega = \sum_j \int_\Omega N_i \rho c N_j \frac{d\hat{\Phi}_j(t)}{dt} d\Omega \tag{63}$$

where it has been assumed that the problem considered is one-dimensional. However, in the case of problems in two or three dimensions, the procedure to follow in this first stage of discretization in space is similar to that described in the chapter dedicated to elliptic type equations, and in fact, in all cases we arrive at a term that can be written, in a more compact way, as

$$\mathbf{C}.\frac{d}{dt}\hat{\mathbf{\Phi}}(t) \tag{64}$$

where

$$C_{ij} = \int_\Omega \rho c N_i N_j d\Omega \tag{65}$$

$$\mathbf{C} = \int_\Omega \mathbf{N}^T.\rho c.\mathbf{N}.d\Omega \tag{66}$$

In this way we arrive at the equation

$$\mathbf{K}.\hat{\mathbf{\Phi}}(t)+\mathbf{C}.\frac{d}{dt}\hat{\mathbf{\Phi}}(t)=\mathbf{f} \tag{67}$$

where

$$\mathbf{K}=\int_{\Omega}\nabla^{T}\mathbf{N}.k.\nabla\mathbf{N}.d \tag{68}$$

$$\mathbf{f}=\int_{\Omega}s.\mathbf{N}.d\Omega+\int_{\Gamma_q}\mathbf{N}.\bar{q}.d\Gamma \tag{69}$$

with the vector $\hat{\mathbf{\Phi}}(t)$ given by:

$$\hat{\mathbf{\Phi}}(t)=\left\{\hat{\Phi}_1(t),\hat{\Phi}_2(t),\hat{\Phi}_3(t),...\right\} \tag{70}$$

It is important to notice that the unknowns $\hat{\Phi}_i(t)$ now depend on time. These expressions are general, and can be used in both 1D, 2D and 3D problems. However, for simplicity, both in the numerical stability analysis and in some examples, the 1D case will be studied.

A simple scheme for the equation (67) is obtained by particularizing it in $t_n$ , and approximating the time derivative by means of forward differences (Forward Euler):

$$\mathbf{K}.\hat{\mathbf{\Phi}}(t)+\mathbf{C}.\frac{d}{dt}\hat{\mathbf{\Phi}}(t)=\mathbf{f} \tag{71}$$

$$\frac{d}{dt}\hat{\mathbf{\Phi}}(t)=\frac{\hat{\mathbf{\Phi}}^{n+1}-\hat{\mathbf{\Phi}}^{n}}{\triangle t} \tag{72}$$

from where we obtain

$$\mathbf{K}.\hat{\mathbf{\Phi}}^{n}+\mathbf{C}\frac{\hat{\mathbf{\Phi}}^{n+1}-\hat{\mathbf{\Phi}}^{n}}{\triangle t}=\mathbf{f}^{n} \tag{73}$$

arriving to

$$\hat{\mathbf{\Phi}}^{n+1}=\hat{\mathbf{\Phi}}^{n}+\triangle t\mathbf{C}^{-1}\left(\mathbf{f}^{n}-\mathbf{K}.\hat{\mathbf{\Phi}}^{n}\right) \tag{74}$$

The scheme is explicit when a diagonal representation of the matrix $\mathbf{C}$ is used so that it is not necessary to invert it, and conditionally stable, oscillations appearing for values of the time increment greater than a certain critical value. In this case, the scheme obtained coincides with the explicit one derived previously in Finite Differences.

- **Example**

Given a bar of length $L$ and constant scross ection $A$ whose left end located at $x=0$ is maintained at a temperature $\Phi=1^{0}C$, the right end being insulated, and the initial temperature being $\Phi(x,0)=1-x/L$ , obtain the finite element equations for the case of the explicit scheme presented in this section, particularizing them for the case where a diagonal representation of the matrix C is used. Three elements of equal length will be used in the analysis.

**Solution:**

The contribution of each element to matrix **C** is obtained first as:

$$C_{ij}^{(e)} = \int_{(e)} \rho c N_i N_j A dx \tag{75}$$

These integrals can be done in a relatively simple way taking into account that the area enclosed by a parabola is equal to 2/3 of the area of the rectangle where it is inscribed, as indicated in Figure10.

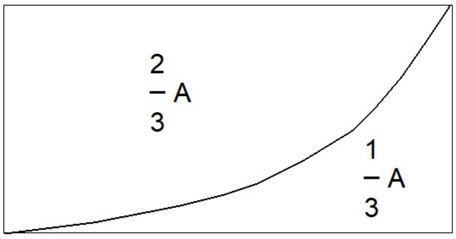


Figure 10: Areas defined by a parabola.

From here, and taking into account that the length of each element is $L/3$, we obtain:

$$C_{11}^{(e)} = C_{22}^{(e)} = \frac{L}{3} A\rho c . \frac{1}{3} = \frac{1}{9} AL\rho c \tag{76}$$

$$C_{12}^{(e)} = C_{21}^{(e)} = \frac{L}{3} A\rho c \frac{2}{3} . \frac{1}{4} = \frac{1}{18} AL\rho c \tag{77}$$

where the factor $1/4$ that appears in the second expression comes from the fact that the product of the shape functions $N_1$ and $N_2$ at the midpoint of the element is $\frac{1}{2} . \frac{1}{2}$

The matrix $\mathbf{C}^{(e)}$ of the element is, then,

$$\mathbf{C}^{(e)} = \frac{1}{18} AL\rho c \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{78}$$

and its diagonal representation $\mathbf{C}_L^{(e)}$ is obtained by concentrating the sum of all the elements of each row on the diagonal:s

$$\mathbf{C}_L^{(e)} = \frac{1}{18} AL\rho c \begin{pmatrix} 2+1 & 0 \\ 0 & 2+1 \end{pmatrix} = \frac{1}{6} AL\rho c \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{79}$$

The global matrix is obtained by assembling the contributions of each element:

$$\mathbf{C} = \frac{1}{18} AL\rho c \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2+2 & 1 & 0 \\ 0 & 1 & 2+2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} = \frac{1}{18} AL\rho c \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \tag{80}$$

where $\mathbf{C}_L$ is

$$\mathbf{C}_L = \frac{1}{6} AL\rho c \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{81}$$

On the other hand, the coefficient matrices of the elements are given by

$$\mathbf{K}^{(e)} = \frac{DA}{L/3} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{82}$$

which, once assembled give

$$\mathbf{K} = \frac{DA}{L/3} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \tag{83}$$

The resulting system is, then,

$$\begin{pmatrix} \hat{\Phi}_1^{n+1} \\ \hat{\Phi}_2^{n+1} \\ \hat{\Phi}_3^{n+1} \\ \hat{\Phi}_4^{n+1} \end{pmatrix} = \begin{pmatrix} \hat{\Phi}_1^{n} \\ \hat{\Phi}_2^{n} \\ \hat{\Phi}_3^{n} \\ \hat{\Phi}_4^{n} \end{pmatrix} + \frac{3\triangle t}{AL\rho c} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} . \tag{84}$$

$$\left\{ \begin{pmatrix} R_1 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \frac{DA}{L/3} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \hat{\Phi}_1^{n} \\ \hat{\Phi}_2^{n} \\ \hat{\Phi}_3^{n} \\ \hat{\Phi}_4^{n} \end{pmatrix} \right\}$$

Taking into account now that $\triangle x = L/3$, the second equation is:

$$\hat{\Phi}_2^{n+1} = \hat{\Phi}_2^{n} + \frac{D\triangle t}{\rho c \triangle x^2} \left( \hat{\Phi}_1^{n} - 2\hat{\Phi}_2^{n} + \hat{\Phi}_3^{n} \right) \tag{85}$$

which coincides with those obtained in the explicit finite difference scheme studied above.

The Finite Element scheme presented above is a simple case of a single step-in-time scheme. In fact, it is a particular case of a general family of single-pass schemes known as generalized Newmark-type schemes. To obtain it, the values of $\hat{\mathbf{\Phi}}$ and its derivative with respect to time are approximated by the following expressions:

$$\hat{\mathbf{\Phi}}^{n+1} = \hat{\mathbf{\Phi}}^{n} + \triangle t . \overset{\bullet}{\hat{\mathbf{\Phi}}}^{n} + \beta \triangle t \triangle \overset{\bullet}{\hat{\mathbf{\Phi}}}^{n} \tag{86}$$

and

$$\overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n+1} = \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} + \triangle \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} \tag{87}$$

where

$$\overset{\bullet}{\hat{\boldsymbol{\Phi}}} = \frac{d\,\hat{\boldsymbol{\Phi}}}{dt} \tag{88}$$

being $\beta$ a parameter ranging from 0 to 1.

$$0 \leq \beta \leq 1 \tag{89}$$

Writing now the discretized equation in space at time $t_{(n+1)}$,

$$\mathbf{K}.\hat{\boldsymbol{\Phi}}^{n+1} - \mathbf{f}^{n+1} + \mathbf{C}.\overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n+1} = \mathbf{0} \tag{90}$$

and substituting the values of $\hat{\boldsymbol{\Phi}}^{n+1}$ and its derivative into it, we arrive at:

$$(\mathbf{C}+\beta \triangle t.\mathbf{K}) \triangle \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} = \mathbf{f}^{n+1} - \left( \mathbf{C}.\overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} + \mathbf{K}.\left( \hat{\boldsymbol{\Phi}}^{n} + \triangle t \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} \right) \right) \tag{91}$$

which can be expressed in a more compact way as:

$$(\mathbf{C}+\beta \triangle t.\mathbf{K}) \triangle \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} = \boldsymbol{\Psi}^{n+1} \tag{92}$$

where $\boldsymbol{\Psi}^{n+1}$ is

$$\boldsymbol{\Psi}^{n+1} = \mathbf{f}^{n+1} - \mathbf{K}.\hat{\boldsymbol{\Phi}}^{n+1,pred} - \mathbf{C}.\overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n+1,pred} \tag{93}$$

and $\hat{\boldsymbol{\Phi}}^{n+1,pred}, \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n+1,pred}$ the approximated values of $\boldsymbol{\Phi}$ and its derivative with respect to time at $t^{n+1}$

$$\hat{\boldsymbol{\Phi}}^{n+1,pred} = \hat{\boldsymbol{\Phi}}^{n} + \triangle t \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} \tag{94}$$

$$\overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n+1,pred} = \overset{\bullet}{\hat{\boldsymbol{\Phi}}}^{n} \tag{95}$$

The resulting scheme is reduced, for certain values of the parameter $\beta$ to schemes such as:

| $\beta$ |  | scheme |
|---|---|---|
| 0 |  | Forward Euler |
| 1 |  | Backward Euler |
| $\frac{1}{2}$ |  | Crank-Nicolson |
| $\frac{1}{3}$ |  | Galerkin (time) |

(96)

For values of the parameter $\beta \geq 0.5$ the scheme is unconditionally stable, being in the other cases conditionally stable, which implies that for values of the time increment greater than a certain critical value, oscillations will appear that will grow with time

In the particular case of $\beta = 0$, the resulting scheme is the explicit Euler one, in which the matrix $\mathbf{C}$ can be easily inverted if a diagonal representation is used, which consists of replacing it with a diagonal matrix whose terms are the sum of those of each row of the original matrix (Lumped mass matrix):

$$C_{Li,i} = \sum_{j=1}^{n} C_{i,j} \tag{97}$$

In this case, it is immediate to invert the matrix $\mathbf{C}_L$ .

To start the algorithm it is necessary to know the value of $\hat{\mathbf{\Phi}}^0$, as well as estimate the value of its derivative with respect to time $\overset{\bullet}{\hat{\mathbf{\Phi}}}{}^0$ , using, for example,

$$\mathbf{K}.\hat{\mathbf{\Phi}}^0 - \mathbf{f}^0 + \mathbf{C}.\overset{\bullet}{\hat{\mathbf{\Phi}}}{}^0 = \mathbf{0} \tag{98}$$

from where

$$\overset{\bullet}{\hat{\mathbf{\Phi}}}{}^0 = \mathbf{C}^{-1}.\left(\mathbf{f}^0 - \mathbf{K}.\hat{\mathbf{\Phi}}^0\right) \tag{99}$$

## 4.2 Finite elements for the 1d linear convective transport problem

So far, we have studied some explicit FD schemes for the 1D convective transport equation. We have studied the stability of the FTCS and the Lax Wendroff schemes, and found that FTCS schemes were unconditionally unstables.

Here we will consider first the problems found when applying the classical Galerkin approximation for the simple 1D convective problem, finding it unconditionally stable. Indeed, when using a 1D mesh of equally spaced nodes both FTCS finite differences and elements will result on the same set of discrete equations.

The objective of this section is to present some Finite Element techniques for convective transport problems. We recommend the material given in the following references: [Cho90], [Far82], [Gui03], [Hir88], [Lev92], [Tor97], [Tor01] y [Zie00b].

**Classical Galerkin approximation: a fundamental problem** First of all, we will present a simple 1D example where we will see how classical Galerkin Finite Elements are unconditionally unstable when applied to the scalar convective transport equation.

The mathematical model describing the convective transport of a magnitude $\phi$ along a 1D channel where the velocity $u$ is constant is

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} = 0 \tag{100}$$

with the initial condition

$$\phi\left(x, 0\right) = 0 \quad 0 \leq x \leq L \tag{101}$$

and the boundary condition

$$\phi\left(0, t\right) = 0 \quad 0 < t \leq T \tag{102}$$

The analytical solution is sketched in Figure 11.

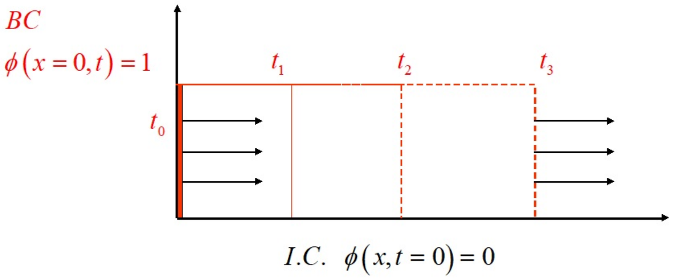

$BC$
$\phi\left(x=0,t\right)=1$
$t_1$
$t_2$
$t_3$
$t_0$
$I.C.\ \ \phi\left(x,t=0\right)=0$

Figure 11: Analytical solution.

We will discretize the problem using the simple mesh consisting of 5 nodes and 4 elements which can be seen in figure 12 below. The element size is constant and equal to $h = L/nelem = L/4$. The time step is denoted by $\Delta t$ as usual.

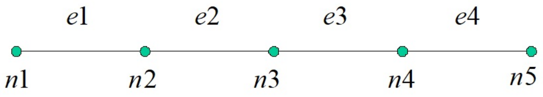


Figure 12: One dimensional finite element mesh for the convective transport problem.

We will define the global shape functions $N_j\left(x\right) \quad j = 1, ..5$, and introduce the nodal variables $\widehat{\phi}_j\left(t\right)$ which will be used to approximate the solution as:

$$\phi\left(x, t\right) \approx \hat{\phi}\left(x, t\right) = \sum_{j=1}^{5} N_j\left(x\right)\, \widehat{\phi}_j\left(t\right) \tag{103}$$

or, in a more compact manner:

$$\hat{\phi}\left(x, t\right) = \mathbf{N}.\widehat{\Phi} \tag{104}$$

where we have introduced the vectors of global shape functions $\mathbf{N}$ and nodal unknowns $\widehat{\Phi}$.

The initial condition reads

$$\widehat{\Phi} = 0 \tag{105}$$

The boundary condition of the proposed example consists, simply, on making

$$\widehat{\phi}_1(t) = 1 \quad 0 < t \tag{106}$$

Following Galerkin method, we will introduce the error or residual $R_\Omega$ as

$$R_\Omega = \frac{\partial \hat{\phi}}{\partial t} + u \frac{\partial \hat{\phi}}{\partial x} \tag{107}$$

and substitute 103 in it

$$R_\Omega = \frac{\partial}{\partial t}\left(\sum_{j=1}^{5} N_j(x)\, \hat{\Phi}_j(t)\right) + u \frac{\partial}{\partial x}\left(\sum_{j=1}^{5} N_j(x)\, \hat{\Phi}_j(t)\right) \tag{108}$$

from where we obtain

$$R_\Omega = \sum_{j\,=\,1}^{5} N_j(x) \frac{d\hat{\Phi}_j}{dt} + \sum_{j\,=\,1}^{5} v\, \hat{\Phi}_j \frac{dN_j}{dx} \tag{109}$$

The Galerkin method consists of obtaining the unknowns $\hat{\Phi}_j(t)$ using the equations:

$$\int_\Omega N_i\, R_\Omega d\Omega = 0 \tag{110}$$

which provides us with the same number of equations than unknowns, taking into account that in node 1 the nodal value is known.

Equation 110 can be interpreted as making the error orthogonal to the subspace where we are building the approximation. When approximating functions, the best approximation to a given function using a certain basis is such that the residual or error is orthogonal to all vectors of the subspace.

If we further develop 110 using 109, we obtain:

$$\int_\Omega N_i \left(\sum_{j=1}^{5} N_j(x) \frac{d\hat{\Phi}_j}{dt}\right) d\Omega = -\int_\Omega N_i \left(\sum_{j=1}^{5} u\, \hat{\Phi}_j \frac{dN_j}{dx}\right) d\Omega \tag{111}$$

from where:

$$\left(\int_\Omega N_i N_j d\Omega\right) \frac{d\hat{\Phi}_j}{dt} = -u \left(\int_\Omega N_i \frac{dN_j}{dx} d\Omega\right) \hat{\Phi}_j \tag{112}$$

This equation can be written in matrix form as:

$$\mathbf{M}\frac{d\hat{\mathbf{\Phi}}}{dt} = -u\,\mathbf{H}\,\hat{\mathbf{\Phi}} \tag{113}$$

The mass matrix $\mathbf{M}$ and the discrete convective matrix $\mathbf{H}$ are obtained by assembling the contributions of all elements in the mesh. All element matrices are equal:

$$\mathbf{M}^{(e)} = \frac{h}{6}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \mathbf{H}^{(e)} = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \tag{114}$$

The assembled matrices are:

$$\mathbf{M} = \frac{h}{6}\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix} \tag{115}$$

and

$$\mathbf{H} = \frac{1}{2}\begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \tag{116}$$

Next, we will discretize in time equation 113 considering a series of time stations $t^0,\ t^1, ..t^n$ with $t^n = t^0 + n\Delta t$

The value of the vector of unknowns at time $n$ will be denoted $\hat{\mathbf{\Phi}}^{(n)}$. We will introduce a simple forwards approximation of the time derivative:

$$\frac{d\hat{\Phi}^n}{dt} = \frac{\hat{\Phi}^{n+1} - \hat{\Phi}^n}{\Delta t} \tag{117}$$

which will be substituted in 113 to yield:

$$\hat{\Phi}^{n+1} = \hat{\Phi}^n - \Delta t\, u\, M^{-1} H \hat{\Phi}^n \tag{118}$$

or

$$\begin{aligned} \hat{\Phi}^{n+1} &= \left(I - \Delta t\, u\, M^{-1} H\right)\hat{\Phi}^n \\ \hat{\Phi}^{n+1} &= A\,\hat{\Phi}^n \end{aligned} \tag{119}$$

where $\mathbf{I}$ is the identity matrix of order 5 and $\mathbf{A}$ the iteration matrix. The scheme is said to be explicit as the matrix of coefficients $\mathbf{M}$ is the mass matrix. Indeed, this problem can be solved using a Jacobi iteration scheme. Usually, a reasonably accurate solution is obtained with 3-5 iterations.

We will start with an initial solution $\hat{\Phi}^0 = (1, 0, 0, 0, 0)^T$ , and choose an increment of time $\Delta t$, obtaining:

$$\begin{aligned} \hat{\Phi}^1 &= A\hat{\Phi}^0 \\ \hat{\Phi}^2 &= A\hat{\Phi}^1 \\ &\dots \\ \hat{\Phi}^{n+1} &= A\hat{\Phi}^n \end{aligned} \tag{120}$$

The results present important oscillations which grow up with time which can be seen in figures 13 and 14.

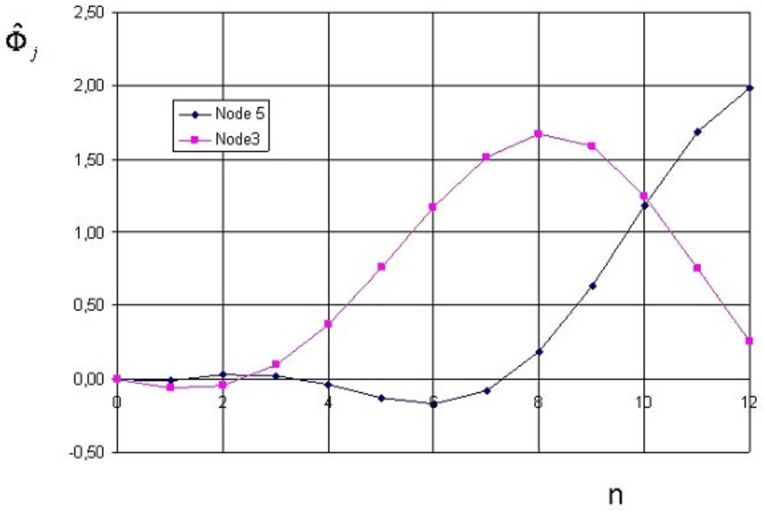


Figure 13: Evolution of concentration at nodes 3 and 5 as a function of time step.

The former shows how the solution evolves with time at two control nodes, while the latter gives the concentration in the domain at two different times. The analytical solution is a step function, located at $x_s = n\Delta t$

This type of behaviour will be obtained no matter the increment of time used.

The reason is that the proposed scheme is unconditionally unstable, i.e., it will not converge for any value of .

A simple proof of why the error is growing can be obtained by considering the scheme given in eqn. 119. As the exact solution will fulfil this equation, we can write:

$$\bar{\hat{\Phi}}^{n+1} = A\bar{\hat{\Phi}}^n \tag{121}$$

where $\bar{\hat{\Phi}}^n$ is the exact value at time $n$. If we subtract from this equation 119, we find that the error at times $n+1$ and $n$ are related by the same numerical scheme we are

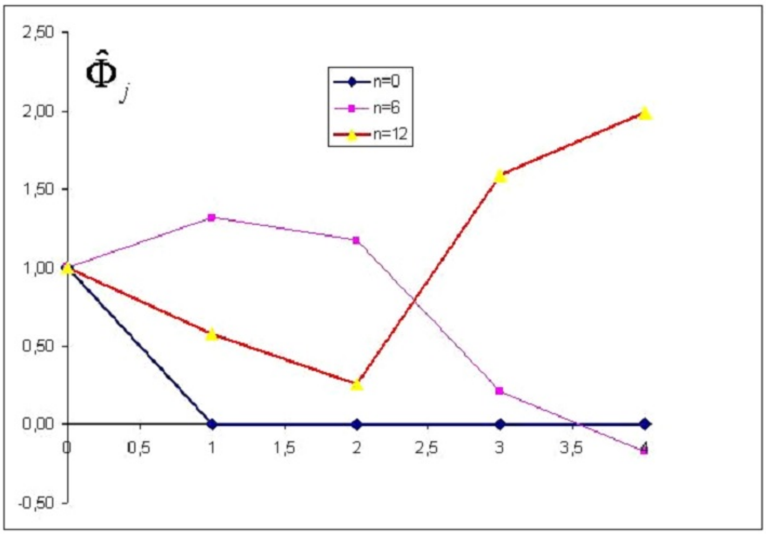


Figure 14: Concentration in the mesh at time steps 0, 6 and 12.

using to obtain the solution

$$\hat{\varepsilon}^{n+1} = A\hat{\varepsilon}^{n} \tag{122}$$

It can be shown that the necessary and sufficient condition for the error not to grow is that the moduli of all the eigenvalues have to be smaller than unity. In the case we are considering here, there are complex eigenvalues with their modulus larger than one.

Sometimes, in order to save computer time, the consistent mass matrix $\mathbf{M}$ is approximated by a diagonal matrix with diagonal terms which are the sum of all the coefficients in the same row. This diagonal matrix is referred to as "lumped mass matrix" or $\mathbf{M}_L$ .

In our case, $\mathbf{M}_L$ is obtained immediately from $\mathbf{M}$ as

$$\mathbf{M}_L = \frac{h}{6}\begin{pmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \tag{123}$$

The iteration matrix $\mathbf{A}$ has the eigenvalues $\{1 \pm 0.3536, \quad 1.0\,(triple)\}]$ and it is given by:

$$A \quad = \quad \left(I - \Delta t\, v\, M_L^{-1} H\right) \; = \tag{124}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} - \frac{u\Delta t}{2h} \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

The iteration equation for node 3 is

$$\phi_j^{n+1} = \phi_j^n - \frac{u\,\Delta t}{h}\left(\frac{\phi_{j+1}^n - \phi_{j-1}^n}{2}\right) \quad \text{with} \quad j = 2 \tag{125}$$

which is exactly the finite difference equation obtained for the scheme FTCS.

**The basic Taylor Galerkin algorithm** There exist today a variety of finite element methods for the discretization of the convective transport equation. Among them, it is worth mentioning the Taylor Galerkin schemes proposed by Donea, Peraire, Zienkiewicz and Morgan, and the Characteristic based Galerkin methods, introduced by Zienkiewicz, Codina and Ortiz . The classic text by Zienkiewicz and Taylor [Zie00b] presents an excellent state of art of these methods.

Here, we will focus in the Taylor-Galerkin method, which we will apply to solve both the linear transport equation and the quasi linear Burger equation.

The method is similar to the above described Lax Wendroff scheme, as it does a Taylor series expansion on time around $t_n$.The difference with the Lax Wendroff method arises in the second part, where the space derivatives are discretized. We will consider here the 1D equation including a source term of intensity $S$ which we have written in conservative form as:

$$\frac{\partial}{\partial t}\phi + \frac{\partial}{\partial x}F \; = S \tag{126}$$

with $F = u\phi$ in the case of the convective transport equation and $F = \left(\phi^2/2\right)$ in the Burger equation.

The Taylor series expansion up to second order results on

$$\phi^{n+1} = \phi^n + \Delta t \left.\frac{\partial \phi}{\partial t}\right|^n + \frac{1}{2}\Delta t^2 \left.\frac{\partial^2 \phi}{\partial t^2}\right|^n \tag{127}$$

Then, we substitute the first order partial derivative with respect to time using the PDE:

$$\left.\frac{\partial \phi}{\partial t}\right|^n = \left.\left(-\frac{\partial}{\partial x}F + S\right)\right|^n \tag{128}$$

and obtain the second order derivative with respect to time as

$$\left.\frac{\partial^2 \phi}{\partial t^2}\right|^n = \frac{\partial}{\partial t}\left(-\frac{\partial}{\partial x}F + S\right)\Bigg|^n \tag{129}$$

which consists of the terms

$$\frac{\partial}{\partial t}\left(-\frac{\partial}{\partial x}F\right) = -\frac{\partial}{\partial x}\frac{\partial}{\partial t}F \tag{130}$$

and,

$$\frac{\partial S}{\partial t} \tag{131}$$

The derivatives of the flux and source terms are:

$$\begin{aligned} \frac{\partial}{\partial t}F &= \frac{\partial F}{\partial \phi}\frac{\partial \phi}{\partial t} = A\frac{\partial \phi}{\partial t} \\ \frac{\partial}{\partial t}S &= \frac{\partial S}{\partial \phi}\frac{\partial \phi}{\partial t} = B\frac{\partial \phi}{\partial t} \end{aligned} \tag{132}$$

where $A$ and $B$ are the derivatives of the flux $F$ and source terms with respect to the unknown $\phi$. ( $A = 0$ or $A = \phi$ in the two cases being analyzed)

Next, we will substitute in (132) the time derivative of $\phi$ :

$$\begin{aligned} \frac{\partial F}{\partial t} &= A\left(-\frac{\partial F}{\partial x} + S\right) \\ \frac{\partial S}{\partial t} &= B\left(-\frac{\partial F}{\partial x} + S\right) \end{aligned} \tag{133}$$

and introduce both expressions in (129):

$$\frac{\partial^2 \phi}{\partial t^2} = \frac{\partial}{\partial x}\left\{A\left(\frac{\partial F}{\partial x} - S\right)\right\} + B\left(-\frac{\partial F}{\partial x} + S\right) \tag{134}$$

From here, substituting (128) and (134) in the Taylor series expansion ( 127) we arrive to:

$$\phi^{n+1} = \tag{135}$$

$$\phi^n - \Delta t\left(\frac{\partial F}{\partial x} - S\right)\Bigg|^n + \frac{\Delta t^2}{2}\left\{\frac{\partial}{\partial x}\left[A\left(\frac{\partial F}{\partial x} - S\right)\right] + B\left(-\frac{\partial F}{\partial x} + S\right)\right\}^n$$

Once we have discretized on time the convective transport equation, arriving to 4.2, the discretization in space is performed using standard Galerkin Finite Elements.

We will present the space discretization of the advective terms, leaving to the reader as an exercise the extension to the case with sources.

We will start by approximating $\phi\left(x,t\right)$ as:

$$\hat{\phi}\left(x,t\right)=N_j\left(x\right)\ \hat{\phi}_j\left(t\right) \tag{136}$$

where $\hat{\phi}_j\left(t\right)$ is the value of the unknown $\phi$ at node $j$ and time $t$, and $N_j\left(x\right)$ is the correspondent shape function. If we introduce the increment of the unknown $\Delta\hat{\phi}_j^n = \hat{\phi}_j^{n+1}-\hat{\phi}_j^n$ the result is:

$$\left(N_i,\Delta\hat{\phi}^n\right)=-\Delta t\ \left(N_i,\left.\frac{\partial F}{\partial x}\right|^n\right)+\frac{\Delta t^2}{2}\left(N_i,\frac{\partial}{\partial x}\left(A\ \left.\frac{\partial F}{\partial x}\right|^n\right)\right) \tag{137}$$

where we have used the notation

$$\left(f,g\right)=\int_\Omega f\,g\,d\Omega \tag{138}$$

From above equation, and taking into account the following expressions

$$\Delta\hat{\phi}\left(x,t\right)=N_j\left(x\right)\ \Delta\hat{\phi}_j\left(t\right) \tag{139}$$

$$\left.\frac{\partial F}{\partial x}\right|^n=\frac{\partial N_j\left(x\right)}{\partial x}\ \widehat{F}_j^n \tag{140}$$

where $\widehat{F}_j^n$ is the value of the flux $F$ at node $j$ and time $t_n$,we arrive, after integrating by parts, to:

$$\begin{aligned}\left(N_i,N_j\right)\Delta\hat{\phi}_j^n \quad = \quad & -\Delta t\ \left(N_i,\frac{\partial N_j\left(x\right)}{\partial x}\right)\widehat{F}_j^n \\ & -\frac{\Delta t^2}{2}\left(\frac{\partial N_i\left(x\right)}{\partial x}A\frac{\partial N_j\left(x\right)}{\partial x}\right)\hat{\phi}_j^n \\ & +\frac{\Delta t^2}{2}\int_{\partial\Omega}N_iA\ \left.\frac{\partial F}{\partial n}\right|^n d\Gamma\end{aligned} \tag{141}$$

where $n$ is the unit normal vector at the boundary $\partial\Omega$.

The term $\left.\frac{\partial F}{\partial n}\right|^n$can be evaluated at every element belonging to the boundary.

The discretized system can be written in matrix form as:

$$\mathbf{M}\,\Delta\hat{\phi}^n=-\Delta t\,\mathbf{H}\,\widehat{F}^n-\frac{\Delta t^2}{2}\mathbf{K}\,\hat{\phi}^n\ +\frac{\Delta t^2}{2}\mathbf{f} \tag{142}$$

where

$$M_{ij}=\left(N_i,N_j\right) \tag{143}$$

$$H_{ij}=\left(N_i,\frac{\partial N_j\left(x\right)}{\partial x}\right) \tag{144}$$

$$f_i = \frac{\Delta t^2}{2} \int_{\partial\Omega} N_i A \left. \frac{\partial F}{\partial n} \right|^n d\Gamma \tag{145}$$

The boundary conditions are directly applied at nodal points belonging to the boundary where the velocity enters the domain.

### 4.2.1 The two steps Taylor Galerkin algorithm

An alternative, which in the case of systems of PDEs reduces the computational cost is the two steps algorithm proposed by Peraire et al. [Per86], which is similar to a 2nd order Runge-Kutta scheme.

In the first step, a Taylor series expansion of first order is performed, and the value of $\phi$ at time $t^{n+1/2}$is obtained as:

$$\phi^{n+\frac{1}{2}} = \phi^n - \frac{\Delta t}{2} \left. \left( \frac{\partial F}{\partial x} - S \right) \right|^n \tag{146}$$

Once $\phi^{n+\frac{1}{2}}$ is known, we can obtain the values of the flux and source terms at $t^{n+1/2}$as:

$$\begin{aligned} F^{n+\frac{1}{2}} &= F^n + \frac{\Delta t}{2} \frac{\partial F}{\partial t} \\ &= F^n + \frac{\Delta t}{2} A \left( -\frac{\partial F}{\partial x} + S \right)^n \end{aligned} \tag{147}$$

and

$$S^{n+1/2} = S^n + \frac{\Delta t}{2} B \left( -\frac{\partial F}{\partial x} + S \right)^n \tag{148}$$

From here, we obtain:

$$A \left( -\frac{\partial F}{\partial x} + S \right)^n = \frac{F_x^{n+\frac{1}{2}} - F_x^n}{\Delta t/2} \tag{149}$$

and

$$B \left( \frac{\partial F}{\partial x} + S \right)^n = \frac{S^{n+\frac{1}{2}} - S^n}{\Delta t/2} \tag{150}$$

which substituted in (4.2), give:

$$\begin{aligned} \phi^{n+1} = \phi^n + \Delta t \left. \left( -\frac{\partial F_x}{\partial x} + S \right) \right|^n \\ + \frac{\Delta t^2}{2} \left( \frac{\partial}{\partial x} \left( - \frac{F^{n+\frac{1}{2}} - F^n}{\Delta t/2} \right) + \left( \frac{S^{n+\frac{1}{2}} - S^n}{\Delta t/2} \right) \right) \end{aligned} \tag{151}$$

from where we obtain:

$$\phi^{n+1} = \phi^n + \Delta t \left\{ -\frac{\partial}{\partial x} F^{n+\frac{1}{2}} + S^{n+\frac{1}{2}} \right\} \tag{152}$$

This equation can be discretized using the method of Galerkin, The result is

$$\left(N_i, \hat{\phi}^{n+1}\right) = \left(N_i, \hat{\phi}^n\right) - \Delta t \left(N_i, \frac{\partial}{\partial x} F^{n+\frac{1}{2}}\right) - \Delta t \left(N_i, S^{n+\frac{1}{2}}\right) \tag{153}$$

From here, integrating by parts the term $-\Delta t \left(N_i, \frac{\partial}{\partial x} F^{n+\frac{1}{2}}\right)$ ,we obtain:

$$\begin{aligned} \left(N_i, \hat{\phi}^{n+1}\right) = \left(N_i, \hat{\phi}^n\right) + \Delta t \left(\frac{\partial}{\partial x} N_i, F^{n+\frac{1}{2}}\right) \\ -\Delta t \left(N_i, S^{n+\frac{1}{2}}\right) + \Delta t \int_{\partial\Omega} N_i . \frac{\partial}{\partial n} F^{n+\frac{1}{2}} d\Gamma \end{aligned} \tag{154}$$

which can be written in a more compact form as,

$$\begin{aligned} \mathbf{M}.\mathbf{\Delta}\hat{\phi}^{n+1} = \\ \Delta t \left\{ \int_\Omega \frac{\partial \mathbf{N}}{\partial x} F^{n+\frac{1}{2}} d\Omega - \int_\Omega \mathbf{N}\, S^{n+\frac{1}{2}} d\Omega + \int_{\partial\Omega} \mathbf{N}. \frac{\partial}{\partial n} F^{n+\frac{1}{2}} d\Gamma \right\} \end{aligned} \tag{155}$$

The systems of equations is of both the 1step and the 2 steps Taylor Galerkin are of the type

$$\mathbf{M}.\mathbf{x} = \mathbf{f} \tag{156}$$

which can be solved using a Jacobi-like iterative method as,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{M}_L^{-1} \left(\mathbf{f} - \mathbf{M}\,\mathbf{x}^{(k)}\right) \tag{157}$$

This scheme converges in very few iterations (3-5), and in consequence, the method is considered as explicit. In above equation, the superindex $(k)$ is the iterations counter. The algorithm uses as starting value $\mathbf{x}^{(0)} = \mathbf{0}$.

# References

[Cho90] A.J. Chorin and J.E. Marsden. *A mathematical Introduction to Fluid Mechanics*, Springer-Verlag, 1992.

[Far82] S.J. Farlow. *Partial Differential Equations for Scientists and Engineers*, John Wiley and Sons, 1982.

[Fle88] C.A. Fletcher. *Computational techniques for fluid dynamics. Volume 1-Fundamental and general techniques. Volume 2-Specific techniques for different flow categories.* Springer-Verlag, Berlin and New York, 1988.

[Gui03] V. Guinot. *Godunov-type schemes. An introduction for engineers*, Elsevier, 2003.

[Hir88] C. Hirsch. *Numerical Computation of Internal and External Flows*, Vol I and II, John Wiley and Sons, 1988.

[Hug87] T.J.R. Hughes. *The Finite Element Method. Linear static and Dynamic Finite Element Analysis*, Prentice-Hall Int. Ed. London, 1987.

[Lev92] R.J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhauser Verlag, 1992.

[Pre86] W.H. Press, B.P. Flannery, S.A. Teuloski and W.T. Vetterling. *Numerical Recipes: The Art of Scientific Computing,* Cambridge University Press, 1986.

[Roa98] P.J. Roache. *Fundamentals of computational fluid dynamics*, Albuquerque, NM: Hermosa Publishers, 1998.

[Smi78] G.D. Smith. *Numerical solution of partial differential equations*, Oxford University Press, Open University Set Book, 1978.

[Str89] J.C. Strikwerda. *Finite difference schemes and partial differential equations*, Wadsworth & Brooks. Cole, Pacific Grove, CA, 1989.

[Tor97] E.F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer Verlag, 1997.

[Tor01] E.F. Toro. *Shock-Capturing Methods for Free-Surface Shallow Flows*, John Wiley and Sons, 2001.

[Zie00a] O.C. Zienkiewicz and R.L. Taylor. *The Finite Element Method, Vol.I,* Butterworth and Heinemann, Oxford, 2000.

[Zie00b] O.C. Zienkiewicz and R.L. Taylor. *The Finite Element Method* Vol.3 Fluid Dynamics, Birkhäuser, 2000.

[Per86] J. Peraire, O.C. Zienkiewicz, K. Morgan. *Shallow water problems: A general explicit formulation* Vol.22, International Journal for Numerical Methods in Engineering, 1986.

# Practical aspects of the finite element method

**Manuel Pastor*, Pablo Mira**, J.A. Fernandez Merodo*****

** E.T.S. Ingenieros de Caminos, Universidad Politécnica de Madrid*
*Profesor Aranguren s/n, 28040 Madrid, SPAIN*
*** Centro de Estudios y Experimentación de Obras Públicas*
*CEDEX, MITMA, Alfonso XII, 3, 28014 Madrid, SPAIN*
*Pablo.Mira@cedex.es*
**** Instituto Geológico y Minero de España*
*Ríos Rosas, 23, 28003 Madrid, SPAIN*

*This paper deals with practical aspects of the use and implementation of the finite element method. These aspects frequently cause serious difficulties specially to new comers to finite element practice and may even be the source of important errors on the final results of computations. The aspects that will be analyzed are: - the computational aspects of bending - the reduced integration - the volumetric locking of incompressible situations - the patch test for mixed formulations - the solver of the system of equations.*

## 1 Introduction

The Finite Element method has become one powerful tool of analysis which is being used all over the world both in the industry and in research. Many young (and not so young) engineers are nowadays familiar with commercial codes such as ANSYS, ABAQVS, COSMOS..., just to mention a few. Beginners face some questions such as (i) Which element should I use? (ii) How many elements and how big (or small)?, (iii) Which material model? (iv) Is there any saving in using reduced integration?, and so on.

While the number of questions and doubts is huge, there are some important aspects of which all of us should be aware when running finite element codes. The purpose of this Chapter is to provide a pocket guide for travellers in this unknown country. Of course, it is just a pocket guide. We do not pretend otherwise! And they have been described in detail in classical guides, such as those of Bathe [Bat96], Hughes [Hug87], Irons and Shrive [IS83], Zienkiewicz and Taylor [ZT00].

Therefore, we will describe some pitfalls and difficulties in doing finite element computations, such as poor bending behaviour, locking and modes of zero energy when using reduced integration. We will also deal with topics as Babuska-Brezzi restrictions (the light version), and will comment on which solvers are easy to program and are efficient.

# 2 The Misteries of Bending

Not all elements perform as we wish when dealing with problems in which bending is important. For instance, linear triangles give much stiffer response of the structure than they should. Moreover, when obtaining natural frequencies of vibration, we will get higher values because of this extra stiffness. To understand the problem, we will consider the simple case of a square $[-1,1]x[-1,1]$ under pure bending conditions. We will discretize the domain with one bilinear quadrilateral, as depicted in Fig.1.

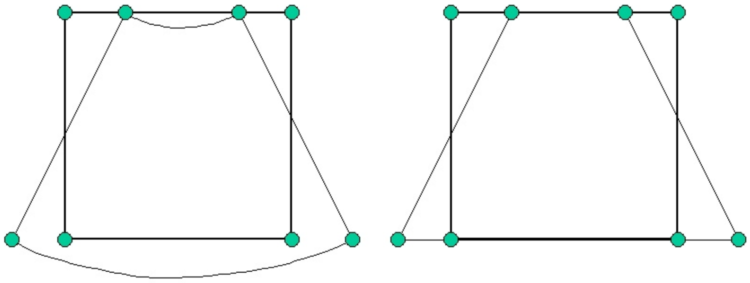


Figure 1: Bending of a bilinear quadrilateral

The solution of this plane stress problem is the displacement field

$$u = -\frac{M}{EI}xy \quad v = -\frac{M}{EI}\left(1 - x^2\right) \tag{1}$$

with horizontal displacements at the nodes given by $\pm u_0 = \frac{M}{EI}$.

$$u = -\frac{M}{EI}xy \quad v = -\frac{M}{EI}\left(1 - x^2\right) \tag{2}$$

The origin of coordinate axes have been taken at the centre of the square. The strain field is given by

$$\varepsilon_x = -\frac{M}{EI}y \;\; \varepsilon_y = 0 \;\; \gamma_{xy} = 0 \tag{3}$$

Should we reproduce the displacement field with the bilinear quadrilateral, we would obtain a constant strain field

$$\varepsilon_x^h = -\frac{M}{EI}y \;\; \varepsilon_y^h = 0 \;\; \gamma_{xy}^h = -\frac{M}{EI}x \tag{4}$$

It is important to note that the discretization has introduced an spurious shear strain which is only zero at the centre. Therefore, the element will be stiffer, and the deformation under a given moment $M$ will be smaller than it should. The reader can verify that if we discretize the square with two linear elements, the situation is the same.

## 3 Risks of Reduced Integration

Reduced integration consists on using an integration rule of smaller degree of precision than required with less integration points. In this way, we get two advanteges (i) The cost of computation -and therefore the time- is reduced. We can analyze larger problems in the same time or we reduce the computer time. (ii) We obtain better performance (sometimes) when computing limit loads.

Two popular reduced integration rules are: (I) One point for bilinear quadrilaterals (ii) Two by two points for 8 noded quadrilaterals. Let us consider the first case. The bilinear quadrilateral has 8 degrees of freedom, and the dimension of the stiffness matrix is $8x8$. The matrix has eigth eigenvectors which are sketched in Fig.2. The eigenvalues of the two translation and rotation modes are zero, but this mode of deformation with zero energy is prevented by boundary conditions avoiding rigid solid motions.

If we integrate the stiffness matrix using just one point of integration (the centre), the situation changes. The stiffness matrix is given by

$$\mathbf{K} = \int_\Omega \mathbf{B}^T.\mathbf{D}.\mathbf{B}\; d\Omega \tag{5}$$

which is computed as:

$$\mathbf{K} = \mathbf{B}_0^T.\mathbf{D}.\mathbf{B}_0\; .W_0 \tag{6}$$

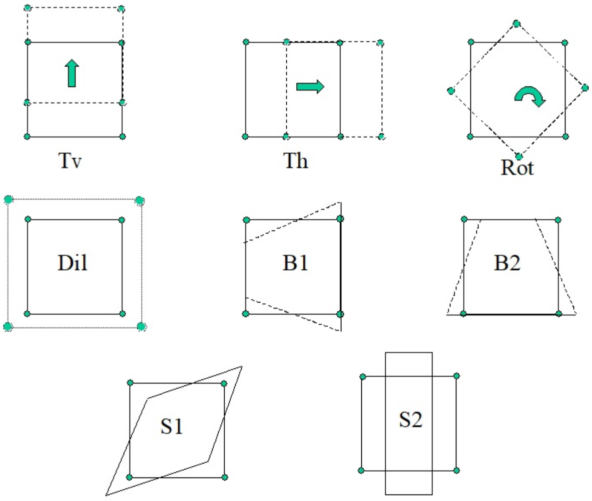


Figure 2: Eigenvectors of the stiffness matrix

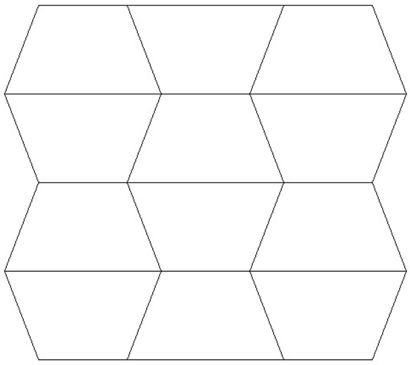


Figure 3: Hourglassing of 4 noded quadrilaterals

where the subindex 0 refers to the integration point. Dimensions of matrix $\mathbf{B}$ are $8x3$, and $\mathbf{D}$ is a $3x3$ matrix. Therefore, **the rank of $K$ is 3**. This means that, in addition to the 3 free energy modes, there are two more now. These modes are the ”bending” modes B1 and B2. A finite element mesh can in some conditions (for instance, poor conditioning of the equations system) exhibit an spurious mode called ”hourglassing” because of the shape of the deformed elements (Fig.3)

Reduced integration of 8 node quadrilaterals produce also another hourglass mode. The shape is sketched in Fig.4. The advantages of this under-integrated element are important, specially in computation of failure loads, and some finite element codes incorporate a ”horglass control” to warn the user this spurious mechanism is present.

# 4 Volumetric locking and failure loads

We will consider now the case of an incompressible material under plane strain conditions. The problem is sketched in Fig. 5

Boundary conditions are: (i) Prescribed zero horizontal and vertical displacements at the left side and the bottom, (ii) traction free rigth side and top, with a point load applied at the corner. If the material is incompressible, the node 4 cannot move in the vertical direction as it belongs to triangle $124$ which otherwise would change its volume, and it cannot move in horizontal as it belongs to triangle $143$. Therefore, the node cannot move. This can be repeated for node 6, which belongs to triangles $256$ and $264$, and for all the remaining nodes in the mesh. All the nodes are ”blocked”, which is unrealistic. This example can be reproduced with plane strain finite elements using Poisson ratios approaching $0.5$ (For instance, $0.49, 0.499, 0.4999$ and so on.

The reader should be aware of this problem when trying to model the behaviour of saturated soils under fast loading or undrained conditions, where values of Poisson ratio close to $0.5$ are usually chosen.

All displacement based finite elements present this problem to a certain extent. The elements performing better are higher order triangles (15 nodes). Other popular alternative is to use quadratical eight noded quadrilaterals with a reduced integration rule of $2x2$ points.

However, the best choices are assumed strain elements (Simo Rifai, for instance), or mixed displacement-pressure formulations. This techniques will be described later on in this book.

The reader should be aware that this problem is also exhibited in Plasticity, because at failure the flow rule imposes an additional condition on the rate of plastic strain (which can be zero). To ilustrate this problem let us analyse the problem of a footing on a vertical slope as sketched in figure 6. The material models used in the analysis are presented in the following table:

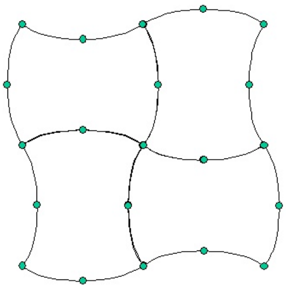


Figure 4: Hourglassing of 8 noded quadrilaterals

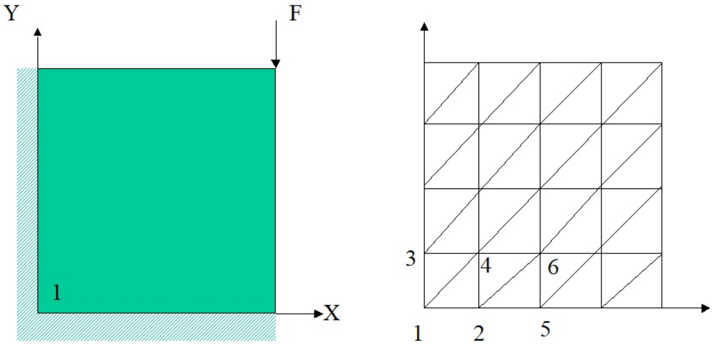


Figure 5: Volumetric Locking

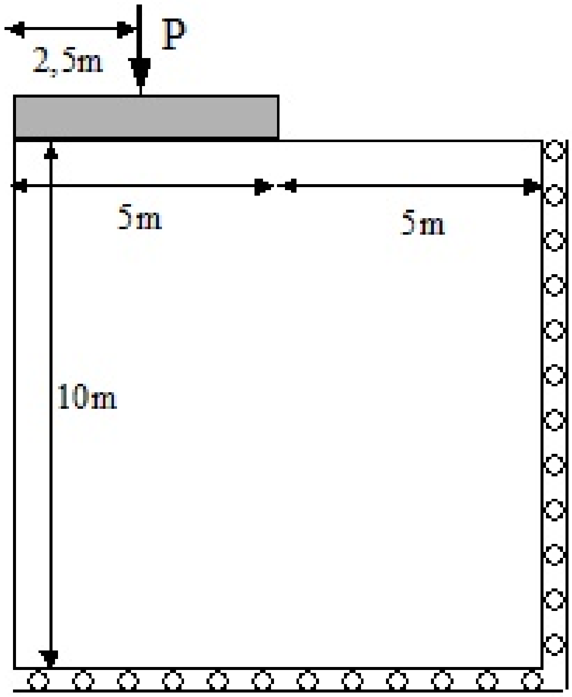


Figure 6: Footing on vertical slope

| | Material type | E(Pa) | $\upsilon$ | $\sigma_y$ (Pa) |
|---|---|---|---|---|
| Soil | Von Mises | 1.0E5 | 0.35 | 200.0 |
| Footing | Linear elastic | 1.0E8 | 0.35 | |

The objective of the analysis is to obtain a failure mechanism and a value for the limit load using different element types and different meshes. The different element types used in the analysis are listed in the following table:

| | |
|---|---|
| Standard 3 node displacement triangle | 3st0 |
| Standard 4 node displacement quadrilateral | 4st0 |
| $\overline{\mathbf{B}}$ 4 node quadrilateral | 4bb0 |
| Simo&Rifai 4 node quadrilaterals with 4 internal modes | 4sr0 |
| Simo&Rifai 4 node quadrilaterals with 5 internal modes | 4sr1 |
| Simo&Rifai 4 node quadrilaterals with 6 internal modes | 6sr4 |
| Standard 6 node displacement triangle | 6st0 |
| $\overline{\mathbf{B}}$ 6 node triangle | 6bb2 |
| Standard 7 node displacement triangle | 7st0 |
| $\overline{\mathbf{B}}$ 7 node triangle | 7bb2 |

A theoretical solution for this problem may be obtained based on limit analysis. The failure mechanism would be a shear band descending leftwards from the bottom right corner of the footing in a $45^o$ angle. Accordingly for each element type two mesh orientations were tested :

1) A so called right orientation (r) in which mesh alignment ran parallel to the direction of the expected failure mechanism

2) A so called wrong orientation in which mesh alignment ran perpendicular to the direction of the expected failure mechanism. (w)

The different meshes used are presented in figure 7.

As will be shown, this test appears to be very demanding. Poor performing elements exhibit not only significant differences in the load-displacement curves and different shear band widths but also significant changes in the failure mechanism direction depending on the mesh orientation. Load displacement diagrams are presented in figures 8 and 9. Failure mechanisms are presented in figures 10 and 11.

The worst performers are as always standard displacement elements. The best performers are Simo-Rifai and 7 node elements.

In the case of the poor performers a wrong mechanism appears which significantly prevails over the correct one. In the case of the good performers both mechanisms appear but the correct one prevails over the wrong one.

As we can see from the graphs the failure mechanism appears to be more important than the load-displacement curve since good performers appear to have more problems in the first one than in the second one.

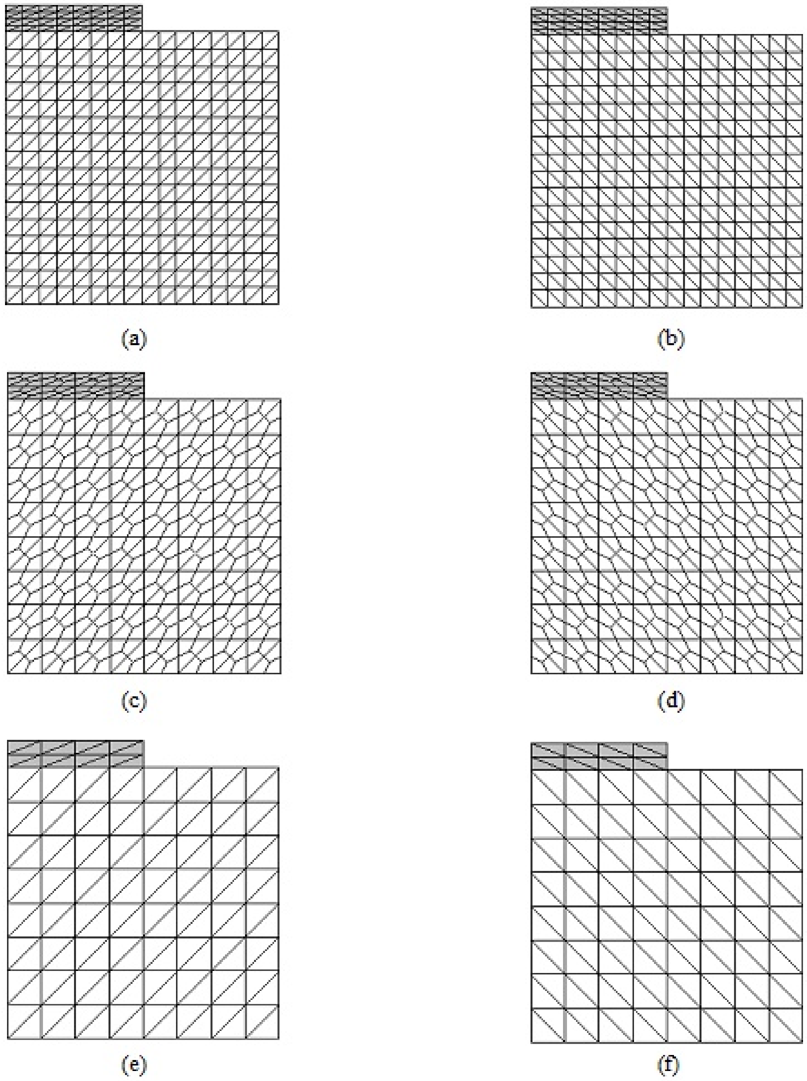


Figure 7: Meshes used to analyse footing on vertical slope depending on the number of nodes per element: (a) 3 nodes right orientation (b) 3 nodes wrong orientation (c) 4 nodes right orientation (d) 4 nodes wrong orientation (e) 6 and 7 nodes right orientation (f) 6 and 7 nodes wrong orientation

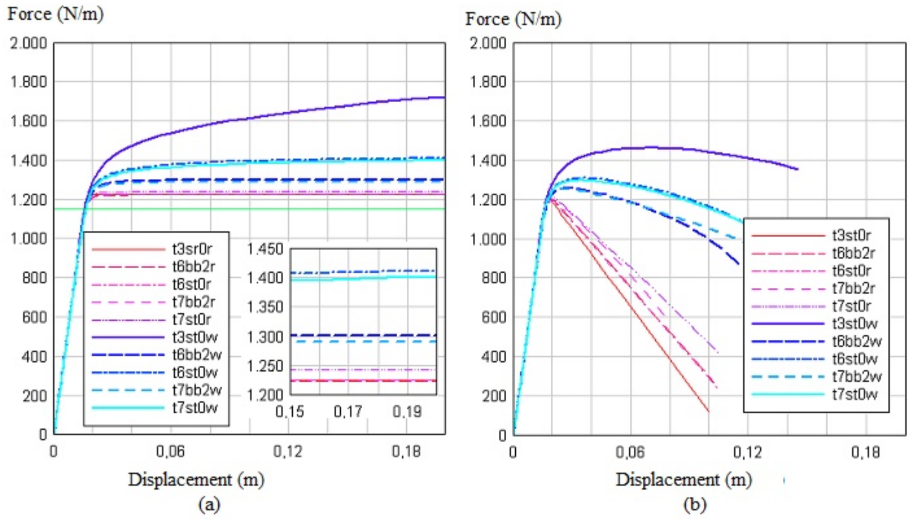


Figure 8: Force displacement diagrams for triangles (a) H=0.0 (b) H=-0.01E

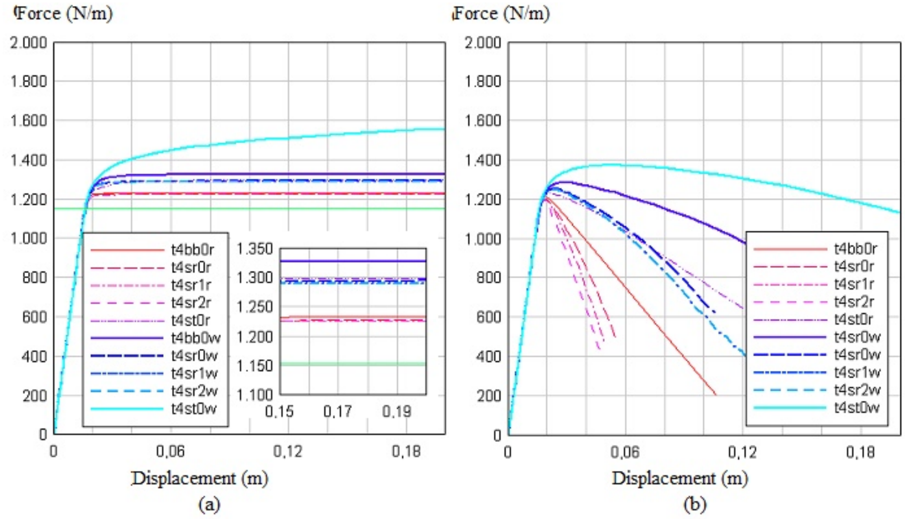
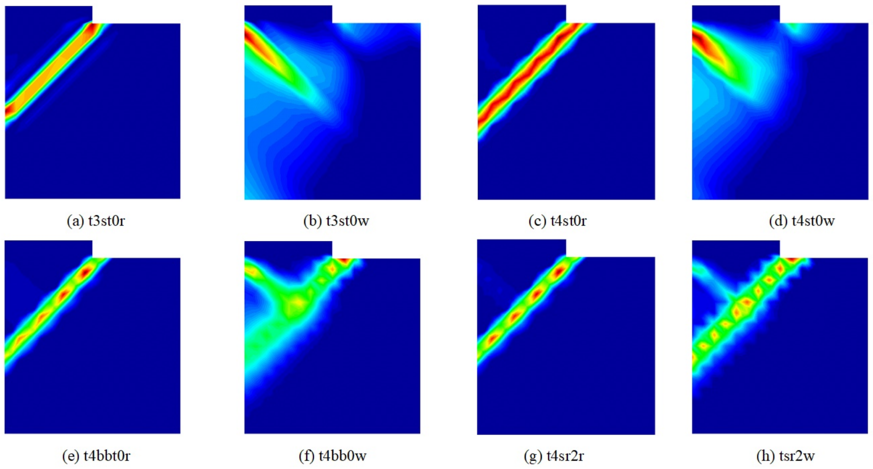


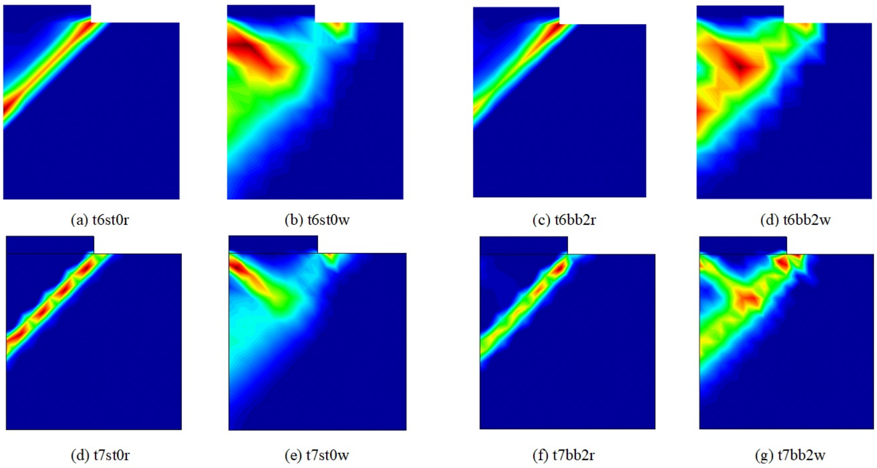
Figure 9: Force displacement diagrams for quadrilaterals (a) H=0.0 (b) H=-0.01E



(a) t3st0r (b) t3st0w (c) t4st0r (d) t4st0w

(e) t4bbt0r (f) t4bb0w (g) t4sr2r (h) tsr2w

Figure 10: Failure mechanisms for footing on vertical slope using 3 and 4 node elements and a perfectly plastic material



(a) t6st0r (b) t6st0w (c) t6bb2r (d) t6bb2w

(d) t7st0r (e) t7st0w (f) t7bb2r (g) t7bb2w

Figure 11: Failure mechanisms for footing on vertical slope using 6 and 7 node elements and a perfectly plastic material

## 5 Why are T3P3 and Q4P4 elements not a good idea for geotechnical analysis: the Zienkiewicz-Taylor patch test

Users of finite element codes for geotechnical analysis often wonder why the choice of elements is so limited. For instance, it is not possible to choose bilinear quadrilaterals or enhanced strain elements like those of Simo-Rifai, which are excellent for bending, quasi incompressibility, failure loads, etc. The reason is that in mixed displacement-pore pressure problems we cannot use the same shape functions for interpolation of both fields - unless we develop special stabilized elements-. Well, we can indeed, but if the permeability is very small (undrained conditions), the system of equations is of the form

$$\begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ -\mathbf{B}^{\,T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \xi \\ \phi \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{\xi} \\ \mathbf{f}\phi \end{bmatrix} \tag{7}$$

This is in detail similar to those found in mixed formulations of incompressible solid mechanics and fluid dynamics problems.

It will be demonstrated next that the system will be singular and present pressure oscillations whenever the number of $\xi$ variables $n_{\xi} \leq n_{\phi}$ ,i.e., the number of $\phi$ variables. Although this condition is not sufficient for stability (and solvability) it is neccessary and it generally excludes equal order of interpolation spaces for pressure and displacements.

A complete mathematical treatment of the problem can be found in Refs. [Bab73] and [Bre74]. However, a much simpler explanation is provided by the patch test for mixed formulations proposed by Zienkiewicz *et al.* [ZQTN86], which will be described later.

To illustrate the problem, let us consider a layer of saturated soil of infinite length and depth $L$ depicted in Fig.12, subjected to a harmonic distributed load on its surface.

The problem has been discretized using a column of 1 m. width with the following boundary conditions:

(i) At the top of the layer $y = L$ :

Prescribed pressure $p_w = 0$

Prescribed traction $\bar{t}_y = 100\exp(-i\omega t)$(Pa)

(ii) on the vertical sides

$$\frac{\partial p_w}{\partial x} = 0$$

$$\bar{t}_y \quad = 0 \qquad \text{and} \qquad u_x = 0$$

(iii) on the bottom

$$\frac{\partial p_w}{\partial y} = 0$$

$$u_x = u_y = 0$$

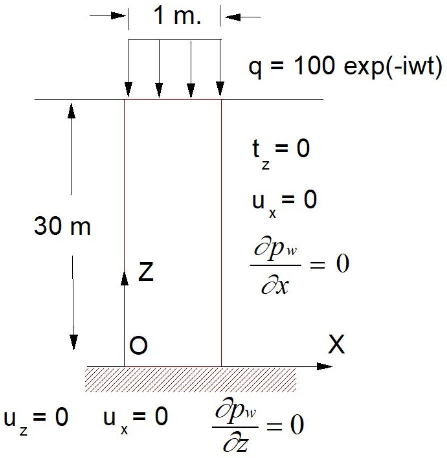


Figure 12: Boundary conditions for saturated soil column with periodic surface load ( $q = t_y = 100 \exp(-i\omega t)$ )

The material parameters chosen for the analysis are the following:

| | |
|---|---|
| $k_w$ | $10^{-7} m/s$ |
| $n$ | 0.333 |
| $E$ | 7.492 $10^8$ (Pa) |
| $\nu$ | 0.2 |
| $\rho_s$ | $2.0 \times 10^3$ $(N/m^3)$ |
| $\rho_w$ | $1.0 \times 10^3$ $(N/m^3)$ |

The height of the column has been taken as $L = 30\ m$ and the excitation frequency chosen is $\omega = 3.379$ rad/s.

The problem was discretized using 20 four node quadrilateral elements with bilinear shape functions for both pressure and displacements. As can be seen in the table, permeability is $10^{-7}$ ms$^{-1}$, so the column is close to undrained conditions.Two different compressibilities of water and solid grains will be considered: $10^4 MPa$ and $10^9 MPa.$ The results have been plotted in Fig.13 where it can be seen how spurious oscillations grow as compressibility (i.e. $1/Q^*$) decreases. If we choose now quadratic polynomials for displacements and bilinear for pressure, no oscillations appear, as it can be seen in Fig.14.

Therefore, even if some compressibility exists, it is not possible to use any combination of shape functions for pressure and displacements, as oscillations can appear as the undrained-incompressible limit is approached.

In the case of quadrilaterals, allowable interpolation functions are bilinear for pressure and quadratic for displacements. Similarly, quadratic displacement triangles with linear pressure are admissible (Fig.15).

To provide a rational explanation of why some combinations work while some others don't, we will follow that given in [ZQTN86] and begin by rewriting system (7) as

$$\begin{bmatrix} \mathbf{K} & -\mathbf{Q} \\ -\mathbf{Q}^{\,T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_u \\ \mathbf{f}_p \end{bmatrix} \tag{8}$$

from which we obtain, using the first equation of (8)

$$\mathbf{u} = \mathbf{K}^{-1}\mathbf{Q}.\mathbf{p} + \mathbf{K}^{-1}\mathbf{f}_u \tag{9}$$

Substituting this into the second equation of (8), we arrive at

$$-\mathbf{Q}^T \left(\mathbf{K}^{-1}\mathbf{Q}.\mathbf{p} + \mathbf{K}^{-1}\mathbf{f}_u\right) = 0$$

or

$$(\mathbf{Q}^T\mathbf{K}^{-1}\mathbf{Q})\mathbf{p} = -\,\mathbf{Q}^T\mathbf{K}^{-1}\mathbf{f}_u \tag{10}$$

which is a system of equations with a matrix of coefficients of dimension $(n_p\text{x}n_p)$ obtained by multiplication of three matrices of dimensions $(n_p\text{x}n_u)$, $(n_u\text{x}n_u)$ and $(n_u\text{x}n_p)$ where $n_u$ and $n_p$ are the number of degrees of freedom of displacements and pressures once suitable boundary conditions have been applied.

If $n_u < n_p$, the system is singular, and spurious oscillations in the pore pressure field will always appear. Therefore, the condition to be fulfilled is

$$n_u \geq n_p \tag{11}$$

for any assembly (or patch) of elements.

It is important to note that this is a necessary but not a sufficient condition and singularity has to be tested in all cases.

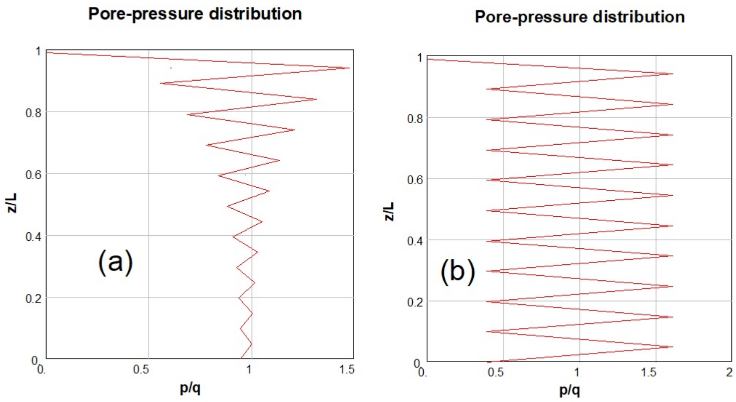


Figure 13: Amplitude of pore pressure for the soil column problem using 20 Q4P4 elements with $k = 10^{-7} m/s$ and (a) $Q^* = 10^4 MPa$ (b) $Q^* = 10^9 MPa$

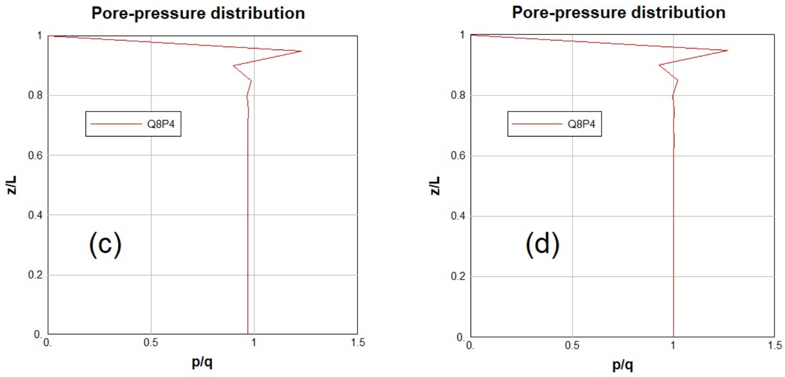


Figure 14: Amplitude of pore pressure for the soil column problem using 20 Q8P4 elements with $k = 10^{-7} m/s$ and (a) $Q^* = 10^4 MPa$ (b) $Q^* = 10^9 MPa$

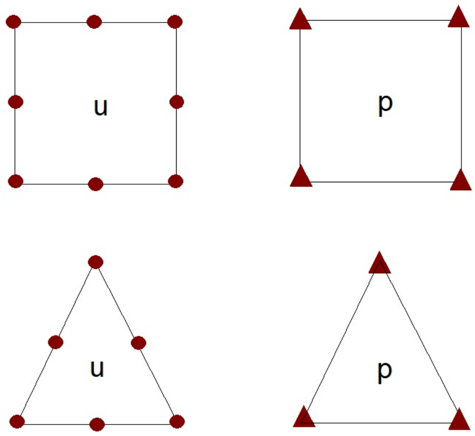


Figure 15: Some allowed elements

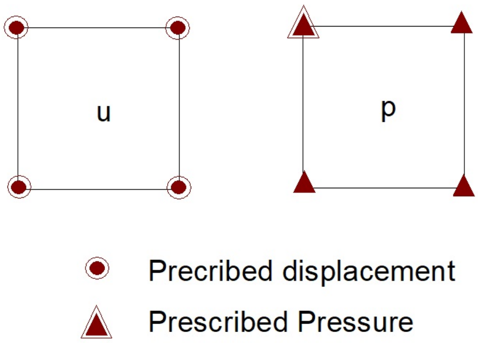


Figure 16: Bilinear quadrilateral with equal order of interpolation of displacement and pressure

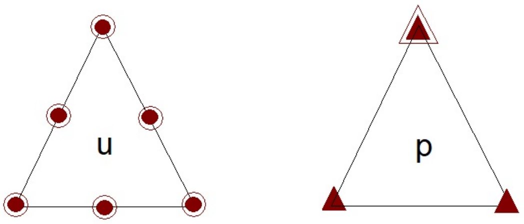


Figure 17: A single element patch of quadratic displacement - linear pressure triangle.

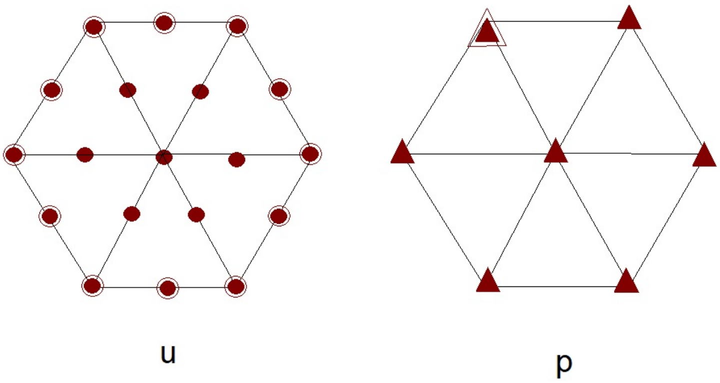


Figure 18: Patch of six T6P3 elements.

In order to illustrate the application of this condition, we will consider next the case of a quadrilateral with bilinear shape functions for both pressure and displacement (Fig.16).

We shall first use a single element patch. If the displacements have been prescribed on the boundary, and pore pressure has been fixed at one point we have zero degrees of freedom for displacements and three for pressures. The element does not pass the count conditions of the patch test as $n_u < n_p$ and the element will present oscillations in the undrained-incompressible limit as was shown in previous example. For larger patches obviously the same factors of the count will continue.

Examining the quadratic displacement triangle shown in Fig.17, a first patch consisting in one element alone does not pass the test, as, again, all displacement degrees of freedom have been fixed, and, therefore,

$$(n_u = 0) < (n_p = 2)$$

However, this a rather uncommon situation, and patches incorporating more elements do pass the patch test.

This is illustrated in Fig.18. Now, there are 7 free nodes in the interior of the patch, and 14 degrees of freedom for displacements, and 6 degrees of freedom for the pressure. Therefore,

$$(n_u = 14) > (n_p = 6)$$

and the count of the patch passes the test.

# 6 Our favourite solver (do we have any?)

Finite elements can give rise to huge systems of equations which we have to solve, perhaps many times if we are analyzing transient problems. If you are thinking of building your own finite element code, this is a crucial question: which solver should I use? And the answer depends on the problem.

In our opinion, iterative solvers like the preconditioned conjugate gradient or the Jacobi are excellent choices because they are really simple to program, especially if we are using a language like FORTRAN 90 or C. Another advantage is that they do not require any renumbering to save memory. We will describe the above-mentioned iterative solvers in the following section.

The choice of iterative methods mentioned above is in our opinion the best for really large problems (tens or hundreds of thousands of degrees of freedom). For large but not so large problems ( just thousands of degrees of freedom) a direct method with a special storage scheme for sparse problems is also a good choice and in general requires less numerical operations. Direct methods such as LU Gauss-Jordan factorization perform very satisfactorily under these circumstances. Additionally, they present the advantage of requiring relatively few changes to solve non symmetric systems, while conjugate gradient methods require special schemes such as GMRES to solve this type of problems. When using direct methods it is decissive to use sparse storage schemes, otherwise the computational cost and the required storage space are not affordable and round off errors reach unacceptable levels. A well known sparse storage scheme is the skyline method which will be presented in section 6.3. A more recent and more efficient storage scheme known as Harwell-Boeing is presented in section 6.4.

## 6.1 Conjugate Gradient Method with Preconditioning

One of the most effective and simple iterative methods (when used with preconditioning) for solving $\mathbf{Ax} = \mathbf{b}$ is the conjugate gradient algorithm. The algorithm is based on the idea that the solution of $\mathbf{Ax} = \mathbf{b}$ minimizes the total potential $\Pi = \frac{1}{2}\mathbf{x}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{b}$. Hence, the task in the iteration is, given an approximate $\mathbf{x}^k$ to $\mathbf{x}$ for wich the potential is $\Pi^k$, to find an improved approximation $\mathbf{x}^{k+1}$ for wich $\Pi^{k+1} < \Pi^k$. However, not only do we want the total potential to decrease in each iteration but we also want $\mathbf{x}^{k+1}$ to be calculated efficiently and the decrease in the total potential to occur rapidly. Then the iteration will converge fast.

In the conjugate gradient method, we use in the $k$th iteration the linearly independent vectors $\mathbf{p}^1, \mathbf{p}^2, \mathbf{p}^3, ..., \mathbf{p}^k$ and calculate the minimum of the potential in the space of the potential in the space spanned by these vectors. This gives $\mathbf{x}^{k+1}$. Also, we establish the additional basis vector $\mathbf{p}^{k+1}$ used in the subsequent iteration.

The algorithm can be summarized as follows.

Choose the starting iteration vector $\mathbf{x}^1$ (frequently $\mathbf{x}^1$ is the null vector).

Calculate the residual $\mathbf{r}^1 = \mathbf{b} - \mathbf{Ax}^1$. If $\mathbf{r}^1 = 0$ , quit.

Else:

Set $\mathbf{p}^1 = \mathbf{r}^1$.

Calculate for $k = 1, 2, ...,$

$$\begin{aligned}
\alpha^k &= \frac{\mathbf{r}^{k^T}\mathbf{r}^k}{\mathbf{p}^{k^T}\mathbf{A}\mathbf{p}^k} \\
\mathbf{x}^{k+1} &= \mathbf{x}^k + \alpha^k\mathbf{p}^k \\
\mathbf{r}^{k+1} &= \mathbf{r}^k - \alpha^k\mathbf{A}\mathbf{p}^k \\
\beta^k &= \frac{\mathbf{r}^{k+1^T}\mathbf{r}^{k+1}}{\mathbf{r}^{k^T}\mathbf{A}\mathbf{r}^k} \\
\mathbf{p}^{k+1} &= \mathbf{p}^{k+1} + \beta^k\mathbf{p}^k
\end{aligned} \tag{12}$$

We continue iterating until $\left\|\mathbf{r}^k\right\| \leq \varepsilon$, where $\varepsilon$ is the convergence tolerance. A convergence criterion on $\left\|\mathbf{x}^k\right\|$ could also be used.

The conjugate gradient algorithm satisfies two important orthogonality properties regarding the direction vectors $\mathbf{p}_i$ and the residual $\mathbf{r}_i$, namely, we have

$$\begin{aligned} \mathbf{p}_i^T \mathbf{A} \mathbf{p}_j^k &= 0 \qquad (13) \\ \mathbf{P}_j^T \mathbf{r}^{k+1} &= \mathbf{0} \qquad (14) \end{aligned}$$

where $\mathbf{P}_j = [\mathbf{p}_1, ..., \mathbf{p}_j]$ .

Convergence to the solution $\mathbf{x}$, in exact arithmetic, is achieved in at most $n$ iterations. Of course, in practice, we want convergence to be reached in much fewer than $n$ iterations.

The rate of convergence of the conjugate gradient algorithm depends on the condition number of matrix $\mathbf{A}$, defined as cond($\mathbf{A}$)=$\lambda_n/\lambda_1$, where $\lambda_1$ is the smallest eigenvalue and $\lambda_n$ is the largest eigenvalue of $\mathbf{A}$. The larger the condition number, the slower the convergence, and in practice, when the matrix is ill-conditioned, convergence can be very slow.

To increase the rate of convergence of the solution algorithm, preconditioning is used. The basic idea is that instead of solving $\mathbf{Ax} = \mathbf{b}$, we solve

$$\widetilde{\mathbf{A}}^{-1}\mathbf{Ax} = \widetilde{\mathbf{A}}^{-1}\mathbf{b} \qquad (15)$$

where $\widetilde{\mathbf{A}}$ is called the preconditioner. The objective with this transformation is to obtain a matrix $\widetilde{\mathbf{A}}^{-1}\mathbf{A}$ with a much improved conditioned number choosing an easy inverting matrix $\widetilde{\mathbf{A}}$. Various preconditioners have been proposed, the choose of the diagonal part of $\mathbf{A}$ results in the Jacobi

Conjugate Gradient method (JCG).

The new algorithm introduces an additional set of vectors $\mathbf{z}^k$ defined by:

$$\mathbf{z}^k = \widetilde{\mathbf{A}}^{-1}\mathbf{r}^k \qquad (16)$$

who modifies the definitions of $\alpha^k, \beta^k, \mathbf{p}^k$ :

$$\begin{aligned} \alpha^k &= \frac{\mathbf{z}^{k^T}\mathbf{r}^k}{\mathbf{p}^{k^T}\mathbf{A}\mathbf{p}^k} \\ \beta^k &= \frac{\mathbf{z}^{k+1^T}\mathbf{r}^{k+1}}{\mathbf{z}^{k^T}\mathbf{r}^k} \qquad (17) \\ \mathbf{p}^{k+1} &= \mathbf{z}^{k+1} + \beta^k\mathbf{p}^k \end{aligned}$$

## 6.2 GMRES iterative method

We considered in the previous section only the case of a symmetric coefficient matrix. It should be noted that finite element models for geomechanical problems frequently present non symmetric coefficient matrices. For non symmetric coefficient matrices, the conjugate gradient method has been generalized and other iterative schemes, notably, the generalized minimal residual (GMRES) method, have been developed and researched.

The GMRES method is an iterative method for the numerical solution of non-symmetric linear equation systems developed by Yousef Saad and Martin H. Schulz in 1986 . The method approximates the solution of the system based on the vector that minimizes the residual within the Krylov subspace associated to each iteration. The Krylov subspace is updated at every iteration. To search for that vector, the method uses Arnoldi iteration. The details may be found in [SS86]

In the context of finite element models for saturated soil mechanics, the use of the full Newton-Raphson method is common. This method involves updating the stiffness matrix at each iteration and solving the conrresponding system of linear equations. In this context, it is decisive to have an efficient method for solving systems of linear equations. This task becomes particularly complicated since it is common for these systems of equations to be poorly conditioned due to the presence of undrained conditions and quasi-incompressibility, as explained in the previous section. The most precise methods would be the direct methods, based on the Gaussian elimination algorithm, but they require a large amount of computer memory. Iterative methods constitute a good alternative, it is necessary to have good preconditioners that allow overcoming the greater sensitivity of iterative methods to poor conditioning of the system. The work of White and Borja [WB11] provides a very efficient preconditioning strategy based on the two blocks of the coupled formulation, the displacement block and the pore pressure block, with a different preconditioning strategy for each of them.

## 6.3 Skyline storage scheme

The skyline scheme consists of storing the stiffness matrix in a vector including the diagonal terms and the off diagonal terms between the non zero off diagonal term which is farthest from the diagonal in each row (or column) and the corresponding diagonal term. Additionally, it will be necessary to use an N component integer vector to store the direction in the stiffness vector where the last component of the ith row ( or column) is stored. Let us assume our stiffness matrix is the following one:

$$\begin{bmatrix} 2 & -2 & 0 & 0 & -1 \\ -2 & 3 & -2 & 0 & 0 \\ 0 & -2 & 5 & -3 & 0 \\ 0 & 0 & -3 & 10 & 4 \\ -1 & 0 & 0 & 4 & 10 \end{bmatrix}$$

Taking advantage of the symmetry of the matrix and storing the upper half would produce the following vector:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{a}(i)$ | 2 | $-2$ | 3 | $-2$ | 5 | $-3$ | 10 | $-1$ | 0 | 0 | 4 | 10 |

The integer vector storing the positions of diagonal terms would be:

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **jdiag**($i$) | 1 | 3 | 5 | 7 | 12 |

Once the matrix has been stored in this fashion, all the operations leading to the solution of the linear equation system are performed on vector $\mathbf{a}$. Factorization will change the contents of $\mathbf{a}$ and back substitution will use these new contents in conjunction with the load vector to produce the solution.

### 6.4 Harwell-Boeing storage scheme

However, the skyline method is not optimal either since it stores many zeros and operates with them. Operations carried out with zeroes are trivial, avoidable. They increase the computational cost without relevant benefits. More recently, the use of storage methods without zeros has spread with the consequent savings in space and operations. One of the best known schemes is the Harwell-Boeing compressed columns. Also called CSC (Compact Storage Columns) and similar to CSR (Compact Storage Rows) except in columns. It requires the storage of a vector with non-zero coefficients and two auxiliary vectors with integers. Using the same matrix as in the previous subsection the following vectors would be required:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **a**($i$) | 2 | $-2$ | 3 | $-2$ | 5 | $-3$ | 10 | $-1$ | 4 | 10 |
| **row**($i$) | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 1 | 4 | 5 |

The integer vector storing the starting positions of each column would be:

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **ColStart**($i$) | 1 | 2 | 4 | 6 | 8 |

## 7 Conclusions

The finite element method is a powerful tool that has been used in many fields of engineering analysis. But it should be used with care. We have discussed some practical aspects that could introduce important errors in the results of computations. One fundamental aspect is the choice of the kind of element to do the analysis. The element type chosen strongly affects the results, and can also be the source of unsatisfactory performance in bending and incompressible situations. For mixed formulations, the element type should also verify some important restrictions (patch test). Concerning the type of algorithm that should be used to solve the system of equations obtained with the FEM the discussion is open. The choice of iterative methods is in our opinion the best for really large problems.

## References

[Bab73] I Babŭska. The finite element method with lagrange multipiers. *Num. Math.*, 20:179–192, 1973.

[Bat96] KJ Bathe. *Finite element Procedures*. Prentice Hall, New Jersey, 1996.

[Bre74] F Brezzi. On the existence, uniqueness and approximation of saddle point problems arising from lagrangian multipliers. *RAIRO*, 8-R2:129–151, 1974.

[Hug87] TJR Hughes. *The Finite Element Method:Linear Static and Dynamic Finite Element Analysis*. Prentice-Hall, Inc.,Englewood Cliffs, New Jersey, 1987.

[IS83] B Irons and N Shrive. *Finite element Primer*. Ellis Horwood Limited, 1983.

[SS86] Y Saad and MH Schultz. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. ScI. STAT. COMPUT.*, 7 No. 3, 1986.

[WB11] JA White and RI Borja. Block-preconditioned newton-krylov solvers for fully coupled flow and geomechanics. *Comput Geosci.*, 15:647–659, 2011.

[ZQTN86] OC Zienkiewicz, S Qu, RL Taylor, and S Nakazawa. The patch test for mixed formulations. *Int.J.Numer.Meth.Eng.*, 23:1873–1883, 1986.

[ZT00] OC Zienkiewicz and RL Taylor. *The Finite Element Method 5th edition*. Butterword-Heinemann, Oxford, 2000.

# The theory of plasticity in constitutive modeling of rate–independent soils

**Claudio Tamagnini**[a], **Kateryna Oliynyk**[a,b]

[a] *University of Perugia, Italy*
[b] *University of Dundee, UK*

*This chapter presents a review of the applications of the theory of plasticity to the modeling of rate–independent geomaterials, starting from classical approaches and covering some of the advanced versions of the theory designed to improve its capabilities in cyclic/dynamic loading conditions as well as to model "environmental" loading effects. In the discussion of the different approaches, particular attention is given to the incremental nature of the constitutive equations, emerging from the need of reproducing the essential features of the history–dependent behavior of soils. The relative merits and limitations of each class of models discussed are outlined with emphasis on those inherent features of their mathematical structure which might be of help in the assessment of their predictive capabilities when applied to practical geotechnical problems. This chapter was first published in the lecture notes of the 2021 ALERT School "Constitutive Modelling in Geomaterials".*

## 1 Introduction

In the application of continuum theories to the analysis of any solid mechanics problem, a fundamental role is played by the *constitutive equations*, which are expected to describe in precise mathematical terms the actual mechanical behavior of the material. Constitutive equations do not represent universal laws of nature. Rather, they can be considered definitions of *ideal materials*, *i.e.*, what is usually referred to as *constitutive models*. Constitutive models may possess the properties of the actual materials they are intended to model only to a limited extent. However, this do not lessen their worth, which is to produce a mathematical tool to predict the behavior of the physical system under any possible circumstance, starting from the limited knowledge gathered in a few experimental observations.

The quality of the predictions depends on the ability to define a suitable idealization for the real material which is capable to capture, from a quantitative point of view, the experimentally observed features which are thought to be of relevance for the practical problem at hand. This is particularly true in computational geomechanics, where the materials under consideration – *i.e.*, soil layers or rock masses – are usually characterized by a complex multi–phase structure and by a highly non–linear, irreversible and history–dependent response to the applied mechanical or "environmental" loading conditions.

The main objective of this chapter is to provide an outline of the different classes of constitutive equations for soils developed within the general framework of the theory of plasticity – from the early, pioneering works in perfect plasticity, to more recent developments in bounding surface and generalized plasticity, as well as in plasticity with generalized hardening laws to capture the effects of "environmental" loading conditions.

The topics covered in the following are not intended to provide a comprehensive review of the enormous amount of work which has been done in the applications of the theory of plasticity to soil mechanics over many decades. For this, the reader is referred, for example, to the following monographs [DS84, DS02, Woo04, Yu06, Bor13, Has17]. Rather, the presentation will be limited to those aspects of the general framework of the theory of which reflect the authors' own experience and interests. In particular, the discussion will be mostly focused – with the only exception of Sect. 9 – on constitutive equations for *rate–independent*, *saturated soils* in *isothermal conditions*, obeying the *principle of effective stress* as stated by Terzaghi [Ter48]. Details on how the constitutive models for saturated soils should be extended to account for partially saturated conditions can be found in the chapter by Jommi [Jom21] in this book. In the presentation of the different classes of models, we will focus on the infinitesimal theory of plasticity, suitable for small deformations and rotations. The extension of the theory to finite deformations is discussed in the chapter by Oliynyk and Tamagnini [OT21] in this book. The constitutive equations for brittle materials – *e.g.*, rocks or concrete – developed in the framework of damage mechanics are deliberately left out of this exposition. Finally, only constitutive equations for *simple materials*, according to Truesdell & Noll [TN65], will be considered in the following. Although non–local or weakly non–local theories for materials with microstructure – such as polar, second gradient or micromorphic materials – have been the subject of a considerable amount of research in geomechanics, mainly in relation to the study of strain localization into shear bands, they are outside the scope of the present work. For this interesting subject, the reader is referred to the books by Vardoulakis and Sulem [VS95, Var19], and references therein.

## 2 Notation

In the following, boldface lower– and upper–case letters are used to represent vector and tensor quantities. The symbols $\mathbf{1}$ and $\boldsymbol{I}^s$ are used for the second–order and fourth–

order identity tensors, with components:

$$(\mathbf{1})_{ij} = \delta_{ij} \qquad\qquad (\boldsymbol{I}^s)_{ijkl} = \frac{1}{2}\,(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{1}$$

The symmetric and skew–symmetric parts of a second–order tensor $\boldsymbol{X}$ are denoted as: $\operatorname{sym}\boldsymbol{X} := (\boldsymbol{X} + \boldsymbol{X}^T)/2$ and $\operatorname{skw}\boldsymbol{X} := (\boldsymbol{X} - \boldsymbol{X}^T)/2$, respectively. The dot product is defined as follows: $\boldsymbol{v} \cdot \boldsymbol{w} := v_i w_i$ for any two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$; $\boldsymbol{X} \cdot \boldsymbol{Y} := X_{ij} Y_{ij}$ for any two second–order tensors $\boldsymbol{X}$ and $\boldsymbol{Y}$. The dyadic product is defined as follows: $[\boldsymbol{v} \otimes \boldsymbol{w}]_{ij} := v_i w_j$ for any two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$; $[\boldsymbol{X} \otimes \boldsymbol{Y}]_{ijkl} := X_{ij} Y_{kl}$ for any two second–order tensors $\boldsymbol{X}$ and $\boldsymbol{Y}$. The quantity $\|\boldsymbol{X}\| := \sqrt{\boldsymbol{X} \cdot \boldsymbol{X}}$ denotes the Euclidean norm of $\boldsymbol{X}$. The usual sign convention of soil mechanics (compression positive) is adopted throughout. In line with Terzaghi's principle of effective stress, all stresses are *effective* stresses, unless otherwise stated. In the representation of stress and strain states, use will sometimes be made of the invariant quantities: $p$ (mean stress), $q$ (deviator stress), and $\theta$ (Lode angle), defined as:

$$p := \frac{1}{3}(\boldsymbol{\sigma} \cdot \mathbf{1})\,; \quad q := \sqrt{\frac{3}{2}}\,\|\boldsymbol{s}\|\,; \quad \sin(3\theta) := \sqrt{6}\,\frac{(\boldsymbol{s}^3) \cdot \mathbf{1}}{[(\boldsymbol{s}^2) \cdot \mathbf{1}]^{3/2}} \tag{2}$$

and: $\epsilon_v$ (volumetric strain), $\epsilon_s$ (deviatoric strain), $\dot{\epsilon}_v$ (volumetric strain rate), and $\dot{\epsilon}_s$ (deviatoric strain rate), defined as:

$$\begin{aligned} \epsilon_v &:= \boldsymbol{\epsilon} \cdot \mathbf{1}\,; \quad \epsilon_s := \sqrt{\frac{2}{3}}\,\|\boldsymbol{e}\|\,; \quad \theta_\epsilon := \sqrt{6}\,\frac{(\boldsymbol{e}^3) \cdot \mathbf{1}}{[(\boldsymbol{e}^2) \cdot \mathbf{1}]^{3/2}} \\ \dot{\epsilon}_v &:= \dot{\boldsymbol{\epsilon}} \cdot \mathbf{1}\,; \quad \dot{\epsilon}_s := \sqrt{\frac{2}{3}}\,\|\dot{\boldsymbol{e}}\| \quad\; \dot{\theta}_\epsilon := \sqrt{6}\,\frac{(\dot{\boldsymbol{e}}^3) \cdot \mathbf{1}}{[(\dot{\boldsymbol{e}}^2) \cdot \mathbf{1}]^{3/2}} \end{aligned} \tag{3}$$

In eqs. (2) and (3), $\boldsymbol{s} := \boldsymbol{\sigma} - p\,\mathbf{1}$ is the deviatoric part of the stress tensor; $\boldsymbol{e} := \boldsymbol{\epsilon} - (1/3)\epsilon_v\,\mathbf{1}$ and $\dot{\boldsymbol{e}} := \dot{\boldsymbol{\epsilon}} - (1/3)\dot{\epsilon}_v\,\mathbf{1}$ are the deviatoric parts of the strain and the strain rate tensors, respectively, while $\boldsymbol{s}^2$ and $\boldsymbol{s}^3$ are the square and the cube of the deviatoric stress tensor, with components $(\boldsymbol{s}^2)_{ij} := s_{ik}s_{kj}$ and $(\boldsymbol{s}^3)_{ij} := s_{ik}s_{kl}s_{lj}$. It is worth noting that in eqs. $(3)_5$ and $(3)_6$, with a slight abuse of notation, the symbols $\dot{\epsilon}_s$ and $\dot{\theta}_\epsilon$ have been employed to denote the second and third invariants of the strain rate tensor, which generally do not coincide with the time rates of $\epsilon_s$ and $\theta_\epsilon$, as defined in in eqs. $(3)_2$ and $(3)_3$.

# 3 History–dependent materials modeling and the need for constitutive equations in rate–form

According to the principles of *determinism* and *local action* [TN65], the most general expression for the constitutive equation of a *simple* material is given by:

$$\boldsymbol{\sigma}(\boldsymbol{x}, t) = \mathop{\boldsymbol{\mathcal{G}}}_{\tau=0}^{\infty}\left[\boldsymbol{F}^{(t)}(\boldsymbol{X}, \tau)\right] \tag{4}$$

where $\boldsymbol{\mathcal{G}}$ is a *functional* of the *history* up to time $t$ of the *deformation gradient* associated with the motion $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X}, t)$ carrying the material point $\boldsymbol{X}$ in the reference configuration to its position $\boldsymbol{x}$ in the current configuration at time $t$, defined as:

$$\boldsymbol{F}^{(t)}(\boldsymbol{X}, s) := \boldsymbol{F}(\boldsymbol{X}, t - s) \qquad \boldsymbol{F}(\boldsymbol{X}, t) := \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{X}}(\boldsymbol{X}, t) \qquad (s \geq 0) \tag{5}$$

Eq. (4) essentially states that the (effective) stress tensor $\boldsymbol{\sigma}$ is a function of the *entire deformation history*, *i.e.*, that the knowledge of the state of strain at a given time $t$ is in general *not sufficient* to determine the stress state. This is an essential feature of inelastic, history–dependent materials such as soils.

A third fundamental principle, the *principle of material frame indifference*, implies the following restriction to the functional $\boldsymbol{\mathcal{G}}$: for every orthogonal tensor function $\boldsymbol{Q}(\tau)$ and every history $\boldsymbol{F}^{(t)}(\boldsymbol{X}, \tau)$, the relation:

$$\boldsymbol{Q}_0 \overset{\infty}{\underset{\tau=0}{\boldsymbol{\mathcal{G}}}} \left[\boldsymbol{F}^{(t)}(\boldsymbol{X}, \tau)\right] \boldsymbol{Q}_0^T = \overset{\infty}{\underset{\tau=0}{\boldsymbol{\mathcal{G}}}} \left[\boldsymbol{Q}(\tau)\boldsymbol{F}^{(t)}(\boldsymbol{X}, \tau)\right] \qquad \boldsymbol{Q}_0 := \boldsymbol{Q}(0) \tag{6}$$

must hold. Conversely, any such functional $\boldsymbol{\mathcal{G}}$ satisfying eq. (6) can be considered as defining the constitutive equation of a particular material.

The fundamental properties of the functional $\boldsymbol{\mathcal{G}}$ should be defined according to our knowledge of the main characteristic of the mechanical behavior of the materials we intend to model. As far as geomaterials – and soils in particular – are concerned, a long standing experimental evidence indicates that the mechanical response of such materials is strongly non–linear and dependent on such factors as current state, previous loading history, load increment size and loading direction. Even the simplest and most common laboratory tests, such as a one–dimensional compression test or a axisymmetric (triaxial) drained compression test, can highlight such features in both fine and coarse–grained soils.

A main consequence of this observation is that the constitutive functional $\boldsymbol{\mathcal{G}}$ must be *non–linear* and *non–differentiable*, see [OW69]. However, working with non–linear, non–differentiable functionals poses formidable mathematical problems, even in the simplest cases. An alternative strategy, which overcomes this difficulty and is commonly adopted in nonlinear solid mechanics, is to avoid formulating the constitutive equation in *global terms*, as in eq. (4), and rather adopt an incremental (or *rate–type*) formulation, in which the (objective) stress rate is given as a *function* of the rate of deformation $\boldsymbol{d} := \operatorname{sym} \nabla \boldsymbol{v}$ ($\boldsymbol{v} := d\boldsymbol{\varphi}/dt \circ \boldsymbol{\varphi}$ being the spatial velocity) and of the current state of the material:

$$\overset{\circ}{\boldsymbol{\sigma}} = \boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{d}\right) \tag{7}$$

In eq. (7), $\overset{\circ}{\boldsymbol{\sigma}}$ denotes a suitable objective stress rate, such as the Jaumann–Zaremba stress rate, defined as:

$$\overset{\nabla}{\boldsymbol{\sigma}} := \dot{\boldsymbol{\sigma}} + \boldsymbol{\sigma}\boldsymbol{\omega} - \boldsymbol{\omega}\boldsymbol{\sigma} \tag{8}$$

where $\boldsymbol{\omega} := \operatorname{skw} \nabla \boldsymbol{v}$ is the spin tensor. In eq. (7), $\boldsymbol{q}$ represents a set of *internal* state variables, which are introduced to account for the effects of the previous loading history. An additional set of rate equations is then required to define the evolution of the internal variables in time. In classical elastoplasticity, these evolution equations are referred to as *hardening laws*.

Restricting our discussion to the infinitesimal theory, the objective stress rate $\overset{\circ}{\boldsymbol{\sigma}}$ can be replaced by the standard objective time rate $\dot{\boldsymbol{\sigma}}$, and the rate of deformation $\boldsymbol{d}$ with the (linearized) strain rate tensor $\dot{\boldsymbol{\epsilon}}$. Thus, eq. (7) can be rewritten as:

$$\dot{\boldsymbol{\sigma}} \ = \boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \dot{\boldsymbol{\epsilon}}\right) \tag{9}$$

Rate–indepence means that a change in the time scale does not affect the material response, *e.g.*, doubling the strain rate doubles the stress rate. More generally:

$$\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \lambda\dot{\boldsymbol{\epsilon}}\right) = \lambda\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \dot{\boldsymbol{\epsilon}}\right) \qquad\qquad \forall\ \lambda > 0 \tag{10}$$

A direct consequence of the above equation is that the function $\boldsymbol{G}$ is *positively homogeneous* of degree one in $\dot{\boldsymbol{\epsilon}}$. This latter property yields the following alternative expression for the constitutive equation (9):

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta}\right)\dot{\boldsymbol{\epsilon}} \tag{11}$$

where $\boldsymbol{D}$ is the (fourth–order) tangent stiffness tensor at the current state, which depends on the strain rate only through its *direction*, defined by the unit tensor $\boldsymbol{\eta} := \dot{\boldsymbol{\epsilon}}/\left\|\dot{\boldsymbol{\epsilon}}\right\|$. Eq. (11) provides a general representation for rate–independent constitutive equations which encompasses as particular cases all the constitutive equations derived within the general framework of the theory of plasticity.

# 4 Non–linearity and incremental non–linearity

Let $(\boldsymbol{\sigma}_0, \boldsymbol{q}_0)$ be the initial state of the material at time $t = 0$. For a given strain path $\mathcal{E}$ from $\boldsymbol{\epsilon}_0$ to $\boldsymbol{\epsilon}(t)$, the state of stress at time $t$, $\boldsymbol{\sigma}(t)$, is obtained by integrating eq. (11):

$$\boldsymbol{\sigma}(t) = \hat{\boldsymbol{\sigma}}\left(\boldsymbol{\sigma}_0, \boldsymbol{q}_0, \mathcal{E}\right) = \boldsymbol{\sigma}_0 + \int_{\mathcal{E}} \boldsymbol{D}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta}\right)\frac{d\boldsymbol{\epsilon}}{ds}\,ds \tag{12}$$

From the above equation, it is immediately apparent that the dependence of the tangent stiffness $\boldsymbol{D}$ on the current state $(\boldsymbol{\sigma}, \boldsymbol{q})$ renders the function $\hat{\boldsymbol{\sigma}}$ *non–linear*, *e.g.*, doubling the strain increment does not result in doubling the stress increment. This is the notion of non–linearity to be invoked when describing a material response for which the observed stress–strain curve (*e.g.*, in a triaxial compression path) is not a straight line.

An independent concept of non–linearity can be defined by considering the functional relation between stress rate and strain rate, as first suggested by Darve [Dar78]. If the constitutive function $\boldsymbol{G}$ is *linear* in $\dot{\boldsymbol{\epsilon}}$, then the material is said to be *incrementally linear*. In this case, the tangent stiffness tensor $\boldsymbol{D}$ does not depend on the strain rate direction $\boldsymbol{\eta}$, and eq. (11) reduces to:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}\left(\boldsymbol{\sigma}, \boldsymbol{q}\right)\dot{\boldsymbol{\epsilon}} \tag{13}$$

While a linear behavior implies incremental linearity, the opposite is not true. That is, incremental linearity does not imply linearity of the stress–strain response over a finite load increment. On the other hand, when $\boldsymbol{G}$ is a *non–linear* function of the strain rate, *i.e.*, for any $\dot{\boldsymbol{\epsilon}}_1$ and $\dot{\boldsymbol{\epsilon}}_2$ and $a, b \in \mathbb{R}$:

$$\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, a\dot{\boldsymbol{\epsilon}}_1 + b\dot{\boldsymbol{\epsilon}}_2\right) \neq a\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \dot{\boldsymbol{\epsilon}}_1\right) + b\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \dot{\boldsymbol{\epsilon}}_2\right) \tag{14}$$

the material behavior is said to be *incrementally non–linear*. In this case, the tangent stiffness $\boldsymbol{D}$ explicitly depends on the strain rate direction, see eq. (11).

From eq. (14) it follows that:

$$\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \dot{\boldsymbol{\epsilon}}\right) \neq -\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, -\dot{\boldsymbol{\epsilon}}\right) \tag{15}$$

which, in turn, implies:

$$\boldsymbol{D}(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta}) \neq \boldsymbol{D}(\boldsymbol{\sigma}, \boldsymbol{q}, -\boldsymbol{\eta}) \tag{16}$$

Equation (16) expresses a fundamental feature of incrementally non–linear models: for any strain rate direction, the reversal of the loading path is always associated with a change in the tangent stiffness $\boldsymbol{D}$. Indeed, such a feature is *necessary* in order to correctly describe irreversible behavior. In fact, although eq. (13) is in general non–integrable, the response of an incrementally linear material remains completely *reversible* in any closed loading–unloading program following the same path in two opposite directions.

When discussing the dependence of $\boldsymbol{D}$ on $\boldsymbol{\eta}$, it is useful to introduce the concept of *tensorial zone*, as defined by Darve [Dar78, Dar90]. A tensorial zone $Z$ is a portion of the strain rate space in which $\boldsymbol{G}$ is a linear function of $\dot{\boldsymbol{\epsilon}}$. Accordingly, in a particular tensorial zone the tangent stiffness is *independent of* $\boldsymbol{\eta}$:

$$\boldsymbol{D}(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta}) = \boldsymbol{D}^Z(\boldsymbol{\sigma}, \boldsymbol{q}) \qquad\qquad \forall\, \boldsymbol{\eta} \in Z \tag{17}$$

As $\boldsymbol{G}$ is positively homogeneous of degree one in $\dot{\boldsymbol{\epsilon}}$, $Z$ is a cone in the strain rate space with the vertex at the origin (*i.e.*, all strain rates $\lambda\dot{\boldsymbol{\epsilon}}$ with $\lambda > 0$ belong to the same tensorial zone as $\dot{\boldsymbol{\epsilon}}$).

Following Darve [Dar90], incrementally non–linear, rate–independent constitutive equations can be classified according to the number of associated tensorial zones. When the number of tensorial zones of $\boldsymbol{G}$ is finite, the constitutive equation is *incrementally multi–linear* (*bi–linear* in the particular case of only two zones). In incrementally multi–linear materials, an important issue is represented by the *continuity* of the response at the boundary between any two tensorial zones [Gud79]. Let $\partial Z_{AB}$ be such a boundary between the tensorial zones $Z_A$ and $Z_B$. If $\dot{\boldsymbol{\epsilon}}^* \in \partial Z_{AB}$, then, continuity of the response requires that:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{Z_A}\dot{\boldsymbol{\epsilon}}^* = \boldsymbol{D}^{Z_B}\dot{\boldsymbol{\epsilon}}^* \qquad \Rightarrow \qquad \left(\boldsymbol{D}^{Z_A} - \boldsymbol{D}^{Z_B}\right)\dot{\boldsymbol{\epsilon}}^* = \boldsymbol{0} \tag{18}$$

Equation $(18)_2$ represents a generalization of the continuity condition established by Green [Gre56] for hypoelastic materials. In the following, we will focus on models with one or two tensorial zones, leaving aside the theories of plasticity with multiple plastic mechanisms and more than two tensorial zones (*multi–surface plasticity*). Interested readers may refer to Ch. 5 of the book by Simo and Hughes [SH97] for a general treatment of this subject.

As opposed to multi–linearity, a *strictly* incrementally non–linear behavior is provided by constitutive models for which a *continuous* dependence of $\boldsymbol{D}$ on $\boldsymbol{\eta}$ is assumed. This is the case of rate–type constitutive models developed within the framework of the theory of hypoplasticity [Kol91, TVC00]. This subject is presented in the chapter by Mašín [Mas21] in this book.

# 5 Linear elasticity, hyperelasticity and hypoelasticity

In the early application of continuum mechanics to geotechnical engineering, the enormous analytical difficulties posed by the design of even simple geotechnical structures led to the traditional distinction between "deformation" and "failure" problems, for which different, very simple constitutive equations could be used, see *e.g.*, [TP48]. The rationale behind this approach is that only some very specific features of soil behavior are of interest for the particular problem at hand, while the others could be neglected without affecting the quality of the prediction in a substantial way. In particular, the only possible constitutive framework for which (analytical) solutions to deformation problems could be obtained at that time – in lack of suitable numerical methods and powerful computer platforms – was provided by the theory of *linear elasticity*. Its successful application then relied on the "proper" selection of the relevant soil constants (in essence, the Young's modulus), which had to be assumed to depend on such primary factors as current stress state, previous stress history, and nature of the applied stress path – in terms of magnitude and, possibly, direction.

Nowadays, the theory of elasticity still plays an important role, as it can be considered a cornerstone of any plasticity theory. For this reason, the main features of elasticity models adopted in the description of soil behavior are briefly recalled in this Section.

## 5.1 Linear elasticity

The simplest linear elastic model is provided by the Hooke's law for isotropic materials:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}\dot{\boldsymbol{\epsilon}}^e \qquad \boldsymbol{D} = K\mathbf{1}\otimes\mathbf{1} + 2G\left(\boldsymbol{I}^s - \frac{1}{3}\mathbf{1}\otimes\mathbf{1}\right) \tag{19}$$

where $\dot{\boldsymbol{\epsilon}}^e$ is the elastic strain rate – coinciding with the total strain rate if there are no irreversible deformations – while $K$ and $G$ are the (constant) bulk and shear moduli of the material. The two elastic constants can be replaced by other, frequently used pairs of alternative elastic properties, such as the Young's modulus and Poisson's ratio

$(E, \nu)$ related to $K$ and $G$ by the relations:

$$K = \frac{E}{3(1-2\nu)} \qquad\qquad G = \frac{E}{2(1+\nu)}$$

or the Lame's constants $(\lambda, \mu)$, linked to $E$ and $\nu$ by the relations:

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \qquad\qquad \mu = \frac{E}{2(1+\nu)} = G$$

Linear isotropic elasticity is still widely used in a number of important geotechnical applications. However, it fails to capture an essential feature of the reversible response of granular material, *i.e.*, global non–linearity, due to the dependence of the stiffness constants on the current stress state. This feature of soils' elastic response originates from the nature of the reversible grain to grain interactions at the microscopic level [CJR13].

## 5.2 Hypoelasticity

Early attempts to incorporate global non–linearity in the elastic response of the soil can be traced back to the works of Kondner & Zelasko [KZ63] and Duncan & Chang [DC70]. In essence, it consists in adopting an isotropic elastic constitutive equation in the form of eq. (19), where the elastic stiffness coefficients are not constants but rather functions of the strain level and/or of the stress state. Generally speaking, all models of this kind are defined as *hypoelastic*, since the quantity:

$$d\boldsymbol{\epsilon}^e = \boldsymbol{C} d\boldsymbol{\sigma} \qquad\qquad \boldsymbol{C} := \boldsymbol{D}^{-1} \tag{20}$$

is not an exact differential, *i.e.*, it is not possible to define a one–to–one correspondence between the stress and strain tensors, and a closed stress cycle might result in the development of residual deformations.

The early hypoelastic formulations adopted an elastic tangent stiffness tensor $\boldsymbol{D}$ of the form:

$$\boldsymbol{D}(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = K_t(p, \epsilon_v)\, \mathbf{1} \otimes \mathbf{1} + 2G_t(p, \epsilon_s) \left( \boldsymbol{I}^s - \frac{1}{3} \mathbf{1} \otimes \mathbf{1} \right) \tag{21}$$

In constitutive models of this class, the dependence of the tangent bulk and shear moduli, $K_t(p, \epsilon_v)$, and $G_t(p, \epsilon_s)$, on the strain invariants is obtained by curve–fitting the observed stress–strain response in standard loading paths, such as drained (or undrained) triaxial compression, and isotropic compression, see for example [JPFB86, JP88, JPSJH91]. For this reason, these constitutive equations are also referred to as *variable–moduli models*.

A main drawback of variable–moduli models is the fact that, in this case, the strain invariants cannot be considered as true state variables, since the reference configuration from which the strains are defined is arbitrary. In this respect, a more sound approach is provided by those hypoelastic models in which the stiffness coefficients depend only on the current stress state, typically through the mean stress $p$:

$$K_t(p) = K_{t0} \left( \frac{p}{p_{\text{atm}}} \right)^{\alpha} \qquad\qquad G_t(p) = G_{t0} \left( \frac{p}{p_{\text{atm}}} \right)^{\beta} \tag{22}$$

where $p_{\text{atm}}$ is the atmospheric pressure, used as a scaling factor for the mean stress, and $K_{t0}$, $G_{t0}$, $\alpha$ and $\beta$ are model constants, determined by empirically fitting stress–strain curves from conventional laboratory test results. The phenomenological nature of the relations (22) implies that the resulting elastic constitutive equation is hypoelastic and cannot be derived from a potential function. Zytynski *et al.* [ZRNW78] have discussed the necessary conditions for the stiffness coefficient to make $d\boldsymbol{\epsilon}^e$ in eq. (20) an exact differential. In particular, they observe that if both $K_t$ and $G_t$ depend only on $p$, as in eq. (22), then the resulting elastic constitutive equation in rate form cannot be integrated and is therefore hypoelastic.

Hypoelastic constitutive equations have been and still are widely used in the formulation of both classical and advanced plasticity theories for soils. However, their use should remain limited to monotonic loading conditions or to situations where the soil undergoes only a small number of cycles, as pointed out in [BTA97].

## 5.3 Hyperelasticity

A material is said to be *hyperelastic* (or *Green elastic* [Ogd97]) when there exists an elastic potential function $\psi(\boldsymbol{\epsilon}^e)$ such that:

$$\boldsymbol{\sigma} = \frac{\partial \psi}{\partial \boldsymbol{\epsilon}^e}(\boldsymbol{\epsilon}^e) \tag{23}$$

Eq. (23) defines a hyperelastic constitutive equation, which implies the existence of a direct functional relation between the stress tensor $\boldsymbol{\sigma}$ and the elastic strain tensor $\boldsymbol{\epsilon}^e$. This relation can be recast in rate form by differentiating both sides of eq. (23), obtaining:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}(\boldsymbol{\epsilon}^e)\dot{\boldsymbol{\epsilon}}^e \qquad\qquad \boldsymbol{D}(\boldsymbol{\epsilon}^e) := \frac{\partial^2 \psi}{\partial \boldsymbol{\epsilon}^e \otimes \partial \boldsymbol{\epsilon}^e}(\boldsymbol{\epsilon}^e) \tag{24}$$

where the elastic tangent stiffness is obtained as the second derivative of $\psi$ with respect of its argument. This time, $\dot{\boldsymbol{\sigma}}$ is an exact differential, and no permanent stress changes may occur in any closed elastic strain cycle.

The dual formulation of the hyperelastic constitutive equation (23) is obtained by postulating the existence of a complementary energy function $g(\boldsymbol{\sigma})$, such that:

$$\boldsymbol{\epsilon}^e = \frac{\partial g}{\partial \boldsymbol{\sigma}}(\boldsymbol{\sigma}) \tag{25}$$

By differentiating eq. (25) we obtain the following complementary hyperelastic constitutive equation in rate–form:

$$\dot{\boldsymbol{\epsilon}}^e = \boldsymbol{C}(\boldsymbol{\sigma})\dot{\boldsymbol{\sigma}} \qquad\qquad \boldsymbol{C}(\boldsymbol{\sigma}) := \frac{\partial^2 g}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}}(\boldsymbol{\sigma}) \tag{26}$$

where $\boldsymbol{C} = \boldsymbol{D}^{-1}$ is the material tangent compliance tensor. The two potentials $\psi$ and $g$ are related one another as $g$ can be considered the Legendre transform of $\psi$, see, *e.g.*, [HP07]. As a consequence of the principle of material frame indifference, $\psi$ and $g$ must depend on their tensorial arguments only through their invariants [SH97], *i.e.*:

$$\psi(\boldsymbol{\epsilon}^e) = \hat{\psi}(\epsilon_v^e, \epsilon_s^e, \theta_\epsilon) = \bar{\psi}(\epsilon_1^e, \epsilon_2^e, \epsilon_3^e) \tag{27}$$

$$g(\boldsymbol{\sigma}) = \hat{g}(p, q, \theta) = \bar{g}(\sigma_1, \sigma_2, \sigma_3) \tag{28}$$

The hyperelastic constitutive equations (23) and (25) allow to describe a non–linear elastic behavior, whenever the two elastic potentials are not quadratic functions of their arguments. However, it is worth noting that hyperelasticity (which includes linear elasticity as a special case) as well as hypoelasticity are both incrementally linear theories, with only one tensorial zone. Examples of hyperelastic formulations for isotropic granular materials are provided in the works of [Hou85, BTA97, HP07].

# 6 Thermodynamics–based approach: the theory of hyperplasticity

The most important case of constitutive equations with two tensorial zones is provided by the classical *theory of plasticity* with a *single plastic mechanism*, and its various generalizations to describe, for example, induced anisotropy and cyclic behavior. The general framework of the theory of plasticity is now well established and a thorough treatment of this subject can be found in many excellent textbooks, *e.g.*, [Lub90, SH97, JB02]. As for plasticity in soil mechanics, good references are provided, *e.g.*, by [DS84, DS02, Yu06, Bor13, Has17]. As compared to those references, in the presentation of the basic principles of the theory we have adopted a slightly different point of view, starting from the basic principles of the thermodynamics of continuous media and following the approach of the so–called *theory of hyperplasticity*, as defined by Houlsby and Puzrin [HP07]. Then, the classical approach is presented as a generalization of the basic concepts of hyperplasticity.

The attempts to derive the evolution equations of the infinitesimal rate–independent plasticity from basic thermodynamics principles can be traced back to the early works of the French school [Mor70, HN75, GNS83]. Important contributions to the understanding of the thermo–mechanics of solid materials have been provided, *e.g.*, in the works of [Zie83, ZW87, Mau92, RM93, HR99]. The advantages of ensuring thermodynamic consistency when dealing with the inelastic behavior of geomaterials have been emphasized by Houlsby [Hou81] and Collins and Houlsby [CH97], in view of the potential drawbacks associated with purely phenomenological modeling of materials featuring stress–dependent stiffness, non–associative behavior and dilatant plastic flow. Significant contributions to the development of infinitesimal elastoplastic models for soils within the framework of continuum thermo–mechanics have been given, for example, by [MLA94, HP00, PH01, CK02, CH02, CM03, EP04, DT05, EHN07, OT20].

## 6.1 Free energy and dissipation functions

In the framework of infinitesimal elastoplasticity, we assume the customary additive decomposition of the total strain tensor into an elastic, reversible part and an inelastic part:

$$\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^e + \boldsymbol{\epsilon}^p \qquad\qquad \dot{\boldsymbol{\epsilon}} = \dot{\boldsymbol{\epsilon}}^e + \boldsymbol{\epsilon}^p \tag{29}$$

By limiting the set of state variables $\mathcal{S}$ to the elastic strain tensor $\boldsymbol{\epsilon}^e$ and to a pseudo–vector of strain–like internal variables $\boldsymbol{\alpha}$ (the components of which could be scalars or second–order tensors), we postulate the existence of a Helmholtz free energy function per unit volume of the form:

$$\psi(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}) = \psi^e(\boldsymbol{\epsilon}^e) + \psi^p(\boldsymbol{\alpha}) \tag{30}$$

This assumption is equivalent to consider the contributions to the free energy function of elastic strains and plastic internal variables as fully uncoupled. This could represent a somewhat restrictive assumption, but it can be considered sufficiently general for the scope of this work.

For isothermal processes, the second principle of thermodynamics requires that the dissipation function $\mathcal{D}$, defined as:

$$\mathcal{D} := \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\epsilon}} - \dot{\psi} \geq 0 \tag{31}$$

is non–negative. Taking into account the definition of the free energy function given in eq. (30), and introducing the set of *generalized stresses* $\overline{\mathcal{K}} = \{\overline{\boldsymbol{\chi}}, \overline{\boldsymbol{\chi}}_\alpha\}$, defined by:

$$\overline{\boldsymbol{\chi}} = \frac{\partial \psi^e}{\partial \boldsymbol{\epsilon}^e} \qquad\qquad \overline{\boldsymbol{\chi}}_\alpha = -\frac{\partial \psi^p}{\partial \boldsymbol{\alpha}} \tag{32}$$

we have:

$$\begin{aligned} \mathcal{D} &= \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\epsilon}} - \left\{ \frac{\partial \psi^e}{\partial \boldsymbol{\epsilon}^e} \cdot \dot{\boldsymbol{\epsilon}}^e + \frac{\partial \psi^p}{\partial \boldsymbol{\alpha}} \dot{\boldsymbol{\alpha}} \right\} \\ &= \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\epsilon}} - \overline{\boldsymbol{\chi}} \cdot (\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p) + \overline{\boldsymbol{\chi}}_\alpha \cdot \dot{\boldsymbol{\alpha}} \\ &= (\boldsymbol{\sigma} - \overline{\boldsymbol{\chi}}) \cdot \dot{\boldsymbol{\epsilon}} + \overline{\boldsymbol{\chi}} \cdot \dot{\boldsymbol{\epsilon}}^p + \overline{\boldsymbol{\chi}}_\alpha \cdot \dot{\boldsymbol{\alpha}} \geq 0 \end{aligned} \tag{33}$$

For this inequality to hold for any possible non–dissipative processes, for which $\dot{\boldsymbol{\epsilon}}^p = \mathbf{0}$ and $\dot{\boldsymbol{\alpha}} = \mathbf{0}$, we must have:

$$\boldsymbol{\sigma} = \overline{\boldsymbol{\chi}} = \frac{\partial \psi}{\partial \boldsymbol{\epsilon}^e} \tag{34}$$

Eq. (34) is the hyperelastic constitutive equation of the material, establishing a functional relation between the stress tensor $\boldsymbol{\sigma}$ and the elastic strain tensor $\boldsymbol{\epsilon}^e$. Substituting this last result into eq. (33), we obtain the following reduced dissipation inequality:

$$\mathcal{D} = \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\epsilon}}^p + \overline{\boldsymbol{\chi}}_\alpha \cdot \dot{\boldsymbol{\alpha}} \geq 0 \tag{35}$$

Eqs. (32) and (35) suggest the following functional dependence for the dissipation function $\mathcal{D}$ on both the set $\mathcal{S}$ of the state variables and the set of dissipative flows $\mathcal{F} := \{\dot{\boldsymbol{\epsilon}}^p, \dot{\boldsymbol{\alpha}}\}$:

$$\mathcal{D}(\mathcal{S}, \mathcal{F}) = \mathcal{D}\left(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}, \dot{\boldsymbol{\epsilon}}^p, \dot{\boldsymbol{\alpha}}\right) \tag{36}$$

To describe the behavior of a rate–independent material, we postulate that the dissipation function $\mathcal{D}$ is homogeneous of degree one in the elements of $\mathcal{F}$. Euler's theorem for homogeneous functions then requires that:

$$\mathcal{D} = \frac{\partial \mathcal{D}}{\partial \dot{\boldsymbol{\epsilon}}^p} \cdot \dot{\boldsymbol{\epsilon}}^p + \frac{\partial \mathcal{D}}{\partial \dot{\boldsymbol{\alpha}}} \cdot \dot{\boldsymbol{\alpha}} \tag{37}$$

By introducing the set of *generalized dissipative stresses* $\mathcal{K} := \{\boldsymbol{\chi}, \boldsymbol{\chi}_\alpha\}$, defined as:

$$\boldsymbol{\chi} = \frac{\partial \mathcal{D}}{\partial \dot{\boldsymbol{\epsilon}}^p} \qquad\qquad \boldsymbol{\chi}_\alpha = \frac{\partial \mathcal{D}}{\partial \dot{\boldsymbol{\alpha}}} \tag{38}$$

eq. (37) can be rewritten as:

$$\mathcal{D} = \boldsymbol{\chi} \cdot \dot{\boldsymbol{\epsilon}}^p + \boldsymbol{\chi}_\alpha \cdot \dot{\boldsymbol{\alpha}} \tag{39}$$

Comparing eqs. (37) and (39) we observe that generalized stresses and generalized dissipative stresses must fulfill the following relation:

$$(\boldsymbol{\chi} - \overline{\boldsymbol{\chi}}) \cdot \dot{\boldsymbol{\epsilon}}^p + (\boldsymbol{\chi}_\alpha - \overline{\boldsymbol{\chi}}_\alpha) \cdot \dot{\boldsymbol{\alpha}} = 0 \tag{40}$$

This equality is trivially satisfied if Ziegler's orthogonality conditions – see [HP07] – are assumed:

$$\boldsymbol{\chi} = \overline{\boldsymbol{\chi}} \qquad\qquad \boldsymbol{\chi}_\alpha = \overline{\boldsymbol{\chi}}_\alpha \tag{41}$$

Eq. (41) is a sufficient condition for eq. (40) to hold, but not a necessary one. Therefore, Ziegler's orthogonality condition must be considered as a (weak) restrictive constitutive assumption, yet compatible with realistic descriptions of many classes of granular materials characterized by frictional dissipation, see *e.g.*, [CH97, HP07].

## 6.2 Yield function and evolution equations

The homogeneity of degree one of $\mathcal{D}$ in the dissipative flows implies that the (degenerate) partial Legendre transformation of $\mathcal{D}$ with respect to the arguments in $\mathcal{F}$ is a scalar function $\hat{f}$, called *yield function*, such that:

$$\dot{\gamma} \hat{f}(\mathcal{S}, \mathcal{K}) := \boldsymbol{\chi} \cdot \dot{\boldsymbol{\epsilon}}^p + \boldsymbol{\chi}_\alpha \cdot \dot{\boldsymbol{\alpha}} - \mathcal{D} = 0 \tag{42}$$

for dissipative processes, *i.e.*, when the elements of $\mathcal{F}$ are non–zero. In the LHS of eq. (42), the scalar $\dot{\gamma} \geq 0$ is the so–called *plastic multiplier*. The set:

$$\mathbb{E} := \left\{ (\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\chi}_\alpha) \in \mathcal{S} \times \mathcal{K} \;\middle|\; \hat{f}(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\chi}_\alpha) < 0 \right\} \tag{43}$$

is the *elastic domain* of the material, where the plastic multiplier is zero and all the processes are non dissipative ($\dot{\boldsymbol{\epsilon}}^p = \mathbf{0}$, $\dot{\boldsymbol{\alpha}} = \mathbf{0}$). The boundary of $\mathbb{E}$:

$$\partial\mathbb{E} := \left\{ (\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\chi}_\alpha) \in \mathcal{S} \times \mathcal{K} \;\middle|\; \hat{f}(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\chi}_\alpha) = 0 \right\} \tag{44}$$

is the *yield surface*, on which $\dot{\gamma}$ may be positive and irreversible processes may occur.

It is worth noting that, due to the orthogonality conditions (41) and the constitutive equations (32), the yield function $f$ can be considered as a function of $\boldsymbol{\epsilon}^e$ and $\boldsymbol{\alpha}$:

$$\hat{f}(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\chi}_\alpha) = f^*(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha}) = 0 \tag{45}$$

*i.e.*, of the elastic strain and the strain–like internal variables. Therefore, the elastic domain and the yield function provided by eqs. (43) and (44) are defined in *strain space*. The stress–space counterparts of $\mathbb{E}$ and $f$ is recovered by noting that the stress tensor $\boldsymbol{\sigma}$ and the stress–like internal variables $\overline{\boldsymbol{\chi}}_\alpha$ are given functions of $(\boldsymbol{\epsilon}^e, \boldsymbol{\alpha})$ through the constitutive equations (32). The yield function in stress space then reads:

$$f(\boldsymbol{\sigma}, \overline{\boldsymbol{\chi}}_\alpha) = f^* \left\{ \boldsymbol{\epsilon}^e(\boldsymbol{\sigma}), \boldsymbol{\alpha}(\overline{\boldsymbol{\chi}}_\alpha) \right\} = 0 \tag{46}$$

From the properties of the Legendre transform of eq. (42) the following *associative flow rules* for the elements of $\mathcal{F}$ can be obtained:

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\gamma}\,\frac{\partial \hat{f}}{\partial \boldsymbol{\chi}} = \dot{\gamma}\,\frac{\partial f}{\partial \boldsymbol{\sigma}} \tag{47a}$$

$$\dot{\boldsymbol{\alpha}} = \dot{\gamma}\,\frac{\partial \hat{f}}{\partial \boldsymbol{\chi}_\alpha} = \dot{\gamma}\,\frac{\partial f}{\partial \overline{\boldsymbol{\chi}}_\alpha} \tag{47b}$$

Eq. (47a) is the standard associative flow rule for the plastic strain rate, while eq. (47b) provides the associative hardening law for the internal variable $\boldsymbol{\alpha}$. It is worth noting that the associativity of the flow rule (47a) holds in the generalized dissipative stress space. Thus, this result does not prevent the possibility of modeling non–associative plastic flow in standard Cauchy stress space for free energy functions different from the one adopted in eq. (30), see [CH97, HP07].

## 6.3 Consistency conditions and constitutive equations in rate–form

The yield function and the plastic multiplier are subjected to the *Kuhn–Tucker complementarity conditions*:

$$\dot{\gamma} \geq 0 \qquad\qquad f(\boldsymbol{\sigma}, \overline{\boldsymbol{\chi}}_\alpha) \leq 0 \qquad\qquad \dot{\gamma} f(\boldsymbol{\sigma}, \overline{\boldsymbol{\chi}}_\alpha) = 0 \tag{48}$$

stating that plastic flow may occur only for stress states on the yield surface (yield state). However, these conditions do not allow to distinguish which deformation processes taking place from a yield state are actually plastic, *i.e.*, cause the development of plastic deformations. Moreover, no information is yet provided on how the plastic multiplier depends on the current state and the imposed deformation rate.

These issues are addressed by the so–called *Prager's consistency condition*, stating that for a plastic process taking place from a state on the yield surface the value of $f$ must remain zero, *i.e.*:

$$\dot{f} = \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \dot{\boldsymbol{\sigma}} + \frac{\partial f}{\partial \overline{\boldsymbol{\chi}}_\alpha} \cdot \dot{\overline{\boldsymbol{\chi}}}_\alpha = 0 \tag{49}$$

From eqs. (32) and (47) we can derive the following expressions for $\dot{\boldsymbol{\sigma}}$ and $\dot{\overline{\boldsymbol{\chi}}}_\alpha$:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^e \left( \dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p \right) = \boldsymbol{D}^e \left( \dot{\boldsymbol{\epsilon}} - \dot{\gamma} \, \frac{\partial f}{\partial \boldsymbol{\sigma}} \right) \qquad \boldsymbol{D}^e := \frac{\partial^2 \psi^e}{\partial \boldsymbol{\epsilon}^e \otimes \partial \boldsymbol{\epsilon}^e} (\boldsymbol{\epsilon}^e) \tag{50a}$$

$$\dot{\overline{\boldsymbol{\chi}}}_\alpha = -\boldsymbol{\Xi} \, \dot{\boldsymbol{\alpha}} = -\dot{\gamma} \, \boldsymbol{\Xi} \, \frac{\partial f}{\partial \overline{\boldsymbol{\chi}}_\alpha} \qquad \boldsymbol{\Xi} := \frac{\partial^2 \psi^p}{\partial \boldsymbol{\alpha} \otimes \partial \boldsymbol{\alpha}} \tag{50b}$$

which, inserted in eq. (49) provide the following expression for the plastic multiplier:

$$\dot{\gamma} = \frac{1}{K_p} \left\langle \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \right\rangle \tag{51}$$

where the McCauley brackets $\langle x \rangle := (x + |x|)/2$ are used to denote the positive part of their argument (as by definition the plastic multiplier cannot be negative) and the positive scalar $K_p$ is given by:

$$K_p := \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \frac{\partial f}{\partial \boldsymbol{\sigma}} + H_p > 0 \qquad H_p := \frac{\partial f}{\partial \overline{\boldsymbol{\chi}}_\alpha} \cdot \boldsymbol{\Xi} \, \frac{\partial f}{\partial \overline{\boldsymbol{\chi}}_\alpha} \tag{52}$$

in which $H_p$ is known as the *plastic modulus*. A positive value of $H_p$ denotes *hardening*, a negative value indicates *softening*, while $H_p = 0$ characterize the special case of *perfect plasticity*. As thoroughly discussed in, *e.g.*, [SH97, JB02], the constitutive assumption that $K_p > 0$ is crucial in the establishment of the correct formulation of the loading/unloading conditions in presence of softening. Its effect is essentially to place a restriction on the amount of allowable softening.

Substituting the expression (51) for the plastic multiplier in eqs. (50a) and (50b), we obtain the following constitutive equations and hardening laws in rate form:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{ep} \dot{\boldsymbol{\epsilon}} \qquad \dot{\overline{\boldsymbol{\chi}}}_\alpha = \boldsymbol{H}^p \dot{\boldsymbol{\epsilon}} \tag{53}$$

where:

$$\boldsymbol{D}^{ep} := \boldsymbol{D}^e - \frac{\mathcal{H}(\dot{\gamma})}{K_p} \left( \boldsymbol{D}^e \, \frac{\partial f}{\partial \boldsymbol{\sigma}} \right) \otimes \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \, \boldsymbol{D}^e \right) \tag{54a}$$

$$\boldsymbol{H}^p := \frac{\mathcal{H}(\dot{\gamma})}{K_p} \left( \boldsymbol{\Xi} \, \frac{\partial f}{\partial \overline{\boldsymbol{\chi}}_\alpha} \right) \otimes \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \boldsymbol{D}^e \right) \tag{54b}$$

where $\mathcal{H}(x)$ denotes the Heaviside step function, equal to one if $x > 0$ and zero otherwise.

The constitutive equations in rate–form given by eqs. (53) are incrementally bi–linear. In fact, according to the expression (51) for the plastic multiplier, the tangent stiffness $\boldsymbol{D}$ assumes two possible values depending on the direction of $\dot{\boldsymbol{\epsilon}}$:

$$\boldsymbol{D} = \begin{cases} \boldsymbol{D}^{ep} & \text{if} \quad (\partial f/\partial \boldsymbol{\sigma}) \cdot \boldsymbol{D}^{e} \dot{\boldsymbol{\epsilon}} > 0 \quad \text{(plastic loading conditions)} \\ \boldsymbol{D}^{e} & \text{if} \quad (\partial f/\partial \boldsymbol{\sigma}) \cdot \boldsymbol{D}^{e} \dot{\boldsymbol{\epsilon}} = 0 \quad \text{(neutral loading conditions)} \\ \boldsymbol{D}^{e} & \text{if} \quad (\partial f/\partial \boldsymbol{\sigma}) \cdot \boldsymbol{D}^{e} \dot{\boldsymbol{\epsilon}} < 0 \quad \text{(elastic unloading conditions)} \end{cases} \tag{55}$$

The continuity at the boundary between the two tensorial zones (the neutral loading conditions) is guaranteed by the fact that $\dot{\gamma} \to 0$ as the neutral loading condition is approached from the plastic loading zone.

# 7 Non–associative phenomenological plasticity

The evolution equations of the theory of hyperplasticity – eqs. (53) – are developed from the knowledge of the two scalar functions $\psi$ (free energy function) and $\mathcal{D}$ (dissipation function), in such a way to guarantee the consistency with the second principle of thermodynamics. In this respect, the name hyperplasticity is adopted to distinguish it from classical phenomenological plasticity in the same way as hyperelasticity is distinguished from hypoelasticity based on the existence of a potential function.

Historically, however, the classical theory of rate–independent plasticity has been developed following a different strategy, in which the various elements of the theory are chosen ad–hoc, based on the available experimental evidence. This phenomenological approach has led to the most successful applications of plasticity to the modeling of the inelastic and history–dependent behavior of soils, and it is, by far, still the most widely used in soil mechanics.

The main assumption of the classical phenomenological theory of plasticity will be presented in the following, pointing out the main differences with hyperplasticity.

## 7.1 Basic assumptions and general formulation

Starting from the additive split of the strain rate into an elastic and a plasic part, eq. (29), the elastic strain rate is linked to the stress rate by assuming a hypoelastic constitutive equation:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{e} (\boldsymbol{\sigma}) \, \dot{\boldsymbol{\epsilon}}^{e} = \boldsymbol{D}^{e} (\boldsymbol{\sigma}) \, (\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^{p}) \tag{56}$$

in which the elastic tangent stiffness tensor $\boldsymbol{D}^e$ generally depends on the current stress state.

Irreversibility is introduced by requiring that the state of the material $(\boldsymbol{\sigma}, \boldsymbol{q})$ belongs to the convex set:

$$\mathbb{E} := \left\{ (\boldsymbol{\sigma}, \boldsymbol{q}) \;\middle|\; f(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0 \right\} \tag{57}$$

defined in terms of a phenomenologically derived yield function $f$, depending on the current stress and on a set of internal variables $\boldsymbol{q}$ which account for the effects of the previous loading history.

The plastic strain rate is prescribed, as in hyperplasticity, by a suitable flow rule:

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\gamma}\,\frac{\partial g}{\partial \boldsymbol{\sigma}}(\boldsymbol{\sigma}, \boldsymbol{q}) \tag{58}$$

in which $g(\boldsymbol{\sigma}, \boldsymbol{q})$ is a prescribed *plastic potential* function, chosen in order to match available experimental observations – *e.g.*, stress–dilatancy relations. In general, the plastic potential function is independent of the yield function $f$. When $g$ and $f$ do not coincide, the flow rule is said to be *non–associative*. Associative plastic flow is recovered when $g \equiv f$, as in eq. (47a).

The evolution of the internal variables is provided by assigning a suitable *hardening law*:

$$\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{q}) \tag{59}$$

where $\boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{q})$ is a prescribed hardening function. Although the structure of the hardening law (59) is similar to the hardening law of hyperplasticity – eq. (47b) – and allows changes in the internal variables to take place only during plastic loading processes (for which $\dot{\gamma} > 0$), eq. (59) is *non–associative*, in the sense that the hardening function $\boldsymbol{h}$ is not derived from $\partial f / \partial \boldsymbol{q}$.

Again, the yield function and the plastic multiplier are subjected to the Kuhn–Tucker complementarity conditions of eq. (48), stating that plastic deformations may occur only for states on the yield surface. The consistency condition for plastic loading processes ($\dot{f} = 0$) allows to derive the following expression for the plastic multiplier:

$$\dot{\gamma} = \frac{1}{K_p}\left\langle \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \right\rangle \tag{60}$$

formally identical to eq. (51) but in which:

$$K_p := \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \frac{\partial g}{\partial \boldsymbol{\sigma}} + H_p > 0 \qquad\qquad H_p := -\frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{h} \tag{61}$$

Substituting eq. (60) in eqs. (56) and (59), we obtain:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{ep}\dot{\boldsymbol{\epsilon}} \qquad\qquad \dot{\boldsymbol{q}} = \boldsymbol{H}^p\dot{\boldsymbol{\epsilon}} \tag{62}$$

where:

$$\boldsymbol{D}^{ep} := \boldsymbol{D}^e - \frac{\mathcal{H}(\dot{\gamma})}{K_p}\left(\boldsymbol{D}^e \frac{\partial g}{\partial \boldsymbol{\sigma}}\right) \otimes \left(\frac{\partial f}{\partial \boldsymbol{\sigma}} \boldsymbol{D}^e\right) \tag{63a}$$

$$\boldsymbol{H}^p := \frac{\mathcal{H}(\dot{\gamma})}{K_p}\boldsymbol{h} \otimes \left(\frac{\partial f}{\partial \boldsymbol{\sigma}}\boldsymbol{D}^e\right) \tag{63b}$$

where $K_p$ is provided by eq. (61).

For the developments of Sect. 8, it is useful to recast the evolution equations (58) and (62) in terms of the unit tensors:

$$\boldsymbol{n}_f := \left\| \frac{\partial f}{\partial \boldsymbol{\sigma}} \right\|^{-1} \frac{\partial f}{\partial \boldsymbol{\sigma}} \qquad\qquad \boldsymbol{n}_g := \left\| \frac{\partial g}{\partial \boldsymbol{\sigma}} \right\|^{-1} \frac{\partial g}{\partial \boldsymbol{\sigma}} \tag{64}$$

providing the *loading direction* and the *plastic flow direction*, respectively. The flow rule, hardening law, plastic multiplier and elastoplastic tangent stiffness then assume the following alternative expressions:

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\lambda}\, \boldsymbol{n}_g \qquad\qquad \dot{\boldsymbol{q}} = \dot{\lambda} \hat{\boldsymbol{h}} \tag{65}$$

$$\dot{\lambda} = \frac{1}{\widehat{K}_p} \left\langle \boldsymbol{n}_f \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \right\rangle = \left\| \frac{\partial g}{\partial \boldsymbol{\sigma}} \right\| \dot{\gamma} \tag{66}$$

$$\boldsymbol{D}^{ep} = \boldsymbol{D}^e - \frac{\mathcal{H}(\dot{\lambda})}{\widehat{K}_p} \left( \boldsymbol{D}^e\, \boldsymbol{n}_g \right) \otimes \left( \boldsymbol{n}_f\, \boldsymbol{D}^e \right) \tag{67}$$

where:

$$\widehat{K}_p := \boldsymbol{n}_f \cdot \boldsymbol{D}^e \boldsymbol{n}_g + \widehat{H}_p \qquad\qquad \widehat{H}_p := \left( \left\| \frac{\partial f}{\partial \boldsymbol{\sigma}} \right\| \left\| \frac{\partial g}{\partial \boldsymbol{\sigma}} \right\| \right)^{-1} H_p \tag{68}$$

are the corresponding plastic moduli.

## 7.2 Perfect plasticity

The particular case in which the set of state variables contains the Cauchy stress only (*i.e.*, $\boldsymbol{q} = \boldsymbol{0}$) is known as *perfect plasticity*. In perfect plasticity the yield function and the plastic potential are given functions of the stress tensor $\boldsymbol{\sigma}$ only. The constitutive equations for perfect plasticity are recovered from eqs. $(62)_1$ and (63a), setting $H_p = 0$ in eq. (61). Due to this specific feature, in perfect plasticity yielding along a predefined stress path occurs at constant stress and constant plastic strain rate:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{0} \qquad\qquad \dot{\boldsymbol{\epsilon}}^e = \boldsymbol{0} \qquad\qquad \dot{\boldsymbol{\epsilon}}^p = \text{const.}$$

*i.e.*, yield states are also *failure states*.

The early applications of perfect plasticity to soil mechanics can be traced back to the various solutions of failure problems for foundations, retaining walls, or slopes obtained through the method of characteristics (slip line theory) [Sok65], or the application of the upper and lower bound theorems of limit analysis [Che76]. In both these approaches, the soil is modelled as a rigid–perfectly plastic medium, with a failure condition provided, *e.g.*, by the so–called *Mohr–Coulomb* yield function:

$$f(\boldsymbol{\sigma}) = (\sigma_1 - \sigma_3) - 2c \cos\phi - (\sigma_1 + \sigma_3) \sin\phi = 0 \tag{69}$$

where $\sigma_1$ and $\sigma_3$ are the maximum and minimum principal stresses, and $c$ and $\phi$ are two material constants defining the *cohesion* and the *friction angle* of the soil, respectively.

In these applications, the emphasis was placed in the determination of the stress field in limit conditions, in order to evaluate the stability of the system with respect to a particular collapse mechanism. The success of this simple and elegant approach to failure problems is witnessed by the fact that most of the design methods currently in use for geotechnical structures are still based on such limit solutions, and specific numerical techniques have been developed to extend limit analysis to those cases for which no sufficiently accurate analytic solution can be found, see, *e.g.*, [SK95, PBZL97, LS02].

Extension of the above concepts to the analysis of more complex deformation problems, such as soil–structure interaction or the modelling of the transition from the small–strain regime up to failure conditions, had to wait until the pioneering application of the finite element method to soil mechanics. Examples of the use of elastic–perfectly plastic models with pressure–dependent yield functions such as those of Mohr–Coulomb and Drucker–Prager models, are given, *e.g.*, in [ZH77, SD83] for shallow foundations, [ZHL75] for slopes, [SH92, SGvW95] for flexible retaining structures, and [RK83, WK91] for tunnels.

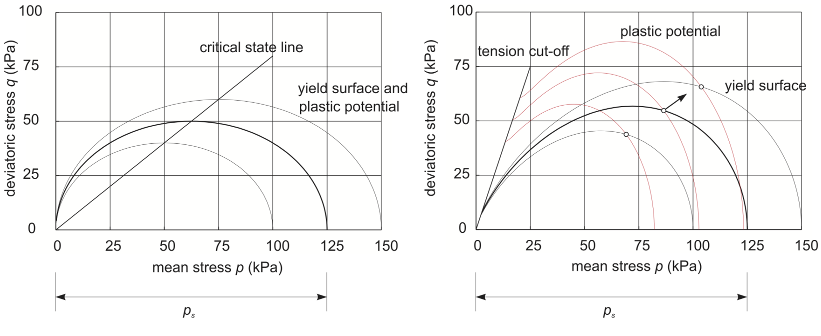
## 7.3 Isotropic hardening plasticity

The experience gathered in using classical perfect plasticity in the analysis of deformation problems has shown how these formulations provide a too crude description of the actual behavior of natural soils in pre–failure conditions. A radical change of perspective in soil plasticity occurred after the pioneering work of Roscoe and coworkers in Cambridge, which lead, in the sixties, to the basic principles of the so–called “Critical State Soil Mechanics” (CSSM) [RB68, SW68]. The practical use of CSSM in geotechnical applications started in the early seventies, when CSSM was interpreted as a particular application of isotropic hardening plasticity, see *e.g.*, [ZN71], and generalized to full six–dimensional stress and strain states. The road was then open to a new approach to geotechnical engineering practice, in which no such distinction between failure and deformation problems, or elastic response and plastic collapse was needed any longer.

Isotropic hardening plasticity is obtained from the general formulation of Sect. 7.1 when all the elements of the pseudo–vector $\boldsymbol{q}$ collecting the internal variables are *scalar quantities*, and, as such do not provide any information about the orientation of the microstructure.

The prototype of isotropic hardening elastoplastic models for cohesive soils is the so–called “Modified Cam–Clay” (MCC) [RB68], which assumes an associative flow rule. The yield surface adopted in the original MCC model is given by:

$$f(p, q, p_s) = p(p - p_s) + \frac{q^2}{M^2} = 0 \tag{70}$$



(a) (b)

Figure 1: Yield surfaces and plastic potentials for isotropic hardening elastoplastic models: a) Modified Cam–Clay [RB68]; b) Sinfonietta Classica [Nov88].

In the $q$:$p$ stress invariant space, it has the shape of an ellipse passing through the origin, with its principal axes parallel to the coordinate axes. The scalar quantity $p_s$ (*preconsolidation pressure*) controls the size of the yield surface, and represents the only internal state variable of the material. For any possible value of $p_s$, eq. (70) describes a family of such ellipses, see fig. 1a. In eq. (70) $M$ is a material constant defining the aspect ratio of the ellipse.

The evolution equation for the preconsolidation pressure is provided by an empirically derived logarithmic law of the type[1]:

$$\dot{p}_s = \rho_s \, p_s \, \dot{\epsilon}_v^p \tag{71}$$

with $\rho_s = \mathrm{const.}$ The hardening law provided by eq. (71) is purely volumetric, *i.e.*, $p_s$ may change only when plastic volumetric strains occur. Positive (contractant) plastic volumetric strains cause an increase in $p_s$ (expansion of the elastic domain), while negative (dilatant) plastic volumetric strains induce a reduction of $p_s$ and the shrinkage of the elastic domain.

In the notation of eqs. (59) and $(61)_2$:

$$\dot{p}_s = \dot{\gamma} \, h_s(p, q, p_s) \qquad h_s := \rho_s \, p_s \, \frac{\partial f}{\partial p} \qquad H_p = -\rho_s \, p_s \, \frac{\partial f}{\partial p} \, \frac{\partial f}{\partial p_s} \tag{72}$$

From eq. (72) it is clear that the "failure" conditions for the material (which occur when $p_s = \mathrm{const.}$ and $H_p = 0$) are characterized by purely distortional plastic strain rates, *i.e.*, the material can be deformed indefinitely at *constant stress and constant volume*. Such particular failure states, the existence of which is experimentally observed in both fine– and coarse–grained soils, are defined *critical states*, and form the basis of almost all subsequent modern treatments of hardening plasticity for soils.

[1] Note that, to avoid using too many different symbols, the notations employed in this work can sometimes be different from the one adopted in the original works cited.

Modifications of MCC to improve its predictive capabilities have been discussed by numerous authors. Among them, we recall the extension of the yield function (70) to include the third stress invariant $\theta$ [ZN73], the adoption of a composite yield surface to include the so called *Hvorslev surface* for yield points on the “supercritical side” of the critical state line [ZN73, HWW84], and the adoption of a hyperelastic constitutive equation [Hou85, BTA97]. The isotropic hardening models known in the literature as *cap models* [DMS71, SDMB76] can be considered essentially as Critical State models with a modified supercritical yield function, the position of which, however, does not change with plastic strains.

In the application of the concepts of isotropic hardening plasticity to coarse–grained soils, two major limitations of classical critical state models have been pointed out. First, the assumption of an associated flow rule is generally not supported by available experimental data on sand dilatancy, see *e.g.*, [PHS66, PHS67]. In addition, the modeling of *static liquefaction* in the *hardening regime*, observed in loose sands under undrained conditions is not possible adopting an associative flow rule [Nov96]. Second, the hypothesis of purely volumetric hardening does not allow to describe the so–called *phase transition* effect – *i.e.*, the transition from contractant to dilatant behavior – typically observed in dense sand under undrained compression.

Non–associative isotropic hardening models for sands have been proposed since the pioneering work of Pooroshasb *et al.* [PHS66, PHS67], who coupled a Cam–Clay type plastic potential with a classical Mohr–Coulomb yield locus. Subsequent improvements were proposed, *e.g.*, by Nova & Wood [NW79] and Kim & Lade [KL88, LK88]. As for the hardening function, Nova [Nov77] and Wilde [Wil77] independently proposed an extension of the volumetric hardening rule (71) which incorporates the effect of deviatoric plastic strain rate:

$$\dot{p}_s = \rho_s \, p_s \, \{\dot{\epsilon}_v^p + \xi_s \dot{\epsilon}_s^p\} \qquad H_p := -\rho_s \, p_s \left\{ \frac{\partial g}{\partial p} + \xi_s \, \frac{\partial g}{\partial q} \right\} \frac{\partial f}{\partial p_s} \tag{73}$$

The scalar quantity $\xi_s$ appearing in eq. (73) can be considered either a constant, as in [Nov77], or a monotonically decreasing function of the accumulated plastic deviatoric strains, as in [Wil77]. In this last case, a critical state is recovered in the ultimate conditions at very large plastic strains.

An example of isotropic hardening models for sands – which combines good predictive capabilities for monotonic loading with a limited number of material constants easily linked to observed material behavior in standard tests – is provided by the model proposed by Nova under the name *Sinfonietta Classica* [Nov88]. For this model, the adopted yield function and plastic potential are given by the following equations:

$$f(p, \boldsymbol{r}, p_s) = 3\beta \, (\gamma - 3) \, \ln\left(\frac{p}{p_s}\right) - \gamma \, \mathrm{tr}\left(\boldsymbol{r}^3\right) + \frac{9}{4} \, (\gamma - 1) \, \mathrm{tr}\left(\boldsymbol{r}^2\right) = 0 \tag{74}$$

$$g(p, \boldsymbol{r}, p_s^*) = 9 \, (\gamma - 3) \, \ln\left(\frac{p}{p_s^*}\right) - \gamma \, \mathrm{tr}\left(\boldsymbol{r}^3\right) + \frac{9}{4} \, (\gamma - 1) \, \mathrm{tr}\left(\boldsymbol{r}^2\right) = 0 \tag{75}$$

where $\boldsymbol{r} := \boldsymbol{s}/p$ is the stress–ratio tensor while $\beta$ and $\gamma$ are material parameters ($\beta \neq 3$ denotes non–associative behavior). The corresponding surfaces in the $q : p$ plane are shown in Fig. 1b. An hardening law with volumetric and deviatoric hardening similar to eq. (73) is assumed for the internal variable $p_s$, which controls the size of the yield surface.

Subsequent developments gave rise to a number of constitutive models which progressively diverged from the basic assumptions of CSSM in the attempt of covering further aspects of experimentally observed soil behavior, as well as to tackle other, more challenging classes of engineering problems. The breath and depth of such scientific production is well portrayed, for example, by the proceedings of the workshops held in Grenoble in 1982 [GDV84], Cleveland in 1988 [SB89], and Horton in 1992 [Kol93]. Another useful source of references is provided by the special volume published on the occasion of the XI ICSMFE [Mur85].

## 7.4 Anisotropic hardening plasticity

Almost all geotechnical materials such as rocks, coarse–grained soils and fine–grained soils are characterized – to a certain extent – by the existence of some preferential orientations at the microstructural level. In granular soils such preferential orientations can be associated to the spatial distributions of the contact normals, to grain shape and to void shape, see [ONNK85]. Moreover, the directional properties of the microstructure might remain more or less stable during the deformation of the solid skeleton (as, *e.g.*, the distribution of grain orientations in the tests performed by Oda *et al.* [ONNK85]), or they might evolve as a consequence of grain rearrangements upon applied loading (as, *e.g.*, the distribution of contact normals [ONNK85]). From this observations, it follows naturally that the macroscopic response of the material – reflecting the properties of the microstructure – can be characterized by a more or less marked *anisotropy*, both in terms of stress–strain response in pre–failure conditions, and in terms of shear strength.

According to the possibility that superimposed loading histories may change the directional properties of the microstructure, two different kind of anisotropy can be distinguished at the macroscopic level, see [CC44]:

– *inherent anisotropy*, “[...] a physical characteristics inherent in the material and entirely independent of the applied strains”;

– *induced anisotropy*, “[...] a physical characteristic due exclusively to the strain associated with the applied stress”.

Inherent anisotropy is usually relevant in hard, heavily overconsolidated soils and stratified rocks, where strong intergranular bonds prevent the occurrence of significant rearrangements of the microstructure, or in coarse–grained soils with strongly non–circular particles, the orientation of which cannot be modified easily unless a substantial amount of grain crushing occurs. On the contrary, induced anisotropy plays a major role in non–cemented granular soils with rounded particles, or in clays where the applied loading can modify and, in some cases, even erase the effects of the previous loading history. Direct and indirect experimental evidence of inherent and induced anisotropy is reported, *e.g.*, in [ABS77] for soft rocks, in [OKH78, WA85, YMH91, YIV98] for sands, and in [DS66, Mit70, TL77, GNL83, SJH92] for clays.

In the framework of the classical theory of plasticity, inherent anisotropy can be dealt with in the formulation of the elastic constitutive equation – as in, *e.g.*, [Boe75, GH83] – and/or in the definition of yield and plastic potential functions. Constitutive equations for inherent anisotropy have been proposed, *e.g.*, by Nova [Nov86], Pastor [Pas91] and Semnani *et al.* [SWB16], based on a approach first suggested by Hill [Hil50]. Essentially, these models are derived from existing isotropic hardening formulations by replacing the standard invariants of the stress tensor with corresponding anisotropic invariants defined by means of suitably chosen (constant) *structure tensors*, which are employed as metric tensors in the construction of the scalar invariants entering in the constitutive functions. An alternative strategy to incorporate inherent anisotropy, based on the use of a microstructure tensor in the definition of the yield surface, has been proposed in [PM00, PLS02, OKKA02].

The description of induced anisotropy – *i.e.*, the evolution of the directional properties of the material with the loading history – requires the set of internal state variables $\boldsymbol{q}$ to include at least one tensor–valued quantity. In most of the existing anisotropic hardening plasticity models, this is usually assumed to be a symmetric second–order tensor, with the character of a *microstructure tensor*. Although this limits the degree of symmetry of the material to *orthotropy*, see [Boe87], it is considered sufficient for most geomaterials of relevant practical interest.

In presence of a symmetric second–order microstructure tensor among the internal variables, the general restrictions imposed to the yield and plastic potential functions by the principle of material frame indifference, as well as the consequences of induced anisotropy on the relative orientation between the principal directions of the stress and the plastic strain tensors are discussed in detail in [BD84]. Plasticity models with anisotropic hardening can be broadly grouped into two different classes, according to the experimental evidence which they were intended to reproduce, namely:

a) constitutive models with *kinematic hardening*, capable of modelling soil behavior under cyclic loading paths, see, *e.g.*, [Woo82] and references therein;

b) constitutive models with *rotational hardening*, which are capable of describing the changes in the orientation of the yield surface with the evolution of plastic strains, as observed, *e.g.*, in [YMH91, TL77, GNL83, SJH92].

Kinematic hardening models for soils originate from the pioneering work of Mroz [Mro67], Iwan [Iwa67], and Dafalias & Popov [DP75]. In such models, a yield function of the form:

$$f(\boldsymbol{\sigma}, \boldsymbol{\alpha}, q_k) = \hat{f}(\hat{\boldsymbol{\sigma}}, q_k) = 0 \qquad\qquad \hat{\boldsymbol{\sigma}} := \boldsymbol{\sigma} - \boldsymbol{\alpha} \tag{76}$$

is assumed, in which the so–called *back–stress* $\boldsymbol{\alpha}$ is the microstructure tensor, responsible for the induced anisotropy, and the scalars $q_k$ $(k = 1, \dots, n)$ denote the

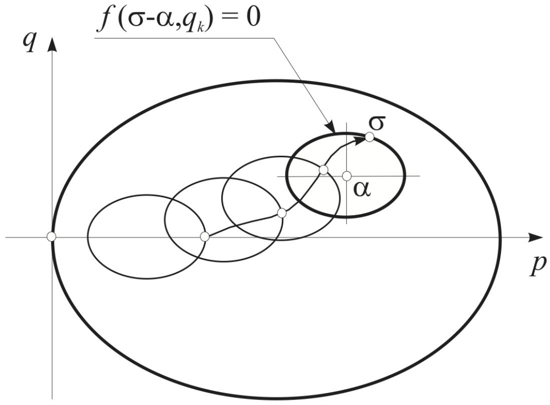


Figure 2: Kinematic hardening inside the Bounding Surface.

other internal variables. As $\boldsymbol{\alpha}$ changes during the loading process, the yield surface is dragged by the stress–path as indicated qualitatively in Fig. 2. The motion of the yield surface in stress space is typically restricted by a larger, outer surface, referred to as *Bounding Surface* (BS), of equation:

$$F(\boldsymbol{\sigma}, \overline{q}_k) = 0 \qquad \{\overline{q}_k\} \subset \{q_k\} \tag{77}$$

In classical anisotropic plasticity, the BS – generally similar in shape to the yield surface – provides a limit to the possible evolution of the back–stress $\boldsymbol{\alpha}$. Models of this kind have been proposed by various authors. Among them we recall the works of Prevost [Pre77, Pre86], Mroz *et al.* [MNZ78, MNZ81], Hashiguchi [Has85, Has88], Wood and coworkers [ATW89, GW99, RW00] and Stallebrass and Taylor [ST97].

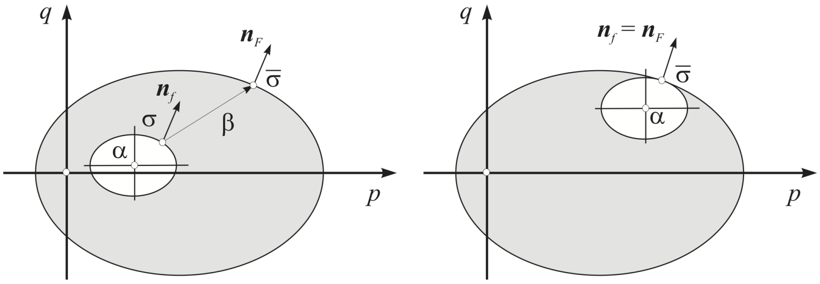
Most of these works represent a straightforward extension of classical Modified Cam–Clay, see Sect. 7.3. As an example, in the model of Al–Tabbaa & Wood [ATW89], the yield and Bounding Surface functions are given by:

$$F(\boldsymbol{\sigma}, p_c) = \frac{3}{2M_\theta}\, \boldsymbol{s} \cdot \boldsymbol{s} + (p - p_c)^2 - p_c^2 = 0 \tag{78}$$

$$f(\boldsymbol{\sigma}, p_c) = \frac{3}{2M_\theta}\, (\boldsymbol{s} - \operatorname{dev} \boldsymbol{\alpha}) \cdot (\boldsymbol{s} - \operatorname{dev} \boldsymbol{\alpha}) + (p - p_\alpha)^2 - R^2 p_c^2 = 0 \tag{79}$$

where $p_c = p_s/2$, $p_\alpha = \operatorname{tr} \boldsymbol{\alpha}/3$ and $R \ll 1$ is a material constant representing the ratio between the sizes of the two surfaces.

In this class of models, the hardening function adopted for $p_c$ (or $p_s$) is similar to the one adopted in critical state models, see eq. (71). As for the tensor $\boldsymbol{\alpha}$, rather than prescribing explicitly the hardening function, the hardening modulus $H_p$ is assigned as a monotonically decreasing function of the distance $\delta$ between the current state and a *image state* $\overline{\boldsymbol{\sigma}}$ on the BS, defined as the point at which the unit normals to $f = 0$ and



(a) (b)

Figure 3: Kinematic hardening models: definition of the image point.

$F = 0$ have the same direction (see Fig. 3a):

$$H_p = \widehat{H}\left(\overline{H}_p, \delta\right) \qquad \frac{\partial \widehat{H}}{\partial \delta} > 0 \qquad \widehat{H}\left(\overline{H}_p, 0\right) = \overline{H}_p \tag{80}$$

In eq. (80), $\delta := \|\overline{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|$ and $\overline{H}_p$ is the plastic modulus at $\overline{\boldsymbol{\sigma}}$:

$$\overline{H}_p := -\frac{\partial F}{\partial p_c} h_c \tag{81}$$

obtained from the consistency condition on the BS:

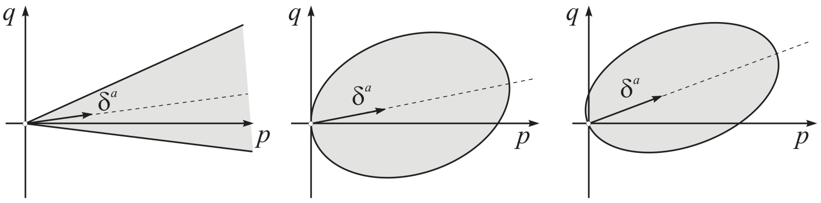
$$\dot{F}(\overline{\boldsymbol{\sigma}}, \overline{q}_k) = 0$$

When the stress–path touches the BS, the two surfaces must share the same tangent, otherwise some admissible states would fall *outside* the BS, see Fig. 3b. As shown by Hashiguchi [Has85], this is obtained through an appropriate definition of the evolution equation for $\boldsymbol{\alpha}$. For the Al–Tabbaa & Wood model [ATW89], the non–intersection condition requires that:

$$\dot{\boldsymbol{\alpha}} = \dot{\overline{\boldsymbol{\alpha}}} + (\boldsymbol{\alpha} - \overline{\boldsymbol{\alpha}}) \frac{\dot{p}_c}{p_c} + \frac{\boldsymbol{n} \cdot \left[\dot{\overline{\boldsymbol{\sigma}}} - (\dot{p}_c/p_c)\, \overline{\boldsymbol{\sigma}}\right]}{\boldsymbol{n} \cdot (\overline{\boldsymbol{\sigma}} - \boldsymbol{\sigma})} (\overline{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) \tag{82}$$

In eq. (82), the first term is related to the translation of the center of the BS, the second represents the effect of the change in size of the BS (and of the yield surface), and the third a net translation in the direction of the tensor $\boldsymbol{\beta} := \overline{\boldsymbol{\sigma}} - \boldsymbol{\sigma}$, see Fig. 3a.

Anisotropic plasticity models with rotational hardening are more suitable for describing the anisotropy induced by loading histories associated to depositional processes in natural deposits, such as one–dimensional compression and, possibly, swelling. These models can be traced back to the pioneering works of Sekiguchi & Ohta [SO77] for clays, or Ghaboussi & Momen [GM82] for sands. Constitutive equations of this kind



(a) (b) (c)

Figure 4: Typical yield surfaces adopted in rotational hardening models.

proposed for sands are usually intended to model irreversible processes associated with deviatoric loading paths, and therefore adopt conical–shaped yield surfaces, *open* towards the range of high mean pressures (fig. 4a). Among them, we recall the models proposed in refs. [GM82, PP85, MD97, GW99, DM04, TD08].

Rotational hardening models for fine–grained soils, on the contrary, adopt *closed* yield surfaces (fig. 4b,c), in order to reproduce the irreversible deformations usually observed in these materials along isotropic or proportional loading paths ($q/p = \mathrm{const.}$). Examples of rotational hardening models for clays are given in the works of [Has79, BY86, AD86, WNKL03, DMP06, TDP10].

Exceptions to this general trend are provided, for example, by the models of di Prisco *et al.* [dPNL93] – actually a generalization of the Sinfonietta Classica model discussed in the previous section – and Pestana & Whittle [PW99], which can be employed for coarse as well as fine–grained materials. It is worth noting that in most of the aforementioned models, the rotational anisotropy is employed in connection to some form of generalized plasticity allowing plastic flow inside the main state boundary surface, which will be discussed in Sect. 8.

Rotational hardening models can be easily derived as generalizations of classical isotropic hardening formulations (*e.g.*, Modified Cam–Clay) by simply replacing the stress invariants entering in the yield and plastic potential functions with appropriate *mixed invariants* which take into due account the microstructure tensor. Possible ways of defining such mixed invariants are provided, for example, by Anandarajah and Dafalias [AD86] (slightly modified):

$$p^a := \frac{1}{3}\,\boldsymbol{\sigma}\cdot\boldsymbol{\delta}^a \qquad q^a := \sqrt{\frac{3}{2}}\,\|\boldsymbol{s}^a\| \qquad \sin(3\theta^a) := \sqrt{6}\,\frac{(\boldsymbol{d}^a)^3\cdot\mathbf{1}}{[(\boldsymbol{d}^a)^2\cdot\mathbf{1}]^{3/2}} \tag{83}$$

where $\boldsymbol{\delta}^a$ is the microstructure tensor, and:

$$\boldsymbol{s}^a := \boldsymbol{\sigma} - p^a\,\boldsymbol{\delta}^a \qquad \boldsymbol{d}^a := \mathrm{dev}\,(\boldsymbol{s}^a) \qquad \boldsymbol{\delta}^a\cdot\boldsymbol{\delta}^a = 3 \tag{84}$$

or by Wheeler *et al.* [WNKL03]:

$$p^a := \frac{1}{3}\,\boldsymbol{\sigma}\cdot\mathbf{1} = p \qquad q^a := \sqrt{\frac{3}{2}}\,\|\boldsymbol{s}^a\| \qquad \sin(3\theta^a) := \sqrt{6}\,\frac{(\boldsymbol{s}^a)^3\cdot\mathbf{1}}{[(\boldsymbol{s}^a)^2\cdot\mathbf{1}]^{3/2}} \tag{85}$$

where $\boldsymbol{\delta}^a$ is a purely deviatoric microstructure tensor, and:

$$\boldsymbol{s}^a := \boldsymbol{s} - p\,\boldsymbol{\delta}^a \qquad \boldsymbol{s} := \operatorname{dev}(\boldsymbol{\sigma}) \qquad \boldsymbol{\delta}^a \cdot \mathbf{1} = 0 \tag{86}$$

In eqs. (83) and (84), the projection of $\boldsymbol{\sigma}$ on the isotropic axis, commonly used to construct the isotropic and deviatoric invariants of the stress tensor, are replaced by the corresponding projection on the microstructure tensor $\boldsymbol{\delta}^a$, now playing the role of the unit tensor – compare eqs. (83) with eqs. (2) – and, in fact, defining the rotation of the surface with respect to the isotropic axis, see fig. 4c.

In eqs. (85) and (86) the yield surface is distorted in the direction of the deviatoric axis, rather than rotated around the origin of the stress space. This effect is obtained by shifting the deviatoric stress by a quantity proportional to the deviatoric microstructure tensor and the current mean stress, see fig. 4b.

Several alternative strategies have been proposed to link the evolution of the microstructure tensor (*i.e.*, the rotation of the yield surface) with the plastic strain rate. All of them must, however, satisfy the orthogonality condition $\dot{\boldsymbol{\delta}}^a \cdot \boldsymbol{\delta}^a = 0$, required by the assumption $(84)_3$, or the requirement set by eq. $(86)_3$. A thorough discussion on the different rotational hardening mechanisms adopted for fine–grained soils has been presented by Dafalias and Taiebat [DT13].

# 8 Bounding Surface models and generalized plasticity

An important limitation of classical elastoplasticity as applied to geomaterials is represented by the assumption of a large elastic domain, inside which the response of the material is purely reversible. In light of the concepts introduced in Sects. 6 and 7, classical elastoplasticity is characterized by an incrementally bi–linear constitutive equation *only* for states *on the yield surface*. All elastic states are, by definition, endowed with an incrementally linear response. However, a large body of experimental evidence suggests that soil behavior can be irreversible and path–dependent *even for strongly preloaded states*, and that plastic yielding is a rather gradual process. Although such effects can be considered of secondary importance in the simulation of monotonic loading paths, it must be noted that a strong dependence of the small–strain stiffness on the loading path direction has been observed, *e.g.*, by [ARS86, Sta90] in heavily overconsolidated soils, and that such a feature of soil behavior – which cannot be reproduced by any incrementally linear model – can be of great importance in all practical applications in which strong variations of the stress–path direction are expected in different zones of the soil mass, *e.g.*, in the analysis of excavations. Moreover, irreversible (plastic) strains occurring well inside the locus of admissible stress states are obviously of great importance in cyclic loading processes, and the accurate description of such phenomena as cyclic mobility or liquefaction under repeated loading (see, *e.g.*, [Woo82]) requires to take them into proper account.

The kinematic hardening models discussed in the previous section – mostly developed during the early '80, in response to the problems posed by the design of structures such as offshore platforms, or by the quantitative prediction of soil response during earthquakes – are certainly capable to deal successfully with this particular issue. However, a number of alternative strategies have also been proposed for the same purpose, which represent genuine generalizations of the classical framework. Among them, definitely worth of mention are the so–called *Bounding Surface models*, originally developed by Dafalias and coworkers, and the models developed in the framework of *Generalized Plasticity*, as defined by Pastor et al. [PZC90].

The key concept in the formulation of a Bounding Surface model is the fact that, as in kinematic hardening elastoplastic models mentioned before, there exists a surface in stress space – the Bounding Surface (BS), defined by an equation similar to eq. (77) – which separates admissible from impossible states. Such a surface is subjected to hardening processes which may change its size, shape and orientation due to the development of plastic strains, exactly as a standard yield surface in classical plasticity. However, such a surface is *not* a yield surface, as plastic strains can occur for stress states located in its interior. In particular, at each admissible state (inside or on the BS), a flow rule identical to eq. $(65)_1$ is assumed, in which the plastic multiplier $\dot{\lambda}$ is replaced by:

$$\dot{\lambda} = \frac{1}{\widetilde{K}_p} \left\langle \boldsymbol{n}_L \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \right\rangle \tag{87}$$

where:

$$\widetilde{K}_p := \boldsymbol{n}_L \cdot \boldsymbol{D}^e \boldsymbol{n}_g + \widetilde{H}_p \tag{88}$$

in which $\boldsymbol{n}_L$ is a unit tensor defining the loading direction, and $\widetilde{H}_p$, by analogy with the standard formulation, plays the role of the plastic modulus. The definition of these last two quantites relies crucially on the possibility of associating to each stress state $\boldsymbol{\sigma}$ inside the BS a corresponding *image state* $\overline{\boldsymbol{\sigma}}$ on the BS, through a non–invertible *mapping rule*.

In the so–called *radial mapping* BS models, see [Daf86], this is accomplished by simply projecting the current stress onto the BS from a given *projection center* $\boldsymbol{\alpha}$, see Fig. 5. Once the image state is found, the loading direction is taken as the gradient of the BS at $\overline{\boldsymbol{\sigma}}$, while the plastic modulus $\widetilde{H}_p$ is assumed to be a monotonically decreasing function of the distance $\delta := \|\overline{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|$ between the current state and the image state, and of the plastic modulus $\overline{H}_p$ at $\overline{\boldsymbol{\sigma}}$:

$$\widetilde{H}_p = \widetilde{H}\left(\overline{H}_p, \delta\right) \qquad \frac{\partial \widetilde{H}}{\partial \delta} > 0 \qquad \widetilde{H}\left(\overline{H}_p, 0\right) = \overline{H}_p \tag{89}$$

The stress–strain relation in rate form is then given by an equation similar to eq. $(62)_1$, with the tangent stiffness $\boldsymbol{D}^{ep}$ provided by eq. (67), the plastic multiplier $\dot{\lambda}$ provided by eq. (66) and $\widehat{K}_p$ replaced by $\widetilde{K}_p$ of eq. (89). The analogies existing between this procedure for defining the loading direction and the plastic modulus and the one outlined for kinematic hardening models in Sect. 7.4 are apparent. As a matter of fact, Dafalias [Daf86] considered kinematic hardening models as a special class of BS models, characterized by a special form of mapping rule.

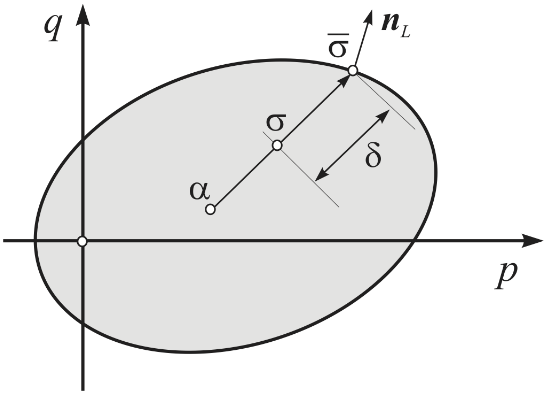


Figure 5: Radial mapping rule in Bounding Surface models.

However, differently than in kinematic hardening plasticity, in radial mapping BS models, no elastic region exists anymore, and the material features an incrementally bi–linear response at *any* state. A comprehensive review of the Bounding Surface concept is provided by Dafalias [Daf86]. Applications of the Bounding Surface Concept to the modelling of clays are reported, *e.g.*, in [ZLP85, DH86, AD86, WK94, LYKT02, DMP06, TDP10], while applications to coarse–grained soils are given, *e.g.*, by [PZL85, Bar86, Cw94, MD97, DM04, TD08].

Starting from the works of Zienkiewicz & Mroz [ZM84], Pastor *et al.* [PZC90] developed the framework of *Generalized Plasticity* as a further generalization of the Bounding Surface concept, where the concepts of plastic potential, yield function and consistency condition are completely abandoned. In the incrementally bi–linear version of the theory, the plastic strain rate is provided by the following equations:

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\lambda}_L\, \boldsymbol{n}_{gL} \qquad \text{if}: \quad \boldsymbol{n}_L \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} > 0 \quad \text{(loading)} \tag{90}$$

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\lambda}_U\, \boldsymbol{n}_{gU} \qquad \text{if}: \quad \boldsymbol{n}_L \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} < 0 \quad \text{(unloading)} \tag{91}$$

$$\dot{\boldsymbol{\epsilon}}^p = \boldsymbol{0} \qquad \text{if}: \quad \boldsymbol{n}_L \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} = 0 \quad \text{(neutral loading)} \tag{92}$$

in which:

$$\dot{\lambda}_L = \frac{1}{\widehat{K}_{p,L}}\, \boldsymbol{n}_L \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \qquad \widehat{K}_{pL} := \boldsymbol{n}_L \cdot \boldsymbol{D}^e \boldsymbol{n}_{gL} + \widehat{H}_{p,L} \tag{93}$$

$$\dot{\lambda}_U = \frac{1}{\widehat{K}_{p,U}}\, \boldsymbol{n}_L \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \qquad \widehat{K}_{pU} := \boldsymbol{n}_L \cdot \boldsymbol{D}^e \boldsymbol{n}_{gU} + \widehat{H}_{p,U} \tag{94}$$

In eqs. (90)–(94), $\boldsymbol{n}_L$, $\boldsymbol{n}_{gL}$ and $\boldsymbol{n}_{gU}$ are three second–order unit tensors representing the loading direction, the plastic flow direction for plastic loading and the plastic flow direction for plastic unloading (*reverse loading*), respectively, while the scalars $\widehat{H}_{p,L}$ and $\widehat{H}_{p,U}$ are the corresponding plastic moduli for (plastic) loading and unloading. All these quantities are considered as prescribed functions of the state variables $(\boldsymbol{\sigma}, \boldsymbol{q})$, and, in general, their definition do not require any yield function, plastic potential or consistency condition to be assumed.

The corresponding expressions for the elastoplastic tangent stiffness tensor are given by:

$$\boldsymbol{D}^{ep} = \begin{cases} \boldsymbol{D}^e - (1/\widehat{K}_{p,L})\ (\boldsymbol{D}^e\, \boldsymbol{n}_{gL}) \otimes (\boldsymbol{n}_L\, \boldsymbol{D}^e) & \text{(plastic loading)} \\ \boldsymbol{D}^e - (1/\widehat{K}_{p,U})\ (\boldsymbol{D}^e\, \boldsymbol{n}_{gU}) \otimes (\boldsymbol{n}_L\, \boldsymbol{D}^e) & \text{(plastic unloading)} \end{cases} \tag{95}$$

It is worth noting that both classical plasticity and Bounding Surface plasticity are recovered from generalized plasticity as special cases, with suitable choices for the constitutive functions $\boldsymbol{n}_L$, $\boldsymbol{n}_{gL}$, $\boldsymbol{n}_{gU}$, $\widehat{H}_{p,L}$ and $\widehat{H}_{p,U}$, see [PZC90] for further details.

# 9 Plasticity with generalized hardening

A last, notable case of incrementally bilinear formulations is provided by the theory of plasticity with generalized hardening – as defined by Tamagnini and Ciantia [TC16] – proposed in the geomechanics context to describe a number of practically relevant aspects of the mechanical behavior of geomaterials. A common, distinctive feature of those constitutive theories is that the size and shape of the yield locus, as well as its evolution with the loading process are assumed to depend, in addition to accumulated plastic strains, on some other non–mechanical state variables, usually of scalar nature. Among them, we recall:

- the thermoplastic models proposed by Nova [Nov86] or Laloui and Cekerevac [LC08] to describe the influence of temperature on the brittle–ductile transition of rocks in geophysical applications, in which the preconsolidation pressure depend on the temperature $T$;
- the elastoplastic models for unsaturated soil (formulated in terms of Bishop effective stresses) in which an explicit dependence of the size of the yield surface on the degree of saturation is assumed to simulate the phenomenon of collapse upon wetting for partially saturated soil, see, *e.g.*, [Jom00];
- the extension of classical elastoplasticity advocated by [Nov00] to describe the effects of weathering on cemented soils or weak rocks, in which some bonding–related internal variables are subject to both mechanical and chemical degradation, described through a normalized, scalar weathering function $X_d$, see [TCN02, NCT03].

These approaches share also some similarities with a number of viscoplastic models based on the concept of a non–stationary yield locus, see *e.g.*, [FN90, Bor92], and to chemoplastic models proposed for early–age concrete [UC96] or clays subject to environmental loading [Hue92, Hue97].

The main features of the theory, as detailed in [TC16] are summarized in the following. Let $\vartheta$ denote the additional (scalar) variable affecting the mechanical response of the material, *i.e.*, temperature, suction or chemical degradation. A first modification of the classical theory to account for the changes in $\vartheta$ is introduced in the elastic constitutive equations, which now reads, in rate–form:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^e \left(\boldsymbol{\sigma}, \vartheta\right) \left(\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p\right) + \boldsymbol{m} \left(\boldsymbol{\sigma}, \vartheta\right) \dot{\vartheta} \tag{96}$$

In eq. (96), $\boldsymbol{m}(\boldsymbol{\sigma}, \vartheta)$ is a coupling coefficient (*e.g.*, thermal stress coefficient for $\vartheta \equiv T$). While the definitions of elastic domain, flow rule and loading/unloading conditions are identical to those of the classical theory – eqs. (57), (58) and (55) – the evolution equation for the internal variables now assumes the following generalized form:

$$\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{q}, \vartheta) + \dot{\vartheta}\boldsymbol{\eta}(\boldsymbol{\sigma}, \boldsymbol{q}, \vartheta) \tag{97}$$

where: $\boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{q}, \vartheta)$ and $\boldsymbol{\eta}(\boldsymbol{\sigma}, \boldsymbol{q}, \vartheta)$ are suitable hardening functions. The first term on the RHS of eq. (97) quantifies the changes in the internal variables due to plastic deformations, while the second term accounts for all non–mechanical hardening/softening processes induced by a change of $\vartheta$.

From the consistency condition $\dot{\gamma}\dot{f}(\boldsymbol{\sigma}, \boldsymbol{q}) = 0$, the elastic constitutive equation (96) and the flow rule (58), the following generalized expression for the plastic multiplier is obtained:

$$\dot{\gamma} = \frac{1}{K_p} \left\langle \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} + \left( \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{\eta} + \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{m} \right) \dot{\vartheta} \right\rangle \tag{98}$$

with $K_p$ given by eq. (61). This in turns provides the following constitutive equations in rate form:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{ep}\,\dot{\boldsymbol{\epsilon}} + \boldsymbol{m}^{ep}\,\dot{\vartheta} \tag{99}$$

$$\dot{\boldsymbol{q}} = \boldsymbol{G}\,\dot{\boldsymbol{\epsilon}} + \boldsymbol{G}_\vartheta\,\dot{\vartheta} \tag{100}$$

in which:

$$\boldsymbol{D}^{ep} := \boldsymbol{D}^e - \frac{\mathcal{H}(\dot{\gamma})}{K_p} \left( \boldsymbol{D}^e \frac{\partial g}{\partial \boldsymbol{\sigma}} \right) \otimes \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \boldsymbol{D}^e \right) \tag{101}$$

$$\boldsymbol{m}^{ep} := \boldsymbol{m} - \frac{\mathcal{H}(\dot{\gamma})}{K_p} \left( \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{\eta} + \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{m} \right) \boldsymbol{D}^e \frac{\partial g}{\partial \boldsymbol{\sigma}} \tag{102}$$

$$\boldsymbol{G} := \frac{\mathcal{H}(\dot{\gamma})}{K_p} \boldsymbol{h} \otimes \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \boldsymbol{D}^e \right) \tag{103}$$

$$\boldsymbol{G}_\vartheta := \frac{\mathcal{H}(\dot{\gamma})}{K_p} \left( \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{\eta} + \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{m} \right) \boldsymbol{h} + \boldsymbol{\eta} \tag{104}$$

According to eq. (98), the plastic multiplier $\dot{\gamma}$ can be considered the sum of the following two terms:

$$\dot{\gamma}_m := \frac{1}{K_p} \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \qquad \dot{\gamma}_\vartheta := \frac{1}{K_p} \left( \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{\eta} + \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{m} \right) \dot{\vartheta} \tag{105}$$

The first, $\dot{\gamma}_m$, coincides with the plastic multiplier of classical elastoplasticity – see eq. (60) – while the second, $\dot{\gamma}_\vartheta$, accounts for the effect of non–mechanical hardening/softening processes. Note that, for plastic loading to occur, only the sum of $\dot{\gamma}_m$ and $\dot{\gamma}_\vartheta$ needs to be positive. In particular, plastic strains may occur even for a trial stress rate $\dot{\boldsymbol{\sigma}}^{\mathrm{tr}} := \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}}$ pointing *inwards* the current yield locus ($\dot{\gamma}_m < 0$), provided that the change in $\vartheta$ gives rise to a reduction in size of the elastic domain sufficiently large to keep the plastic multiplier positive, as, for example, in the case of chemical degradation.

Examples of application of this general framework to the modelling of mechanical and chemical degradation processes in weak rocks or bonded soils are provided in, *e.g.*, [NC01, TCN02, NCT03, CdP16, TC16].

# 10 Concluding remarks

In this chapter, the basic principles of the theory of plasticity have been presented, starting from the basic thermodynamic foundations of the theory of hyperplasticity and moving to classical (perfect, isotropic and anisotropic hardening) phenomenological plasticity, in which the main ingredients of the theory are selected ad–hoc, based on the available experimental evidence. Some of the most relevant extensions of the classical theory – such as Bounding Surface plasticity, generalized plasticity and plasticity with generalized hardening laws, developed to improve its predictive capabilities for complex loading conditions including cyclic loading and environmental loading – have also been discussed to provide an overview of the capabilities of advanced plasticity formulations as applied to particular geotechnical problems.

One important aspect of mathematical modeling of soil behavior which has been thoroughly discussed is the need to distinguish between the non–linearity of the stress–strain response *for finite stress or strain increments* and the concept of *incremental non–linearity*. While a non–linear soil model can be obtained with a simple hypoelastic constitutive equation, the modeling of irreversible and history–dependent behavior requires the constitutive equation to be formulated in rate–form and the use of *incrementally non–linear* relations between the stress and the strain rates.

The theory of plasticity represents the earliest and perhaps simplest approach to incremental non–linearity, achieved through the introduction of the loading/unloading conditions in the incremental response. Its appeal throughout the decades since its early applications to geotechnical problems stems from the ease with which some of its basic concepts (the elastic response, the yield surface, the plastic potential) could find a physical interpretation in the examination of classical laboratory test results.

While classical perfect plasticity is still widely used in the analysis of failure problems in geotechnical engineering, the more advanced versions of the theory have been mostly developed in the attempt of making more accurate numerical predictions in terms of performance of the geotechnical structures under complex loading conditions – *i.e.*, relevant displacement and deformations.

It is worth noting that, as the models increase in their predictive capabilities, they also necessarily require the introduction of more material constants as well as of a larger pool of history–dependent internal variables. This creates two different order of problems whenever the use of these advanced tools is required:

a) A large pool of experimental data, gathered from tests exploring different loading paths, is required to calibrate models with a large set of material constants.

b) In presence of one or more internal variables, some of which could be second–order tensors, the definition of their initial values at the beginning of the loading process is necessary, in the same way as the definition of the initial stress state is required in order to start the evolution process governed by the constitutive equations in rate–form.

As for point (a), a desirable feature of the model would be that the calibration does not require complex testing procedures to be performed with non–standard experimental devices (*e.g.*, true triaxial cell, hollow cylinder apparatus, simple shear devices). The calibration of a relatively large set of material constants has always been considered one of the main drawbacks of advanced plasticity models, and has motivated a number of studies aimed at devising calibration algorithms for the automatic identification of the model constants from a set of experimental data.

However, it is usually point (b) which poses the most challenging task. In fact, it is sufficient to consider how difficult could be to make a reasonable estimate of the coefficient of earth pressure at rest, $K_0$, for a heavily overconsolidated soil deposit, even in simple geometric conditions (horizontal ground surface, horizontal contacts between soil layers), to have an idea on how hard is to estimate the initial values of a structure tensor when no information is available on the details of the geological history of the site, or the ground surface is not horizontal and simple geostatic conditions do not apply. In some cases, the definition of the initial state in terms of stress and internal variables fields could require the simulation of the entire geological history of the deposit and could represent a significant part of the numerical modeling activities for the design of a geotechnical structure.

# References

[ABS77] D. Allirot, J. P. Boehler, and A. Sawczuk. Irreversible deformations of anisotropic rock under hydrostatic pressure. *Int. J. Rock Mech. Min. Sci. & Geomech. Abstr.*, 14:77–83, 1977.

[AD86] A. Anandarajah and Y. F. Dafalias. Bounding surface plasticity. III: Application to anisotropic cohesive soils. *J. Engng. Mech., ASCE*, 112(12):1292–1318, 1986.

[ARS86] J. H. Atkinson, D. Richardson, and S. E. Stallebrass. Effect of recent stress history on the stiffness of overconsolidated clay. *Géotechnique*, 40(4):531–540, 1986.

[ATW89] A. Al-Tabbaa and D. M. Wood. An experimentally based bubble model for clay. In S. Pietrusczczak and G. N. Pande, editors, *NUMOG III*, pages 91–99. Elsevier Applied Science, 1989.

[Bar86] J. P. Bardet. Modelling of sand behaviour with bounding surface plasticity. In G. N. Pande and W. F. van Impe, editors, *Numerical Models in Geomechanics*, pages 131–150. Jackson and Sons., 1986.

[BD84] R. Baker and C. S. Desai. Induced anisotropy during plastic straining. *Int. J. Num. Anal. Meth. Geomech.*, 8:167–185, 1984.

[Boe75] J. P. Boehler. Sur les formes invariantes dans les sous–groupe orthotrope de revolution. *Zeits. Angew. Math. Mech.*, 55:609–611, 1975.

[Boe87] J. P. Boehler. *Application of tensor functions in solid mechanics*. CISM Courses and Lectures n. 292. Springer Verlag, New York, 1987.

[Bor92] R. I. Borja. Generalized creep and stress–relaxation model for clays. *J. Geotech. Engng., ASCE*, 118(11):1765–1786, 1992.

[Bor13] R. I. Borja. *Plasticity: modeling & computation*. Springer Science & Business Media, 2013.

[BTA97] R. I. Borja, C. Tamagnini, and A. Amorosi. Coupling plasticity and energy-conserving elasticity models for clays. *J. Geotech. Geoenv. Engng., ASCE*, 123(10):948–957, 1997.

[BY86] P. K. Banerjee and N. B. Yousif. A plasticity model for the mechanical behavior of anisotropically consolidated clay. *Int. J. Num. Anal. Meth. Geomech.*, 10:521–541, 1986.

[CC44] A. Casagrande and N. Carillo. Shear failure of anisotropic materials. *Proc. Boston Soc. Civil Engrs.*, 31:74–87, 1944.

[CdP16] M. O. Ciantia and C. di Prisco. Extension of plasticity theory to debonding, grain dissolution, and chemical damage of calcarenites. *Int. J. Num. Anal. Meth. Geomech.*, 40(3):315–343, 2016.

[CH97] I.F. Collins and G. T. Houlsby. Application of thermomechanical principles to the modeling of geotechnical materials. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 453:1975–2000, 1997.

[CH02] I. F. Collins and T. Hilder. A theoretical framework for constructing elastic/plastic constitutive models of triaxial tests. *Int. J. Num. Anal. Meth. Geomech.*, 26(13):1313–1347, 2002.

[Che76] W. F. Chen. *Limit Analysis and Soil Plasticity*. Elsevier, 1976.

[CJR13] B. Cambou, M. Jean, and F. Radjaï. *Micromechanics of granular materials*. John Wiley & Sons, 2013.

[CK02] I. F. Collins and P. A. Kelly. A thermomechanical analysis of a family of soil models. *Géotechnique*, 52(7):507–518, 2002.

[CM03] I. F. Collins and B. Muhunthan. On the relationship between stress–dilatancy, anisotropy, and plastic dissipation for granular materials. *Geotechnique*, 53(7):611–618, 2003.

[Cw94] R. S. Crouch and J. P. wolf. Unified 3d critical state bounding–surface plasticity model for soils incorporating continuous plastic loading under cyclic paths. Part I: constitutive relations. *Int. J. Num. Anal. Meth. Geomech.*, 18:735–758, 1994.

[Daf86] Y. F. Dafalias. Bounding surface plasticity. I: Mathematical foundation and hypoplasticity. *J. Engng. Mech., ASCE*, 112(9):966–987, 1986.

[Dar78] F. Darve. *Une formulation incrémentale non–lineaire des lois rhéologiques; application aux sols*. Thèse d'Etat, Grenoble, 1978.

[Dar90] F. Darve. The expression of rheological laws in incremental form and the main classes of constitutive equations. In F. Darve, editor, *Geomaterials: Constitutive Equations and Modelling*, pages 123–148. Elsevier, 1990.

[DC70] J. M. Duncan and C. Y. Chang. Nonlinear analysis of stress and strain in soils. *J. Soil Mech. Found. Div., ASCE*, 96(SM5):1629–1653, 1970.

[DH86] Y. F. Dafalias and L. R. Herrmann. Bounding surface plasticity. II: Application to isotropic cohesive soils. *J. Engng. Mech., ASCE*, 112(12):1263–1291, 1986.

[DM04] Y. F. Dafalias and M. T. Manzari. Simple plasticity sand model accounting for fabric change effects. *J. Engng. Mech., ASCE*, 130(6):622–634, 2004.

[DMP06] Y. F. Dafalias, M. T. Manzari, and A. G. Papadimitriou. SANICLAY: simple anisotropic clay plasticity model. *Int. J. Num. Anal. Meth. Geomech.*, 30(12):1231–1257, 2006.

[DMS71] F. L. Di Maggio and I. S. Sandler. Material model for granular soil. *J. Engng. Mech. Div., ASCE*, 97:935–950, 1971.

[DP75] Y. F. Dafalias and E. V. Popov. A model of non–linearly hardening materials for complex loading. *Arch. Mech.*, 21:173–192, 1975.

[dPNL93] C. di Prisco, R. Nova, and J. Lanier. A mixed isotropic–kinematic hardening constitutive law for sand. In D. Kolymbas, editor, *Modern Approaches to Plasticity*, pages 83–124. Elsevier, 1993.

[DS66] J. M. Duncan and H. B. Seed. Anisotropy and stress reorientation of clays. *J. Soil Mech. Found. Div., ASCE*, 92(SM5):21–50, 1966.

[DS84] C. S. Desai and H. J. Siriwardane. *Constitutive Laws for Engineering Materials, with Emphasis on Geologic Materials*. Prentice–Hall, 1984.

[DS02] R. O. Davis and A. P. S. Selvadurai. *Plasticity and geomechanics*. Cambridge university press, 2002.

[DT05] A. DeSimone and C. Tamagnini. Stress–dilatancy based modelling of granular materials and extensions to soils with crushable grains. *Int. J. Num. Anal. Meth. Geomech.*, 29(1):73–101, 2005.

[DT13] Y. F. Dafalias and M. Taiebat. Anatomy of rotational hardening in clay plasticity. *Géotechnique*, 63(16):1406–1418, 2013.

[EHN07] I. Einav, G. T. Houlsby, and G. D. Nguyen. Coupled damage and plasticity models derived from energy and dissipation potentials. *Int. Journal of Solids and Structures*, 44(7-8):2487–2508, 2007.

[EP04] I. Einav and A. M. Puzrin. Continuous hyperplastic critical state (chcs) model: derivation. *Int. Journal of Solids and Structures*, 41(1):199–226, 2004.

[FN90] E. Flavigny and R. Nova. Viscous properties of geomaterials. In F. Darve, editor, *Geomaterials: Constitutive Equations and Modelling*, pages 27–54. Elsevier, 1990.

[GDV84] G. Gudehus, F. Darve, and I. Vardoulakis. *Constitutive Relations for Soils*. Balkema, Rotterdam, 1984.

[GH83] J. Graham and G. T. Houlsby. Anisotropic elasticity of a natural clay. *Géotechnique*, 33(2):165–180, 1983.

[GM82] J. Ghaboussi and H. Momen. Modelling and analysis of cyclic behavior of sands. In G. N. Pande and O. C. Zienkiewicz, editors, *Soil Mechanics: Transient and Cyclic Loads*, pages 313–342. Wiley, New York, 1982.

[GNL83] J. Graham, M. L. Noonan, and K. V. Lew. Yield states and stress–strain relationship in a natural plastic clay. *Can. Geotech. J.*, 20:502–516, 1983.

[GNS83] P. Germain, Q. S. Nguyen, and P. Suquet. Continuum thermodynamics. *J. Appl. Mechanics, ASME*, 50:1010–1020, 1983.

[Gre56] A. E. Green. Hypoelasticity and plasticity. *Archive for Rational Mechanics and Analysis*, 5:725–734, 1956.

[Gud79] G. Gudehus. A comparison of some constitutive laws for soils under radially symmetric loading and unloading. In Wittke, editor, *3*$^{\text{rd}}$ *Int. Conf. Num. Meth. Geomech.,* Aachen, pages 1309–1324. Balkema, Rotterdam, 1979.

[GW99] A. Gajo and D. M. Wood. A kinematic hardening constitutive model for sands: the multiaxial formulation. *Int. J. Num. Anal. Meth. Geomech.*, 23:925–965, 1999.

[Has79] K. Hashiguchi. Constitutive equations of granular media with an anisotropic hardening. In *III Int. Conf. Num. Meth. in Geomechanics*, pages 435–439, Aachen, Germany, 1979. Balkema, Rotterdam.

[Has85] K. Hashiguchi. Two– and three–surface models of plasticity. In *V Int. Conf. Num. Meth. in Geomechanics*, pages 285–292, Nagoya, Japan, 1985. Balkema, Rotterdam.

[Has88] K. Hashiguchi. Mathematically consistent formulation of elastoplastic constitutive equations. In *VI Int. Conf. Num. Meth. in Geomechanics*, pages 467–472, Innsbruck, Austria, 1988. Balkema, Rotterdam.

[Has17] K. Hashiguchi. *Foundations of elastoplasticity: subloading surface model*. Springer, 2017.

[Hil50] R. Hill. *The Mathematical Theory of Plasticity*. Oxford University Press, Oxford, 1950.

[HN75] B. Halphen and Q. S. Nguyen. Sur les matériaux standards généralisés. *Journal de Mécanique*, 14:39–63, 1975.

[Hou81] G. T. Houlsby. *A study of plasticity theories and their applicability to soils*. PhD thesis, Cambridge University, 1981.

[Hou85] G. T. Houlsby. The use of a variable shear modulus in elastic–plastic models for clays. *Comp. & Geotechnics*, 1:3–13, 1985.

[HP00] G. T. Houlsby and A. M. Puzrin. A thermomechanical framework for constitutive models for rate-independent dissipative materials. *Intl. J. of Plasticity*, 16(9):1017–1047, 2000.

[HP07] G. T. Houlsby and A. M. Puzrin. *Principles of hyperplasticity: an approach to plasticity theory based on thermodynamic principles*. Springer Science & Business Media, 2007.

[HR99] W. Han and B. D. Reddy. *Plasticity: Mathematical Theory and Numerical Analysis*. Springer Verlag, New York, 1999.

[Hue92] T. Hueckel. Water mineral interactions in hygromechanics of clay exposed to environmental load: a mixture theory approach. *Can. Geotech. J.*, 29:1071–1086, 1992.

[Hue97] T. Hueckel. Chemo–plasticity of clays subjected to stress and flow of a single contaminant. *Int. J. Num. Anal. Meth. Geomech.*, 21:43–72, 1997.

[HWW84] G. T. Houlsby, C. P. Wroth, and D. M. Wood. Prediction of the results of laboratory tests on a clay using a critical state model. In G. Gudehus, F. Darve, and I. Vardoulakis, editors, *Constitutive Relations for Soils*. Balkema, Rotterdam, 1984.

[Iwa67] W. D. Iwan. On a class of models for the yield behaviour of continuous and composite systems. *J. Appl. Mechanics, ASME*, 34:612–617, 1967.

[JB02] M. Jirasek and Z. P. Bazant. *Inelastic Analysis of Structures*. Wiley, Chichester, 2002.

[Jom00] C. Jommi. Remarks on the constitutive modelling of unsaturated soils. In *Experimental evidence and theoretical approaches in unsaturated soils*, pages 139–153. Balkema, Rotterdam, 2000.

[Jom21] C. Jommi. General overview on modelling the coupled behaviour of unsaturated soils. In C. Tamagnini and D. Mašín, editors, *Constitutive modelling of soils*. ALERT Geomaterials, 2021. This volume.

[JP88] R. J. Jardine and D. M. Potts. Hutton tension leg platform foundations: an approach to the prediction of driven pile behaviour. *Géotechnique*, 38(2):231–252, 1988.

[JPFB86] R. J. Jardine, D. M. Potts, A. B. Fourie, and J. B. Burland. Studies of the influence of non-linear stress-strain characteristics in soil-structure interaction. *Géotechnique*, 36(3):377–396, 1986.

[JPSJH91] R. J. Jardine, D. M. Potts, H. D. St. John, and D. W. Hight. Some practical applications of a non–linear ground model. In AGI, editor, *X ECSMFE,* Firenze, volume 1, pages 223–228. Balkema, Rotterdam, 1991.

[KL88] M. K. Kim and P. V. Lade. Single hardening constitutive model for frictional materials. I: Plastic potential function. *Comp. & Geotechnics*, 5:307–324, 1988.

[Kol91] D. Kolymbas. An outline of hypoplasticity. *Archive of Applied Mechanics*, 61:143–151, 1991.

[Kol93] D. Kolymbas. *Modern Approaches to Plasticity*. Elsevier, 1993.

[KZ63] R. L. Kondner and J. S. Zelasko. A hyperbolic stress–strain formulation for sands. In *II Pan.–Am. Conf. SMFE*, volume 1, pages 289–324, 1963.

[LC08] L. Laloui and C. Cekerevac. Non-isothermal plasticity model for cyclic behaviour of soils. *Int. J. Num. Anal. Meth. Geomech.*, 32(5):437–460, 2008.

[LK88] P. V. Lade and M. K. Kim. Single hardening constitutive model for frictional materials. II: Yield criterion and plastic work contours. *Comp. & Geotechnics*, 6:13–29, 1988.

[LS02] A. V. Lyamin and S. W. Sloan. Upper bound limit analysis using linear finite elements and nonlinear programming. *Int. J. Num. Anal. Meth. Geomech.*, 26:181–216, 2002.

[Lub90] J. Lubliner. *Plasticity Theory*. Mac Millan, London, 1990.

[LYKT02] H. I. Ling, D. Yue, V. N. Kaliakin, and N. J. Themelis. Anisotropic elastoplastic bounding surface model for cohesive soils. *J. Engng. Mech., ASCE*, 128(7):748–758, 2002.

[Mas21] D. Masin. Hypoplasticity and other incrementally non–linear modelling approaches. In C. Tamagnini and D. Mašín, editors, *Constitutive modelling of soils*. ALERT Geomaterials, 2021. This volume.

[Mau92] G. A. Maugin. *Thermomechanics of plasticity and fracture*. Cambridge University Press, 1992.

[MD97] M. T. Manzari and Y. F. Dafalias. A critical state two–surface plasticity model for sands. *Géotechnique*, 47(2):255–272, 1997.

[Mit70] R. J. Mitchell. On the yielding and mechanical strength of leda clays. *Can. Geotech. J.*, 7:297–312, 1970.

[MLA94] L. Modaressi, L. Laloui, and D. Aubry. Thermodynamical approach for camclay–family models with roscoe–type dilatancy rules. *Int. J. Num. Anal. Meth. Geomech.*, 18:133–138, 1994.

[MNZ78] Z. Mroz, V. A. Norris, and O. C. Zienkiewicz. An anisotropic hardening model for soils and its application to cyclic loading. *Int. J. Num. Anal. Meth. Geomech.*, 2:203–221, 1978.

[MNZ81] Z. Mroz, V. A. Norris, and O. C. Zienkiewicz. An anisotropical critical state model for soils subject to cyclic loading. *Géotechnique*, 31(4):451–469, 1981.

[Mor70] J. J. Moreau. Sur les lois de frottement, de viscosité et de plasticité. *C. R. Acad. Sci.*, 271:608–611, 1970.

[Mro67] Z. Mroz. On the description of anisotropic work–hardening. *Journal of the Mechanics and Physics of Solids*, 15:163–175, 1967.

[Mur85] S. Murayama. *Constitutive laws of soils*. Japanese Society of Soil Mechanics and Foundation Engineering, 1985.

[NC01] R. Nova and R. Castellanza. Modelling weathering effects on the mechanical on the mechanical behaviour of soft rocks. In *Int. Conf. on Civil Engineering*, pages 157–167, Bangalore, India, 2001. Interline Publishing.

[NCT03] R. Nova, R. Castellanza, and C. Tamagnini. A constitutive model for bonded geomaterials subject to mechanical and/or chemical degradation. *Int. J. Num. Anal. Meth. Geomech.*, 27(9):705–732, 2003.

[Nov77] R. Nova. On the hardening of soils. *Archiwum Mechaniki Stosowanej*, 29:445–458, 1977.

[Nov86] R. Nova. Soil models as a basis for modelling the behaviour of geophysical materials. *Arch. Mech.*, 64:31–44, 1986.

[Nov88] R. Nova. Sinfonietta classica: an exercise on classical soil modelling. In Saada and Bianchini, editors, *Constitutive Equations for Granular Non–Cohesive Soils*, Cleveland, 1988. Balkema, Rotterdam.

[Nov96] R. Nova. Modelling: classical elastoplastic models. In R. Chambon, editor, *8th ALERT School on Bifurcation and Localization in Geomaterials*. ALERT Geomaterials, 1996.

[Nov00] R. Nova. Modelling the weathering effects on the mechanical behaviour of granite. In D. Kolymbas, editor, *Constitutive Modelling of Granular Materials*, Horton, Greece, 2000. Springer, Berlin.

[NW79] R. Nova and D. M. Wood. A constitutive model for sand in triaxial compression. *Int. J. Num. Anal. Meth. Geomech.*, 3:255–278, 1979.

[Ogd97] R. W. Ogden. *Non–linear elastic deformations*. Dover, 1997.

[OKH78] M. Oda, I. Koishikawa, and T. Higuchi. Experimental study of anisotropic shear strength of sand by plane strain test. *Soils and Foundations*, 18(1):25–38, 1978.

[OKKA02] F. Oka, S. Kimoto, H. Kobayashi, and T. Adachi. Anisotropic behavior of soft sedimentary rock and a constitutive model. *Soils and foundations*, 42(5):59–70, 2002.

[ONNK85] M. Oda, S. Nemat-Nasser, and J. Konishi. Stress–induced anisotropy in granular masses. *Soils and Foundations*, 25(3):85–97, 1985.

[OT20] K. Oliynyk and C. Tamagnini. Finite deformation hyperplasticity theory for crushable, cemented granular materials. *Open Geomechanics*, 2:1–33, 2020.

[OT21] K. Oliynyk and C. Tamagnini. Finite deformation plasticity. In C. Tamagnini and D. Mašín, editors, *Constitutive modelling of soils*. ALERT Geomaterials, 2021. This volume.

[OW69] D. R. Owen and W. O. Williams. On the time derivatives of equilibrated response functions. *Archive for Rational Mechanics and Analysis*, 33(4):288–306, 1969.

[Pas91] M. Pastor. Modelling of anisotropic sand behaviour. *Comp. & Geotechnics*, 11:173–208, 1991.

[PBZL97] I. D. S. Pontes, L. A. Borges, N. Zouain, and F. R. Lopes. An approach to limit analysis with cone–shaped yield surfaces. *Int. J. Num. Meth. Engng.*, 40:4011–4032, 1997.

[PH01] A. M. Puzrin and G. T. Houlsby. Fundamentals of kinematic hardening hyperplasticity. *International journal of solids and structures*, 38(21):3771–3794, 2001.

[PHS66] H. B. Pooroshasb, I. Holubec, and A. N. Sherbourne. yielding and flow of sand in triaxial compression (part I). *Can. Geotech. J.*, 3:179–190, 1966.

[PHS67] H. B. Pooroshasb, I. Holubec, and A. N. Sherbourne. yielding and flow of sand in triaxial compression (part II). *Can. Geotech. J.*, 4:376–397, 1967.

[PLS02] S. Pietruszczak, D. Lydzba, and J.-F. Shao. Modelling of inherent anisotropy in sedimentary rocks. *Int. Journal of Solids and Structures*, 39(3):637–648, 2002.

[PM00] S. Pietruszczak and Z. Mroz. Formulation of anisotropic failure criteria incorporating a microstructure tensor. *Computers and Geotechnics*, 26(2):105–112, 2000.

[PP85] H. B. Pooroshasb and S. Pietruszczak. On the yielding and flow of sand: a generalized two–surface model. *Comp. & Geotechnics*, 1:33–58, 1985.

[Pre77] J. H. Prevost. Mathematical modelling of monotonic and cyclic undrained clay behaviour. *Int. J. Num. Anal. Meth. Geomech.*, 1:195–216, 1977.

[Pre86] J. H. Prevost. Constitutive equations for pressure–sensitive soils: theory, numerical implementation, and examples. In R. Dungar and J. A. Studer, editors, *Geomechanical Modelling in Engineering Practice*, pages 331–350. Balkema, Rotterdam, 1986.

[PW99] A. Pestana and A. J. Whittle. Formulation of a unified constitutive model for clays and sands. *Int. J. Num. Anal. Meth. Geomech.*, 23:1215–1243, 1999.

[PZC90] M. Pastor, O. C. Zienkiewicz, and A. H. C. Chan. Generalized plasticity and the modelling of soil behaviour. *Int. J. Num. Anal. Meth. Geomech.*, 14:151–190, 1990.

[PZL85] M. Pastor, O. C. Zienkiewicz, and K. H. Leung. Simple model for transient soil loading in earthquake analysis. II: non–associative model for sands. *Int. J. Num. Anal. Meth. Geomech.*, 9:477–498, 1985.

[RB68] K. H. Roscoe and J. B. Burland. On the generalised stress–strain behaviour of 'wet' clay. In J. Heyman and F. A. Leckie, editors, *Engineering Plasticity*, pages 535–609. Cambridge Univ. Press, Cambridge, 1968.

[RK83] R. K. Rowe and G. J. Kack. A theoretical examination of the settlements induced by tunnelling: four case histories. *Can. Geotech. J.*, 20:299–314, 1983.

[RM93] B.D. Reddy and J.B. Martin. Internal variable formulations of problems in elastoplasticity: constitutive and algorithmic aspects. *Applied Mech. Review*, 47:429–456, 1993.

[RW00] M. Rouainia and D. M. Wood. A kinematic hardening constitutive model for natural clays with loss of structure. *Géotechnique*, 50(2):153–164, 2000.

[SB89] A. S. Saada and G. F. Bianchini. *Constitutive equations for granular non–cohesive soils*. Balkema, Rotterdam, 1989.

[SD83] H. J. Siriwardane and C. S. Desai. Computational procedures for nonlinear three–dimensional analysis with some advanced constitutive laws. *Int. J. Num. Anal. Meth. Geomech.*, 7:143–171, 1983.

[SDMB76] I. S. Sandler, F. L. Di Maggio, and G. Y. Baladi. Generalised cap model for geologic materials. *J. Soil Mech. Found. Div., ASCE*, 102:683–697, 1976.

[SGvW95] I. Sharour, S. Ghorbanbeigi, and P. A. von Wolffersdordff. Three–dimensional finite element analysis of diaphragm wall construction. *Revue Francaise de Géotechnique*, 71:39–47, 1995.

[SH92] I. M. Smith and D. K. H. Ho. Influence of construction technique on the performance of a braced excavation in marine clay. *Int. J. Num. Anal. Meth. Geomech.*, 16:845–867, 1992.

[SH97] J. C. Simo and T. J. R. Hughes. *Computational Inelasticity*. Springer, 1997.

[SJH92] P. R. Smith, R. J. Jardine, and D. W. Hight. The yielding of Bothkennar clay. *Géotechnique*, 42(2):257–274, 1992.

[SK95] S. W. Sloan and P. W. Kleeman. Upper bound limit analysis with discontinuous velocity fields. *Comp. Meth. Appl. Mech. Engng.*, 127:293–314, 1995.

[SO77] H. Sekiguchi and H. Ohta. Induced anisotropy and time dependency in clays. In *IX ICSMFE, Specialty Session 9*, pages 229–238. Balkema, Rotterdam, 1977.

[Sok65] V. V. Sokolowski. *Statics of Granular Media*. Pergamon, Oxford, 1965.

[ST97] S. E. Stallebrass and R. N. Taylor. The development and evaluation of a constitutive model for the prediction of ground movements in overconsolidated clay. *Géotechnique*, 47(2):235–253, 1997.

[Sta90] S. E. Stallebrass. *Modelling the effect of recent stress history on the behaviour of overconsolidated soils*. PhD thesis, The City University, London, 1990.

[SW68] A. N. Schofield and C. P. Wroth. *Critical State Soil Mechanics*. McGraw–Hill, London, 1968.

[SWB16] S. J. Semnani, J. A. White, and R. I. Borja. Thermoplasticity and strain localization in transversely isotropic materials based on anisotropic critical state plasticity. *Int. J. Num. Anal. Meth. Geomech.*, 40(18):2423–2449, 2016.

[TC16] C. Tamagnini and M. O. Ciantia. Plasticity with generalized hardening: constitutive modeling and computational aspects. *Acta Geotechnica*, 11(3):595–623, 2016.

[TCN02] C. Tamagnini, R. Castellanza, and R. Nova. A generalized backward euler algorithm for the numerical integration of an isotropic hardening elastoplastic model for mechanical and chemical degradation of bonded geomaterials. *Int. J. Num. Anal. Meth. Geomech.*, page submitted for publication., 2002.

[TD08] M. Taiebat and Y. F. Dafalias. SANISAND: Simple anisotropic sand plasticity model. *Int. J. Num. Anal. Meth. Geomech.*, 32(8):915–948, 2008.

[TDP10] M. Taiebat, Y. F. Dafalias, and R. Peek. A destructuration theory and its application to SANICLAY model. *Int. J. Num. Anal. Meth. Geomech.*, 34(10):1009–1040, 2010.

[Ter48] K. Terzaghi. *Theoretical Soil Mechanics*. John Wiley, New York, 1948.

[TL77] F. Tavenas and S. Leroueil. Effects of stresses and time on on yielding of clays. In *IX ICSMFE*, volume 1, pages 319–326. Balkema, Rotterdam, 1977.

[TN65] C. A. Truesdell and W. Noll. The non–linear field theories of mechanics. In S. Flügge, editor, *Encyclopedia of Physics*, volume III/3. Springer, Berlin, 1965.

[TP48] K. Terzaghi and R. B. Peck. *Soil Mechanics in Engineering Practice*. John Wiley, New York, 1948.

[TVC00] C. Tamagnini, G. Viggiani, and R. Chambon. A review of two different approaches to hypoplasticity. In D. Kolymbas, editor, *Constitutive Modelling of Granular Materials*, pages 107–145. Springer, Berlin, 2000.

[UC96] F. J. Ulm and O. Coussy. Strength growth as chemo–plastic hardening in early age concrete. *J. Engng. Mech., ASCE*, 122(12):1123–1132, 1996.

[Var19] I. Vardoulakis. *Cosserat continuum mechanics*. Springer, 2019.

[VS95] I. Vardoulakis and J. Sulem. *Bifurcation Analysis in Geomechanics*. Blackie Acad. & Professional, New York, 1995.

[WA85] R. K. S. Wong and J. R. F. Arthur. Induced and inherent anisotropy in sand. *Géotechnique*, 35(4):471–481, 1985.

[Wil77] P. Wilde. Two invariants depending models of granular media. *Arch. Mech. Stos.*, 29:799–809, 1977.

[WK91] R. C. K. Wong and P. K. Kaiser. Performance assessment of tunnels in cohesionless soils. *J. Geotech. Engng., ASCE*, 117(12):1880–1901, 1991.

[WK94] A. J. Whittle and M. J. Kavvadas. Formulation of MIT–E3 constitutive model for overconsolidated clays. *J. Geotech. Engng., ASCE*, 120(1):173–198, 1994.

[WNKL03] S. J. Wheeler, A. Näätänen, M. Karstunen, and M. Lojander. An anisotropic elastoplastic model for soft clays. *Can. Geotech. J.*, 40(2):403–418, 2003.

[Woo82] D. M. Wood. Laboratory investigations of the behaviour of soils under cyclic loading: a review. In G. N. Pande and O. C. Zienkiewicz, editors, *Soil Mechanics – Cyclic and Transient Loads*, pages 513–582. Wiley, Chichester, 1982.

[Woo04] D. M. Wood. *Geotechnical modelling*, volume 1. Spon press, 2004.

[YIV98] M. Yoshimine, K. Ishihara, and W. Vargas. Effects of principal stress direction and intermediate principal stress on undrained shear behavior of sand. *Soils and Foundations*, 38(3):179–188, 1998.

[YMH91] N. Yasufuku, H. Murata, and M. Hyodo. Yield characteristics of anisotropically consolidated sand under low and high stresses. *Soils and Foundations*, 31(1):95–109, 1991.

[Yu06] H.-S. Yu. *Plasticity and geotechnics*, volume 13. Springer Science & Business Media, 2006.

[ZH77] O. C. Zienkiewicz and C. Humpheson. Viscoplasticity: a generalized model for soil behaviour. In C. A. Desai and J. T. Christian, editors, *Numerical Models in Geotechnical Engineering*. McGraw–Hill, New York, 1977.

[ZHL75] O. C. Zienkiewicz, C. Humpheson, and R. W. Lewis. Associated and non–associated visco–plasticity and plasticity in soil mechanics. *Géotechnique*, 25(4):671–689, 1975.

[Zie83] H. Ziegler. *An introduction to thermomechanics*. North Holland, 1983.

[ZLP85] O. C. Zienkiewicz, K. H. Leung, and M. Pastor. Simple model for transient soil loading in earthquake analysis. I: basic model and its application. *Int. J. Num. Anal. Meth. Geomech.*, 9:453–476, 1985.

[ZM84] O. C. Zienkiewicz and Z. Mroz. Generalized plasticity formulation and applications to geomechanics. In C. S. Desai and R. H. Gallagher, editors, *Mechanics of Engineering Materials*. Wiley, 1984.

[ZN71] O. C. Zienkiewicz and D. J. Naylor. An adaptation of critical state soil mechanics theory for use in finite elements. In R. H. G. Parry, editor, *Stress–Strain Behaviour of Soils (Roscoe Mem. Symp.)*. Foulis, Henley–on–Thames, 1971.

[ZN73] O. C. Zienkiewicz and D. J. Naylor. Finite element studies of soils and porous media. In J. T. Oden and E. R. de Arantes, editors, *Lect. Finite Elements in Continuum Mechanics*. UAH Press, 1973.

[ZRNW78] M. Zytynski, M. F. Randolph, R. Nova, and C. P. Wroth. On modelling the unloading-reloading behaviour of soils. *Int. J. Num. Anal. Meth. Geomech.*, 2(1):87–93, 1978.

[ZW87] H. Ziegler and C. Wehrli. The derivation of constitutive relations from the free energy and the dissipation function. *Advances in applied mechanics*, page 183, 1987.

# Modelling the influence of time on the mechanical behaviour of geomaterials

**Claudio di Prisco*, Luca Flessati**, Matteo Zerbi***

**Politecnico di Milano*
*** Technical University of Delft*

*In literature, the time-dependent mechanical behaviour of geomaterials is well established and explained in the light of various phenomena, ranging from purely mechanical processes, evolving with time, to those resulting from thermo/hydro/chemo/mechanical coupling. This chapter provides a concise overview of the most prominent constitutive modelling approaches, developed in recent decades, incorporating time variable and based on rate-dependent elastic-plasticity.*

## 1 Introduction

Numerous experimental test results show that the behaviour of geomaterials (both soils and rocks) is influenced by (i) the loading rate and (ii) the duration of the perturbation applied. Related to the time dependence of the mechanical behaviour of geomaterials is a wide variety of phenomena, affecting natural events and the performance of geo-structures. Among the others, well known are:

- the creeping landslides, gravitational movements, evolving with time and subject to transient periods of acceleration and deceleration;
- the aging in clayey materials, associated with the progressive accumulation of strains, not induced by any apparent perturbation of the effective stress state, causing an evolution of the material microstructure and its over-consolidation;
- squeezing processes in deep tunnels excavated in rocks, causing a progressive increase in the stresses acting in permanent and provisional liming;
- solid-fluid and viceversa regime transitions, taking place in granular materials either flowing or depositing, when fast landslides occur.

According to both geo-material microstructure and type of perturbation applied, the time dependency of the mechanical behaviour observed at the macro-scale (the one corresponding to the Representative Elementary Volume – REV) is the result of many very different micro-mechanical processes, ranging from micro-inertial effects in granular media to thermo-hydro-chemo-mechanical processes in structured soils and rocks.

A very synthetic measure of the mechanical behaviour time dependency is the so-called characteristic time, a sort of parameter describing for each material its slowness in reacting to mechanical perturbations. When the reaction is instantaneous, as is assumed in case, for instance, in elastic-plastic constitutive models, the characteristic time is nil. The characteristic time concept is very similar to the one of characteristic length, employed to describe localisation processes: when characteristic length nullifies the mechanical response is local and the constitutive relationship is independent of the mechanical response of the points located in the neighbourhood.

To account for rate/time dependency, popular is the use of viscous constitutive relationships and nowadays the most common is the approach based on delayed plasticity (or viscoplasticity), theory originally conceived by Perzyna [Per63]. This approach has been successfully employed, in case of granular media, to account for (i) “micro-inertial effects”, associated with microstructural fabric rearrangement due to mechanical perturbations (§2), (ii) progressive failure in structured/bonded materials, caused by the time-dependent propagation of micro-cracks in either, according to the cases, grains or in intergranular bonds (§4). The widespread popularity of elastic-visco-plasticity of Perzyna type is mainly due to its simplicity for numerical implementation, since it is a straightforward extension of standard elastic-plasticity. Under suitable simplifying assumptions, for very slow perturbations (or for sufficiently small material characteristic times), the viscoplastic solution converges to the corresponding rate-independent elastic-plastic one, and the same can be said for the conditions of mechanical instability (§3).

In contrast, when the time-dependent material response is due to coupled thermo-hydro-chemo-mechanical processes (§5), the material characteristic time is governed by the coupled processes occurring at the microscale and the approach employed is not of Perzyna type, since the micro-structural temporal evolution is governed by additional variables.

Strain rate dependency is also observed in granular materials subject to very large strain rates or when long-lasting force chains collapse and energy in the medium is mainly dissipated by collisions (§6**).** When collisions become the predominant mechanism between particles at the microscale, with respect to force chains, granular materials do not behave like solids, but like highly compressible fluids. In this case, time-dependency is ruled by both current void ratio and granular temperature, a measure of the material agitation [Gol08].

## 2 Micro-inertial effects in granular media

When granular materials, under either dry or saturated conditions, are subject to a rapid perturbation in stresses (as it is, for instance, in creep tests), the material response can be delayed, in particular if the perturbation applied is sufficiently severe to cause an irreversible microstructural fabric rearrangements. This delay results in macroscopic irreversible strains, progressively increasing with time. According to [dPI96], this time dependency is the result of numerous local grain movements, associated with force chain collapses and reconstructions. It is about of a probabilistic evolution of the microstructure from the initial stable configuration to the final one, passing through numerous intermediate fabric configurations: the evolution rate is ruled by particle micro-inertia ([SV95]), since the fabric evolution at the micro-scale is associated with micro-dynamic phenomena with some grains accelerating and others decelerating.

To predict at the macro-scale the material behaviour, the delayed plasticity theory proposed in [P63] can be suitably employed: the strain rate tensor is decomposed into an instantaneous/reversible contribution and a delayed/irreversible one ($\dot{\varepsilon}_{ij}^{irr}$), calculated as it follows:

$$\dot{\varepsilon}_{ij}^{irr} = \gamma \Phi \frac{\partial g}{\partial \sigma'_{ij}}\,, \tag{1}$$

being $g$ the plastic potential (usually chosen to obtain gradient $\frac{\partial g}{\partial \sigma'_{ij}}$ to be non-dimensional), $\sigma'_{ij}$ the effective stress tensor, $\gamma$ [ $s^{-1}$] the fluidity parameter (positive by definition) and $\Phi$ the viscous nucleus, substituting, together with $\gamma$ the plastic multiplier employed in standard elastic-plasticity. $\Phi$ is defined by some authors as the distance (measured through a suitably conceived mapping rule) between the current yield locus $f = 0$ and the current state of stress, but, in most of the cases, $\Phi$ is simply a non-dimensional, non-negative and increasing function of $f$ (suitably expressed in a non-dimensional form).

Equation (1) allows us to introduce time dependency, since irreversible strains may develop even if the effective stress is not varied, as it is, for instance, in creep tests in the period of time in between two subsequent load perturbations.

In elastic-viscoplasticity, consistency rule, employed in standard elastic-plasticity to calculate the plastic multiplier, is abolished and the current effective stress image point may be either inside or outside the yield locus, that is:

$$f = f\left(\sigma'_{ij}, \alpha_{ij}\right) \lesseqgtr 0. \tag{2}$$

where $\alpha_{ij}$ stands for hardening variables.

In case $\Phi = \Phi\,(f)$, $f$ may be interpreted as a scalar measure of the probability of occurrence of fabric rearrangements and, consequently, of irreversible strain development. In most of the cases, $\Phi$ is assumed to be nil for $f$<0 and this implies the elastic domain existence. Nevertheless, in general, irreversible strains may be assumed to develop even for $f$<0 ([dPSZ07]).

$\gamma$, generally identified with $t_c^{-1}$ (being $t_c$ the material characteristic time) governs, together with $\Phi$ and $f$, the material rate dependence: when $\gamma \to \infty$, irreversible strain rate becomes infinite and the material response becomes instantaneous. Moreover if $\Phi = 0$ for $< 0$ , when $\gamma \to \infty$ standard elastic-plasticity is recovered. In standard elastic-viscoplasticity, hardening rules are defined as in standard elastic-plasticity. Nevertheless, since irreversible strains develop with time and are delayed, even the evolution of hardening variables take place with time: in case of creep tests, this implies that, in between two successive stress increments, hardening variables evolve and yield locus evolves.

The choice of $\Phi(\mathrm{f})$ is fundamental in determining the temporal material mechanical response [LSdPP19]. The simplest choice is to assume $\Phi(\mathrm{f})$ linearly dependent on $f$. In this case, if f is also chosen to be linearly dependent on $\sigma'_{ij}$, visco-plastic approach reduces to the standard Maxwell viscous model [Max67].

Non-linear expressions for $\Phi(\mathrm{f})$ have been also proposed. The most common is bilinear:

$$\Phi(\mathrm{f}) = \langle f^{\tilde{\alpha}} \rangle \tag{3}$$

with the non-dimensional parameter $\tilde{\alpha}$ equal to unity. In case $\tilde{\alpha} > 1$, nonlinearity is lost even for $f > 0$. In equation (3), Macaulay brackets ensure that for negative $f$ values, $\Phi = 0$. In this case, yield locus coincides with the elastic domain boundary.

An alternative non-linear definition for $\Phi(\mathrm{f})$ is

$$\Phi(\mathrm{f}) = \mathrm{e}^{\tilde{\alpha} f}, \tag{4}$$

which implies the viscous nucleus to be always positive. This implies that visco-plastic strains can develop even for $f < 0$. This allows us to more suitably simulate ageing phenomena, ruled by the shape of the negative branch of $\Phi(\mathrm{f})$, associated with very small strain rates.

The exponential expression defined in equation (4), on the other side, suggests a very rapid response of the material when large strain rates are taken into account, corresponding to large $f$ values and occurring when dynamic perturbations are applied. To capture the time dependency of the mechanical behaviour of granular materials in case of both very small and very large strain rates, alternative expressions for $\Phi(\mathrm{f})$ have been proposed, like the one reported here below [dPI03]:

$$\Phi(\mathrm{f}) = \begin{cases} \mathrm{e}^{\tilde{\alpha} f} & if\ f \leq f_0 \\ \beta \sqrt{\log(\zeta f)} & if\ f > f_0 \end{cases} \tag{5}$$

According to which the exponential function is substituted for $f \geq f_0$ (where $f_0$ is a constitutive parameter) by a logarithmic dependence ruled by two additional non-dimensional parameters ( $\beta$, $\zeta$ ). Obviously, the choice of the expression for $f \geq f_0$ is each time tailored on the expression employed for $f$, that in the specific case here mentioned was inspired to the one originally introduced in Cam-Clay model [SW68].

Under dynamic impulsive perturbations, like those illustrated in Figure 1, characterized by the same maximum value (it is about a biaxial test at constant $\sigma_{xx}^{\square}$ (Figure 1a)), the response of the specimen is totally different according to the strain rate imposed (Figure 1b): for very large strain rates, the response becomes more rigid and reversible, whereas the opposite occurs when the impulse increases (Figure 1c).

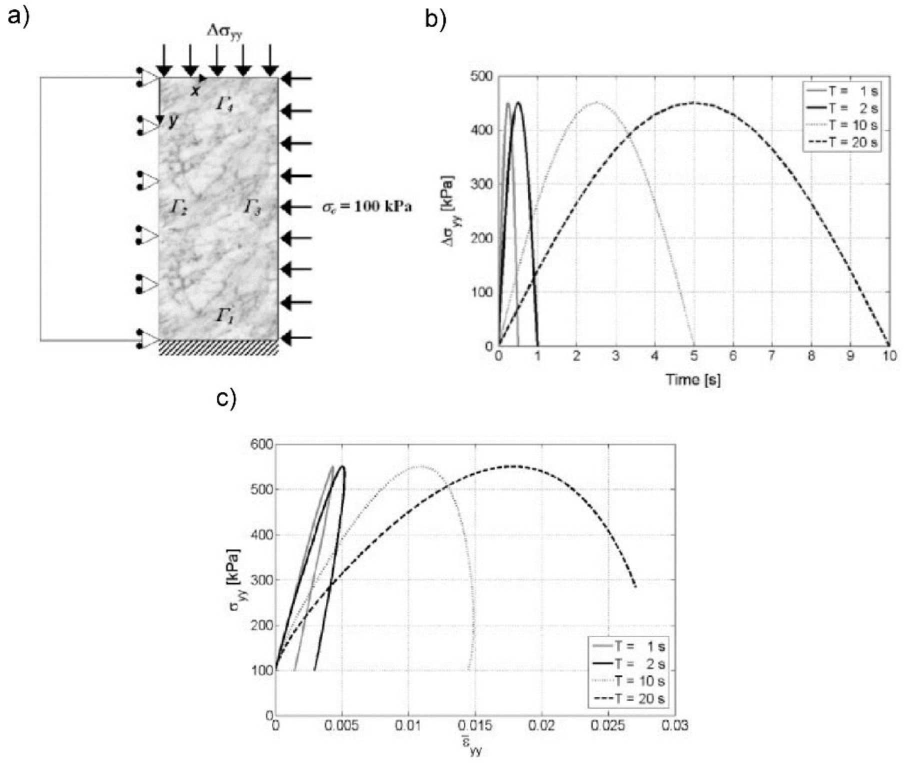


Figure 1: Numerical biaxial tests performed by [dPSZ07]: a) scheme of the tests, (b) different loading time histories and (c) numerical stress–strain curves obtained for four different loading time periods (from [dPSZ07]).

Under cyclic reverse tests, this implies that elastic-viscoplastic models are capable of capture the dependence of damping parameter on frequency, at least for sufficiently small values of frequencies (Figure 2). Indeed, if an elastic-plastic constitutive relationship is adopted, damping ratio depends on the perturbation amplitude, but not its frequency (Hysteretic damping), whereas if an elastic-viscoplastic constitutive relationship is adopted the dependency on the perturbation amplitude is still well captured and damping progressively reduces with frequency. This is not confirmed by the experimental test results, that conversely are characterized by non-monotonic dependence ([Shi95]) on frequency (Figure 2).

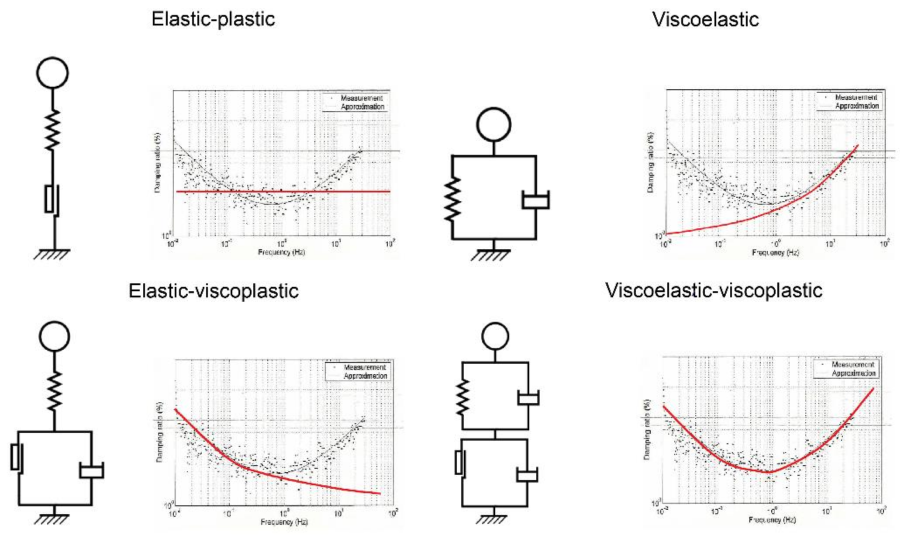


Figure 2: dependence of damping ratio on load frequency: comparison between experimental measurement obtained by [Men03] (interpolated by black lines) and different constitutive model predictions (red lines)

This trend is the result of two antagonistic effects: (i) the former one associated with the previously mentioned delayed plasticity (dominant at low frequencies), (ii) the latter to the linear viscosity (dominant at higher frequencies) not associated with both an evolution of the micro-structure and any accumulation of irreversible strains. Linear viscosity is the result at macro-scale of the local agitation of grains dissipating energy at contacts for sliding, but not implying any force chain collapse or grain dislocation: under dynamic conditions the material is capable of dissipating energy even without getting rise to any yielding.

From a numerical point of view, elastic-viscoplastic models have been successfully employed in the literature to both regularize numerical solutions, in particular under dynamic conditions, and to implement non-local approaches for the simulation of localization processes in many different types of geomaterials [LP91, WSdB98, LSS15]. In this last case, the thickness of the localized zone has been demonstrated to be a function not only of the material characteristic length but also of fluidity parameter and rate of the perturbation applied.

## 3 Stability analysis for elastic-viscoplastic materials

As is well-known, when geomaterials are tested under constant effective loads (for instance, when standard triaxial creep tests are performed), strain rate evolves with time. After the load increment, if either the applied load is sufficiently large or the current stress level (Figure 3) sufficiently high, an initial strain deceleration (primary creep) is followed by a constant strain rate branch (secondary creep) and by a subsequent severe strain acceleration (tertiary creep).

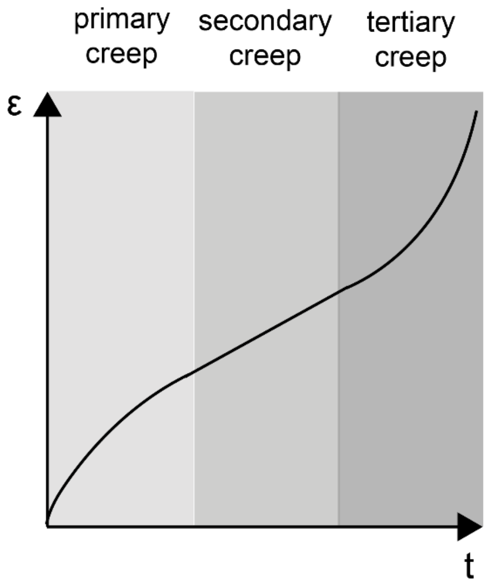


Figure 3: Transition from primary to tertiary creep.

As was already mentioned, incremental constitutive relationships are not suitable for describing the temporal accumulation of strains taking place at constant effective stress and, thus, not even for theoretically justifying the onset of instability, instability that in case of incremental constitutive relationships is discussed in the framework of either bifurcation ([SV95]) or controllability ([N94]) theories.

In case of creep tests, the hardening (Figure 4) of the yield function (represented in the triaxial $q - p'$ plane, where $q = \sigma'_a - \sigma'_r$ and $p' = (\sigma'_a + 2\,\sigma'_r)/3$ being $\sigma'_a$ and $\sigma'_r$ the effective axial and radial stress, respectively) is associated with a reduction in strain rates (primary creep), whereas its softening with strain acceleration (tertiary creep). To discuss material instability, in case of elastic-viscoplastic constitutive relationships, is, thus, very convenient expressing the constitutive relationship in terms of acceleration, that in case of creep tests, since $\dot{\sigma}'_{ij} = 0$, can be written as follows:

$$\ddot{\varepsilon}_{ij} = \ddot{\varepsilon}_{ij}^{irr} = \gamma\dot{\Phi}\frac{\partial g}{\partial \sigma'_{ij}} + \gamma\Phi\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial \sigma'_{ij}}\right) \tag{6}$$

where:

$$\dot{\Phi} = \frac{\partial\Phi}{\partial f}\frac{\partial f}{\partial t} = \frac{\partial\Phi}{\partial \mathrm{f}}\left(\frac{\partial f}{\partial\sigma'_{ij}}\dot{\sigma}'_{ij} - H\,\Phi(f)\right) = -\frac{\partial\Phi}{\partial \mathrm{f}}H\,\Phi(f) \tag{7}$$

being $H$ the hardening modulus, defined as follows.

$$H = -\frac{\partial f}{\partial\alpha_{ij}^{\square}}\frac{\partial\alpha_{ij}^{\square}}{\partial\varepsilon_{rs}^{irr}}\frac{\partial g}{\sigma'_{rs}} \tag{8}$$

Since by definition both $\Phi(f) \geq 0$ and $\frac{\partial\Phi}{\partial \mathrm{f}} \geq 0$, during creep tests, the sign of $\dot{\Phi}$ is governed by $H$. As a consequence, if $\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial\sigma'_{ij}}\right) = 0$ (condition satisfied by a numerous category of elastic-viscoplastic constitutive relationships), a deceleration in the accumulation of irreversible strains is observed (stable response, primary creep, Figure 3), when $H > 0$ and an acceleration, when $H < 0$. (tertiary creep, Figure 3). When state variable evolution stops ($H = 0$), the velocity of accumulation of irreversible strains is constant with time (secondary creep). If a change in sign of $H$, from positive to negative, is observed, a transition from primary to tertiary creep is also observed.

By substituting equations (7) and (1) into equation (6) and by assuming $\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial\sigma'_{ij}}\right) = 0$, we obtain:

$$\ddot{\varepsilon}_{ij}^{\square} = \gamma\dot{\Phi}\frac{\partial g}{\partial\sigma'_{ij}} =-\frac{\partial\Phi}{\partial f}H\gamma\,\Phi(f)\frac{\partial g}{\partial\sigma'_{ij}} = -\frac{\partial\Phi}{\partial f}H\,\dot{\varepsilon}_{ij}^{\square}, \tag{9}$$

that, in case $\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial\sigma'_{ij}}\right) = 0$ is not satisfied, has to be modified as in [PdP16].

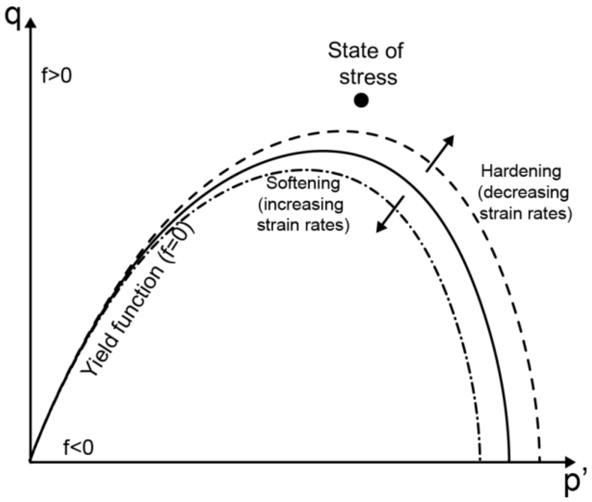

q
f>0
State of stress
Hardening (decreasing strain rates)
Softening (increasing strain rates)
Yield function (f=0)
f<0
p'

Figure 4: Evolution of yield function after an instantaneous stress increment: schematic representation in the triaxial plane

In a general case, that is for a general type of control (generalized creep tests), the onset of instability can be discussed by employing the approach proposed in [PdP16], combining Lyapunov theory of stability [L92] with controllability theory [N94], that is an extension to any type of test control of the method simply described here above. According to this approach, under quasi-static (inertia contributions are neglected) mixed stress-strain control conditions, once derived over time, the constitutive relationship can be written as it follows:

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{F} \tag{10}$$

being **X** a vector containing the rate of response variables ($\mathbf{X} = \frac{\mathrm{d}^{\square}}{dt^{\square}}\begin{bmatrix}\varepsilon_\alpha \\ \sigma_\beta\end{bmatrix}$, changing according to the test control), $\mathbf{F} = \begin{bmatrix}F_\alpha \\ F_\beta\end{bmatrix}$ a forcing term related to controlled variables (($\sigma_\alpha$) and ($\varepsilon_\beta$) ) and their first and second time derivatives, whereas matrix **A** depends on both constitutive relationship and controlled variable rates.

In case of "generalized creep tests" (i.e. when both rate and acceleration of controlled variables are nil), **F=0** and **A** only depends on the constitutive relationship. The eigenvalues of **A** can be employed to define the system stability: a stable response is obtained when all the eigenvalues of **A** are negative.

In contrast, instability takes place, when at least one eigenvalue becomes non-negative, corresponding with condition $H \leq H_\chi$, being $H_\chi$ the controllability modulus ([BDdP11]), already defined, by using standard controllability theory in ase of standard elastic-plastic constitutive relationships ([BDdP11]).This implies that in case of elastic-viscoplastic constitutive relationships even instability may be delayed with time but from a mechanical point of view its occurrence is governed by the same condition already found for elastic-plastic models.

To clarify what inferred here above, in [dPP17] this theory was applied to interpret the onset of instability in infinitely long slopes, that in case of Simple Shear (SS) conditions. Under SS conditions, both normal and shear stresses are controlled, whereas, along direction 2 of Figure 5 and out- of plane, normal strains are imposed to be nil.

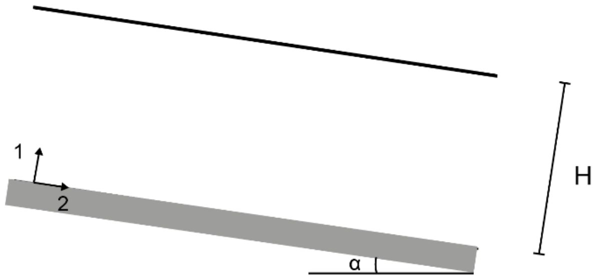


Figure 5: Infinitely long slope, representative of simple shear conditions

Under SS conditions, in a general form, that is for $\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial \sigma'_{ij}}\right) \neq 0$:

$$A_{\alpha\alpha} = -\frac{\delta\Phi}{\delta f}(H - H_L)I_{\alpha\alpha} \tag{11}$$

$$A_{\beta\beta} = -\frac{\delta\Phi}{\delta f}(H - H_L)I_{\beta\beta} - \Phi D_{22}\frac{\delta^2 g}{\delta^2\sigma_{\beta\beta}} \tag{12}$$

$$A_{\beta\alpha} = 0 \tag{13}$$

$$A_{\beta\alpha} = \Phi\left(\frac{\delta^2 g}{\delta\sigma_\alpha \otimes \delta\sigma_\beta} - C_{\alpha\beta}C_{\beta\beta}^{-1}\frac{\delta^2 g}{\delta\sigma_\beta \otimes \delta\sigma_\beta}\right) \tag{14}$$

where $\alpha = 1,4,\ \beta = 2,3,\ C^{el}$ is the elastic compliance matrix ($C_{\alpha\beta}$ indicates the partition implied by the specific loading programme), $I_{\beta\beta}$ and $I_{\alpha\alpha}$ identity matrices and $D_{22}$ stands for the slope elastic term corresponding to the direction parallel to the slope. Since the derivative of the viscous nucleus with respect to the yield function is positive by definition, whereas $\frac{\delta^2 g}{\delta\sigma_\beta \otimes \delta\sigma_\beta} > 0$ , due to $g$ convexity, instability may occur if and only if $H \leq H_L$.

Under SS conditions, since with time both $H$ and $H_L$ evolve with time, if the current stress level is sufficiently high, after applying the perturbation, the response may be stable (primary creep in Figure 3) and only in the following instability occurs (tertiary creep).

# 4 Progressive failure in bonded/cemented soils

In the scientific literature, the progressive failure term is commonly associated with the spatial propagation, under load-controlled conditions, of the damaged zone, and identified with process that leads with time to the eventual failure of the entire system. Progressive failure is associated with an increase in the displacement rate of the unstable soil or rock mass. In laboratory samples, in particular, in case of bonded geomaterials and granular materials under very large stresses, this phenomenon has been observed and thus described by numerous authors [SU69, ASY71, CV74, ALL04, EKTS16] and is notably commonly associated with the transition from primary, secondary and tertiary creeps. This behaviour becomes extremely relevant in the context of creeping landslides, which are generally characterized by slow movements, but, under particular conditions, may evolve in rapid and unexpected collapses.

This process takes place in overconsolidated clayey soils, structured and cemented granular materials and, in general, in geomaterials characterised by a high level of fragility. In these materials, progressive failure is strictly related to the process known as 'sub-critical crack growth' [WFT80, Atk84, Fre84, OA01, OA07], eventually leading to the breakage of bonds or particles ('grain crushing'). At a microscopic scale, the propagation of cracks over time results in a gradual spatial rearrangement of the microstructure, testified by the accumulation of irreversible strains and a gradual reduction in material strength. Therefore, in bonded geomaterials, progressive failure is typically the consequence of the subcritical crack growth [Atk84] in bonds. Accordingly, some authors [Ken05, OA07, ZB17] suggested to integrate concepts of fracture mechanics to describe, at the macroscale, the temporal evolution of the mechanical response of structured media.

The term bonded geomaterials (used for instance for calcarenites and chalks) indicates all those natural materials, that at the microscale show the presence of interconnected grains by means of inter-granular chemical bonds, causing an increase in stiffness and a non-negligible tensile strength.

In the literature, the mechanical behaviour of bonded geomaterials is generally interpreted within a continuum mechanics framework and according to two alternative approaches:

- One (Nova-Gens approach) inspired to modified elastic-plastic models, based on critical state theory [GN93], according to which the effect of bonding is taken into account by employing additional internal hardening variables related to bonds [ANN00, NCT03].
- The second one consisting in defining binary mixture type models, interpreting the material mechanical behaviour as the combination of two contributions, one related to bonds and the other one to grains. The two contributions are assumed to behave in parallel and suitably assembled by respecting either micro-mechanical equilibrium or energy balance [AK97, YPU98].

As far as the first approach is concerned, starting from the '90s ([GN93], [LN95], [NCT03]), under the hypothesis of material isotropy, the most popular models assume both yield function and plastic potential to depend on only two additional hardening variables: $p_m$ and $p_t$ (Figure 7), independent of $p_s$, describing this latter the size of the yield locus for the equivalent unbonded material. According to this approach, during diagenesis or natural cementation, both $p_m$ and $p_t$ increase with time as a result of chemo/mechanical processes. In contrast, in case a mechanical perturbation is applied and yielding takes place, at the microscale microcracks propagate in space, resulting in a progressive time dependent damage of bonds and in a progressive reduction in both $p_m$ and $p_t$. When both $p_m$ and $p_t$ nullify, the material is assumed to behave as a residual "granular soil" with no cohesion.

To account for time dependency, the Perzyna approach can be applied, without distinguishing time dependency related to granular fabric rearrangement and bon degradation. Although this approach may seem theoretically unacceptable, its use is quite interesting, since it can justify, even in case of standard triaxial tests the change from primary to tertiary creep and vice versa.

This can be easily justified in the light of Eq.9, since during creep tests the stability of the response is governed by the sign of *H* and this, in case of bonded geomaterials, according to Nova-Gens approach, can be written as the combination of three terms, differently evolving with time:

$$H = -\left(\frac{\partial f}{\partial p_s^{\square}}\frac{\partial p_s^{\square}}{\partial \varepsilon_{rs}^{irr}}\frac{\partial g}{\sigma'_{rs}} + \frac{\partial f}{\partial p_m^{\square}}\frac{\partial p_m^{\square}}{\partial \varepsilon_{rs}^{irr}}\frac{\partial g}{\sigma'_{rs}} + \frac{\partial f}{\partial p_t^{\square}}\frac{\partial p_t^{\square}}{\partial \varepsilon_{rs}^{irr}}\frac{\partial g}{\sigma'_{rs}}\right) = H_1 + H_2 + H_3 \ . \tag{15}$$

In Figure 7 the temporal evolution of axial strains during a creep test, obtained by using this approach is illustrated and compared with the evolution with time of *H* (Figure 7a) and *H<sub>i</sub>* (Figure 7b).

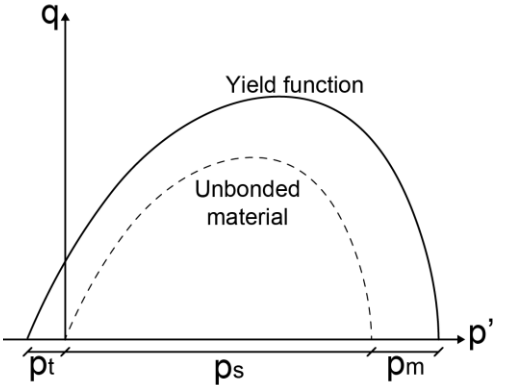


Figure 6: Schematic representation of an isotropic yield locus, modified to account for bond resistance.

Alternatively, by following the second approach, a parallel scheme can be assumed (Figure 7): one contribution refers to bonded medium and the other to the unbonded one. Two yield loci, two plastic potentials and two different rules have to be defined. It is worth mentioning that this approach allows us to assign two distinct fluidity parameters, suitably distinguishing the characteristic times related to micro-cracks propagation ($t_m^*$) and to fabric re-arrangement ($t_s^*$). Additionally, since the global elastic stiffness is given by the sum of the two contributions, it becomes straightforward predicting the on-going reduction in the elastic properties, consequent to bond degradation.

The model is thus capable of predicting very different temporal evolutions according to the current stress state, since the combination of the two contributions and the role of the two fluidity parameters may vary severely. For the sake of clarity, the numerical simulation of a standard triaxial compression creep test, performed by employing an in-parallel model is illustrated in Figure 8.

This type of models are also suitable for capturing the mechanical response of bonded materials, when subject to impulsive perturbations (Figure 9). If the stress level is sufficiently high and/or the impulsive perturbation sufficiently severe, the mechanical response may be unstable for a very long period of time, largely more extended then the perturbation time (Δt), followed by a re-stabilization (testified by a change in sign of the second derivative over time of strains) (Figure 10). In Figure 10 the predicted mechanical response for two impulses characterized by the same load amplitude $\bar{\sigma}'$, but different duration $\Delta t$ is reported. If a threshold duration value Δt is overcome, instability occurs, and the mechanical response is similar to that predicted for creep tests.

The unstable response corresponding to Δt = 140 min is due to a rapid damage of bonds, caused by the impulse, leading to a progressive hardening of the granular matrix: stability is regained, when the hardening of unbonded contribution compensates the progressive damage of the bonded one.

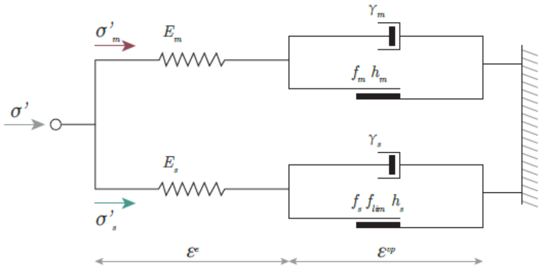


Figure 7: Conceptual structure of the double fluidity elastic-viscoplastic model (from [For23]).

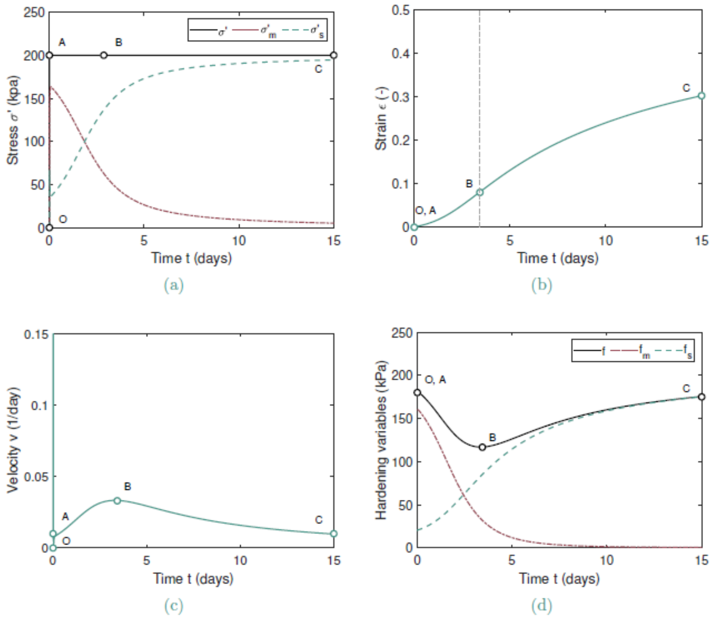


Figure 8: double fluidity elastic-viscoplastic model; predictions of a creep standard compression triaxial test on an isotropically consolidated sample (cell pressure equal to 100 kPa, load increment equal to 200 kPa); temporal evolution of axial total, bond and unbonded stresses (a), axial strain (b), axial strain rate (c) and gardening moduli (d) (from [For23]).

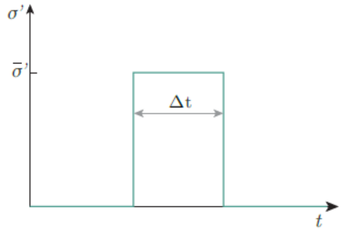


Figure 9: single square impulsive stress perturbation (from [For23]).

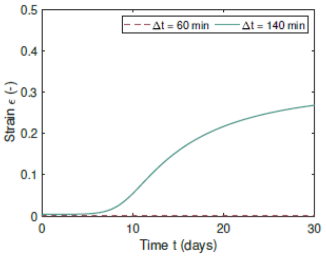


Figure 10: elastic-viscoplastic double fluidity model prediction, corresponding to an initial axial stress of 175 kPa, an impulsive stress of 75 kPa and two different durations Δt (from [For23]).

What observed in case of bonded geomaterials is also valid in case granular materials are loaded under very high confining pressures, since under these conditions grain crushing occurs (YL93). The macroscopic consequence of grain-crushing is the evolution of particle size distribution (Figure 11), but as is testified by many experimental tests this material transformation is time dependent.

As is suggested in the sketch of Figure 12, even in this case, the evolution with time of irreversible strains is due to the spatial propagation of microcracks in grains. Particularly interesting, in case of grain crushing, are the approaches based on breakage mechanics (Ein07a, Ein07b), since in these models time dependence stems from a suitable upscaling procedure [ZB17].

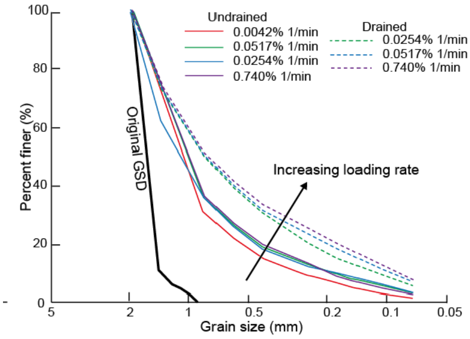


Figure 11: Evolution of particle size distribution due to grain crushing in granular soils and its dependency on strain rate imposed (from [YL93])

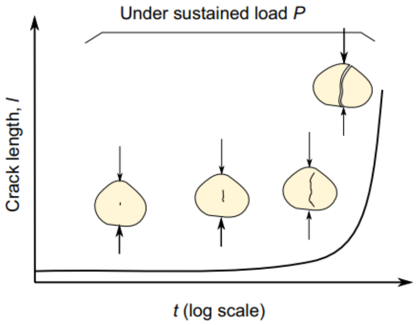


Figure 12: Schematic representation of grain-crushing (from [ZB17])

# 5 Thermo/hydro/chemo/mechanical processes

As was mentioned in the introduction, the time dependency of the mechanical behaviour of geomaterials in some cases is related to the activation of thermo/hydro/chemo-mechanical coupled processes, particularly evident in clays, peats, and naturally/artificially cemented soils when, for instance, these latter are interested by weathering phenomena. To model this kind of time dependence, common is the use of standard elastic-plastic models, including additional hardening functions, putting in relation thermo/hydro/chemical control to hardening variables, in which control variables evolve with time according to prescribed rules, implying both $f$ and $g$ to be dependent on both accumulated irreversible strains and chemical processes.

In most of the cases, the temporal evolution of the mechanical consequences of chemical reactions is assigned ([ZC74], [AO82], [BK85], [DZ87], [KS92], [NA01]) phenomenologically and not derived from micro to macro upscaling procedures.

According to these approaches, yield function may evolve without any either mechanical perturbation (i.e., at constant effective stress) or increase in irreversible strains. Macroscopic evidence of material degradation may be absent, until the shrinkage of $f$ becomes severe enough to get rise to yielding and to the progressive accumulation of irreversible strains, whose size is governed by both consistency rule and temporal evolution of hydro/chemo/mechanical processes ([GCH15], [CdP16]).

An example of simplified micro to macro approach to describe time dependency in calcarenites due to weathering induced by material saturation is in [CdP16]. In this model, $p_m$ and $p_t$ (Figure 6) are assumed to be correlated between each other via a non-dimensional constant, and related, through a simplified upscaling procedure, to the intact material's tensile strength σ¯, and the mean diameter of intergranular bonds Y¯, evolving with time, due to the dissolution of calcium carbonate into water. This process is assumed to induce a change in the normalized calcite mass ξ and ξ to be related to $p_t$as follows:

$$p_t = \bar{\sigma} X \bar{Y}(\xi), \tag{16}$$

being X an upscaling parameter. In the case of calcarenites, where the intergranular bonds are made of calcite, the evolution rule for ξ is derived from the rate of dissolution of calcium carbonate into water solutions ([CH13]):

$$\dot{\xi} = K_b \left( C - [Ca^{2+}]^{\frac{1}{2}} [CO_3^{2-}]^{\frac{1}{2}} \right) (1 + \phi \, \varepsilon_v^{pl}), \tag{17}$$

where $K_b$ is a dissolution parameter, C is an equilibrium constant, $\varepsilon_v^{pl}$ the volumetric plastic strains, $[Ca^{2+}]$ and $[CO_3^{2-}]$ the ionic concentration values, and a chemo-mechanical coupling parameter, accounting for the increase in wet surface due to the development of microcracks within the intergranular bonds.

As is evident from equations (16) and (17), the temporal evolution of $p_t$ is therefore ruled by the velocity of chemical reactions (chemo/mechanical coupling) whereas what is not described here, for the sake of brevity, is the damage induced by mechanical perturbations, affecting the $\bar{\sigma}$ value.

In the previous sections, the importance in determining fluidity parameter $\gamma$ was mentioned and, in particular, the authors observed that $\gamma$ is a function, at the macro-scale, of the processes taking place at the micro-scale. In this perspective, to take thermo-mechanical coupling in clays into account, [SWH21] suggested to simulate temperature dependence of viscous properties by assuming $\gamma$ not to be constant but depending on temperature $T$ as follows:

$$\gamma(T) = B\, exp\left(-\frac{m}{T}\right), \tag{18}$$

where $B$ is a material non dimensional parameter and $m$ a material parameter expressed in degrees Celsius. Equation (18), according to the previously mentioned authors, is justified in the light of Arrhenius equation, suggesting that chemical reaction rates increase with any increase in temperature [MS05].

# 6 Response under large strain rates

The most important peculiarity of granular materials consists in their double nature: they behave like solids or fluids, according to the applied strain rate, current porosity and confinement. In the former case, at the microscale, grains interact mainly through long lasting force chains, whereas, in the latter one, mainly through instantaneous inelastic collisions.

At the macroscale, according to the prevailing mechanism of interaction among particles, granular matters behave differently and, when collisional interactions govern the material mechanical response, this becomes very time dependent.

To visualize regimes and strain rate dependency of the mechanical behaviour of granular materials, in Figure 13 numerical results obtained by performing discrete element numerical constant volume simple shear tests on a dry monodisperse granular assembly are illustrated in the non-dimensional plane $\tau$*-$\dot{\gamma}^*$(being $\tau$*= $\tau\frac{d}{k_n}$, $\dot{\gamma}^* = \dot{\gamma}\, d\sqrt{\rho_p d/k_n}$ , whereas $\rho_p$ is the particle density. $d$ the particle diameter and $k_n$ the stiffness of the spring describing the elastic interaction among particles).

When the material is sufficiently loose and the shear rate sufficiently low (collisional regime in Figure 13), particles interact mainly through inelastic collisions, energy is mainly stored as kinetic fluctuating and the behaviour is fluid-like and the depends of $\tau$ on $\dot{\gamma}$ is quadratic. When clusters are present in the system the regime is called correlated (Figure 14a). whereas, when the material is sufficiently rarefied, uncorrelated (Figure 14b). Within clusters, collisions are absent, but clusters fluctuate colliding against either other clusters or single particles. The correlation reduces the rate of dissipated energy, since the number of collisions reduce.

The regime is quasi-static, and, at steady state, the mechanical behaviour practically rate independent, when the granular material is sufficiently dense and the strain rate sufficiently small. In this regime, the behaviour of the material is solid-like and long lasting force chains are the predominant mechanism of interaction among particles and energy is stored mainly as elastic.
If the shear rate is increased, the regime becomes hybrid with the simultaneous presence of collisions and force chains. The behaviour remains solid-like. The kinetic fluctuating energy stored is comparable with the elastic but remains lower.

Finally, independently of the porosity value, if the shear rate is sufficiently high (Figure 13), the regime becomes inertial and shear stress, under simple shear conditions, is proportional to shear strain rate: the behaviour is fluid-like and the stored elastic energy approximately coincides with the kinetic fluctuating one.

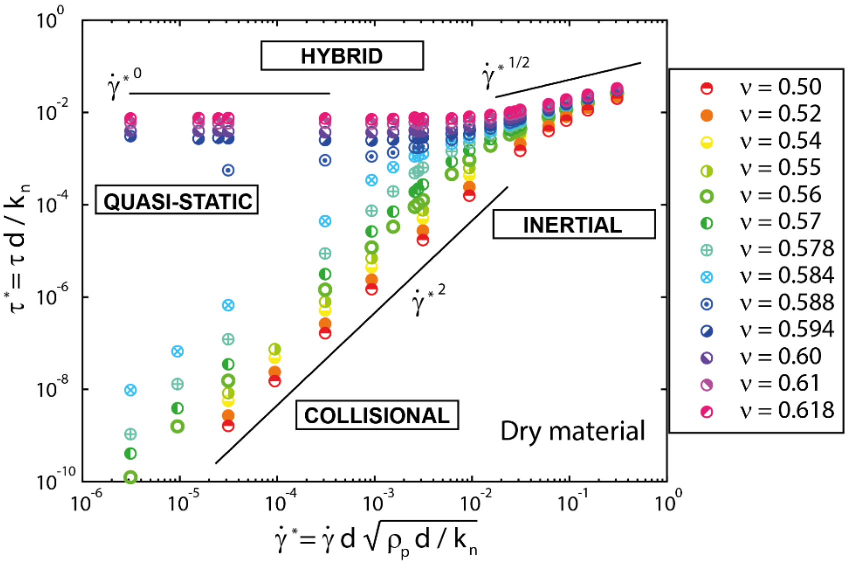


Figure 13: Definition of regimes for dry granular materials under simple shear conditions where $\tau$ is the shear stress, $\dot{\gamma}$ is the shear strain rate, $\rho_p$ is the particle density, $d$ is the particle diameter and $k_n$ is the contact normal stiffness (from [Zer24])

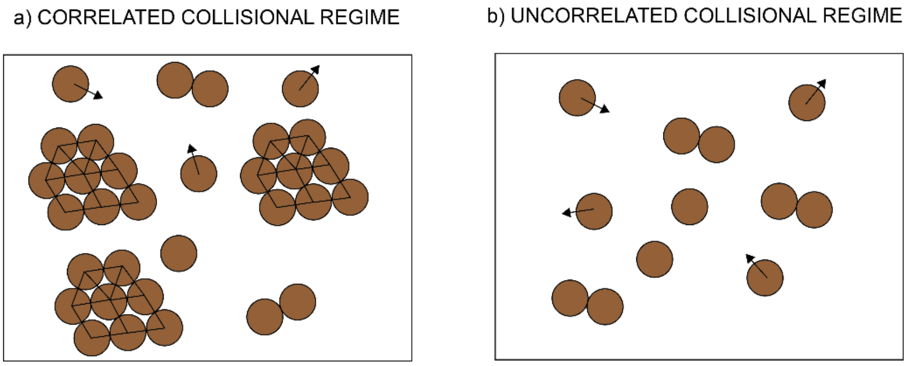

a) CORRELATED COLLISIONAL REGIME
b) UNCORRELATED COLLISIONAL REGIME

Figure 14: Definition of collisional regimes a) correlated and b) uncorrelated (from [Zer24])

In case of saturated conditions, according to the type of dependency, the regime may be (i) quasi-static, (ii) inertial, (iii) Newtonian or (iv) Bagnoldian (Figure 15):

- both quasi-static and inertial regimes are defined as it is under dry conditions;
- Newtonian regime (characterized by a linear dependence of $\tau$ on $\dot{\gamma}$) takes place for sufficiently small values of both solid concentration and shear rate. In this regime;
- finally, for higher shear rates, Bagnoldiam regime, in which the shear stress is proportional to the square of the shear rate. occurs.

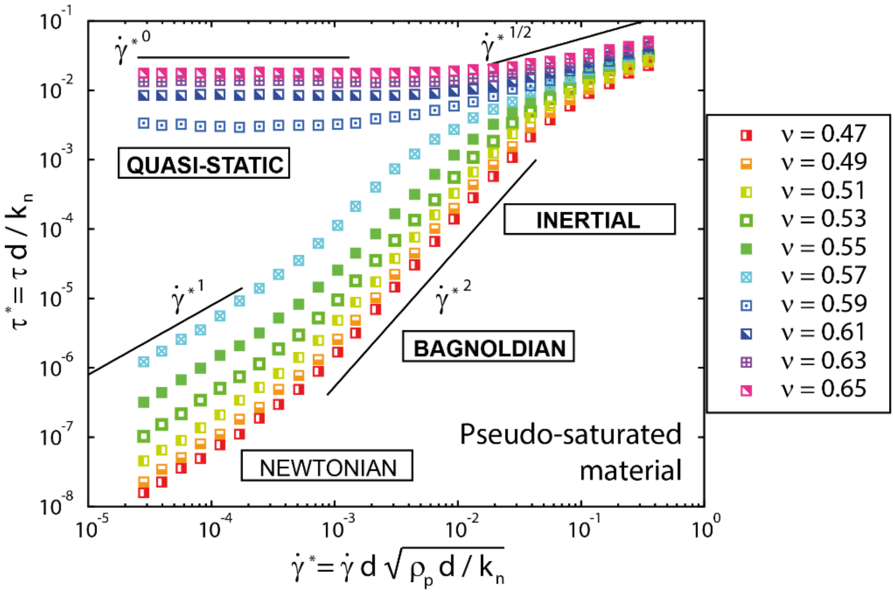


Figure 15: Definition of regimes for saturated granular materials under simple shear conditions where $\tau$ is the shear stress, $\dot{\gamma}$ is the deviatoric strain rate, $\rho_p$ is the particle density, $d$ is the particle diameter and $k_n$ is the contact normal stiffness (from [Zer24])

From an historical point of view, it is worth mentioning that constitutive approaches simulating quasi-static and rime independent mechanical behaviour of granular materials have been proposed in the geotechnical community, whereas models for fluid-like regimes have been developed in the hydraulic/physicist community (kinetic theories of granular gasses). These, in addition to void ratio dependency, account for the role of granular temperature [Gol08]: a scalar measure of particle velocity fluctuations:

$$T = \frac{1}{3}\langle|\breve{\boldsymbol{u}}_{\boldsymbol{P}} \cdot \breve{\boldsymbol{u}}_{\boldsymbol{P}}|\rangle \tag{19}$$

where $\breve{\boldsymbol{u}}_{\boldsymbol{P}} = \boldsymbol{u}_{\boldsymbol{p}} - \boldsymbol{u}_{\boldsymbol{g}}$ is the fluctuating velocity vector of each particle P**,** $\boldsymbol{u}_{\boldsymbol{p}}$ the particle velocity vector, $\boldsymbol{u}_G$ the local mean velocity vector of the granular phase ($\boldsymbol{u}_{\boldsymbol{G}} = \langle\boldsymbol{u}_{\boldsymbol{p}}\rangle$, where symbol ⟨⟩ stands for the average at the macroscopic scale). $T$ is a statistical measure of material agitation.

Only recently, unified constitutive models, suitable of describing the behaviour of granular materials under both quasi-static and dynamic conditions (e.g. [DK15, AME21, GPWW21]) have been proposed. In particular, the Milan research group ([BdPV11, VdPB13, RdPV16, VMdP20, MRdP22, MZdP24] has proposed a multiphase multi-regime model based on: (i) the adoption of the two-phase mixture theory ([TT60]) (ii) the elastic-plastic theory, (iii) the critical state concept, (iv) the kinetic theory of granular gasses and (v) the assumption of an in-parallel response for collisional and quasi-static contributions (Figure 16). In particular, the total stress tensor is defined as follows (Figure 16):

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^{qs} + \boldsymbol{\sigma}^{col} + \boldsymbol{\sigma}^{l} \tag{20}$$

where $\boldsymbol{\sigma}^{qs}$ and $\boldsymbol{\sigma}^{col}$ represent solid quasi-static and dynamic contributions, respectively, whereas $\boldsymbol{\sigma}^{l}$ the liquid phase contribution. From an energetic perspective, the medium involves two storage mechanisms (springs and masses in Figure 16) and two dissipation mechanisms (frictional slider and dashpot in Figure 16). As grains are deformable, elastic energy is stored both in long-lasting force chains and during collisions. The two central in series dashpots in Figure 16 simulate the material's response in the collisional regime: one simulates the energy dissipated by the system in case spheres were rigid, while the other, in parallel with a spring, incorporates the particle deformability.

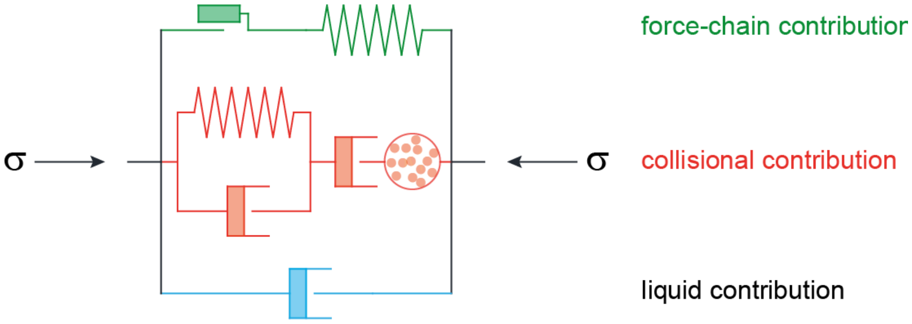


Figure 16 Rheological scheme of the multi-phase/multi-regime model

The contribution of force chains $\boldsymbol{\sigma}^{qs}$ is modelled by employing a non-associated anisotropic strain hardening elastic-plastic constitutive relationship, in which as state variables void ratio and fabric tensor have been chosen. The collisional contribution is instead modelled according to the kinetic theories of granular gasses. When this contribution prevails, the granular material behaves like a very compressible fluid, with a viscosity dependent on both void ratio $e$ and $T$. Under simple shear conditions, the constitutive relationship can be expressed as follows:

$$\sigma^{col} = F_1(e,T)\, T \tag{21}$$

$$\tau^{col} = F_2(e,T)\, T^{1/2}\, \dot{\gamma}\,, \tag{22}$$

where $F_1$ and $F_2$ are non-linear functions, defined according to [BJ15], increasing with $T$ and reducing with $e$, dependence typical for ideal gases, but not for incompressible fluids, in which shear viscosity is due to intermolecular forces [LL80].

As is schematically illustrated under triaxial conditions in Figure 17, in the solid-like regime, total stress coincides with quasi-static contribution and belong to yield locus (Figure 17a); in the hybrid one, the two contributions are equivalent and total stress is outside the yield locus (Figure 17b). Finally, in the collisional regime (Figure 17c), total stress practically coincides with the collisional one and yield locus shrikes into a point coinciding with the origin of the axes.

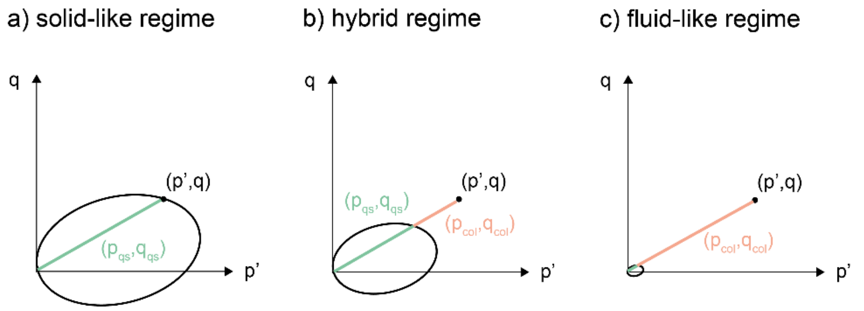


Figure 17 schematic representation in the triaxial plane of solid-like (a), hybrid (b) and fluid-like (c) regimes

Under saturated conditions, the mechanical behaviour of the mixture becomes more complex, since the liquid phase contribution must take into consideration ([VMdP20, MZdP24]) (i) the deviation, near grains, of liquid streamlines from their original paths; (ii) the lubrication effect, occurring when two or more particles come close together, causing the liquid between them to be squeezed out due to the local increase in pressure; (iii) the damping effect of water on the fluctuating motion of particles.

Under the assumption of no turbulence, $\boldsymbol{\sigma}^l$ is modelled according to a Newtonian rheology:

$$\boldsymbol{\sigma_L} = u_w \boldsymbol{I} + 2\eta(e)\dot{\boldsymbol{\varepsilon}}_{\boldsymbol{L}}^{\boldsymbol{d}} \tag{23}$$

where $u_w$ is the isotropic pore pressure and $\eta$ the macroscopic viscosity, function of void ratio $e$ for the reasons summarised here above. For sufficiently large values of void ratio, $\eta \rightarrow \eta_0$ ($\eta_0$ liquid molecular viscosity) according to the well-known Einstein formula [Ein05].

# 7 Concluding remarks

The time and rate dependency of geomaterials is due to various hydro/chemo/mechanical processes occurring at the microscopic level. In this chapter, focussing on granular and cemented geomaterials, the authors have presented various strategies, in the framework of elastic-plastic theory, to theoretically capture and numerically simulate material time dependence. Nowadays, the current research is aimed at justifying the time-dependent mechanical response by employing suitable upscaling approaches, useful for both conceiving constitutive assumptions and calibrating constitutive paramters.

## References

[AK97] Abdulla, A. A. and Kiousisr, P. D. Behavior of cemented sands-ii. modelling. International journal for numerical and analytical methods in geomechanics, 21:549–568, 1997.

[AO82] Adachi, T., & Oka, F. Constitutive equations for normally consolidated clay based on elasto-viscoplasticity. Soils and foundations, 22(4), 57-70, 1982

[AME21] Alaei, E., Marks, B., and Einav, I. A hydrodynamic-plastic formulation for modelling sand using a minimal set of parameters. Journal of the Mechanics and Physics of Solids, 151(September 2020):104388, 2021.

[Atk84] Atkinson, B. K. Subcritical crack growth in geological materials. J. Geophys. Res.: Solid Earth 89, No. B6, 4077–4114, 1984.

[ALL04] Augustesen, A., Liingaard, M. & Lade, P. V. Evaluation oftime-dependent behavior of soils. Int. J. Geomech. 4, No. 3,137–156, 2004.

[ASY71] Arulanandan, K., Shen, C. K. & Young, R. B. Undrainedcreep behaviour of a coastal organic silty clay. Géotechnique 21,No. 4, 359–375, 1971.

[AMN00] Asaoka, A., Nakano, M., and Noda, T.. Superloading yield surface conceptfor highly structured soil behavior. Soils and Foundations, 40:99–110, 2000.

[BdPV11] Berzi, D., di Prisco, C. & Vescovi, D. Constitutive relations for steady, dense granular flows. Physical Review E 84, 031301, 2011

[BJ15] Berzi, D., Jenkins, J.: Steady shearing flows of deformable, inelastic spheres. Soft Matter 11(24), 4799–4808, 2015

[BK85] Borja, R. I., & Kavazanjian, E. A constitutive model for the stress–strain–time behaviour of ‘wet’clays. Geotechnique, 35(3), 283-298. 1985

[BDdP11] Buscarnera, G., Dattola, G., & di Prisco, C. Controllability, uniqueness and existence of the incremental response: A mathematical criterion for elastoplastic constitutive laws. International Journal of Solids and Structures, 48(13), 1867–1878, 2011.

[CV74] Campanella, R. G. & Vaid, Y. P. Triaxial and plane straincreep rupture of an undisturbed clay. Can. Geotech. J. 11, No. 1,1–10, 1974.

[CdP16] Ciantia, M. O., & di Prisco, C. Extension of plasticity theory to debonding, grain dissolution, and chemical damage of calcarenites. International Journal for Numerical and Analytical Methods in Geomechanics, 40(3), 315-343, 2016

[CH13] Ciantia, M. O., & Hueckel, T. Weathering of submerged stressed calcarenites: chemo-mechanical coupling mechanisms. Géotechnique, 63(9), 768-785, 2013.

[DZ87] Desai, C. S., & Zhang, D. Viscoplastic model for geologic materials with generalized flow rule. International Journal for Numerical and Analytical Methods in Geomechanics, 11(6), 603-620, 1987.

[dPI96] di Prisco, C., & Imposimato, S. Time dependent mechanical behaviour of loose sands. Mechanics of Cohesive-frictional Materials: An International Journal on Experiments, Modelling and Computation of Materials and Structures, 1(1), 45-73, 1996.

[dPSZ07] di Prisco, C., Stupazzini, M., & Zambelli, C. Nonlinear SEM numerical analyses of dry dense sand specimens under rapid and dynamic loading. International Journal for Numerical and Analytical Methods in Geomechanics, 31(6), 757-788, 2007.

[DK15] Dunatunga, S. and Kamrin, K. Continuum modelling and simulation of granular flows through their many phases. Journal of Fluid Mechanics, 779, 483–513, 2015.

[Ein07a] Einav, I. Breakage mechanics—part I: theory. Journal of the Mechanics and Physics of Solids, 55(6), 1274-1297, 2007.

[Ein07b] Einav, I. Breakage mechanics—Part II: Modelling granular materials. Journal of the Mechanics and Physics of Solids, 55(6), 1298-1320, 2007.

[Ein05] Einstein, A. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. Annalen der Physik 322, No. 8, 549–560 (in German), 1905.

[EKTS16] Enomoto, T., Koseki, J., Tatsuoka, F. & Sato, T. Creepfailure of natural gravelly soil and its simulation.Géotechnique 66, No. 11, 865–877, 2016.

[For23] Fortunato, M. Multi-scale elasto-viscoplastic constitutive models for bonded geomaterials. Master thesis, Politecnico di Milano, 2023.

[Fre84] Freiman, S. W. Effects of chemical environments on slowc rack growth in glasses and ceramics. J. Geophys. Res.: SolidEarth 89, No. B6, 4072–4076, 1984

[GD99] Garzó, V., Dufty, J.W.: Dense fluid transport for inelastic hard spheres. Phys. Rev. E 59(5), 895–5911, 1999

[GCH15] Gajo, A., Cecinato, F., & Hueckel, T. A micro-scale inspired chemo-mechanical model of bonded geomaterials. International Journal of Rock Mechanics and Mining Sciences, 80, 425-438, 2015.

[GN93] Gens, A., & Nova, R. (1993). Conceptual bases for a constitutive model for bonded soils and weak rocks. Geotechnical engineering of hard soils-soft rocks, 1(1), 485–494

[GPWW21] Guo, X., Peng, C., Wu, W., and Wang, Y. Unified constitutive model for granular-fluid mixture in quasi-static and dense flow regimes. Acta Geotechnica, 16(3):775–787, 2021.

[Gol08] Goldhirsch, I. Introduction to granular temperature. Powder Technology, 182(2):130–136, 2008.

[Kem05] Kemeny, J. Time-dependent drift degradation due to the progressive failure of rock bridges along discontinuities. International Journal of Rock Mechanics and Mining Sciences, 42:35–46, 2005.

[KS92] Kutter, B.L., and Sathialingam, N. Elastic–viscoplastic modelling of the rate-dependent behaviour of clays. Géotechnique,42(3): 427–441, 1992.

[LN95] Lagioia, R. & Nova, R. An experimental and theoretical study of the behaviour of a calcarenite in triaxial compression. Géotechnique 45, No. 4, 633–648, 1995.

[LL80] Landau, L. D., and E. M. Lifshitz. Statistical Physics, Part 1. Vol. 5, 3rd ed., Butterworth-Heinemann, 1980.

[LSS15] Lazari, M., Sanavia, L., and Schrefler, B. A. Local and non-local elasto-viscoplasticity in strain localization analysis of multiphase geomaterials. International Journal for Numerical and Analytical Methods in Geomechanics, 39: 1570–1592, 2015.

[LSdPP19] Lazari M, Sanavia L, di Prisco C, Pisanò F. Predictive potential of Perzyna viscoplastic modelling for granular geomaterials. International Journal for Numerical and Analytical Methods in Geomechanics. 43: 544–567, 2019.

[LP91] Loret, B. & Prevost, J.H. Dynamic Strain Localization in Fluid-Saturated Porous Media. Journal of Engineering Mechanics, 117(4), 907-922, 1991

[Lya92] Lyapunov, A. M. The general problem of the stability of Q13 motion. Kharkovskoye Matematicheskoe Obshchestvo, 1892.

[Men03] Meng, J. The influence of loading frequency on dynamic soil properties. PhD Thesis, Georgia Institute of Technology, 2003.

[MRdP22] Marveggio, P., Redaelli, I., and di Prisco, C. G. Phase transition in monodisperse granular materials: How to model it by using a strain hardening visco-elastic-plastic constitutive relationship. International Journal for Numerical and Analytical Methods in Geomechanics, 46(13):2415–2445, 2022.

[MZdP24] Marveggio, P., Zerbi, M., & Di Prisco, C. A multi-phase/multi-regime modelling approach for saturated granular media. Computers and Geotechnics, 2024.

[MS05] Mitchell, J.K. and Soga, K. Fundamentals of Soil Behavior. 3rd Edition, John Wiley & Sons, Hoboken, 2005

[NA01] Navarro, V., & Alonso, E. E. Secondary compression of clays as a local dehydration process. Géotechnique, 51(10), 859-869, 2001

[Nov82] Nova, R. A viscoplastic constitutive model for normally consolidated clay. In proceedings IUTAM conf. Deformation and Failure of Grnaular Materials 287-295, 1982.

[Nov94] Nova, R. Controllability of the incremental response of soil specimens subjected to arbitrary loading programmes. J. Mech. Behav. Mater. 5, No. 2, 193–202, 1994.

[NCT03] Nova, R., Castellanza, R., & Tamagnini, C. (2003). A constitutive model for bonded geomaterials subject to mechanical and/or chemical degradation. International Journal for Numerical and Analytical Methods in Geomechanics, 27(9), 705–732

[OA01] Oldecop, L. A. & Alonso, E. E. (2001). A model for rockfill compressibility. Géotechnique 51, No. 2, 127–139

[OA07] Oldecop, L. A. & Alonso, E. E. (2007). Theoretical investigation of the time-dependent behaviour of rockfill. Géotechnique 57, No. 3, 289–301

[Per63] Perzyna, P. (1963). The constitutive equations for rate sensitive plastic materials. Quarterly of applied mathematics, 20(4), 321-332.

[PdP16] Pisanò, F., & di Prisco, C., (2016). A stability criterion for elasto-visco-plastic constitutive relationships. International Journal for Numericaland Analytical Methods in Geomechanics, 40(1), 141–156.

[RdP19] Redaelli, I., di Prisco, C.: Three dimensional steady-state locus for dry monodisperse granular materials: DEM numerical results and theoretical modelling. Int. J. Numer. Anal. Meth. Geomech. 43(16), 2525–2550 (2019)

[RdPV16] Redaelli, I., di Prisco, C., Vescovi, D. (2016): A visco-elasto-plastic model for granular materials under simple shear conditions. Int. J. Numer. Anal. Meth. Geomech. 40(1), 80–104

[RMdP21] Redaelli, I., Marveggio, P., & di Prisco, C. (2021). Three-Dimensional Constitutive Model for Dry Granular Materials Under Different Flow Regimes. Volume 125, Pages 548 – 555

[SU69] Saito, M. & Uezawa, H. Forecasting time of slope failure bytertiary creep. Proceedings of the 7th international conference onsoil mechanics and foundation engineering, Mexico city, Mexico,vol. 2, pp. 677–683, 1969

[SWH21] Shi, Z., Wood, D. M., & Huang, M. Interpreting temperature effects in soils using thermally-enhanced viscoplastic model. Computers and Geotechnics, 136, 104208, 2021.

[Shi95] Shibuya, S., Toshiyuki, T., Fukuda, F. and Degoshi, T. (1995): "Strain rate effects on shear modulus and damping of normally consolidated clay," ASTM Geotech. Testing J., Vol. 18, No. 3, pp. 365-375.

[SW68] Schofield A. N., Wroth C.P. Critical state soil mechanics. McGraw-Hill., London, 1968

[SV95] Sulem, J., & Vardoulakis, I. G. Bifurcation analysis in geomechanics. CRC Press, 1995.

[TT60] Truesdell, C., & Toupin, R. The Classical Field Theories. In S. Flügge (A c. Di), Principles of Classical Mechanics and Field Theory / Prinzipien der Klassischen Mechanik und Feldtheorie: Vol. 2 / 3 / 1 (pp. 226–858). Springer Berlin Heidelberg, 1960.

[VdPB13] Vescovi, D., di Prisco, C. G., and Berzi, D. From solid to granular gases: the steady state for granular materials. International Journal for Numerical and Analytical Methods in Geomechanics, 37:2937–2951, 2013.

[VMdP20] Vescovi, D., Marveggio, P., & di Prisco, C. G. Saturated granular flows: constitutive modelling under steady simple shear conditions. Géotechnique, 70(7), 608-620, 2020.

[WFT80] Wiederhorn, S. M., Fuller, E. R. & Thomson, R. Micromechanisms of crack growth in ceramics and glasses in corrosive environments. Met. Sci. 14, No. 8–9, 450–458, 1980.

[WSdB98] Wang, W. M., Sluys, L. J., de Borst, R.. Viscoplasticity for instabilities due to strain softening and strain-rate softening. International Journal for Numerical Methods in Engineering, 40, 3839-3864, 1997.

[YL93] Yamamuro, J. A., & Lade, P. V. Effects of strain rate on instability of granular soils. Geotechnical Testing Journal, 16(3), 304-313, 1993.

[YPU98] Yu, Y., Pu, J., and Ugai, K.. A damage model for soil-cement mixture. Soils and Foundations, 38:1–12, 1998.

[Zer24] Zerbi, M. Numerical investigation of impacts on/of granular masses. PhD Thesis, Politecnico di Milano. 2024.

[ZB17] Zhang, Y.D. and Buscarnera, G. A rate-dependent breakage model based on the kinetics of crack growth at the grain scale. Géotechnique 67, No. 11, 953–967, 2017.

[ZC74] Zienkiewicz, O.C., and Cormeau, I.C. Visco-plasticity, plasticity and creep in elastic solids: a unified numerical solution approach. International Journal for Numerical and Analytical Methods in Geomechanics, 8(2): 821–845, 1974.

# Modelling of unsaturated soils with Generalized Plasticity

**Diego Manzanal, Manuel Pastor, J.A. Fernández-Merodo, Miguel Martín Stikle, Pedro Navas, Ángel Yagüe, Pablo Mira**

*ETS de Ingenieros de Caminos, Canales y Puertos de Madrid*
*Universidad Politécnica de Madrid, Spain.*
*d.manzanal@upm.es*

*The Chapter is devoted to the constitutive modelling of geo-materials based on Generalized Plasticity Theory. The aim is to provide the reader with an overview of the generalized plasticity state parameter-based model to reproduce the hydro-mechanical behavior of unsaturated soils. The proposed model is based on two pairs of stress-strain variables and a suitable hardening law taking into account the bonding—debonding effect of suction and degree of saturation.*

## 1 Introduction

Constitutive modeling of unsaturated soils is a relatively new field compared to the modeling of saturated soils. The effective stress principle is widely accepted to explain the fundamental behavior of saturated soils. In most testing devices, when pore pressures are generated, they are measured and recorded to determine the effective stress. The situation is different for unsaturated soils. Even today, researchers use various alternative stress measures in their models. The initial efforts to understand the behavior of unsaturated soils were made by Haines [Hai25] and Fisher [Fis26]. They examined an assembly of monodisperse spheres and the stabilizing interparticle forces exerted by water menisci. From their results, it is possible to derive a relationship, $f(s)$, between the increments of stabilizing hydrostatic stress at a given suction $s$ and at zero suction.

$$f\left(s\right)=\frac{3}{4}\left\{2-\frac{1}{2s}\left[-\frac{3T_s}{R}+\sqrt{\left(\frac{3T_s}{R}\right)^2+\frac{8T_s}{R}s}\right]\right\} \tag{1}$$

where $T_s$ is the surface tension and $R$ the radius of the spherical particles. There are two limit cases when suction tends to zero and infinity. In the former, $f(s)=1$ and in the latter, when suction tends to infinity, $f(s)=3/2$.

Bishop introduced a generalization of Terzaghi's effective stress principle to explain the mechanical behavior of unsaturated soils using a single stress variable.

$$\sigma'_{ij}=\sigma_{ij}-p_a\cdot\delta_{ij}+\chi\cdot(p_a-p_w)\cdot\delta_{ij} \tag{2}$$

where $\sigma_{ij}$ is the total stress tensor, $p_a$ is the pore air pressure, $p_w$ is the pore water pressure, $p_a-p_w$ is the matrix suction $s$, $\delta_{ij}$ is the Kronecker delta and $\chi$ is a parameter varying between zero and one, which is often referred to as Bishop's effective stress parameter. The stress tensor can be decomposed into a net stress tensor ($\bar{\sigma}_{ij}=\sigma_{ij}-p_a$) and the suction term.

Even though Bishop's effective stress predicted the shear strength of unsaturated soils, it could not reproduce collapse during wetting paths. This limitation prompted Bishop and Blight to introduce the so-called bi-tensorial formulations based on net stress and matric suction tensors. This bi-tensorial approach was employed in the first generation of constitutive models for unsaturated soils. Alonso et al. [AGJ90] introduced a model known today as the Barcelona Basic Model (BBM), which provided a framework for understanding many fundamental aspects of unsaturated soil behavior.

The work of Houlsby [Hou97] who analyzed the work input to unsaturated granular materials, demonstrated that two sets of work conjugated variables were necessary to explain unsaturated soil behaviour. Therefore, a second generation of models for unsaturated soils based both on the effective stress and suction was produced [Jom00].

In this chapter, we present a Generalized Plasticity constitutive model to reproduce the main features of unsaturated soil behavior from a state parameter point of view [MPM11]. The model is formulated using two sets of stress–strain variables, the modified effective stress and suction as stress variables and the strain of solid skeleton and degree of saturation as strain variables, coupling the hydraulic and the mechanical behaviour of unsaturated soils within a Generalized Plasticity framework. The hardening and softening effect due to the change of saturation conditions are expressed in terms of the suction and degree of saturation using a bonding parameter.

## 2 Generalized Plasticity Framework

There are excellent texts and state of the art papers devoted to describing constitutive models and their use in geotechnical engineering. We can mention the classic texts of

Cambou and Di Prisco [CdP00], Zienkiewicz et al. [ZCP$^+$99] among others, and the references provided therein. In this chapter, we will focus on the Generalized Plasticity models which can reproduce behaviour of geomaterials under both monotonic and cyclic loading.

Generalized Plasticity Theory (GPT) introduced by [ZM84] and elaborated by Zienkiewicz and Pastor [ZLP85], [PZC90], as it provides a framework within which accurate models can be developed to describe more relevant phenomena of soil behaviour under monotonic and cyclic loading. Models within GPT have been developed for bonded geomaterials and collapsible soils [FMPM$^+$04], for saturated and unsaturated soils [MMP11], [MPM11] and recently to anisotropic materials [GGMP24] based on state parameter (MPZ model). The MPZ model has been integrated with an explicit scheme within the u-pw formulation in finite element code GeHoMadrid to study several boundary value problems such as marine foundations [MFMM$^+$18], dynamic liquefaction induced by earthquakes [MBLQ$^+$21], and subsidence due to groundwater withdrawal [FMEMea21] and with an implicit integration scheme in the finite element code Plaxis to reproduce failures in mine tailing dams [LMS21], [LSM22].

Generalized Plasticity Theory introduces the dependence of the constitutive tensor relating increments of stress and strain on the direction of the increment of stress via a unit tensor $\mathbf{n}$ which discriminates the states of “loading” and “unloading”

$$\begin{aligned} d\boldsymbol{\varepsilon} &= \mathbf{C}_L : d\boldsymbol{\sigma} \quad for\ \mathbf{n} : d\boldsymbol{\sigma}^e > 0 \\ d\boldsymbol{\varepsilon} &= \mathbf{C}_U : d\boldsymbol{\sigma} \quad for\ \mathbf{n} : d\boldsymbol{\sigma}^e < 0 \end{aligned} \tag{3}$$

where $d\boldsymbol{\sigma}^e$ is the increment of stress which would be produced if the behaviour were elastic, $d\boldsymbol{\sigma}^e = \mathbf{D}^e : d\boldsymbol{\varepsilon}$ , and $\mathbf{D}^e$ is the elastic constitutive tensor.

After imposing the condition of continuity between loading and unloading states, we arrive to

$$\begin{aligned} \mathbf{C}_L &= \mathbf{C}^e + \frac{1}{H_L}\mathbf{n}_{gL} \otimes \mathbf{n} \\ \mathbf{C}_U &= \mathbf{C}^e + \frac{1}{H_U}\mathbf{n}_{gU} \otimes \mathbf{n} \end{aligned} \tag{4}$$

In above equations, subindices $L$ and $U$ refer to “loading” and “unloading”. The scalars are called loading and unloading plastic modulii, and unit tensors give the direction of plastic flow during loading and unloading.

The limit case $\mathbf{n} : d\boldsymbol{\sigma}^e = 0$, is called “neutral loading”, and with the assumption done in equation 4, it can be seen that the response is continuous as:

$$\begin{aligned} d\boldsymbol{\varepsilon}_L &= \mathbf{C}_L : d\boldsymbol{\sigma} = \mathbf{C}^e : d\boldsymbol{\sigma} \\ d\boldsymbol{\varepsilon}_U &= \mathbf{C}_U : d\boldsymbol{\sigma} = \mathbf{C}^e : d\boldsymbol{\sigma} \end{aligned} \tag{5}$$

The strain increment can be decomposed into two parts: elastic and plastic as:

$$d\boldsymbol{\varepsilon} = d\boldsymbol{\varepsilon}^e + d\boldsymbol{\varepsilon}^p \tag{6}$$

with

$$d\boldsymbol{\varepsilon}^e = \mathbf{C}^e : d\boldsymbol{\sigma} \tag{7}$$

and

$$d\boldsymbol{\varepsilon}^p = \frac{1}{H_{L/U}} \mathbf{n}_{gL/U} \otimes \mathbf{n} : d\boldsymbol{\sigma} \tag{8}$$

The increment of strain for unsaturated behavior is assumed to be:

$$d\varepsilon = \mathbf{C}^e : d\boldsymbol{\sigma}' + \frac{1}{H_{L/U}} \cdot \mathbf{n}_{gL/U} \otimes \mathbf{n} : d\boldsymbol{\sigma}' + \frac{1}{H_b} \cdot \mathbf{n}_{gL/U} \cdot ds \tag{9}$$

where the first two terms represent the elastic and plastic strains that have already been described, and the last term represents the plastic strain developed during wetting–drying cycles and $\sigma'$ is effective stress for unsaturated soil, which will be explained later.

The main advantage of the Generalized Plasticity Theory is that all ingredients can be postulated without introducing any yield or plastic potential surface. Moreover, both classical plasticity and Bounding Surface Plasticity models are special cases of the GPT.

Therefore, to fully characterize the non-linear irreversible behavior of soils within a generalized plasticity approach, the following items are necessary: (i) the elastic constitutive tensor $\mathbf{C}^e$ ; (ii) the unit tensor $\mathbf{n}$ discriminating loading and unloading conditions; (iii) the unit tensor describing the direction of plastic flow $\mathbf{n}_{gL/U}$ in loading and unloading; and (iv) the loading and unloading plastic modulus $H_{L|U}$ [PZL85], [PZC90] and a suitable definition of effective stress $\sigma'$ for unsaturated soil behavior [MPM11].

# 3 Principal aspects of unsaturated behaviour

The main features of unsaturated behaviour are the following:

Concerning the **volumetric behaviour** of unsaturated soils, the most important aspects are observed in a series of representative tests:

(i) Constant net confining stress *p = (p$_a$- p$_w$)* varying s tests (or the alternative constant void ratio and variable suction), aiming to obtain relations between the $S_r$ and $s$ which are referred to as water retention curves or soil water characteristic curves. Usually, the results are plotted on the log $s$ – $S_r$ plane, and show, when drying, a primary drying curve. Similarly, a primary wetting curve can be defined for soil which, starting from zero saturation, is wetted. Both primary drying and wetting curves are different, i.e., the soil presents hydraulic hysteresis. This phenomenon is also observed when applying drying-wetting cycles to unsaturated soils. Therefore, there exists hysteresis, as primary drying and wetting curves differ. These two curves act as boundary curves of all possible states of the soil, which in the wetting and drying path follows the secondary branches. Also, the secondary branches present hysteresis, though smaller and often are referred to as "scanning curves". Even at very high suctions, there exists a minimum or residual degree of saturation $S_{r0}$, which corresponds to adsorbed water.These aspects will be illustrated in the next section.

In the primary drying branch, there is a point where the air breaks into the pores, which is referred to as the air entry value $\mathrm{s_{ae}}$. At suctions smaller than the air entry value, the air distribution in the pores is called insular, and the air exists as insulated bubbles. At much larger suction values, the structure tends to pendular, with water only in the menisci. Khogo et al.[KNM94] have proposed an intermediate structure which takes into account the pore and grain size distribution, for which they have proposed the name diffuse.

Suitable laws for wetting-drying primary and intermediate curves have been proposed [vGM80] in the past. It is important to notice that even if these tests are often referred to as purely hydraulic, they present an important variation of effective stress, and what it is observed is coupled hydro-mechanical behaviour.

(ii) Collapse tests decrease the suction by wetting the specimen of unsaturated soil, while keeping the net stress constant. They are referred to as collapse because of the sudden decrease of volume observed. It is important to notice that the effective confining stress $p\prime = p - p_a + s \cdot S_r$ decreases as $\bar{p} = p - p_a$ is constant, and the product $s \cdot S_r$ decreases at the moment the collapse is observed. This observation is an apparent paradox within the framework of plasticity-based only on effective stress, as no plastic deformation should be observed when the stress path is directed towards the interior of the yield surface, and indeed it leads to introducing bitensorial formulations.

(iii) Isotropic consolidation tests (varying net stress under constant suction), where it can be observed that normal consolidation lines depend on the suction (higher suction implies higher pre-consolidation stresses), which can be normalized by using the cementation variable [GGSV03]: $\xi = f(s) \cdot (1 - S_r)$ where the function $f(s)$ is the ratio between the stabilizing pressure at a given suction $s$ and at zero suction introduced by Haines [Hai25] and Fisher [Fis26]. The normalization consists of relating the void ratios at suction $s$ and at saturation at a given effective confining pressure.

(iv) Isotropic mixed paths combining isotropic compression at constant suction and wetting-drying cycles show a coupling between hydraulics and mechanic behaviour. Indeed, a isotropic compression test after a wetting-drying cycle shows a smaller preconsolidation pressure than expected [VRJ00].

Concerning their **shear behaviour**, unsaturated soils show a non-linear increase of strength with suction which has been explained in the past using expressions of the type $q = M \cdot \bar{p} + M_{ss}$ where the term $M_{ss}$ can be interpreted as a "cohesion". This kind of law does disagree that residual states should exhibit zero cohesion. It would be important to verify if using the effective stress introduced for unsaturated soils, it would be possible to recover the CSL expression as $q = M \cdot p'$, where $M$ is the critical state line in deviator stress vs mean effective stress plane. The uniqueness of the CSL for unsaturated soils was studied in [Man08].

We can conclude that a Critical State based model for unsaturated soils should incorporate the following ingredients:

- The model should be formulated in terms of the effective stress tensor $\sigma\prime$
- Analysis of the work input to an unsaturated soil [Hou97] shows that, in addition to effective stress and strain, there are two work conjugated variables, s and $S_r$ which must be taken into account when formulating a constitutive model. Use of only the effective stress and strain results in severe limitations.
- The model should implement an accurate description of wetting-drying curves, accounting for hysteresis and state parameter dependency. The two characteristic values (i) air entry pressure and (ii) the residual degree of saturation related to adsorbed water should be taken into account.
- In the case of classical plasticity models, the size of the yield surface controlled by $p_c$ (known as yield stress or the pre consolidation pressure) depends on both the volumetric plastic strain $\varepsilon_v^p$ and $n \cdot S_r$ (or $s$). For instance, Tamagnini [Tam04] has proposed:

$$\frac{\partial p_c}{\partial \varepsilon_v^p} = \frac{1+e}{\kappa - \lambda} \varepsilon_v^p p_c \tag{10}$$

$$\frac{\partial p_c}{\partial S_r} = -b p_c \tag{11}$$

  Therefore, the yield surface will depend on the degree of saturation or the suction. This explains the collapse observed in isotropic compression tests.

  Classical CS models not accounting for the second pair of work conjugated variables (suction as stress variable and degree of saturation as strain variable) will not be able to predict plastic volumetric collapse. Indeed, soil behaviour depends on both effective stress $p'$ and suction s (or Sr) and yield surfaces have to be defined as functions of both variables.

- The Normal Consolidation Line is kept, with providing a suitable normalization such as [GWK03] is used.
- The concept of Critical State Line is kept, but some modifications can be required to analyze and model experimental results, as it will be described later.

# 4 Generalized Plasticity model for saturated and unsaturated soils

## 4.1 State variables

The state of the material is characterized by a set of state variables $\{\sigma,\ e,\ \zeta_{\max}, \xi, \xi_{dev}\}$, where $e$ is the void ratio, $\zeta_{\max}$ is the maximum reached value of a mobilized stress function and $\xi$ is bonding parameter for unsaturated state to be defined later, and $\xi_{dev} = \int \parallel \dot{\varepsilon}^{\mathbf{p}}_{\mathbf{d}} \parallel$ is the accumulated deviatoric plastic strain. The function of state variables $\psi = e - e_{cs}[p]$ is defined, where $e_{cs}[p]$ is the critical state void ratio

$$e_{cs} = e_{\Gamma} - \ \lambda^* \ \left(\frac{p}{p_{ref}}\right)^{\zeta_c} \tag{12}$$

and $e_{\Gamma}$, $\lambda^*$ and $\zeta_c$ are material parameters. $p$ is the mean effective stress.

## 4.2 Effective stress

The model is formulated using two sets of stress–strain work conjugated variables [Hou97] coupling the hydraulic and the mechanical behaviour of unsaturated soils within a Generalized Plasticity framework. Stress variables are the effective stress tensor and the matrix suction $s$, and Strain variables are the soil skeleton strain and the degree of saturation. The effective stress is given by

$$\sigma\prime_{ij} = \sigma_{ij} - p_a \cdot \delta_{ij} + S_{re} \cdot (p_a - p_w) \cdot \delta_{ij} \tag{13}$$

where $\sigma_{ij}$ is the total stress tensor, $p_a$ is the pore air pressure, $p_w$ is the pore water pressure, $p_a - p_w$ is the matrix suction $s$, $\delta_{ij}$ is the Kronecker delta and $S_{re}$ is the relative degree of saturation which is given by

$$S_{re} = \frac{S_r - S_{r0}}{1 - S_{r0}} \tag{14}$$

where $S_{r0}$ is the residual degree of saturation. We found an important dispersion on the experimental data even when we used the effective stress definition introduced by Schrefler [Sch84] with a modified scalar factor of Bishop effective stress defined by $\chi = S_r$. The improvement obtained by using $S_{re}$ in the effective stress definition can be seen in Figure 1 which shows the predictive and experimental shear strength with both approaches, $\chi = S_{re}$ and $\chi = S_r$, for the experimental data described in Toll [Tol90] and Sivakumar [Siv93].

## 4.3 Bonding variable

The first key component of this model is the function of the state variable (e) known as the state parameter $\psi$, defined in the previous section and based on the critical state

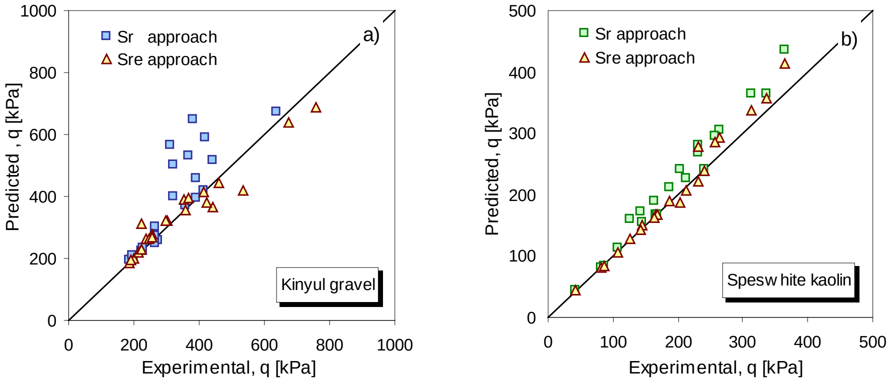


Figure 1: Comparison between predicted and experimental deviatoric stress for a) kinyul gravel (Experimental data from [Tol90] and b) speswhite kaolin (Experimental data from [Siv93].

line (CSL). For unsaturated soils, the CSL depends on suction, making it necessary to define this dependency. The bonding variable $\xi$ defined by [GGSV03], is related to the ratio between the stabilizing pressure at a given suction s and at zero suction $f(s)$, as shown in equation 1 and degree of saturation $S_r$, introduced by Haines [Hai25] and Fisher [Fis26]:

$$\xi = f(s) \cdot (1 - S_r) \tag{15}$$

To relate CSL for saturated states and CLS for different suction, we link the values of $p'$ at saturation and a given suction for a fixed void ratio:

$$\frac{p\prime_{CS}^{unsat}}{p\prime_{CS}^{sat}} = 1 + g\left(\xi\right) \tag{16}$$

where

$$g\left(\xi\right) = a \cdot \left[exp\left(b \cdot \xi\right) - 1\right] \tag{17}$$

and $\xi$ is the bonding variable. The function $g(\xi)$ depends on the degree of saturation and suction and takes a zero value at saturation. The parameters $a$ and $b$ are calibrated from experimental data as shown by [MPM11]. In Fig. 2, we have depicted the CSL for saturated and unsaturated states on the plane and the normalization effect of the function $g(\xi)$.

Combining equations 16 and 17 with a suitable definition of a CSL for saturated states will generalise the critical state line to unsaturated states. We provide in fig. 3 an

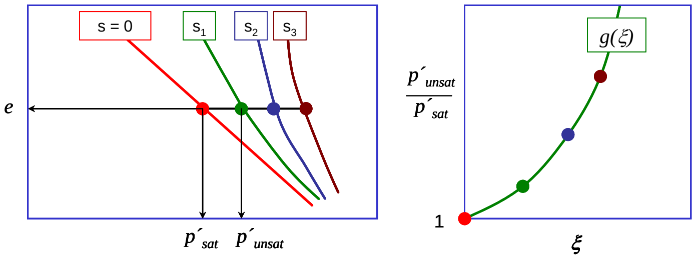


Figure 2: CSLs for saturated and unsaturated state.

example using the experimental data described in Sivakumar [Siv93] which illustrate the effectiveness of the proposed approach.

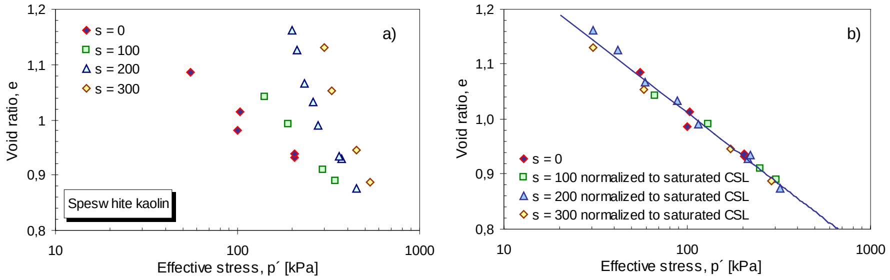


Figure 3: a) Critical state for speswhite kaolin at different suctions b) Normalization of CSLs (Experimental data from [Siv93].

## 4.4 Formulation

Taking into account all experimental facts described above, it is possible to develop a model within the Generalized Plasticity Theory developing the five ingredients that we state in section 2 as follows:

Firstly, the direction of plastic flow, $\mathbf{n}_g$ in the $(p', q)$ plane is postulated as:

$$\mathbf{n}_g^T = (n_{gv}, n_{gs}) \tag{18}$$

with

$$n_{gv} = \frac{d_g}{\sqrt{1 + d_g^2}} \tag{19}$$

$$n_{gs} = \frac{1}{\sqrt{1 + d_g^2}} \tag{20}$$

where the dilatancy $d_g$, which is defined as the ratio between the increments of plastic volumetric and shear strain is given by:

$$d_g = \frac{d_0}{M_g} \cdot (M_g \cdot (m\psi) - \eta) \tag{21}$$

where $d_0$ and m are model constants. $\psi$ is the state parameter; $\eta$ is the stress ratio and $M_g$ is the Critical State Line in the plot q – $p'$.

The loading-unloading discriminating relation $\mathbf{n}$ is obtained in a similar way:

$$\mathbf{n}^T = (n_v, n_s) \tag{22}$$

with

$$n_v = \frac{d_f}{\sqrt{1 + d_f^2}} \tag{23}$$

$$n_s = \frac{1}{\sqrt{1 + d_f^2}} \tag{24}$$

where

$$d_f = \frac{d_0}{M_f} \cdot (M_f \cdot (m\psi) - \eta) \tag{25}$$

In above equations, $M_f$ could be material parameter as in the basic PZ model [ZCP$^+$99] or a function of void ratio [MPM11] as:

$$\frac{M_f}{M_g} = h_1 - h_2 \cdot \left(\frac{e}{e_c}\right)^\beta \tag{26}$$

where $\beta$ , $h_1$ and $h_2$ are model constants. For granular materials, the ratio $\frac{e}{e_c}$ varies between $\frac{e_{min}}{e_{max}}$ and $\frac{e_{max}}{e_{min}}$. When $\frac{e}{e_{CS}}$ reaches its lower limit, $M_f/M_f$ ratio is close to one.

The third ingredient is the plastic modulus $H_L$ and $H_b$.

$H_L$ is defined for loading and unloading. During loading, we will assume:

$$H_L = H_0 \cdot \sqrt{p' \cdot p_a} \cdot H_{DM} \cdot H_f \cdot (H_v + H_s) \tag{27}$$

$H_0$ has been assumed to depend on the state parameter. Here we have chosen the law:

$$H_0 = H_0' \cdot \exp\left[-\beta_0' \left(e/e_c\right)^\beta\right] \tag{28}$$

where $H_0'$ and $\beta_0'$ are additional model parameters and

$$H_f = \left(1 - \frac{\eta}{\eta_f}\right)^4 \tag{29}$$

In above equations, $\eta_f$ acts as a limit for the possible states,

$$\eta_f = \left(1 + \frac{1}{\alpha}\right) M_f \tag{30}$$

The volumetric and deviatoric components of the plastic modulus $H_v$ are assumed to be of the form:

$$H_v = H_{v0} \cdot [\eta_p - \eta] \longrightarrow \eta_p = M_g \cdot \exp\left(-\beta_v \cdot \psi\right) \tag{31}$$

where $H_{v0}$ and $\beta_v$ are model parameters. It can be easily verified that $\eta_p < M_g$ for loose states while $\eta_p > M_g$ for dense states, and $H_s$ depends on the accumulated deviatoric strain $\xi_{dev}$

$$H_s = \beta_0 \beta_1 \exp\left(-\beta_0 \xi_{dev}\right) \tag{32}$$

where

$$d\xi_{dev} = \left(de^p : de^p\right)^{1/2} \tag{33}$$

where $de^p$ is the increment of the plastic deviatoric strain tensor.

Finally, the plastic modulus $H_b$ is given by

$$H_b = w\left(\xi\right) \cdot H_0 \cdot \sqrt{p\prime \cdot p_a} \cdot H_{DM} \cdot H_f \tag{34}$$

where $H_{DM}$ keeps track of the maximum stress level reached by the material:

$$H_{DM} = \left(\frac{\zeta_{\max} \cdot J_s}{\zeta}\right)^\gamma \tag{35}$$

where

$$\zeta = p'. \left\{1 - \left(\frac{1+\alpha}{\alpha}\right)\frac{\eta}{M}\right\}^{1/\alpha} \tag{36}$$

and $\gamma$ is a new material constant. $J_s$ is a discrete memory function incorporating the effect of the suction and degree of saturation,

$$J_s = \exp\left(c.g\left(\xi\right)\right) \tag{37}$$

where $c$ is a model parameter and $g(\xi)$ is defined by equation 17 as a function of the bonding parameter $\xi$. To illustrate the role of $J_s$, we will consider the case of a saturated soil consolidated at $p_{c0}$ and then dried, its suction increasing from zero. We have depicted in Figure 4 the variation of $\zeta_{max} \cdot J_s$ with the bonding parameter. It can be interpreted within the framework of classical plasticity as the increase in size of the yield surface with suction. The value of $\zeta_{max}$ at saturated conditions can be obtained by $r_0 \cdot p_{cs}^{SAT}$ where $r_0$ is a material parameter and $p_{cs}^{SAT}$ is the mean effective stress at a critical state. Indeed, this new parameter in equation 35 is equivalent to the OCR used in classical plasticity models.

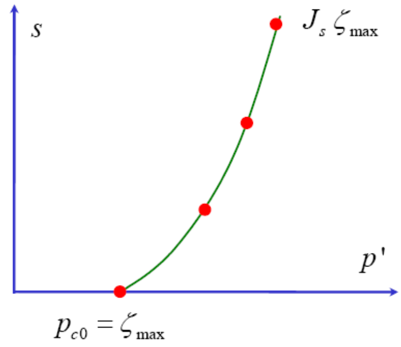


Figure 4: Effect of the bonding parameter on the hardening law.

$w(\xi)$ incorporates the effect of the bonding parameter defined above.

$$w(\xi) = \left\{ \begin{array}{c} -\left\{1 - \exp\left[g(\xi)\right]^2\right\}^2 \\ 1 \end{array} \right\} \tag{38}$$

The model is completed with a suitable hydraulic equation that considers both the hydraulic hysteresis during a drying–wetting cycle and its dependency on past history. We have chosen a modified version of the water retention curve proposed by Fredlund & Xing [FX94]:

$$Sr = Sr_0 + (1 - Sr_0) \cdot \left\{ \ln \left[ \exp(1) + \left( \frac{e^{\Omega} \cdot s}{a_w \cdot p_0} \right)^{n} \right] \right\}^{-m} \tag{39}$$

where $\Omega$, $a_w$, $n$ and $m$ are model parameters, $e$ is the void ratio and $s$ the matrix suction. The main wetting and drying curves are obtained by assuming different values for $a_w$,$n$ and $m$.

Therefore, the non-linear irreversible behaviour of unsaturated soil can be fully characterized within the Generalized Plasticity framework by adding a plastic modulus in wetting and drying paths $H_b$ and a bonding parameter $\xi$ to the State Parameter based model [MMP11]. Coupling with a state-dependent WRC allows the reproduction of the irreversible response in wetting–drying paths and the mechanical effect on the hydraulic behaviour (Figure 5).

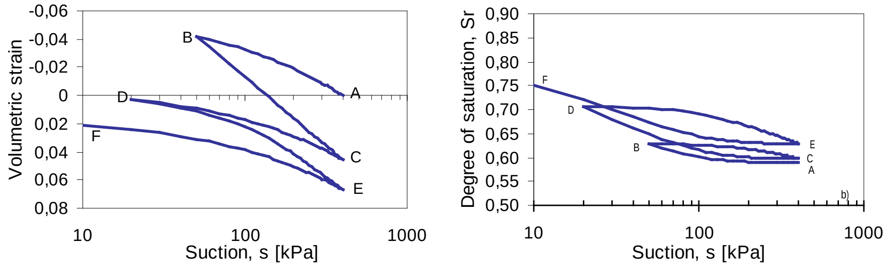


Figure 5: Wetting and drying test.

In order to show the influence of the wetting - drying cycle on the mechanical soil behaviour, we have chosen an experiment performed by Sharma [Sha98]. The test consists of a constant suction isotropic compression loading/unloading cycle (a-b-c), followed by wetting – drying cycle (c-d-e) and a second constant suction isotropic reloading and unloading cycle (e-f-g). Figure 6 provides the model predictions and experimental data on compacted bentonite - kaolin sample in (i) net confining pressure vs void ratio, (ii) degree of saturation vs net stress (iii) degree of saturation vs suction, and (iv) stress path followed during the test. Details about parameter calibration can be found in Manzanal [Man08].The parameters are reported in Manzanal et al. [MPM11].

With the proposed model, it is also possible to reproduce an interesting case: the effect of suction on the undrained behaviour of unsaturated fine-grained soil. We have depicted in Figure 7 constant volume triaxial tests for different suctions. The hardening effect can be observed due to increasing the suction on a loose sample, which in saturated conditions ($s = 0$) arrives to liquefaction. If the soil is sheared at a large initial suction (point A) at constant deviatoric stress, decreasing the suction, the stress path will become unstable and fail catastrophically. This phenomenon has been modelled by [BN11]. It is important to remark that this is just a qualitative example, a complete analysis based on the method proposed by Darve et al. [DL00, DL01] being necessary to understand the process fully.

The Generalized Plasticity unsaturated model was applied for evaluating the mechanical behaviour of a natural volcanic silty soil from steep slopes [CMM$^+$18]. Thus, so-called wetting collapse and static liquefaction may occur during rainfall. Significant issues are posed once the slides turn into flows with high destructive potential. The model is calibrated for 37 saturated/unsaturated laboratory tests. Figure 8 shows the performance of the model in wetting tests performed in Suction-controlled Oe-

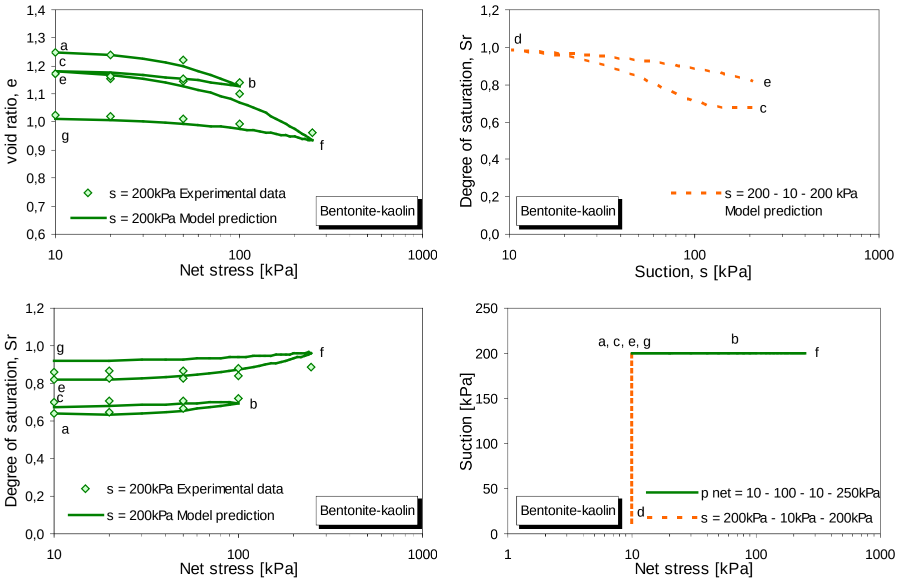


Figure 6: Comparison between model predictions and experimental data of isotropic loading/unloading tests at s = 200kPa with a wetting-drying cycle at pnet = 10kPa on the bentonite-kaolin sample (Experimental data from [Sha98].

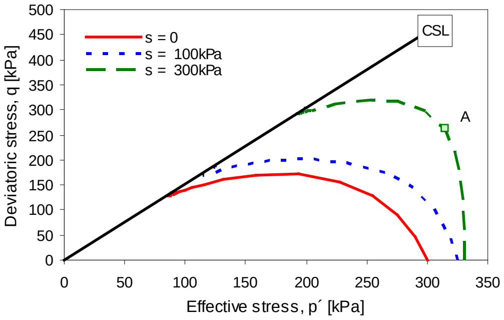


Figure 7: Constant volume triaxial test for different suction.

dometer (ESA), Standard Oedometer (ESL) and Suction-controlled Triaxial (UPS). The parameters are reported in Cuomo et al. [CMM$^+$18].

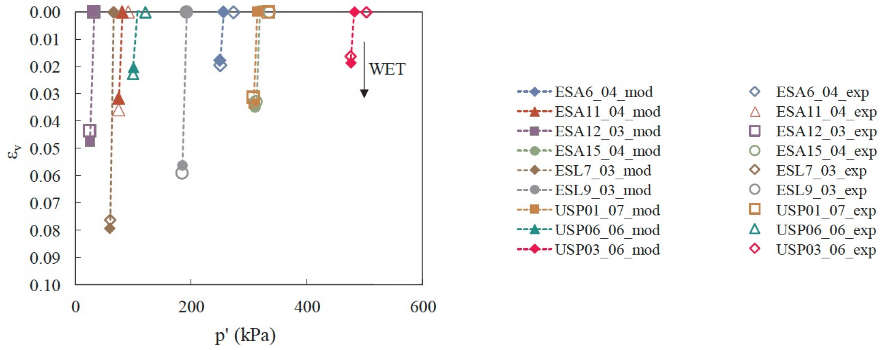


Figure 8: Model Performance in wetting tests performed in Suction-controlled Oedometer (ESA), Standard Oedometer (ESL) and Suction-controlled Triaxial (UPS): experimental versus numerical results in $\varepsilon_v - p'$ plane [CMM$^+$18].

Regarding static liquefaction, the MPZ model has been implemented as a user model of Plaxis to evaluate the vulnerability of the tailing dam to liquefaction [LMS21, LSM22]. Figure 9 shows the stress paths of selected Gauss points for the three actions: a) surface load, b) toe contraction and c) raise in phreatic surface. For points A to F, the resulting paths for both the surface load and the toe contraction are included, whereas, for Point G located at the toe of the dam, only the stress path resulting from the rise in the phreatic surface is shown. The so-called instability line was defined by direct observation of the peak deviatoric stresses for points B, C, A, and F, and it is only included to aid with the interpretation of the stress paths and delineate the boundary between the zone of stable and unstable equilibrium. As the surface load grows, points A, B, C, and E, located in the stable equilibrium zone at the right of the instability line, experience an increase in shear stress until reaching a stress state at the instability line.

# 5 Conclusions

This chapter provides an overview of the hierarchical and versatile formulation of Generalized Plasticity Theory (GPT) developed by Zienkiewicz and Pastor in the middle of the eighties, along with the latest advancements based on state parameter modeling (MPZ model):

(i) Extension of GPT to Unsaturated Behavior: The model integrates the Bishop effective stress and suction as stress variables, while using the strain of the solid skeleton and degree of saturation as deformation variables. It also accounts for the void ratio and hydraulic hysteresis effects on hydraulic behavior.

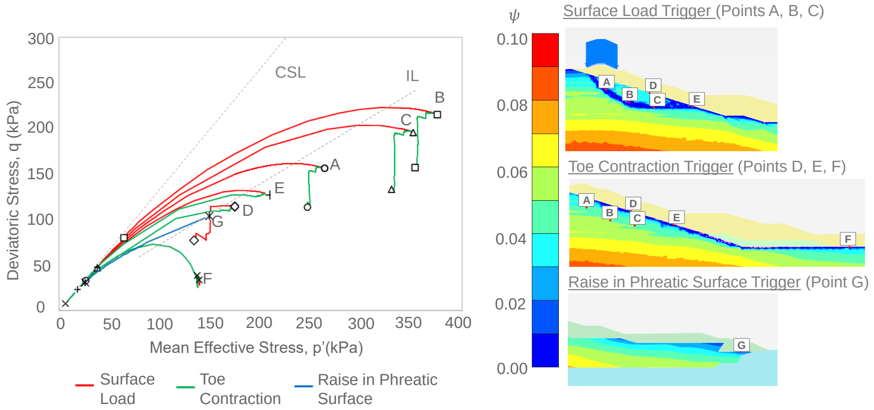


Figure 9: Stress paths for different actions over a tailing dam.[LSM22].

(ii) Generalization of Critical State for Varying Suctions: A novel approach to the critical state concept as a function of a bonding parameter is introduced, enabling the extension of the state parameter concept to partially saturated soils. This allows the model to accurately reproduce the stress-strain behavior of unsaturated soils across different densities, confining pressures, and suctions using unified material constants.

For further details on the extended formulation and calibration, please refer to [Man08, PMM$^+$10, MMP11, MPM11].

### Acknowledgment

The authors acknowledged the European Commission's financial support -H2020 MSCA-RISE 2020 Project DISCO2-STORE, Grant Agreement No 101007851.

## References

[AGJ90] E. Alonso, A. Gens, and A. Josa. A constitutive model for partially saturated soils. *Géotechnique*, 40(3):405–430, 1990.

[BN11] G. Buscarnera and R. Nova. Modelling instabilities in triaxial testing on unsaturated soil specimens. *International Journal for Numerical and Analytical Methods in Geomechanics*, 35(2):179–200, 2011.

[CdP00] B. Cambou and C. di Prisco. *Constitutive Modelling of Geomaterials*. Hermes, 2000.

[CMM$^+$18] S. Cuomo, M. Moscariello, D. Manzanal, M.Pastor, and V. Foresta. Modelling the mechanical behaviour of a natural unsaturated pyroclastic soil within generalized plasticity framework. *Computers and Geotechnics*, 99:191–202, 2018.

[DL00] F. Darve and F. Laouafa. Instabilities in granular materials and application to landslides. *Mechanics of Cohesive-Frictional Materials*, 5:627–652, 2000.

[DL01] F. Darve and F. Laouafa. Modelling of slope failure by a material instability mechanism. *Computer and Geotechnics*, 29:301–325, 2001.

[Fis26] R.A. Fisher. On the capillary forces in an ideal soil; correction of formulas by w.b. haines. *Journal of Agricultural Science*, 16:492–505, 1926.

[FMEMea21] J.A. Fernández-Merodo, P. Ezquerro, D. Manzanal, and et al. Modeling historical subsidence due to groundwater withdrawal in the alto guadalentín aquifer-system (spain). *Engineering Geology*, 283:105998, 2021.

[FMPM$^+$04] J. A. Fernandez Merodo, M. Pastor, P. Mira, L. Tonni, M. I. Herreros, E. Gonzalez, and R. Tamagnini. Modelling of diffuse failure mechanisms of catastrophic landslides. *Computer Methods in Applied Mechanics and Engineering*, 193:2911–2939, 2004.

[FX94] D. Fredlund and A. Xing. Equations for the soil-water characteristic curve. *Canadian Geotechnical Journal*, 31:521–532, 1994.

[GGMP24] M. García-García, D. Manzanal, and M. Pastor. Anisotropy state variable based on phase transformation for generalized plasticity constitutive model. *Acta Geotech.*, 19:899–916, 2024.

[GGSV03] D. Gallipoli, A. Gens, R. Sharma, and J. Vaunat. An elastoplastic model for unsaturated soil incorporating the effects of suction and degree of sturation on mechanical behaviour. *Géotechnique*, 53:123–135, 2003.

[GWK03] D. Gallipoli, S.J. Wheeler, and M. Karstunen. Modelling the variation of degree of saturation in a deformable unsaturated soil. *Géotechnique*, 53:105–112, 2003.

[Hai25] W.B. Haines. Studies in the physical properties of soils. a note on the cohesion developed by capillary forces in an ideal soil. *Journal of Agricultural Science*, 15:529–535, 1925.

[Hou97] G.T. Houlsby. The work input to an unsaturated granular material. *Géotechnique*, 47:193–196, 1997.

[Jom00] C. Jommi. Remarks on the constitutive modelling of unsaturated soils. In Rotterdam: Trento Italy Tarantino, Mancuso Eds. Balkema, editor, *Experimental evidence and theoretical approaches in unsaturated soils.*, pages 139–153. 2000.

[KNM94] Y. Kohgo, M. Nakano, and T. Miyazaki. Theoretical aspects of constitutive modelling for unsaturated soils. *Soil and Foundation*, 33:49–63, 1994.

[LMS21] O. Ledesma, D. Manzanal, and A. Sfriso. Formulation and numerical implementation of a state parameter-based generalized plasticity model for mine tailings. *Computers and Geotechnics*, 135:104158, 2021.

[LSM22] O. Ledesma, A. Sfriso, and D. Manzanal. Procedure for assessing the liquefaction vulnerability of tailings dams. *Computers and Geotechnics*, 144:104632, 2022.

[Man08] D. Manzanal. *Constitutive model based on Generalized Plasticity incorporating state parameter for saturated and unsaturated sand (Spanish)*. PhD thesis, School of Civil Engineering, Technical University of Madrid, 2008.

[MBLQ$^+$21] D. Manzanal, S. Bertelli, S. Lopez-Querol, T. Rossetto, and P. Mira. Influence of fines content on liquefaction from a critical state framework: the christchurch earthquake case study. *Bulletin of Engineering Geology and the Environment*, 80(6):4871–4889, 2021.

[MFMM$^+$18] P. Mira, J.A. Fernández-Merodo, M. Pastorand D. Manzanal, M.M. Stickle, A. Yagüe, and et al. A methodology for the 3d analysis of foundations for marine structures. In *Numerical Methods in Geotechnical Engineering IX, Volume 1*. CRC Press, 2018.

[MMP11] D. Manzanal, J.A. Fernández Merodo, and M. Pastor. Generalized plasticity state parameter-based model for saturated and unsaturated soils. part 1: Saturated state. *International Journal for Numerical and Analytical Methods in Geomechanics*, 35(12):1347–1362, 2011.

[MPM11] D. Manzanal, M. Pastor, and J.A. Fernández Merodo. Generalized plasticity state parameter-based model for saturated and unsaturated soils. part ii: Unsaturated soil modeling. *International Journal for Numerical and Analytical Methods in Geomechanics*, 35(18):1899–1917, 2011.

[PMM$^+$10] M. Pastor, D. Manzanal, J.A. Fernández Merodo, P. Mira, and et al. From solids to fluidized soils: diffuse failure mechanisms in geostructures with applications to fast catastrophic landslides. *Granular Matter*, 12(3):211–228, 2010.

[PZC90] M. Pastor, O.C. Zienkiewicz, and A.H.C. Chan. Generalized plasticity and the modelling of soil behaviour. *International Journal for Numerical and Analytical Methods in Geomechanics*, 14:151–190, 1990.

[PZL85] M. Pastor, O. C. Zienkiewicz, and K. H. Leung. Simple model for transient soil loading in earthquake analysis ii: Non-associative models for sands. *International Journal for Numerical and Analytical Methods in Geomechanics*, 9:477–498, 1985.

[Sch84] B.A. Schrefler. *The finite element method in soil consolidation (with applications to surface subsidence)*. PhD thesis, 1984.

[Sha98] R.S. Sharma. *Mechanical behaviour of unsaturated highly expansive clays*. PhD thesis, 1998.

[Siv93] V. Sivakumar. *A critical state framework for unsaturated soil*. PhD thesis, 1993.

[Tam04] R. Tamagnini. An extended cam-clay model for unsaturated soils with hydraulic hysteresis. *Géotechnique*, 54:223–228, 2004.

[Tol90] D.G. Toll. A framework for unsaturated soil behaviour. *Géotechnique*, 40:31–44, 1990.

[vGM80] van Genuchten MT. A closed-form equation for predicting the hydraulic conductivity of unsaturated soil. *Science Society American Journal*, 44:892–898, 1980.

[VRJ00] J. Vaunat, E. Romero, and C. Jommi. An elastoplastic hydro-mechanical model for unsaturated soils. In Rotterdam: Trento Italy Tarantino, Mancuso Eds. Balkema, editor, *Experimental evidence and theoretical approaches in unsaturated soils.*, pages 121–138. 2000.

[ZCP+99] O.C. Zienkiewicz, A.H. Chan, M Pastor, B. A Schrefler, and T. Shiomi. *Computational Geomechanics*. John Wiley & Sons, 1999.

[ZLP85] O. C. Zienkiewicz, K. H. Leung, and M. Pastor. Simple model for transient soil loading in earthquake analysis. i: Basic model and its application. *International Journal for Numerical and Analytical Methods in Geomechanics*, 9:453–476, 1985.

[ZM84] O. C. Zienkiewicz and Z. Mroz. *Generalized plasticity formulation and applications to geomechanics*. John Wiley & Sons, 1984.

# Coupling equations for variably saturated geomaterials

## A mathematical model for non-isothermal multiphase porous materials: fundamentals and formulation

**Bernhard A. Schrefler, Lorenzo Sanavia**

*Department of Civil Environmental and Architectural Engineering, University of Padua, Italy*

*A mathematical model for a fully saturated and partially saturated non-isothermal porous medium in dynamics is presented. The porous material is treated as a multiphase continuum with the pores of the solid skeleton filled by one or more fluids, e.g. liquid water and gas phase, which may be either water vapour alone or a mixture of dry air and water vapour. The governing equations at macroscopic level are derived in a spatial setting using averaging theories from balance equations developed at microscopic level. Finite kinematics is included in the model. The solid skeleton of the medium can undergo large elastic or inelastic deformations described in the framework of hyperelastoplasticity.*

## 1 Introduction

This paper presents a mathematical model for a non-isothermal variably saturated porous material developed within the Hybrid Mixture Theory in both the non-linear geometric and material settings.

Mechanics of porous materials has a wide spectrum of engineering applications and hence, in recent years, several porous media models and their numerical solutions have appeared in the literature (see [Lew98], [Ocz99], [Boe00] for a comprehensive state of the art). Most of these models are restricted to fluid saturated materials and have been developed using small strain assumptions. Conditions of partial saturation are of importance in engineering practice because many porous materials are in this natural state or can reach this state during deformations. Some simple examples can be found in soils or in concrete and in biological tissues, which can contain air or other gases in the pores together with liquids. For soils, this is the case of the zones above the free surface, or of deep reservoirs of hydrocarbon gas. The partially saturated state can also be reached during the deformation due, for instance, to earthquake in an earth dam or during the particular case of strain localization of dense sands under globally undrained conditions, where negative water pressures are measured and cavitation of the pore water was observed [Var95], [Mok98]. Large strains also can be important. They result when ultimate or serviceability limit state is reached, as for example during slope instability or during the consolidation process in compressible clays. In laboratory, this can be the case of drained or undrained biaxial tests of sands, where axial logarithmic strains of the order of $0.12 - 0.15$ are reached [Var95], [Mok98], or the case of triaxial tests of peats, where axial strains of the order of $0.15$ are measured.

In the model developed in this chapter, the porous medium is treated as an non-isothermal multiphase continuum with the pores of the solid skeleton filled by water and air. The governing equations at macroscopic level are derived in Section 2 in a spatial setting in dynamics and are based on averaging procedures (Hybrid Mixture Theory). This model follows from [Lew98], where the interested reader can find further details and remarks. Solid displacements, temperature, water and gas pressures are the primary variables. Water and gas are assumed to obey Darcy's law. In the partially saturated state, the degree of saturation and the relative permeability of water and gas are dependent on the capillary pressure by experimental functions. Within the formulation developed, the elasto-plastic behaviour of the solid skeleton can be described by the multiplicative decomposition of the deformation gradient into an elastic and a plastic part, as shown in the next chapter of this volume and, with more details, in [San02]. In this case, the modified effective stress in partially saturated conditions (Bishop like stress) in the form of Kirchhoff measure of the stress tensor and the logarithmic principal strains are used in conjunction with an hyperelastic free energy function. The effective stress state can be limited by a suitable yield surface.

As notation and symbols are concerned, bold-face letters denote tensors; capital or lower case letters are used for tensors in the reference or in actual configuration. The symbol '$\cdot$' denotes the scalar product between two vectors (e. g. $\mathbf{a} \cdot \mathbf{b} = a_i\, b_i$), while the symbol ':' denotes a double contraction of (adjacent) indices of two tensors of rank two or/and higher (e. g. $\mathbf{c} : \mathbf{d} = c_{ij}\, d_{ij}, \mathbf{e} : \mathbf{f} = e_{ijkl}\, f_{kl}$). Cartesian co-ordinates are used throughout the paper.

## 2 Mathematical model of thermo-hydro-mechanical behaviour of geomaterials

The full mathematical model necessary to simulate thermo-hydro-mechanical transient behaviour of fully and partially saturated porous media is developed in [Lew98] using averaging theories following Hassanizadeh and Gray [Has79a], [Has79b], [Has80]. The underlying physical model, thermodynamic relations and constitutive equations for the constituents, as well as governing equations are summarized in the following.

The partially saturated porous medium is treated as multiphase system composed of

$\pi = 1, \dots, k$ ($k = 3$) constituents with the voids of the solid skeleton ($s$) filled with water ($w$) and gas ($g$) (see Figure 1). The latter is assumed to behave as an ideal mixture of two species: dry air (non-condensable gas, $ga$) and water vapour (condensable one, $gw$).

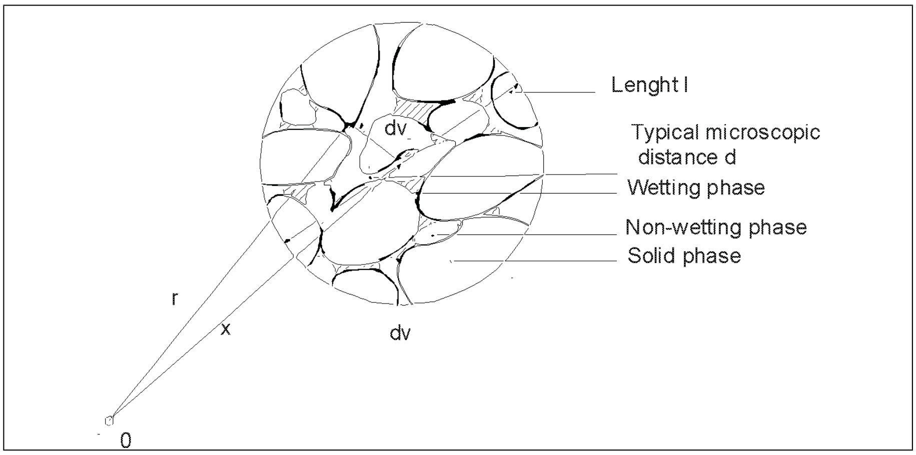


Figure 1: Averaging volume $\mathrm{d}v(\mathbf{x},\, t)$ of a three phase porous medium

## 2.1 Microscopic balance equations

At the microscopic level, the balance equation of any $\pi$-phase may be described by the classical continuum mechanics. At the interfaces with other constituents, the material properties and thermodynamic quantities may present step discontinuities (see e. g. [Gra01] or [Sch02] for the jump conditions to be fulfilled). For a thermodynamic property $\psi$, the balance equation within the $\pi$-phase may be written as [Has79a] and [Has79b]

$$\frac{\partial \rho \psi}{\partial t} + \operatorname{div}(\rho\, \psi \dot{\mathbf{r}}) - \operatorname{div} \mathbf{i} - \rho \mathbf{b} = \rho \mathbf{G} \tag{1}$$

where $\dot{\mathbf{r}}$ is the local value of the velocity field of the $\pi$-phase in a fixed point in space, $\mathbf{i}$ is the flux vector associated with $\psi$, $\mathbf{b}$ the external supply of $\psi$ and $\mathbf{G}$ is the net production of $\psi$. Fluxes are positive as outflows. The thermodynamic property $\psi$ for the different balance equations and the values assumed by the quantities of (1) are listed in Table 1 following [Has79b],

where $E$ is the specific intrinsic energy, $\lambda$ the specific entropy, $\mathbf{t}_m$ the microscopic stress tensor, $\mathbf{q}$ a heat flux vector, $\Phi$ the entropy flux, $\mathbf{g}$ the external momentum supply related to gravitational effects, $h$ the intrinsic heat source, $S$ an intrinsic entropy source and $\varphi$ denotes an increase of entropy. The angular momentum balance equation has

Table 1: Thermodynamic properties for the microscopic mass balance equations

| Quantity | $\psi$ | $\mathbf{i}$ | $\mathbf{b}$ | $\mathbf{G}$ |
|---|---|---|---|---|
| Mass | $1$ | $0$ | $0$ | $0$ |
| Momentum | $\dot{\mathbf{r}}$ | $\mathbf{t}_m$ | $\mathbf{g}$ | $0$ |
| Energy | $E + 0.5\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}$ | $\mathbf{t}_m\dot{\mathbf{r}} - \mathbf{q}$ | $\mathbf{g} \cdot \dot{\mathbf{r}} + h$ | $0$ |
| Entropy | $\lambda$ | $\Phi$ | $S$ | $\varphi$ |

been omitted because the constituents are assumed to be microscopically non polar (the interested reader is referred to [Ehl98] regarding a saturated or empty porous media model with a polar solid skeleton).

Using spatial averaging operators [Has79a] defined over a representative elementary volume R. E. V. (of volume $\mathrm{d}v(\mathbf{x},\, t)$ in the deformed configuration $B_t \subset \mathbb{R}^3$, see Figure 1, where $\mathbf{x}$ is the vector of the spatial co-ordinates and $t$ is the current time), the microscopic equations are integrated over the R. E. V. giving the macroscopic balance equations.
As a conseguence, at the macroscopic level the multiphase porous material results modelled by a substitute continuum of volume $B_t$ with boundary $\partial B_t$ that fills the entire domain simultaneously, instead of the real fluids and the solid which fill only a part of it. In these substitute continuum each constituent $\pi$ has a reduced density which is obtained through the volume fraction $\eta^\pi(\mathbf{x},\, t) = \mathrm{d}v^\pi(\mathbf{x},\, t)/\mathrm{d}v(\mathbf{x}, t)$ with the constraint

$$\sum_{\pi=1}^{k} \eta^\pi = 1 \tag{2}$$

where $\mathrm{d}v^\pi(\mathbf{x},\, t)$ is the $\pi$-phase volume inside the R. E. V. in the actual placement $\mathbf{x}$.

In the present formulation heat conduction, vapour diffusion, heat convection, water flow due to pressure gradients or capillary effects and latent heat transfer due to water phase change (evaporation and condensation) inside the pores are taken into account. The solid is deformable and non-polar, and the fluid, the solid, and the thermal fields are coupled. All fluids are in contact with the solid phase. The constituents are assumed to be isotropic, homogeneous, immiscible except for dry air and vapour, and chemically non-reacting. Local thermal equilibrium between solid matrix, gas, and liquid phases is assumed so that the temperature is the same for all the constituents. The state of the medium is described by water pressure $p^w$, gas pressure $p^g$, temperature $\theta$, and the displacement vector of the solid matrix $\mathbf{u}$.

Before summarizing the macroscopic balance equations, we specify the kinematics introducing the notion of initial and current configuration (Figure 2). In the following, the stress is defined as tension positive for the solid phase, while pore pressure is defined as compressive positive for the fluids.

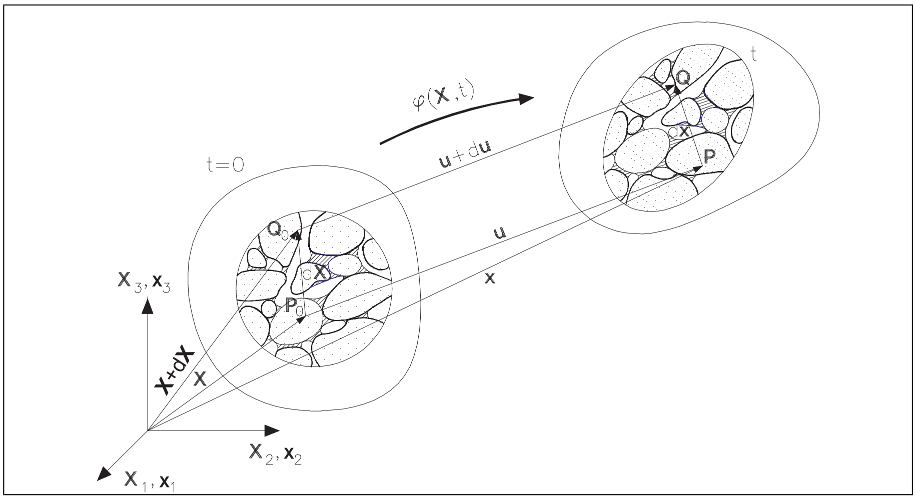


Figure 2: Initial and current configuration of a multiphase medium

## 2.2 Kinematic equations

At the macroscopic level the multiphase medium is described as the superposition of all $\pi$-phases, whose material points $X^\pi$ with co-ordinates $\mathbf{X}^\pi$ in the reference configuration $B_0^\pi \subset \mathbb{R}^3$ at time $t = t_0$ can occupy simultaneously each spatial point $\mathbf{x}$ in the deformed configuration $B_t \subset \mathbb{R}^3$ at time $t$. In the *Lagrange*an description of the motion in terms of material co-ordinates the position of each material point in the actual configuration $\mathbf{x}$ is a function of its placement $\mathbf{X}^\pi$ in a chosen reference configuration $B_0^\pi$ and of the current time $t$:

$$\mathbf{x} = \boldsymbol{\chi}^\pi(\mathbf{X}^\pi, t) \tag{3}$$

with $\mathbf{x} = \mathbf{x}^\pi$, or it is given by the sum of the reference position $\mathbf{X}^\pi$ and the displacement $\mathbf{u}^\pi = (\mathbf{X}^\pi, t)$ at time $t$

$$\mathbf{x} = \mathbf{X}^\pi + \mathbf{u}^\pi(\mathbf{X}^\pi, t) \tag{4}$$

In (3), $\boldsymbol{\chi}^\pi(\mathbf{X}^\pi, t)$ is a continuous and bijective motion function (deformation map) of each phase because the Jacobian $J^\pi$ of each motion function

$$J^\pi = \det \frac{\partial \boldsymbol{\chi}^\pi(\mathbf{X}^\pi, t)}{\partial \mathbf{X}^\pi} > 0 \tag{5}$$

is restricted to be a positive value. The deformation gradient $\mathbf{F}^\pi(\mathbf{X}^\pi,\, t)$ is defined as

$$\mathbf{F}^\pi \,=\, \mathrm{Grad}^\pi \boldsymbol{\chi}^\pi(\mathbf{X}^\pi,\, t) \tag{6}$$

where the differential operator ' $\mathrm{Grad}^\pi$' denotes partial differentiation with respect to the reference position $\mathbf{X}^\pi$. Hence, from (5), $J^\pi \,=\, \det \mathbf{F}^\pi$.

The velocity and the acceleration of each constituent are given as

$$\mathbf{V}^\pi = \frac{\partial \boldsymbol{\chi}^\pi(\mathbf{X}^\pi,\, t)}{\partial t}\,, \qquad \mathbf{A}^\pi = \frac{\partial^2 \boldsymbol{\chi}^\pi(\mathbf{X}^\pi,\, t)}{\partial t^2} \tag{7}$$

Due to the non-singularity of the Lagrangean relationship (3), the existence of its inverse function leads to the description of the motion in terms of spatial co-ordinates:

$$\mathbf{X}^\pi = (\boldsymbol{\chi}^\pi)^{-1}(\mathbf{x},\, t) \tag{8}$$

The inverse $(\mathbf{F}^\pi)^{-1}(\mathbf{x},\, t)$ of the deformation gradient is given by

$$(\mathbf{F}^\pi)^{-1} \,=\, \mathrm{grad}\, \mathbf{X}^\pi(\mathbf{x},\, t) \tag{9}$$

where the differential operator 'grad' is now referred to spatial co-ordinates $\mathbf{x}$. The spatial parametrization of the velocity is given by

$$\mathbf{v}^\pi \,= \mathbf{v}^\pi(\mathbf{x},\, t) = \mathbf{V}^\pi \circ (\boldsymbol{\chi}^\pi)^{-1} \tag{10}$$

where '$\circ$' denotes the composition of functions. The parametrization of the spatial acceleration is related to the spatial velocity by the application of the chain rule to (10):

$$\mathbf{a}^\pi \,= \mathbf{a}^\pi(\mathbf{x},\, t) = \frac{\partial \mathbf{v}^\pi}{\partial t} + (\mathrm{grad}\, \mathbf{v}^\pi)\, \mathbf{v}^\pi \,= \mathbf{A}^\pi \circ (\boldsymbol{\chi}^\pi)^{-1} \tag{11}$$

Since the individual constituents follow in general different motions, different material time derivatives must be formulated. For an arbitrary scalar-valued function $f^\pi(\mathbf{x},\, t)$, its material time derivative following the velocity of the constituents $\pi$ is defined by [Lew98]

$$\frac{\mathrm{D}^\pi f^\pi}{\mathrm{D}t} = \frac{\partial f^\pi}{\partial t} + \,\mathrm{grad}\, f^\pi \cdot \mathbf{v}^\pi \tag{12}$$

where $f^\pi(\mathbf{x},\, t)$ must be substituted by $\mathbf{f}^\pi(\mathbf{x},\, t)$ in case of vector or tensor valued function $\mathbf{f}^\pi(\mathbf{x},\, t)$. Thus, $\mathbf{a}^\pi = \mathrm{D}^\pi \mathbf{v}^\pi/\mathrm{D}t$.

In the theory of multiphase materials it is common to assume the motion of the solid as a reference and to describe the fluids in terms of motion relative to the solid. This means that a fluid relative velocity and the material time derivative with respect to the solid are introduced. The solid motion can be described both in terms of material or spatial co-ordinates. The second approach is now presented because the most natural numerical formulation of the elasto-plastic initial-boundary-value problem is based on the weak form of the balance equations in the spatial setting.

The fluid relative velocity $\mathbf{v}^{\pi s}(\mathbf{x},\, t)$ in spatial parametrization or diffusion velocity is given by

$$\mathbf{v}^{\pi s}(\mathbf{x},\, t) = \mathbf{v}^{\pi}(\mathbf{x},\, t) - \mathbf{v}^{s}(\mathbf{x},\, t) \tag{13}$$

and the material time derivative of $f^\pi(\mathbf{x},\, t)$ with respect to the moving solid phase $(s)$ is given by

$$\frac{\mathrm{D}^s f^\pi}{\mathrm{D}t} = \frac{\mathrm{D}^\pi f^\pi}{\mathrm{D}t} + \operatorname{grad} f^\pi \cdot \mathbf{v}^{s\pi} \quad \text{with} \quad \mathbf{v}^{s\pi} = -\mathbf{v}^{\pi s} \tag{14}$$

For the section closure, the spatial velocity gradient $\mathbf{l}^s(\mathbf{x},\, t)$ of the solid will be recalled, which is defined as the gradient of the velocity (10) with respect to spatial co-ordinates, i. e.

$$\mathbf{l}^s = \operatorname{grad} \mathbf{v}^s = \frac{\partial \mathbf{F}^s}{\partial t}\,(\mathbf{F}^s)^{-1} = \mathbf{d}^s + \mathbf{w}^s \tag{15}$$

where $\mathbf{d}^s(\mathbf{x},\, t)$ and $\mathbf{w}^s(\mathbf{x},\, t)$ are the symmetric and the skew-symmetric part of $\mathbf{l}^s(\mathbf{x},\, t)$, also called spatial rate of deformation tensor and spin tensor, respectively.

All strain measures and strain rates for each constituent follow similarly to classical non-linear continuum mechanics, but are not reported here because they are not useful for the approach developed in the sequel (see e. g. [Mar83]).

## 2.3 Mass balance equations

The averaged macroscopic balance equation for the solid phase is

$$\frac{\mathrm{D}^s \rho_s}{\mathrm{D}t} + \rho_s \operatorname{div} \mathbf{v}^s = \frac{\partial \rho_s}{\partial t} + \operatorname{div}(\rho_s\, \mathbf{v}^s) = 0 \tag{16}$$

where $\mathbf{v}^s(\mathbf{x},\, t)$ is the mass averaged solid velocity, $\rho_s(\mathbf{x}, t)$ is the averaged density of the solid related to the intrinsic averaged density $\rho^s(\mathbf{x}, t)$ by the volume fraction $\eta^s(\mathbf{x},\, t)$.

For the generic $\pi$-phase the relationship between the averaged density and the intrinsic averaged density is

$$\rho_\pi(\mathbf{x},\, t) = \eta^\pi(\mathbf{x},\, t)\, \rho^\pi(\mathbf{x},\, t) \tag{17}$$

where the intrinsic density $\rho^\pi(\mathbf{x},\, t)$ is also named real or true density in the so-called Theory of Porous Media, e. g. [Boe00].

The mass balance equation for the liquid water is

$$\frac{\mathrm{D}^w \rho_w}{\mathrm{D}t} + \rho_w \operatorname{div} \mathbf{v}^w = \frac{\partial}{\partial t}(n\, S_w\, \rho^w) + \operatorname{div}(n\, S_w\, \rho^w\, \mathbf{v}^w) = \rho_w\, e^w \tag{18}$$

where $\rho_w\, e^w(\mathbf{x},\, t)$ is the quantity of liquid water per unit time and volume lost through evaporation. The corresponding equations for dry air and water vapour are, respectively,

$$\frac{\mathrm{D}^{ga} \rho_{ga}}{\mathrm{D}t} + \rho_{ga} \operatorname{div} \mathbf{v}^{ga} = \frac{\partial}{\partial t}(n\, S_g\, \rho^{ga}) + \operatorname{div}(n\, S_g\, \rho^{ga}\, \mathbf{v}^{ga}) = 0 \tag{19}$$

$$\frac{\mathrm{D}^{gw} \rho_{gw}}{\mathrm{D}t} + \rho_{gw} \operatorname{div} \mathbf{v}^{gw} = \frac{\partial}{\partial t}(n\, S_g\, \rho^{gw}) + \operatorname{div}(n\, S_g\, \rho^{gw}\, \mathbf{v}^{gw}) = \rho_{gw}\, e^{gw} \tag{20}$$

where $n(\mathbf{x},\, t)$ is the porosity of the medium, defined as

$$n = \frac{\mathrm{d}v^w + \mathrm{d}v^g}{\mathrm{d}v} = \frac{\mathrm{d}v^{\text{voids}}}{\mathrm{d}v} = 1 - \eta^s \tag{21}$$

and $S_w$ and $S_g$ are the water and gas degrees of saturation. The following relationships hold:

$$\begin{aligned} \eta^w &= n\, S_w \qquad \text{with} \qquad S_w = \frac{\mathrm{d}v^w}{\mathrm{d}v^w + \mathrm{d}v^g}\,, \\ \eta^g &= n\, S_g \qquad \text{with} \qquad S_g = \frac{\mathrm{d}v^g}{\mathrm{d}v^w + \mathrm{d}v^g} \end{aligned} \tag{22}$$

with the saturation constraint $S_w + S_g = 1$. The right-hand side of (18) and (20) sum to zero.

## 2.4 Linear momentum balance equations

The linear momentum balance equations for the solid and the $\pi$-fluid are

$$\operatorname{div} \mathbf{t}^s + \rho_s (\mathbf{g} - \mathbf{a}^s) + \rho_s \hat{\mathbf{t}}^s = \mathbf{0} \tag{23}$$

and

$$\operatorname{div} \mathbf{t}^\pi + \rho_\pi (\mathbf{g} - \mathbf{a}^\pi) + \rho_\pi (\mathbf{e}^\pi + \hat{\mathbf{t}}^\pi) = \mathbf{0} \tag{24}$$

respectively, where $\mathbf{t}^\pi(\mathbf{x},\, t)$ is the partial *Cauchy* stress tensor defined via the constitutive equation presented in Section 2.7. $\hat{\mathbf{t}}^\pi(\mathbf{x},\, t)$ accounts for the exchange of momentum due to mechanical interaction with other phases, $\rho_\pi\, \mathbf{a}^\pi$ for the volume density of the inertial force, $\rho_\pi\, \mathbf{g}$ for the volume density of gravitational force, and $\mathbf{e}^\pi(\mathbf{x},\, t)$ takes into account the momentum exchange due to averaged mass supply or mass exchange between the fluid and the gas phases and the change of density. The linear momentum balance equations of the multiphase medium are subjected to the constraint [Lew98]:

$$\sum_{\pi=1}^{k} \rho_\pi (\mathbf{e}^\pi + \hat{\mathbf{t}}^\pi) = \mathbf{0} \tag{25}$$

## 2.5 Angular momentum balance equation

All the phases are considered microscopically non-polar and hence at macroscopic level the angular momentum balance equation states that the partial stress tensor is symmetric [Lew98]:

$$\mathbf{t}^\pi = (\mathbf{t}^\pi)^T \tag{26}$$

and that the sum of the coupling vectors of angular momentum between the phases vanishes.

## 2.6 Energy balance equation and entropy inequality

The energy balance equation for the $\pi$-phase may be written as [Lew98]

$$\rho_\pi \frac{\mathrm{D}^\pi E^\pi}{\mathrm{D}t} = \mathbf{t}^\pi : \mathbf{d}^\pi + \rho_\pi\, h^\pi - \operatorname{div} \mathbf{q}^\pi + \rho_\pi\, R^\pi \tag{27}$$

where $\rho_\pi\, R^\pi$ represents the exchange of energy between the $\pi$-phase and other phases of the medium due to phase change and mechanical interaction, $\mathbf{q}^\pi$ is the internal heat flux, $h^\pi$ results from the heat sources, and $\mathbf{d}^\pi$ is the spatial rate of the deformation tensor. $E^\pi$ accounts for the specific internal energy of the volume element.

The entropy inequality for the mixture, useful for the development of the constitutive equations [Sch02] is

$$\sum_\pi \left( \rho_\pi \frac{\mathrm{D}^\pi \lambda^\pi}{\mathrm{D}t} + \rho_\pi\, e^\pi\, \lambda^\pi + \operatorname{div} \frac{\mathbf{q}^\pi}{\theta^\pi} - \frac{\rho_\pi\, h^\pi}{\theta^\pi} \right) \geq 0 \tag{28}$$

where $\theta^\pi$ is the absolute temperature, $\lambda^\pi$ is the specific entropy of the constituent $\pi$, and $e^\pi\, \lambda^\pi$ the entropy supply due to mass exchange.

## 2.7 Constitutive equations

The momentum exchange term $\rho_\pi\, \hat{\mathbf{t}}^\pi$ of the linear momentum balance equations of the fluids can be expressed as [Lew98]

$$\rho_\pi\, \hat{\mathbf{t}}^\pi = -\mu^\pi\, (\eta^\pi)^2\, \mathbf{k}^{-1}\, \mathbf{v}^{\pi s} + p^\pi\, \operatorname{grad} \eta^\pi \qquad \text{with} \qquad \pi = g,\, w \tag{29}$$

Here, $\mathbf{k} = k^{r\pi}\, \mathbf{k}^\pi$, where $\mathbf{k}^\pi(\mathbf{x},\, t) = \mathbf{k}^\pi(\rho^\pi,\, \eta^\pi,\, T)$ is the intrinsic permeability tensor of dimension $[\mathrm{L}^2]$ depending, in the isotropic case, on the porosity of the medium; $k^{r\pi}(S_\pi)$ is the relative permeability parameter and $\mu^\pi$ is the dynamic viscosity. The relative permeability is a function of the $\pi$-phase degree of saturation $S_\pi$ and is determined in laboratory tests (see e. g. [Lew98], [Ocz99] and [Cha22]).

The partial stress tensor in the fluid phase of the linear momentum balance equations (24) is related to the macroscopic pressure $p^\pi(\mathbf{x},\, t)$ of the $\pi$-phase

$$\mathbf{t}^\pi = -\eta^\pi\, p^\pi\, \mathbf{1} \tag{30}$$

where $\mathbf{1}$ is the second order unit tensor.

From the entropy inequality it can also be shown that the spatial solid stress tensor $\mathbf{t}^s(\mathbf{x},\, t)$ of the linear momentum balance equations (23) is decomposed as follows:

$$\mathbf{t}^s = \eta^s\, (\mathbf{t}^s_e - p^s\, \mathbf{1}) \tag{31}$$

and that the effective *Cauchy* stress tensor $\boldsymbol{\sigma}'(\mathbf{x},\, t)$, which is responsible for all major deformation in the solid skeleton, is

$$\boldsymbol{\sigma}' = \eta^s \, \mathbf{t}_e^s \tag{32}$$

In (31), $\mathbf{t}_e^s(\mathbf{x},\, t)$ is the dissipative part [Gra91], [Gra01] or effective stress tensor of the solid phase, while $p^s(\mathbf{x},\, t)$ is the equilibrium part, also called solid pressure, with $p^s = S_w \, p^w + S_g \, p^g$.

From the previous equations, it follows that the total *Cauchy* stress tensor $\boldsymbol{\sigma} = \mathbf{t}^s + \mathbf{t}^w + \mathbf{t}^g$ can be written in the usual form used in soil mechanics

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}' - (S_w \, p^w + S_g \, p^g) \; \mathbf{1} \tag{33}$$

or in terms of $p^g(\mathbf{x},\, t)$ and the capillary pressure, $p^c(\mathbf{x},\, t)$, using equation (37) and the saturation condition $S_w + S_g = 1$.

The elasto-plastic behaviour of the solid skeleton at finite strain can be based on the multiplicative decomposition of the deformation gradient $\mathbf{F}^s(\mathbf{X}^s,\, t)$ into an elastic and plastic part originally proposed by Lee [Lee69] for crystals (see e. g. [San02] for details concerning also the numerical algorithm):

$$\mathbf{F}^s = \mathbf{F}^{se} \, \mathbf{F}^{sp} \tag{34}$$

This decomposition states the existence of an intermediate stress free configuration and its validity has been suggested for cohesive-frictional soils by Nemat-Nasser [Nem83], where the plastic part of the deformation gradient is viewed as an internal variable related to the amount of slipping, crushing, yielding, and, for plate like particles, plastic bending of the granules comprizing the soil.

The pressure $p^g(\mathbf{x},\, t)$ is given in the sequel. For a gaseous mixture of dry air and water vapour, the ideal gas law is introduced because the moist air is assumed to be a perfect mixture of two ideal gases. The equation of state of perfect gas (the *Clapeyron* equation) and *Dalton*'s law applied to dry air $(ga)$, water vapour $(gw)$ and moist air $(g)$, yield:

$$p^{ga} = \rho^{ga} \, R\,\theta / M_a \quad , \qquad p^{gw} = \rho^{gw} \, R\,\theta / M_w \tag{35}$$

$$p^{g} = p^{ga} + p^{gw} \quad , \qquad \rho^{g} = \rho^{ga} + \rho^{gw} \tag{36}$$

In the partially saturated zones, water is separated from its vapour by a concave meniscus (capillary water). Due to the curvature of this meniscus, the sorption equilibrium equation gives, at equilibrium, the capillary pressure equation, i.e. the relationship between the capillary pressure $p^c(\mathbf{x},\, t)$ and the gas $p^g(\mathbf{x},\, t)$ and water pressure $p^w(\mathbf{x},\, t)$ [Gra91]:

$$p^c = p^g - p^w \tag{37}$$

The equilibrium water vapour pressure $p^{gw}(\mathbf{x},\, t)$ can be obtained from the *Kelvin-Laplace* equation:

$$p^{gw} = p^{gws}(\theta)\, \exp\left(\frac{p^c\, M_w}{\rho^w\, R\, \theta}\right) \tag{38}$$

where the water vapour saturation pressure $p^{gws}$, depending only upon the temperature $\theta(\mathbf{x},\, t)$, can be calculated from the *Clausius-Clapeyron* equation or from an empirical correlation.

The saturation $S_\pi(\mathbf{x},\, t)$ is an experimentally determined function of the capillary pressure $p^c$ and the temperature $\theta$:

$$S_\pi = S_\pi(p^c,\, \theta) \tag{39}$$

The equation of state for the liquid water can be written as

$$\rho^w = \rho^{w0}\, \exp[-\beta_w \theta + C_w(p^w - p^{w0})] \tag{40}$$

where the superscript zero indicates an initial steady state at standard conditions; $\beta_w$ is the thermal expansion coefficient and $C_w$ the compressibility coefficient. This equation can be simplified by retaining the first order terms of the series expansion obtaining

$$\rho^w = \rho^{w0}[1 - \beta_w \theta + C_w(p^w - p^{w0})] \tag{41}$$

and

$$\frac{1}{\rho^{w0}}\, \frac{\mathrm{D}^w \rho^w}{\mathrm{D}t} = \frac{1}{K_w}\, \frac{\mathrm{D}^w p^w}{\mathrm{D}t} - \beta_w\, \frac{\mathrm{D}^w \theta}{\mathrm{D}t} \tag{42}$$

where $K_w = C_w^{-1}$ is the bulk modulus of the liquid water.

For the binary gas mixture of dry air and water vapour, *Fick*'s law gives the following relative velocities $\mathbf{v}_g^\pi = \mathbf{v}^\pi - \mathbf{v}^g$ ($\pi = ga,\, gw$) of the diffusing species:

$$\mathbf{v}_g^{ga} = -\frac{M_a\,M_w}{M_g^2}\,\mathbf{D}_g \text{ grad}\left(\frac{p^{ga}}{p^g}\right) = -\mathbf{v}_g^{gw} \tag{43}$$

where $\mathbf{D}_g$ is the effective diffusivity tensor and $M_g$ is the molar mass of the gas mixture:

$$\frac{1}{M_g} = \frac{\rho^{gw}}{\rho^g}\,\frac{1}{M_w} + \frac{\rho^{ga}}{\rho^g}\,\frac{1}{M_a} \tag{44}$$

## 2.8 Initial and boundary conditions

For model closure it is necessary to define the initial and boundary conditions. The initial conditions specify the full fields of gas pressure, water pressure, temperature, displacements, and velocity:

$$p^g = p_0^g\,, \quad p^w = p_0^w\,, \quad \theta = \theta_0\,, \quad \mathbf{u} = \mathbf{u}_0\,, \quad \dot{\mathbf{u}} = \dot{\mathbf{u}}_0 \quad \text{at} \quad t = t_0 \tag{45}$$

The boundary conditions can be imposed values on $\partial B_\pi$ or fluxes on $\partial B_\pi^q$, where the boundary is $\partial B = \partial B_\pi \cup \partial B_\pi^q$ . The imposed values on the boundary for gas pressure, water pressure, temperature, and displacements are as follows:

$$\begin{array}{rclcl rclcl}
p^g & = & \hat{p}^g & \text{on} & \partial B_g\,, & p^w & = & \hat{p}^w & \text{on} & \partial B_w\,, \\
\theta & = & \hat{\theta} & \text{on} & \partial B_\theta\,, & \mathbf{u} & = & \hat{\mathbf{u}} & \text{on} & \partial B_u \quad \text{for} \quad t \geq t_0
\end{array} \tag{46}$$

The (volume average) flux boundary conditions for dry air and water species conservation equations and the energy equation to be imposed at the interface between the porous media and the surrounding fluid (the natural boundary conditions) are the following:

$$\begin{array}{rcll}
\left(\rho^{ga}\,\mathbf{v}^g - \rho^g\,\mathbf{v}_g^{gw}\right)\cdot\mathbf{n} & = & q^{ga} & \text{on } \partial B_g^q \\
\left(\rho^{gw}\,\mathbf{v}^g + \rho^w\,\mathbf{v}^w + \rho^g\,\mathbf{v}_g^{gw}\right)\cdot\mathbf{n} & = & \beta_c\,(\rho^{gw} - \rho_\infty^{gw}) + q^{gw} + q^w & \text{on } \partial B_c^q \\
-\left(\rho^w\,\mathbf{v}^w\,\Delta h_{\text{vap}} - \lambda_{\text{eff}}\,\nabla\,\theta\right)\cdot\mathbf{n} & = & \alpha_c\,(\theta - \theta_\infty) + q^\theta & \text{on } \partial B_\theta^q
\end{array} \tag{47}$$

for $t \geq t_0$, where $\mathbf{n}(\mathbf{x},\, t)$ is the vector perpendicular to the surface of the porous medium, pointing towards the surrounding gas, $\rho_\infty^{gw}(\mathbf{x},\, t)$ and $\theta_\infty(\mathbf{x},\, t)$ are, respectively, the mass concentration of water vapour and temperature in the undisturbed gas phase distant from the interface, $\alpha_c(\mathbf{x},\, t)$ and $\beta_c(\mathbf{x},\, t)$ are convective heat and mass transfer coefficients, while $q^{ga}(\mathbf{x},\, t)$, $q^{gw}(\mathbf{x},\, t)$, $q^w(\mathbf{x},\, t)$, and $q^\theta(\mathbf{x},\, t)$ are the imposed dry air flux, imposed vapour flux, imposed liquid flux, and imposed heat flux, respectively.

The traction boundary conditions for the displacement field related to the total *Cauchy* stress tensor $\boldsymbol{\sigma}(\mathbf{x},\, t)$ are

$$\boldsymbol{\sigma}\,\mathbf{n} = \bar{\mathbf{t}} \quad \text{on} \quad \partial B_u^q \tag{48}$$

where $\bar{\mathbf{t}}(\mathbf{x},\, t)$ is the imposed *Cauchy* traction vector.

# 3 General field equations

The macroscopic balance laws are now transformed and the constitutive equations introduced to obtain the general field equations, which are suitable to be implemented in a finite element code. These equations are now summarized for sake of completeness. The interested reader may refer to [Lew98] for the step by step derivation of the equations of this Section.

- Mass balance equation of the solid:

$$\frac{(1-n)}{\rho^s}\,\frac{\mathrm{D}^s\rho^s}{\mathrm{D}t} - \frac{\mathrm{D}^s n}{\mathrm{D}t} + (1-n)\,\mathrm{div}\,\mathbf{v}^s = 0 \tag{49}$$

- Mass balance equation of the water species (liquid water and water vapour):

$$\begin{aligned}
&\left[\rho^w\left(\frac{1-n}{K_s}\,S_w^2 + \frac{nS_w}{K_w}\right) + \rho^{gw}\frac{1-n}{K_s}S_gS_w\right]\frac{\mathrm{D}^s p^w}{\mathrm{D}t} + nS_g\frac{\mathrm{D}^s}{\mathrm{D}t}\rho^{gw} - \beta_{swg}\frac{\mathrm{D}^s\theta}{\mathrm{D}t}\\
&+\left[\frac{(1-n)}{K_s}S_g(\rho^{gw}S_g+\rho^wS_w)\right]\frac{\mathrm{D}^s p^{gw}}{\mathrm{D}t} - \,\mathrm{div}\left[\rho^g\frac{M_aM_w}{{M_g}^2}\mathbf{D}_g\,\mathrm{grad}\left(\frac{p^{gw}}{p^g}\right)\right]\\
&\quad+\left[\frac{1-n}{K_s}\left(\rho^{gw}S_gp^c+\rho^wS_wp^w-\rho^wS_wp^c\right)+n(\rho^w-\rho^{gw})\right]\frac{\mathrm{D}^s S_w}{\mathrm{D}t}\\
&\qquad+(\rho^{gw}S_g+\rho^wS_w)\,\mathrm{div}\mathbf{v}^s + \,\mathrm{div}\,(n\,S_w\,\mathbf{v}^{ws}) + \,\mathrm{div}\,(n\,S_g\,\mathbf{v}^{gs}) = 0
\end{aligned} \tag{50}$$

with $\beta_{swg} = \beta_s(1-n)(S_g\rho^{gw} + \rho^wS_w) + n\beta_w\rho^wS_w$, where $\beta_s$ and $\beta_w$ are the thermal expansion coefficient for the solid grains and the liquid water.

- Mass balance equation of the dry air:

$$\frac{1-n}{K_s} S_g S_w \frac{\mathrm{D}^s p^w}{\mathrm{D}t} + \frac{nS_g}{\rho^{ga}} \frac{\mathrm{D}^s \rho^{ga}}{\mathrm{D}t} + \frac{1-n}{K_s} S_g^2 \frac{\mathrm{D}^s p^{ga}}{\mathrm{D}t} - \left( \frac{1-n}{K_s} S_g p^c + n \right) \frac{\mathrm{D}^s S_w}{\mathrm{D}t} - \frac{1}{\rho^{ga}} \operatorname{div} \left[ \rho^g \frac{M_a M_w}{{M_g}^2} \mathbf{D}_g \operatorname{grad} \left( \frac{p^{ga}}{p^a} \right) \right] + \frac{1}{\rho^{ga}} \operatorname{div} \left( n\, S_g \rho^{ga}\, \mathbf{v}^{gs} \right) - \beta_s (1-n) S_g \frac{\mathrm{D}^s \theta}{\mathrm{D}t} = 0 \quad (51)$$

- Enery balance equation of the mixture:

$$(\rho C_p)_{eff} \frac{\partial \theta}{\partial t} + \left( \rho_w C_p^w \mathbf{v}^w + \rho_g C_p^g \mathbf{v}^g \right) \cdot \operatorname{grad} \theta - \operatorname{div}(\chi_{eff} \operatorname{grad} \theta) = -\dot{m} \Delta H_{vap} \quad (52)$$

with $(\rho C_p)_{eff} = \rho_s C_p^s + \rho_w C_p^w + \rho_g C_p^g$, $\chi_{eff} = \chi^s + \chi^w + \chi^g$ and $\Delta H_{vap}$ the latent heat of evaporation.

- Linear momentum balance equations of the mixture:

$$\operatorname{div} \boldsymbol{\sigma} + \rho (\mathbf{g} - \mathbf{a}^s) - n S_w \rho^w (\mathbf{a}^{ws} + \mathbf{v}^{ws} \cdot \operatorname{grad} \mathbf{v}^w) - n S_g \rho^g (\mathbf{a}^{gs} + \mathbf{v}^{gs} \cdot \operatorname{grad} \mathbf{v}^g) = \mathbf{0} \quad (53)$$

in which $\rho = (1-n)\rho^s + nS_w\rho^w + nS_g\rho^g$ is the density of the mixture.

- Linear balance equation of the fluids:

$$n\, S_\pi\, \mathbf{v}^\pi = -\frac{\mathbf{k}\, k^{r\pi}}{\mu^\pi} \left( \operatorname{grad} p^\pi - \rho^\pi \left( \mathbf{g} - \mathbf{a}^s - \mathbf{a}^{\pi s} \right) \right) \quad (54)$$

# 4 Conclusions

A mathematical formulation for the thermo-hydro-mechanical behaviour of variably saturated porous material has been presented. This model is suitable for the numerical discretisation via the finite element method, as will be shown in the contributions *”A finite element model for variably saturated geomaterials”* and *”Finite element analysis of non-isothermal multiphase porous media in quasi-statics and dynamics”* of this volume. For the interested reader, further details of the use of the model in environmental geomechanics and soil dynamics can be found in the textbooks [Lew98], [Ocz99] and [Cha22] and in the references contained in. Moreover, a thermodynamic framework of the model is developed in [Sch02], where interfacial phenomena between the constituents are taken into account. The extension of the model presented here for the simulation of concrete structure at high temperature is developed, e.g., in [Gaw03]. The development of the model which considers the air dissolved in the liquid water and its desorption at lower water pressures in quasi-statics loading conditions is presented in [Gaw09]. The model illustrated in this chapter can also be derived from the more advanced averaging theory TCAT - Thermodynamically Constrained Averaging Theory and its references listing the journal papers on this topic [Gra14].

## References

[Boe00] de Boer R., *Theory of Porous Media: Highlight in Historical Development and Current State*, Springer-Verlag, Berlin, 2000.

[Cha22] Chan A., Pastor M., Schrefler B. A.Shiomi T. , Zienkiewicz O. C. *Computational Geomechanics. Theory and Applications*, John Wiley & Sons, Chichester, 2022.

[Ehl98] Ehlers W., Volk W., On theoretical and numerical methods in the theory of porous media based on polar and non-polar elasto-plastic solid materials, *Int. J. Solids and Structures* 35, 4597–4617, 1998.

[Gaw03] Gawin D., Pesavento F., Schrefler B.A., Modelling of hygro-thermal behaviour of concrete at high temperature with thermo-chemical and mechanical material degradation, *Computer Methods in Applied Mechanics and Engineering* 192, 1731–1771, 2003.

[Gaw09] Gawin D., Sanavia L.., A unified approach to numerical modelling of fully and partially saturated porous materials by considering air dis-solved in water, *CMES: Computer Modeling in Engineering & Sciences* 53, 255–302, 2009.

[Gra91] Gray W. G., Hassanizadeh M., Unsaturated Flow Theory including Interfacial Phenomena, *Water Resources Res.*, 27, 8, 1855–1863, 1991.

[Gra01] Gray W. G., Schrefler B.A., Thermodynamic approach to effective stress in partially saturated porous media, *Eur. J. Mech. A/Solids*, 20, 521–538, 2001.

[Gra14] Gray W. G., Miller C. T., , *Introduction to the Thermodynamically Constrained Averaging Theory for porous medium systems*, Springer, , 2014.

[Has79a] Hassanizadeh M., Gray W. G., General Conservation Equations for Multiphase System: 1. Averaging technique, *Adv. Water Res.*, 2, 131–144, 1979.

[Has79b] Hassanizadeh M., Gray W. G., General Conservation Equations for Multi-Phase System: 2. Mass, Momenta, Energy and Entropy Equations, *Adv. Water Res.*, 2, 191–201, 1979.

[Has80] Hassanizadeh M., Gray W. G., General Conservation Equations for Multi-Phase System: 3. Constitutive Theory for Porous Media Flow, *Adv. Water Res.*, 3, 25–40, 1980.

[Lee69] Lee E. H., Elastic-Plastic Deformation at Finite Strains, *J. Appl. Mech.*, 1, 6, 1969.

[Lew98] Lewis R. W., Schrefler B. A., *The Finite Element Method in the Static and Dynamic Deformation and Consolidation of Porous Media*, John Wiley & Sons, Chichester, 1998.

[Mar83] Marsden J. E., Hughes T. J. R., *Mathematical Foundations of Elasticity*, Prentice-Hall, Englewood Cliffs, 1983.

[Mok98] Mokni M., Desrues J., Strain Localization Measurements in Undrained Plane-strain Biaxial Tests on Hostun RF Sand, *Mech. Cohes-Frict. Mater.*, 4, 419–441, 1998.

[Nem83] Nemat-Nasser S., On Finite Plastic Flow of Crystalline Solids and Geomaterials, *Transactions of ASME* 50, 1114 –1126, 1983.

[San02] Sanavia L., Schrefler B. A., Steinmann P., A formulation for an unsaturated porous medium undergoing large inelastic strains, *Computational Mechanics*, 28, 25–40, 2002.

[Sch02] Schrefler B.A., Mechanics and thermodynamics of saturated-unsaturated porous materials and quantitative solutions, *Appl. Mech. Rev.*, 55, 351–388, 2002.

[Var95] Vardoulakis J., Sulem J., *Bifurcation Analysis in Geomechanics*, Blakie Academic and Professional, London, 1995.

[Ocz99] Zienkiewicz O. C., Chan A., Pastor M., Schrefler B. A., Shiomi T., *Computational Geomechanics with special Reference to Earthquake Engineering*, John Wiley & Sons, Chichester, 1999.

# A finite element model for variably saturated geomaterials

## A space and time discretisation for a multiphase porous material model at large elasto-plastic strain

**Lorenzo Sanavia, Bernhard A. Schrefler**

*Department of Civil Environmental and Architectural Engineering, University of Padua, Italy*

*A finite element formulation for an isothermal saturated and partially saturated porous medium undergoing large elastic or inelastic deformations is presented. This model is derived from the general thermo-hydro-mechanical model for porous materials developed in the previous contribution from the authors to this volume. The porous medium is treated as a multiphase continuum with the pores of the solid skeleton filled by water and air, this last one at constant pressure. The governing equations at macroscopic level are derived in a spatial setting. Solid grains and water are assumed to be incompressible at the microscopic level for simplicity. The consistent linearisation of the fully non linear coupled system of equations is derived. A spatial finite element formulation of the governing equations conclude this chapter.*

## 1 Introduction

This paper presents the finite element model for a saturated and partially saturated porous material capable to sustain large elastic or elasto-plastic strains, extending the previous work of Sanavia et al. [San02a], [San02b].

The porous medium is treated as an isothermal multiphase continuum with the pores of the solid skeleton filled by water and air, this last one at constant pressure (passive air phase assumption). This pressure may either be the atmospheric pressure or the cavitation pressure (isothermal monospecies approach). Quasi-static loading conditions are considered. The governing equations at macroscopic level are described in Section 2 in a spatial setting. This model follows from the general thermo-hydro-mechanical model developed in [Lew98], which is described in the contribution *"Coupling equations for variably saturated geomaterials"* from the authors to these lecture notes.

Solid displacements and water pressures are the primary variables. The solid grains and water are assumed to be incompressible at microscopic level. The elasto-plastic behaviour of the solid skeleton is described by the multiplicative decomposition of the deformation gradient into an elastic and a plastic part, as shown in Section 3. The generalized effective stress in partially saturated conditions (*Bishop* like stress) in the form of *Kirchhoff* measure of the stress tensor and the logarithmic principal strains are used in conjunction with an hyperelastic free energy function. The effective stress state is limited by the *von Mises* or the *Drucker-Prager* yield surface for simplicity. Water is assumed to obey *Darcy*'s law. In the partially saturated state, the water degree of saturation and the relative permeability are dependent on the capillary pressure by experimental functions. The spatial weak form of the governing equations, the temporal integration of the mixture mass balance equation, which is time dependent because of the seepage process of water, and the consistent linearization are described in Sections 4, 5, and 6, respectively. In particular, the generalized trapezoidal method is used for the time integration. Finally, the finite element discretization in space is obtained by applying a *Galerkin* procedure in the spatial setting, using different shape functions for solid and water (see Section 7). The interested reader is referred to [San02b] for all the computational details.

## 2 Balance equations for an isothermal variably saturated medium

In this section the macroscopic balance equations of the simplified model that we shall use in the sequel are obtained. In this model, the main features are that water pressure and solid displacements are chosen as the primary variables and that the elastoplastic behaviour of the solid skeleton is developed in the framework of the hyperelastoplasticity. Moreover, it permits to outline the main guidelines used in modern computational mechanics.

The following assumptions are now introduced in the general model previously presented:

- All the processes are isothermal. This means that the energy balance equation is no more necessary and the phase changed are neglected.
- Gas phase is assumed to remain at constant pressure and flows without resistance in the partially saturated zone ([Ocz99], [Cha22]). This means that the mass balance equations for dry air and vapour are neglected. The gas pressure may either be the atmospheric pressure or the cavitation pressure at a certain temperature (e. g. the ambient temperature). The first case is a common assumption in soil mechanics because in many cases occurring in practice the air pressure is close to the atmospheric pressure as the pores are interconnected [Ocz99], [Cha22]. The second case can be derived from the experimental observations [Mok98] and the obtained model is also called *Isothermal Monospecies Approach*, which can be used to simulate cavitation at localization in initially

water saturated dense sands under globally undrained conditions, as first developed in [Sch96] for the geometrically linear case. In fact, in this situation, neglecting air dissolved in water, only two fluid phases are present after cavitation: liquid water and water vapour at cavitation pressure, which is then considered constant and is neglected because of its small value.

- At the micro level, the porous medium is assumed to consist of incompressible solid and water constituents. The averaged intrinsic density $\rho^{\pi}(\mathbf{x},\, t)$ $(\pi = s,\, w)$ is hence constant, while the averaged density $\rho_{\pi}(\mathbf{x},\, t)$ can vary due to the volume fraction $\eta^{\pi}(\mathbf{x},\, t)$. Consequently, the density of the mixture $\rho(\mathbf{x},\, t)$ (12) and the porosity $n(\mathbf{x},\, t)$ can change during the deformation of the porous medium.
- The process is considered as quasi-static, so the solid and fluids accelerations are neglected.

The formulation in terms of spatial co-ordinates is now presented.

## 2.1 Mass balance equation

Taking into account the incompressibility constraint of the solid and water constituents in eq. (16) and eq. (18) of [Sch02], the mass balance equation for the solid and water phases becomes

$$\frac{\partial}{\partial t}\,(1-n) + \text{ div }[(1-n)\,\mathbf{v}^{s}] = 0 \tag{1}$$

$$\frac{\partial}{\partial t}\,(n\,S_{w}) + \text{ div }(n\,S_{w}\,\mathbf{v}^{w}) = 0 \tag{2}$$

where the definition of the phase average density $\rho_{\pi}(\mathbf{x},\, t) = \eta^{\pi}(\mathbf{x},\, t)\,\rho^{\pi}(\mathbf{x},\, t)$, $(\pi = s,\, w)$ has been introduced, thus eliminating the intrinsic (constant) average density $\rho^{\pi}(\mathbf{x},\, t)$. Using the concept of the material time derivative, (1) is rewritten as

$$\frac{\mathrm{D}^{s}}{\mathrm{D}t}\,(1-n) + (1-n)\,\text{ div}\,\mathbf{v}^{s} = 0 \tag{3}$$

where the classical relationship

$$\text{div}\,\mathbf{v}^{s} = \frac{\mathrm{D}^{s}J^{s}}{\mathrm{D}t}\,(J^{s})^{-1} \tag{4}$$

can be introduced for the solid deformation [Mar83]. The time integration of (3) gives the evolution law for the porosity $n(\mathbf{x},\, t)$ related to the determinant $J^{s}(\mathbf{X}^{s},\, t)$ of the deformation gradient $\mathbf{F}^{s}(\mathbf{X}^{s},\, t)$:

$$n = 1 - (1 - n_0)\ (J^s)^{-1} \tag{5}$$

where $n_0(\mathbf{X}^s)$ is the porosity in the reference configuration at $t = t_0$ (or initial porosity). Because of the relation $\eta^s(\mathbf{x},\, t) = 1 - n(\mathbf{x},\, t)$, (5) can be rewritten as

$$\eta^s = \eta_0^s\, (J^s)^{-1} \tag{6}$$

where $\eta_0^s(\mathbf{X}^s)$ is the solid volume fraction in the reference configuration at $t = t_0$.

The sum of the mass balance equation of the two constituents (1) and (2) produces the following mass balance equation for the mixture under consideration:

$$\frac{\partial}{\partial t}\,(1 - n + nS_w) + \operatorname{div}\ [(1 - n)\ \mathbf{v}^s + n\, S_w\, \mathbf{v}^w] = 0 \tag{7}$$

Introducing the water velocity relative to the solid, i. e. $\mathbf{v}^{ws} = \mathbf{v}^w - \mathbf{v}^s$ and the definition of material time derivative with respect to the solid, the mixture mass balance equation (7) becomes

$$n\frac{\mathrm{D}^s}{\mathrm{D}t}\,(S_w) + S_w \operatorname{div}\mathbf{v}^s + \operatorname{div}\ (n\, S_w\, \mathbf{v}^{ws}) = 0 \tag{8}$$

The term $n\, S_w\, \mathbf{v}^{ws}(\mathbf{x},\, t)$ represent the filtration water velocity. The water velocity relative to the solid is related to the water pressure by the linear momentum balance equation for water phase, which gives *Darcy*'s law (15), as will be demonstrated in the sequel.

In case of fully saturated conditions, $S_w = 1$ and hence the previous equation is reduced to the one of the saturated model.

## 2.2 Linear momentum balance equations in statics

Neglecting the inertial term in eq. (23) and eq. (24) of [Sch02], the linear momentum balance equations in statics for the solid and water constituents are respectively

$$\operatorname{div}\mathbf{t}^s + (1 - n)\ \rho^s\, \mathbf{g} + (1 - n)\ \rho^s\, \hat{\mathbf{t}}^s = \mathbf{0} \tag{9}$$

$$\operatorname{div}\mathbf{t}^w + n\, S_w\, \rho^w\, \mathbf{g} + n\, S_w\, \rho^w\, (\hat{\mathbf{t}}^w + \mathbf{e}^w) = \mathbf{0} \tag{10}$$

The equilibrium equations for the mixture

$$\operatorname{div}\left(\mathbf{t}^{s}+\mathbf{t}^{w}\right)+\rho\,\mathbf{g}=\mathbf{0} \tag{11}$$

is obtained by summation of (9) and (10), taking into account the constraint (25) of [Sch02].

In (11), $\rho(\mathbf{x},\,t)$ is the density of the mixture:

$$\rho=(1-n)\;\rho^{s}+n\,S_{w}\,\rho^{w} \tag{12}$$

and $\mathbf{t}^{s}+\mathbf{t}^{w}=\boldsymbol{\sigma}$ is the total *Cauchy* stress, which can be decomposed into the effective and pressure (equilibrium) parts following the principle of effective stress

$$\boldsymbol{\sigma}=\boldsymbol{\sigma}'-S_{w}\,p^{w}\,\mathbf{1} \tag{13}$$

where $\boldsymbol{\sigma}'(\mathbf{x},\,t)$ is the modified effective *Cauchy* stress tensor, also called Generalized or *Bishop*'s like or *Schrefler*'s stress tensor in soil mechanics. The equilibrium equations of the mixture in terms of total *Cauchy* stress assumes the form

$$\operatorname{div}\boldsymbol{\sigma}+\rho\,\mathbf{g}=\mathbf{0} \tag{14}$$

Using the constitutive equation (29) of [Sch02] for $n\,S_{w}\,\rho^{w}\,\hat{\mathbf{t}}^{w}$ and the definition of $\mathbf{t}^{w}$ (eq.(30) of [Sch02]), the linear momentum balance equation for water (10) gives *Darcy*'s law

$$n\,S_{w}\,\mathbf{v}^{w}=-\frac{\mathbf{k}\,k^{rw}}{\mu^{w}}\,(\operatorname{grad}p^{w}-\rho^{w}\,\mathbf{g}) \tag{15}$$

for the water, where $k^{rw}=k^{rw}(\mathbf{x},\,t)$ is the relative permeability which is an experimentally determined function of the capillary pressure. This law is valid for the transport of the fluid in slow phenomena when the thermal effects are negligible. Moreover, the equilibrium equation for the fluid pressures ($p^{c}=p^{g}-p^{w}$) is simplified as follows because of the assumption on the gas phase

$$p^{c}\cong-p^{w} \tag{16}$$

which states that capillary pressures can be approximated as pore water tractions. Hence, the water pressure can change sign, which means that a partially saturated zone is developing in the porous medium. The effect of the capillary pressure on the stiffness of the medium is taken into account by the constitutive laws for $S_w(p^c)$ and $k^{rw}(p^c)$.

As a consequence of the above assumptions, the independent fields of the model are the solid displacements $\mathbf{u}(\mathbf{x},\, t)$ and the water pressure $p^w(\mathbf{x},\, t)$.

In case of fully saturated conditions, $S_w = 1$ and $k^{rw} = 1$ and hence (12), (13), and (15) are reduced to those of the saturated model.

# 3 Constitutive equations

The constitutive equations necessary for the model are those related to the solid skeleton and the water in the partially saturated zones. In particular, the structure of the developed model can describe the elasto-plastic behaviour of the solid skeleton at finite strain based on the multiplicative decomposition of the deformation gradient $\mathbf{F}^s(\mathbf{X}^s,\, t)$ into an elastic and plastic part

$$\mathbf{F}^s = \mathbf{F}^{se}\,\mathbf{F}^{sp} \tag{17}$$

In the following, the treatment of the isotropic elasto-plastic behaviour for the solid skeleton based on the product formula algorithm proposed for the single phase material by Simo [Sim98] will be briefly summarized for the interested readers. Experienced readers or non interested readers may wish to turn directly to Section 4. The geometrically linear case can be found e. g. in [Lew98] and [Ocz99]. The spatial formulation is used in this section, accordingly to [San02b] is now presented. The interested reader can refer to [San02b] also for the Lagrangian counterpart in terms of material coordinates.

In this section, the superscript ‘ $^s$ ’ will be neglected and the symbol ‘ $\dot{}$ ’ will be used for the material time derivative with respect to the solid skeleton instead of $\mathrm{D}^s/\mathrm{D}t$ (as well as in the remaining part of the chapter).

The effective *Kirchhoff* stress tensor $\boldsymbol{\tau}'(\mathbf{x},\, t) = J\boldsymbol{\sigma}'(\mathbf{x},\, t)$ and the logarithmic principal values of the elastic left *Cauchy-Green* strain tensor $\boldsymbol{\epsilon}_A(\mathbf{x},\, t)$ are used. In the present sub-section also the prime ‘ $'$ ’ for the effective stress tensor will be neglected. The yield function restricting the stress state is developed in the form of *von Mises* and *Drucker-Prager* for simplicity, to take into account the behaviour of clays under undrained conditions and the dilatant/contractant behaviour of dense or loose sands, respectively. The return mapping and the consistent tangent operator can be developed, solving the singular behaviour of the *Drucker-Prager* yield surface in the zone of the apex using the concept of multisurface plasticity (see [San02b]).

The elastic behaviour of the solid skeleton is assumed to be governed by an hyperelastic free energy $\psi(\mathbf{x},\, t)$ function in the form

$$\psi = \psi(\mathbf{b}^e,\, \xi) \tag{18}$$

dependent on the elastic left *Cauchy-Green* strain tensor $\mathbf{b}^e(\mathbf{x},\, t) = \mathbf{F}^e(\mathbf{F}^e)^{-1}$ and the internal strain like variable $\xi(\mathbf{x}, t)$, the equivalent plastic strain. The second law of thermodynamics yields, under the restriction of isotropy, the constitutive relations

$$\boldsymbol{\tau} = 2\, \frac{\partial \psi}{\partial \mathbf{b}^e}\, \mathbf{b}^e\,, \quad q = -\frac{\partial \psi}{\partial \xi} \tag{19}$$

and the dissipation inequality

$$-\frac{1}{2}\,\boldsymbol{\tau}\, :\, \left[(L_v\, \mathbf{b}^e)\ (\mathbf{b}^e)^{-1}\right] + q\,\dot{\xi} \geq 0\,, \tag{20}$$

where $L_v\, \mathbf{b}^e = \dot{\mathbf{b}}^e - \mathbf{l}\,\mathbf{b}^e - \mathbf{b}^e\,\mathbf{l}^T$ is the *Lie* derivative of the elastic left *Cauchy-Green* strain tensor $\mathbf{b}^e(\mathbf{x},\, t)$.

The evolution equations for the rate terms of the dissipation inequality (20) can be derived from the postulate of the maximum plastic dissipation in the case of associative flow rules

$$-\frac{1}{2}\, L_v\, \mathbf{b}^e = \dot{\gamma}\, \frac{\partial F}{\partial \boldsymbol{\tau}}\, \mathbf{b}^e \tag{21}$$

$$\dot{\xi} = \dot{\gamma}\, \frac{\partial F}{\partial q}\,, \tag{22}$$

subjected to the classical loading-unloading conditions in *Kuhn-Tucker* form:

$$\dot{\gamma} \geq 0, \quad F = F(\boldsymbol{\tau},\, q) \leq 0, \quad \dot{\gamma}\, F = 0 \tag{23}$$

where $\dot{\gamma}$ is the plastic multiplier and $F = F(\boldsymbol{\tau},\, q)$ the isotropic yield function.

Simple examples for the yield functions are those of *Drucker-Prager* and *von Mises* with linear isotropic hardening, in the form, respectively, of

$$F(p,\, \mathbf{s},\, \xi) = 3\, \alpha_F\, p + \|\mathbf{s}\| - \beta_F\, \sqrt{\frac{2}{3}}\ (c_0 + h\, \xi) \tag{24}$$

and

$$F(\mathbf{s},\,\xi) = \|\mathbf{s}\| - \sqrt{\frac{2}{3}}\,(\sigma_0 + h\,\xi) \tag{25}$$

in which $p = \frac{1}{3}(\boldsymbol{\tau} : \mathbf{1})$ is the mean effective *Kirchhoff* pressure, $\|\mathbf{s}\|$ is the $L_2$-norm of the deviator effective *Kirchhoff* stress tensor $\boldsymbol{\tau}$, $c_0$ is the initial apparent cohesion of the *Drucker-Prager* model, $\alpha_F$ and $\beta_F$ are two parameters related to the friction angle $\phi$ of the soil,

$$\alpha_F = 2\,\frac{\sqrt{\frac{2}{3}}\,\sin\phi}{3-\sin\phi}\,, \quad \beta_F = \frac{6\,\cos\phi}{3-\sin\phi}\,, \tag{26}$$

$h$ the hardening/softening modulus, and $\sigma_0$ is the yield stress in the *von Mises* law.

**Remarks**: In the present contribution, the effect of the capillary pressure $p^c$ on the evolution of the yield surface is not taken into account. The interested reader can refer to [Lew98] for a constitutive relationship function of the effective stress and the capillary pressure and to the chapter in this volume by Manzanal et al. Moreover, other models are also cited in the last chapter of this volume.

## 3.1 Algorithmic formulation

The problem of the calculation of $\mathbf{b}^e, \xi$ and $\boldsymbol{\tau}$ is solved by an operator split into an elastic predictor and plastic corrector [Sim98]. The calculation of the trial elastic state $(\bullet)^{tr}$ is based on freezing the plastic flow at time $t_{n+1}$. The $[\mathbf{b}^e_{n+1}]^{tr}$ is hence the push forward of $\mathbf{b}^e_n$ by means of the relative deformation gradient $\mathbf{f}_{n+1}$, i.e.

$$[\mathbf{b}^e_{n+1}]^{tr} = \mathbf{f}_{n+1}\mathbf{b}^e_n\mathbf{f}^T_{n+1} \tag{27}$$

with $\xi^{tr}_{n+1} = \xi_n$ , where $\Delta\mathbf{u}_{n+1}$ is the incremental displacement in the time interval $[t_n, t_{n+1}]$.
The same value can also be obtained from the reference configuration by the push forward of $[\mathbf{C}^p_n]^{-1}$ by means of $\mathbf{F}_{n+1}$

$$[\mathbf{b}^e_{n+1}]^{tr} = \mathbf{F}_{n+1}[\mathbf{C}^p_n]^{-1}\mathbf{F}^T_{n+1} \tag{28}$$

The corresponding trial elastic stress is obtained from the hyperelastic free energy function as

$$\boldsymbol{\tau}^{tr}_{n+1} = 2\left[\frac{\partial \psi}{\partial \mathbf{b}^e}\mathbf{b}^e\right]_{\mathbf{b}^e=[\mathbf{b}^e_{n+1}]^{tr}} = 2\frac{\partial \psi}{\partial \mathbf{b}^e}\bigg|_{\mathbf{b}^e=[\mathbf{b}^e_{n+1}]^{tr}} [\mathbf{b}^e_{n+1}]^{tr} \tag{29}$$

If this trial state is admissible, it does not violate the inequality $F^{tr}_{n+1} \leq 0$ and the stress state is hence already computed.

Otherwise the return mapping or plastic corrector algorithm is applied to satisfy the condition $F_{n+1} = 0$. Since during this phase the spatial position is held fixed, the evolution equation for the elastic left Cauchy-Green strain tensor can be computed as in [San02b] and

$$\mathbf{b}^e_{n+1} \cong \exp\left(-2\Delta\gamma \frac{\partial F}{\partial \boldsymbol{\tau}}\right)\bigg|_{n+1} [\mathbf{b}^e_{n+1}]^{tr}. \tag{30}$$

can be derived. It should now be noted that $\mathbf{b}^e_{n+1}$ commutes with $\boldsymbol{\tau}_{n+1}$ due to the assumption of isotropy and that $[\mathbf{b}^e_{n+1}]^{tr}$ and its principal axis are held fixed during the return mapping; the spectral decomposition of $[\mathbf{b}^e_{n+1}]^{tr}$, $\mathbf{b}^e_{n+1}$ and $\boldsymbol{\tau}_{n+1}$ can hence be written with the same eigenbases

$$\begin{gathered} [\mathbf{b}^e]^{tr} = \sum_{A=1}^{3} [\lambda^{tr}_{Ae}]^2 \mathbf{n}^{tr}_A \otimes \mathbf{n}^{tr}_A \qquad \mathbf{b}^e = \sum_{A=1}^{3} [\lambda_{Ae}]^2 \mathbf{n}^{tr}_A \otimes \mathbf{n}^{tr}_A \\ \boldsymbol{\tau} = \sum_{A=1}^{3} \tau_A \mathbf{n}^{tr}_A \otimes \mathbf{n}^{tr}_A \end{gathered} \tag{31}$$

Using (31) the product formula (30) can be written in principal values in the form

$$[\lambda_{Ae}]^2 = \exp\left(-2\Delta\gamma \frac{\partial F}{\partial \tau_A}\right)\bigg|_{n+1} [\lambda^{tr}_{Ae}]^2. \tag{32}$$

Taking the logarithm of (32) the fundamental *additive* decomposition of the $log$ strain measure in elastic and plastic parts is obtained [Sim98]

$$\varepsilon^{tr}_{Ae_{n+1}} = \varepsilon_{Ae_{n+1}} + \Delta\gamma \frac{\partial F}{\partial \tau_A}\bigg|_{n+1} \tag{33}$$

in which $\varepsilon_{Ae}$ are the principal logarithmic elastic strain $\varepsilon_{Ae} = ln\lambda_A$. This is a very important consequence of the utilised model because it permits to use the return mapping of the elasto-plasticity developed for the linear case [Sim98]. From the knowledge of

$\Delta\gamma$ the equivalent plastic strain is computed by the backward Euler integration of eq. (22)

$$\xi_{n+1} \cong \xi_n + \Delta\gamma \left.\frac{\partial F}{\partial q}\right|_{n+1} \tag{34}$$

The principal Kirchhoff stress components are then computed by the hyperelastic constitutive law

$$\tau_A = 2\lambda_{Ae} \frac{\partial \psi}{\partial \lambda_{Ae}} = \frac{\partial \psi}{\partial \varepsilon_{Ae}} \tag{35}$$

where the free energy $\psi = \hat{\psi}(\varepsilon_{Ae}, \xi)$ is now written as function of the principal elastic logarithmic strain components and the equivalent plastic strain (for isotropic linear hardening).

The return mapping algorithm for the Drucker-Prager model with non-associated volumetric/deviatoric plastic flow is presented in [San02b], where a special treatment of the corner region using the concept of multi-surface plasticity is also formulated.

# 4 Weak form: Variational approach

The weak form of the spatial governing equations presented in the previous section is now derived obtaining the variational equations formally equivalent to the initial-boundary-value problem given by the governing equation and the boundary conditions. This means that the governing equations (7) and (11) are multiplied by independent weighting functions that vanish on the boundary in which *Dirichlet* boundary conditions are applied and are then integrated over the spatial domain $B$ with boundary $\partial B$. The linear momentum balance equation of the binary porous media (11) is hence weighted on the domain by the test function $\delta\mathbf{u}_s$ corresponding to the solid displacement (or virtual displacement) in the form

$$\int\limits_B (\operatorname{div}\boldsymbol{\sigma} + \rho\,\mathbf{g}) \cdot \delta\mathbf{u}_s \,\mathrm{d}v = 0 \quad \forall \;\; \delta\mathbf{u}_s \tag{36}$$

Applying partial integration and *Green*'s theorem in the form (e. g. [Mar83])

$$\int\limits_B \operatorname{div}\boldsymbol{\sigma} \cdot \delta\mathbf{u}_s \,\mathrm{d}v = -\int\limits_B \boldsymbol{\sigma} : \operatorname{grad}\delta\mathbf{u}_s \,\mathrm{d}v + \int\limits_{\partial B} \bar{\mathbf{t}} \cdot \delta\mathbf{u}_s \,\mathrm{d}s \tag{37}$$

to the divergence part of (36) and taking into account the boundary conditions, this equation is transformed into the weak form

$$
\begin{aligned}
&-\int\limits_B (\boldsymbol{\sigma}' - S_w\, p^w \mathbf{1}) :\ \mathrm{grad}\, \delta\mathbf{u}_s\, \mathrm{d}v\ + \int\limits_B \rho\, \mathbf{g} \cdot \delta\mathbf{u}_s\, \mathrm{d}v + \\
&\int\limits_{\partial B} \bar{\mathbf{t}} \cdot \delta\mathbf{u}_s\, \mathrm{d}s = 0 \quad \forall\ \ \delta\mathbf{u}_s
\end{aligned} \tag{38}
$$

where the effective stress principle (13) has been introduced. Using the relation $\mathrm{div}\, \delta\mathbf{u}_s =\ \mathrm{grad}\, \delta\mathbf{u}_s : \mathbf{1}$, the previous weak form is transformed into

$$
\begin{aligned}
&-\int\limits_B \boldsymbol{\sigma}' :\ \mathrm{grad}\, \delta\mathbf{u}_s\, \mathrm{d}v + \int\limits_B S_w\, p^w\, \mathrm{div}\, \delta\mathbf{u}_s\, \mathrm{d}v + \int\limits_B \rho\, \mathbf{g} \cdot \delta\mathbf{u}_s\, \mathrm{d}v + \\
&\int\limits_{\partial B} \bar{\mathbf{t}} \cdot \delta\mathbf{u}_s\, \mathrm{d}s = 0 \quad \forall\ \ \delta\mathbf{u}_s
\end{aligned} \tag{39}
$$

The weak form of the mixture mass balance equation (7) is obtained in a similar way, introducing *Darcy*'s law (15) and using the test function $\delta p^w$ corresponding to $p^w$ (or virtual water pressure):

$$
\begin{aligned}
&\int\limits_B n \frac{\mathrm{D}^s}{\mathrm{D}t}\, (S_w)\, \delta p^w\, \mathrm{d}v + \int\limits_B S_w\, \mathrm{div}\, \mathbf{v}^s\, \delta p^w\, \mathrm{d}v + \\
&\int\limits_B \mathrm{div} \left[ \frac{\mathbf{k}\, k^{rw}}{\mu^w}\, (-\ \mathrm{grad}\, p^w + \rho^w\, \mathbf{g}) \right]\, \delta p^w\, \mathrm{d}v = 0\ \ \forall\, \delta p^w
\end{aligned} \tag{40}
$$

Applying *Green*'s theorem to the last integral term of the previous equation, the following is obtained

$$
\begin{aligned}
&\int\limits_B n \frac{\mathrm{D}^s}{\mathrm{D}t}\, (S_w)\, \delta p^w\, \mathrm{d}v + \int\limits_B S_w\, \mathrm{div}\, \mathbf{v}^s\, \delta p^w\, \mathrm{d}v + \\
&\int\limits_B \left[ \frac{\mathbf{k}\, k^{rw}}{\mu^w}\, (\mathrm{grad}\, p^w - \rho^w\, \mathbf{g}) \right] \cdot\ \mathrm{grad}\, \delta p^w\, \mathrm{d}v + \\
&\int\limits_{\partial B} q^w\, \delta p^w\, \mathrm{d}s = 0 \quad \forall\ \ \delta p^w
\end{aligned} \tag{41}
$$

where $q^w(\mathbf{x}, t)$ is the water flow draining through the surface $\partial B$.

**Remarks**: It can be observed that the weak forms (39) and (41) are very similar to those of the geometrically linear theory, e. g. [Lew98], by substituting the deformed integration domain $B$ with the undeformed one $B_0$. Moreover, in the small strain theory $\operatorname{div} \mathbf{v}^s = \dot{\boldsymbol{\varepsilon}} : \mathbf{1}$, where $\boldsymbol{\varepsilon}$ is the small strain tensor of the solid skeleton, while in finite strain $\operatorname{div} \mathbf{v}^s = \dot{J}^s / J^s$. In the small strain theory the additive decomposition of the strain tensor $\boldsymbol{\varepsilon}$ in elastic $\boldsymbol{\varepsilon}^e$ and plastic $\boldsymbol{\varepsilon}^p$ parts is also possible, thus rendering the computation of the constitutive tangent operator in the linearization of the weak form particularly easy.

# 5 Time discretization

Time integration of the weak form of the mass balance equation (41) over a finite time step $\Delta t = t_{n+1} - t_n$ is necessary because of the time dependent terms $\operatorname{div} \mathbf{v}^s$ and $\mathrm{D}^s S_w / \mathrm{D}t$.

The generalized trapezoidal method is used here, as shown for instance in [Lew98] and in the second chapter of this volume by Pastor and coworkers. Because of the dependence of the integration domain on time, we rewrite the weak forms (39) and (41) with respect to the undeformed domain as follows:

$$\begin{aligned} &\int_{B_0} (\boldsymbol{\tau}' - J^s\, S_w\, p^w\, \mathbf{1}) : \operatorname{grad} \delta\mathbf{u}_s \, \mathrm{d}V - \int_{B_0} \rho_0\, \mathbf{g} \cdot \delta\mathbf{u}_s \, \mathrm{d}V - \\ &\int_{\partial B_0} \bar{\mathbf{T}} \cdot \delta\mathbf{u}_s \, \mathrm{d}A = 0 \quad \forall \;\; \delta\mathbf{u}_s \end{aligned} \tag{42}$$

$$\begin{aligned} &\int_{B_0} J^s\, S_W \, \operatorname{div} \mathbf{v}^s \, \delta p^w \, \mathrm{d}V + \int_{B_0} \left[ J^s \, \frac{\mathbf{k}\, k^{rw}}{\mu^w} \, (\operatorname{grad} p^w - \rho^w\, \mathbf{g}) \right] \cdot \, \operatorname{grad} \delta p^w \mathrm{d}V + \\ &\int_{\partial B_0} Q^w \, \delta p^w \, \mathrm{d}A + \int_{B_0} J^s\, N \, \frac{\mathrm{D}^s}{\mathrm{D}t} \, (S_W) \, \delta p^w \, \mathrm{d}V = 0 \quad \forall \;\; \delta p^w \end{aligned} \tag{43}$$

where $\boldsymbol{\tau}'$ is the modified effective *Kirchhoff* stress tensor and $\bar{\mathbf{T}} = \mathbf{P}\,\mathbf{N}$ and $Q^w = N\, S_W\, \bar{\mathbf{V}}^{ws} \cdot \mathbf{N}$ are, respectively, the traction vector and the water flow computed with respect to the undeformed configuration. The form of (42) and (43) is also useful for the subsequent linearization because it will be easily performed with respect to the undeformed (fixed) domain.

Equation (43) is now rewritten at time $t_{n+1}$ using the relationships

$$\dot{J}^s_{n+\beta} \;=\frac{J^s_{n+1}-J^s_n}{\Delta t}\,, \quad (\dot{S}_w)_{n+\beta} \;=\frac{S_{w\,n+1}-S_{w\,n}}{\Delta t} \tag{44}$$

$$(\cdot)_{n+\beta}=(1-\beta)\,(\cdot)_n+\beta(\cdot)_{n+1}=(\cdot)_n+\beta\,[(\cdot)_{n+1}-(\,\cdot\,)_n] \tag{45}$$

with $\beta\,\in\,[0,\,1]$, obtaining

$$\begin{aligned}&\int\limits_{B_0}(S_w)_{n+\beta}\left(J^s_{n+1}-J^s_n\right)\delta p^w\,\mathrm{d}V-\Delta t\int\limits_{B_0}\left(J^s\,\mathbf{v}^D\cdot\;\mathrm{grad}\,\delta p^w\right)_{n+\beta}\mathrm{d}V+\\&\int\limits_{B_0}(J^s\,n)_{n+\beta}\;[S_{w\,n+1}-S_{w\,n}]\;\delta p^w\,\mathrm{d}V+\;\Delta t\int\limits_{\partial B_0}Q^w_{n+\beta}\,\delta p^w\,\mathrm{d}A=0\quad\forall\;\;\delta p^w\end{aligned} \tag{46}$$

where $\mathbf{v}^D\;=-\mathbf{k}\,k^{rw}/\mu^w\;(\,\mathrm{grad}\,p^w\,-\,\rho^w\,\mathbf{g})$ is *Darcy*'s velocity of the water.

The weak form of the linear momentum balance equation (42) is directly written at time $t_{n+1}$ because it is time independent

$$\begin{aligned}&\int\limits_{B_0}[(\boldsymbol{\tau}'-J^s\,S_w\,p^w\,\mathbf{1})\,:\;\mathrm{grad}\,\delta\mathbf{u}_s]_{\,n+1}\,\mathrm{d}V-\int\limits_{B_0}\rho_{0_{n+1}}\,\mathbf{g}\cdot\delta\mathbf{u}_s\,\mathrm{d}V-\\&\int\limits_{\partial B_0}\bar{\mathbf{T}}_{n+1}\cdot\delta\mathbf{u}_s\,\mathrm{d}A=0\quad\forall\;\;\delta\mathbf{u}_s\end{aligned} \tag{47}$$

Linearized analysis of accuracy and stability suggest the use of $\beta\geq\frac{1}{2}$. In the examples section, implicit one-step time integration has been performed ($\beta=1$).

The weak forms (46) and (47) represent a non-linear coupled equations system where the non-linearities are introduced by the finite kinematics and the constitutive laws.

# 6 Consistent linearization

The non-linear equation system (46) and (47) can be written in the following compact form

$$\mathbf{G}(\boldsymbol{\chi},\,\boldsymbol{\eta})=\mathbf{0}\,,\quad\text{with}\quad\boldsymbol{\chi}=(\boldsymbol{\chi}^s,\,p^w)^{\,T}\quad\text{and}\quad\boldsymbol{\eta}=(\delta\mathbf{u}_s,\,\delta p^w)^{\,T} \tag{48}$$

where $\boldsymbol{\chi}^s(\mathbf{X}^s,\, t)$ is the motion function (deformation map) of the solid. For its numerical solution, iterative methods have to be employed and the linearization at $\bar{\boldsymbol{\chi}}$ is hence necessary

$$\mathbf{G}(\bar{\boldsymbol{\chi}},\, \boldsymbol{\eta},\, \Delta\mathbf{u}) \cong \mathbf{G}(\bar{\boldsymbol{\chi}},\, \boldsymbol{\eta}) + \mathrm{D}\mathbf{G}(\bar{\boldsymbol{\chi}},\, \boldsymbol{\eta}) \cdot \Delta\mathbf{u} \cong \mathbf{0} \tag{49}$$

where $\Delta\mathbf{u} = (\Delta\mathbf{u}_s,\, \Delta p^w)^T$ and $\mathrm{D}\mathbf{G} \cdot \Delta\mathbf{u} = \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{G}(\bar{\boldsymbol{\chi}} + \alpha\Delta\mathbf{u})|_{\alpha=0}$ is the directional derivative or *Gateaux* derivative of $\mathbf{G}$ at $\bar{\boldsymbol{\chi}}$ in the direction of $\Delta\mathbf{u}$ (e. g. [Mar83, Wri93] for single-phase material). Since the equation system $\mathbf{G}$ is composed of the weak form of the linear momentum balance equation $(G_{\mathrm{LBE}})$ and of the mass balance equation $(G_{\mathrm{MBE}})$, then

$$\mathrm{D}\mathbf{G} \cdot \Delta\mathbf{u} = \begin{bmatrix} \mathrm{D}G_{\mathrm{LBE}} \cdot \Delta\mathbf{u}_s + \mathrm{D}G_{\mathrm{LBE}} \cdot \Delta p^w \\ \mathrm{D}G_{\mathrm{MBE}} \cdot \Delta\mathbf{u}_s + \mathrm{D}G_{\mathrm{MBE}} \cdot \Delta p^w \end{bmatrix} \tag{50}$$

Using the symbol $(\cdot)^{k+1}_{n+1}$ to indicate the current iteration in the current time step, the linearization on the configuration $(\cdot)^{k}_{n+1}$ is written as

$$\mathrm{D}\mathbf{G}^k_{n+1} \cdot \Delta\mathbf{u}^{k+1}_{n+1} = -\mathbf{G}^k_{n+1} \tag{51}$$

and the solution vector $\mathbf{u} = (\mathbf{u}_s,\, p^w)^T$ is then updated by the incremental relationship

$$\mathbf{u}^{k+1}_{n+1} = \mathbf{u}^k_{n+1} + \Delta\mathbf{u}^{k+1}_{n+1} \tag{52}$$

For an efficient numerical performance of the scheme (51), the consistent linearization is applied [Wri93] in which the linearization of the integrated constitutive equation plays a central role (this concept was first pointed out in [Sim85] for the geometrically linear case).

The linearization of (46) and (47), performed in the undeformed configuration $B_0$ and then pushed forward in the deformed configuration $B$, gives the following result:

- For the equilibrium equations:

$$\begin{gathered}\int\limits_B \left(\operatorname{grad}\delta\mathbf{u}_s : \mathbf{c}^{ep} : \operatorname{sym}\left(\operatorname{grad}\Delta\mathbf{u}_s\right) + \boldsymbol{\sigma}' : \operatorname{grad}^T\delta\mathbf{u}_s \operatorname{grad}\Delta\mathbf{u}_s\right) \mathrm{d}v + \\ \int\limits_B S_w\, p^w \operatorname{grad}\delta\mathbf{u}_s : \left(\operatorname{grad}^T\Delta\mathbf{u}_s - \operatorname{div}\Delta\mathbf{u}_s\,\mathbf{1}\right) \mathrm{d}v - \\ \int\limits_B \rho^w\, S_w\, \delta\mathbf{u}_s \cdot \mathbf{g} \operatorname{div}\Delta\mathbf{u}_s \,\mathrm{d}v - \int\limits_B \left(p^w \frac{\partial S_w}{\partial p^w} + S_w\right) \operatorname{div}\delta\mathbf{u}_s\,\Delta p^w \,\mathrm{d}v\end{gathered} \tag{53}$$

- For the mass balance equation (in case of isotropic permeability):

$$\begin{gathered}\int\limits_B \delta p^w \left(1 + S_{w\,n+\beta} + \beta\Delta S_w\right) \operatorname{div}\Delta\mathbf{u}_s \,\mathrm{d}v + \\ \beta\,\Delta t \int\limits_B \frac{k\,k^{rw}}{\mu^w} \operatorname{grad}\delta p^w \cdot \operatorname{grad}\Delta p^w \,\mathrm{d}v + \\ \beta\,\Delta t \int\limits_B \operatorname{grad}\delta p^w \cdot \left[\left(\frac{1-n}{k}\frac{\partial k}{\partial n} + 1\right) \frac{k\,k^{rw}}{\mu^w} (\operatorname{grad} p^w - \right. \\ \left. - \rho^w\,\mathbf{g}) \operatorname{div}\Delta\mathbf{u}_s\right] \mathrm{d}v - \\ \beta\,\Delta t \int\limits_B \operatorname{grad}\delta p^w \cdot \left[\frac{2\,k\,k^{rw}}{\mu^w} \operatorname{sym}\left(\operatorname{grad}\Delta\mathbf{u}_s\right) \operatorname{grad} p^w\right] \mathrm{d}v + \\ \beta\,\Delta t \int\limits_B \operatorname{grad}\delta p^w \cdot \left(\frac{k\,k^{rw}}{\mu^w}\,\rho^w \operatorname{grad}\Delta\mathbf{u}_s\,\mathbf{g}\right) \mathrm{d}v + \\ \beta\,\Delta t \int\limits_B \frac{k}{\mu^w} \frac{\partial k^{rw}}{\partial p^w} \operatorname{grad} p^w \cdot \operatorname{grad}\delta p^w \Delta p^w \,\mathrm{d}v - \\ \int\limits_B \delta p^w \frac{\left(\beta\,\Delta J + J_{n+\beta}\, n_{n+\beta}\right)}{J} \frac{\partial S_w}{\partial p^c} \Delta p^w \,\mathrm{d}v\end{gathered} \tag{54}$$

In the directional derivative $\mathrm{D}G_{\mathrm{LBE}} \cdot \Delta\mathbf{u}_s$ the term

$$\int\limits_B \left(\operatorname{grad}\delta\mathbf{u}_s : \mathbf{c}^{ep} : \operatorname{sym}\left(\operatorname{grad}\Delta\mathbf{u}_s\right) + \boldsymbol{\sigma}' : \operatorname{grad}^T\delta\mathbf{u}_s \operatorname{grad}\Delta\mathbf{u}_s\right) \mathrm{d}v \tag{55}$$

contains $\mathbf{c}^{ep}$, the spatial constitutive operator following the linearization of the computed effective stress

$$
\begin{aligned}
\mathbf{c}^{ep}_{n+1} = & \sum_{A=1}^{3}\sum_{B=1}^{3} a^{ep}_{AB_{n+1}} \left(\mathbf{n}^{\mathrm{tr}}_A \otimes \mathbf{n}^{\mathrm{tr}}_A\right) \otimes \left(\mathbf{n}^{\mathrm{tr}}_B \otimes \mathbf{n}^{\mathrm{tr}}_B\right) + \\
& 2\sum_{A=1}^{3} \tau_{A_{n+1}} \mathbf{c}^{\mathrm{tr}(A)}_{n+1}
\end{aligned}
\tag{56}
$$

It is useful to remark that in (56) only the second order tensor $\mathbf{a}^{ep} = \partial\tau_A/\partial\varepsilon^{\mathrm{tr}}_B$ depends on the specific model of plasticity and the structure of the return mapping algorithm in principal stretches, while the tensors $\mathbf{c}^{\mathrm{tr}(A)}_{n+1}$ and $\mathbf{n}^{\mathrm{tr}}_A \otimes \mathbf{n}^{\mathrm{tr}}_A$ are independent of the specific plastic model in use. Moreover, it is easy to proof that the moduli $\mathbf{a}^{ep}$ have a form identical to the algorithmic elasto-plastic tangent moduli of the infinitesimal theory [Sim98]. The expression for $\mathbf{c}^{\mathrm{tr}(A)}_{n+1}$ can be obtained by linearization of the eigenbases dyadic $\mathbf{n}^{\mathrm{tr}}_A \otimes \mathbf{n}^{tr}_A$ in the spatial setting:

$$
\mathbf{c}^{\mathrm{tr}(A)}_{n+1} = \frac{\partial(\mathbf{n}^{\mathrm{tr}}_A \otimes \mathbf{n}^{\mathrm{tr}}_A)}{\partial \mathbf{g}}
\tag{57}
$$

where $\mathbf{g}$ is the spatial metric, or by pull-back [Mar83] of $\mathbf{n}^{\mathrm{tr}}_A \otimes \mathbf{n}^{\mathrm{tr}}_A$, subsequent to linearization in the material setting and then by push-forward of the linearization in spatial setting.

The expressions for the algorithmic moduli $\mathbf{a}^{ep}$ of the *Drucker-Prager* model with non-associated volumetric/deviatoric plastic flow are derived in [San02b], where a special treatment of the apex region of the *Drucker-Prager* model using the concept of multi-surface plasticity is also derived. Hereafter, the algorithmic moduli $\mathbf{a}^{ep}$ for the non-apex zone are recalled for sake of completeness

$$
\begin{aligned}
& \mathbf{a}^{ep}_{n+1} = c_1 K \mathbf{1}\otimes\mathbf{1} + 2G\left[\mathbf{I} - \frac{1}{3}\mathbf{1}\otimes\mathbf{1}\right]\left[1 - \frac{2G\Delta\gamma_{n+1}}{||\mathbf{s}^{tr}_{n+1}||}\right] \\
& -\frac{6\alpha_Q K G}{c_2}\mathbf{1}\otimes\mathbf{n}^{tr}_{n+1} - \frac{6\alpha_F K G}{c_2}\mathbf{n}^{tr}_{n+1}\otimes\mathbf{1} \\
& -4G^2\left[\frac{1}{c_2} - \frac{\Delta\gamma_{n+1}}{||\mathbf{s}^{tr}_{n+1}||}\right]\mathbf{n}^{tr}_{n+1}\otimes\mathbf{n}^{tr}_{n+1}
\end{aligned}
\tag{58}
$$

where the coefficients $c_1$ and $c_2$ are

$$
c_1 = \left[1 - \frac{9\alpha_F\alpha_Q K}{c_2}\right]
$$

$$
c_2 = 9\alpha_F\alpha_Q K + 2G + \beta_F h\sqrt{\frac{2}{3}[1+3\alpha_Q^2]}
$$

# 7 Finite element discretization in space

The suitable spatial finite element formulation is derived by applying the well known *Galerkin* procedure, in which the weighting functions are approximated by the same shape functions used to approximate the driving variables (isoparametric finite elements). This means that the geometry $\mathbf{X}^s$, the current configuration $\mathbf{x}$, the displacement field $\mathbf{u}_s$, the water pressure $p^w$, the incremental generalized displacement $\Delta\mathbf{u} = (\Delta\mathbf{u}_s,\, \Delta p^w)^T$ and the variations $\boldsymbol{\eta} = (\delta\mathbf{u}_s,\, \delta p^w)^T$, are interpolated within a finite element by the same type of functions. In the present setting, different shape functions are chosen for quantities associated respectively to the solid and the fluid, thus satisfying the LBB condition (*Ladyzhenskaya-Babuška-Brezzi* condition) for the locally undrained case. Standard procedures have been applied, following any text books on FEM. With respect to the small strain case, the discretization of the spatial form of the linearized system of equations is made taking into account that each quantity is referred to the spatial co-ordinates $\mathbf{x}$, instead of the co-ordinates of the undeformed configuration $\mathbf{X}^s$. The solid displacement $\mathbf{u}_s(\mathbf{x},\, t)$ and the water pressure $p^w(\mathbf{x},\, t)$ are hence expressed in the whole domain by global shape function matrices $\mathbf{N}_u(\mathbf{x})$ and $\mathbf{N}_w(\mathbf{x})$ and the nodal value vectors $\bar{\mathbf{u}}(t)$ and $\bar{\mathbf{p}}(t)$:

$$\mathbf{u} = \mathbf{N}_u\bar{\mathbf{u}}\,, \qquad p^w = \mathbf{N}_w\bar{\mathbf{p}} \tag{59}$$

The linearized system of equations (51) in matrix form can be expressed as

$$\begin{bmatrix} \mathbf{K}_T + \mathbf{K}_{sw}^{\text{geom}} & -c_{sw}\,\mathbf{Q}_{sw} \\ \mathbf{Q}_{ws} - \beta\,\Delta t\,\mathbf{Q}_{sw}^{\text{geom}} & \beta\,\Delta t\,\mathbf{H} \end{bmatrix} \begin{bmatrix} \Delta\bar{\mathbf{u}} \\ \Delta\bar{\mathbf{p}} \end{bmatrix} = - \begin{bmatrix} \mathbf{G}^u \\ \mathbf{G}^p \end{bmatrix} \tag{60}$$

which is non-symmetric (details concerning the implementation as well as the matrices and the residuum vectors of (60) will be described in a future paper). Owing to the strong coupling between the mechanical and the pore fluid problem, a monolithic solution of (60) is preferred using a *Newton* scheme.

# 8 Conclusions

A finite element formulation for the hydro-mechanical behaviour of variably saturated porous materials has been presented. This model is obtained as a result of a research in progress on the thermo-hydro-mechanical model for multiphase geomaterials undergoing large inelastic strains. For the interested reader, further finite element models as well as the corresponding numerical codes can be found in [Lew98], [Ocz99] and [Cha22].

## References

[Cha22] Chan A., Pastor M., Schrefler B. A.Shiomi T. , Zienkiewicz O. C. *Computational Geomechanics. Theory and Applications*, John Wiley & Sons, Chichester, 2022.

[Lew98] Lewis R. W., Schrefler B. A., *The Finite Element Method in the Static and Dynamic Deformation and Consolidation of Porous Media*, John Wiley & Sons, Chichester, 1998.

[Mar83] Marsden J. E., Hughes T. J. R., *Mathematical Foundations of Elasticity*, Prentice-Hall, Englewood Cliffs, 1983.

[Mok98] Mokni M., Desrues J., Strain Localization Measurements in Undrained Plane-strain Biaxial Tests on Hostun RF Sand, *Mechanics of Cohesive-frictional materials*, 4, 419–441, 1998.

[San02a] Sanavia L., Schrefler B. A., Steinmann, P., A mathematical and numerical model for finite elastoplastic deformations in fluid saturated porous media. In: *Modeling and Mechanics of Granular and Porous Materials*, Series of Modeling and Simulation in Science, Engineering and Technology, G.Capriz, V.N. Ghionna and P. Giovine eds., Birkhäuser, Boston, 297–346, 2002.

[San02b] Sanavia L., Schrefler B. A., Steinmann P., A formulation for an unsaturated porous medium undergoing large inelastic strains, *Computational Mechanics*, 28, 25–40, 2002.

[Sch96] Schrefler B. A., Sanavia L., Majorana, C. E., A Multiphase Medium Model for Localization and Postlocalization Simulation in Geomaterials, *Mechanics of Cohesive-frictional materials*, 1, 95–114, 1996.

[Sch02] Schrefler B.A., Sanavia L., *Coupling equations for water saturated and partially saturated geomaterials*, Lecture notes of the ALERT course 2024, 2024.

[Sim98] Simo J. C., Hughes T. J. R., *Computational Inelasticity*, Springer-Verlag, 1998.

[Sim85] Simo J. C., Taylor R., Consistent Tangent Operators for Rate-Independent Elastoplasticity, *Comp. Meth. In Applied Mech. Eng.*, 48, 101–118, 1985.

[Ocz99] Zienkiewicz O. C., Chan A., Pastor M., Schrefler B. A., Shiomi T., *Computational Geomechanics with special Reference to Earthquake Engineering*, John Wiley & Sons, Chichester, 1999.

[Wri93] Wriggers P., *Continuum Mechanics, Non-linear Finite Element Techniques and Computational Stability*. In Stein, E. (ed.): Progress in computational Analysis of Inelastic Structures, CISM 321, Springer-Verlag, Wien, 1993.

# Computational plasticity (I): non linear analysis techniques

**Pablo Mira*, Manuel Pastor****

** Centro de Estudios y Experimentación de Obras Públicas
CEDEX, MITMA, Alfonso XII, 3, 28014 Madrid, SPAIN
Pablo.Mira@cedex.es
** E.T.S. Ingenieros de Caminos, Universidad Politécnica de Madrid
Profesor Aranguren s/n, 28040 Madrid, SPAIN*

*Solid mechanics problems in general, and specifically in the case of geomaterials, are very often significantly non linear. Although linear numerical models may be used as rough approximations to the problem solution, non linearities are very often too significant to be neglected. To solve these problems special non linear strategies are necessary. A series of non linear analysis techniques are presented in the context of the finite element method. These techniques are based on the Newton-Raphson method, a very well known root finding algorithm. The most typical versions of the method are initially presented: full Newton-Raphson with load control, modified Newton-Raphson, quasi Newton methods and Newton-Raphson with displacement control. The final sections of this chapter introduce two non-standard versions of the method including advanced techniques that can overcome convergence problems in special situations: Arc-length control and line searches.*

## 1 Introduction

Solid mechanics problems in general, and specifically in the case of geomaterials, are very often significantly non linear. Although linear numerical models may be used as rough approximations to the problem solution, non linearities are very often too significant to be neglected. Under those circumstances it is therefore difficult to remain satisfied with a solution based on a linear model. To solve these problems special non linear strategies are necessary. Since different type of non linearities require different strategies and there is no such thing as a generally valid non linear algorithm it is advisable to include several types of non linear algorithms in the model covering the most usual type of non linear situations. There are several very good texts that cover this topic extensively, such as Bathe [Bat96], Belytschko, Liu and Moran [BLM00], Crisfield [Cri91] and [Cri97], Simo and Hughes [SH98]and Zienkiewicz and Taylor [ZT89].

The non linearities in a problem may have different causes. A first attempt to classify these causes would divide them into two big groups : geometrical and material non linearities. Geometrical non linearities arise when some of the geometrical parameters of the model change during the loading process. A typical example of this would be a buckling problem where small displacement theory ceases to be valid and large displacement theory is required. Another example of geometrical no linearity is a contact problem where boundary conditions change during the loading process. On the other hand material non linearities arise when some of the constitutive parameters of the model change during the loading process. Typical examples of this include non linear elastic, elastoplastic or viscoplastic material behaviour, damage models, etc.

Numerical solution of non linear problems requires as expected many more operations than linear problems. Special numerical techniques are necessary not only to reach a solution but also to do so in an efficient manner.

Traditionally, non linear finite element problems have been solved using the Newton-Raphson method in any of its different versions. It is therefore the basic tool that virtually every non linear finite element program should include. However, there are many different ways to implement this algorithm and often the ”typical” or ”standard” versions of the Newton-Raphson method are not sufficient to solve certain problems. Although which techniques may be referred to as ”standard” and which may not is a subjective matter an attempt to do so will be made in this book. This chapter initially presents the so called standard techniques such as the following versions of the Newton-Raphson method: load control, displacement control, modified Newton Raphson methods and Quasi Newton methods.

However, ”standard” non linear analysis techniques as the ones initially presented in this chapter are often not sufficient to solve a specific problem. An example of this are limit load problems with a strongly descending post-peak behavior. In such cases, special nonlinear analysis techniques are necessary. The final sections of the chapter will be devoted to advanced or nonstandard techniques such as arc-length control and line searches.

## 2 The Newton-Raphson method

The Newton-Raphson method, also called Newton’s method is a classical one dimensional root-finding algorithm [PTVF92]. Assuming the problem is stated as finding the value or values of $x$ such that the following equation is satisfied:

$$f(x) = 0$$

the main distinguishing feature of this method from other root-finding methods is the fact that it requires the evaluation of both the function $f(x)$ and the derivative $f'(x)$ in each iteration. The method consists in geometrical terms of extending the tangent line at a current point $x_i$ until it crosses zero, then setting the next guess $x_{i+1}$ to the abscissa of that zero crossing as sketched in figure 1:

$$x_{i+1} = x_i - \frac{f(x_{i)}}{f'(x_i)} \tag{1}$$

Algebraically the method derives from the Taylor expansion series in the neighborhood of a point:

$$f(x + \delta) = f(x) + f'(x)\delta + \frac{1}{2}f''(x)\delta^2 + ...... \tag{2}$$

Given $x_i$ such that $f(x_i) \neq 0$ and using a two term expansion based on (2), a value $x_{i+1} = x_i + \delta$ is looked for such that:

$$f(x_i + \delta) = f(x_i) + f'(x_i)\delta = 0 \tag{3}$$

Solving for $\delta$:

$$\delta = -\frac{f(x_{i)}}{f'(x_i)}$$

Correcting $x_i$ with the value $\delta$ of to obtain $x_{i+1}$ would lead to expression (1).

It is also possible to use this type of analysis to obtain an estimate of the iterative error $\varepsilon$. Expressing $x_{i+1}$ $x_i$ in terms of the root $x$ and iterative errors $\varepsilon_{i+1}$ and $\varepsilon_i$:

$$\begin{aligned} x_{i+1} &= x + \varepsilon_{i+1} \\ x_i &= x + \varepsilon_i \end{aligned}$$

and substituting these expressions into equation (1) a relationship between consecutive iterative errors is obtained:

$$\varepsilon_{i+1} = \varepsilon_i - \frac{f(x_i)}{f'(x_i)} \tag{4}$$

Estimating $f(x_i)$ and $f'(x_i)$ through Taylor series expansions in the neighbourhood of the root $x$ and taking into account that $f(x) = 0$ we obtain:

$$f(x_i) = f(x + \varepsilon_i) = f(x) + f'(x)\varepsilon_i + \frac{1}{2}f''(x)\varepsilon_i^2 + ...... \simeq f'(x)\varepsilon_i + \frac{1}{2}f''(x)\varepsilon_i^2$$

$$f'(x_i) = f'(x+\varepsilon_i) = f'(x) + f''(x)\varepsilon_i + ...... \simeq f'(x) + f''(x)\varepsilon_i$$

Substituting these expressions into equation (1):

$$\begin{aligned} \varepsilon_{i+1} &= \varepsilon_i - \frac{f(x_i)}{f'(x_i)} = \varepsilon_i - \frac{f'(x)\varepsilon_i + \frac{1}{2}f''(x)\varepsilon_i^2}{f'(x)+f''(x)\varepsilon_i} \\ &= \frac{\varepsilon_i(f'(x)+f''(x)\varepsilon_i) - f'(x)\varepsilon_i - \frac{1}{2}f''(x)\varepsilon_i^2}{f'(x)+f''(x)\varepsilon_i} \\ &= -\frac{f''(x)\varepsilon_i^2}{2(f'(x)+f''(x)\varepsilon_i)} \simeq -\varepsilon_i^2\frac{f''(x)}{2f'(x)} \end{aligned} \tag{5}$$

Equation (5) is a recursive expression between consecutive iterative errors saying that the ”new” iterative error $\varepsilon_{i+1}$ is proportional to the square of the ”old” one. In other words, the Newton-Raphson procedure converges *quadratically*. This is a very powerful feature of this procedure, but in order to preserve it we have to make sure we use a good estimate of $f'(x_i)$. Frequently, rough estimates of $f'(x_i)$ are used and quadratic convergence is lost.

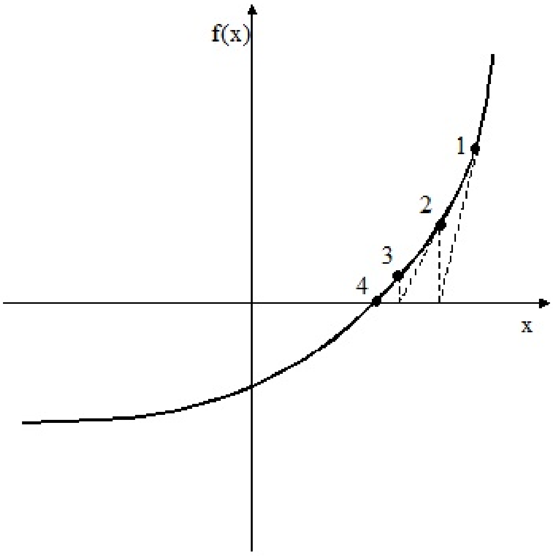


Figure 1: One dimensional Newton-Raphson method.

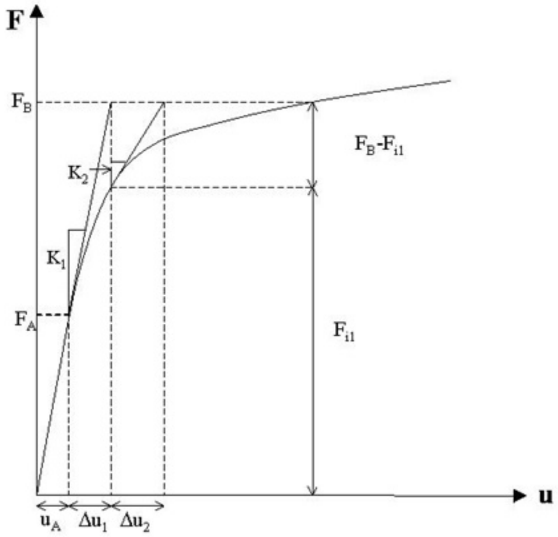


Figure 2: The Newton-Raphson method in Finite Elements.

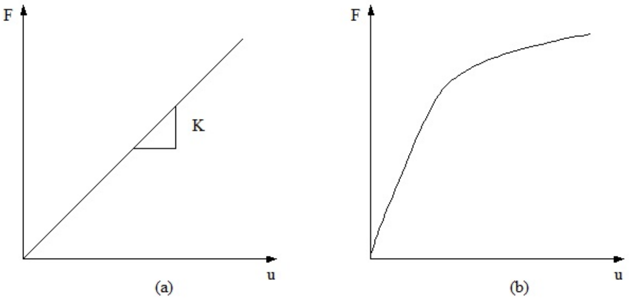


Figure 3: (a) Linear problem (b) Non linear problem.

## 2.1 The Newton-Raphson method in a finite element context

The use of the Newton-Raphson procedure is not limited to one dimensional problems as the one presented at the beginning of this section. The procedure is perfectly applicable to multidimensional problems such as the ones arising from the application of the finite element method.

Let us consider a solid mechanics problem defined in a domain $\Omega$ to be solved through a displacement formulation of the finite element method based on small deformation theory. Let us assume also that the problem is non linear, that is, the relationship between force and displacement is, in graphical terms of the type sketched in figure 3.b as opposed to a linear relationship of the type presented in figure 3.a. In figure 3.b, as the loading process advances and additional gauss points in different elements enter the non linear range the curve becomes less and less steep. In more rigorous and mathematical terms, the objective of the problem could be stated as finding a displacement vector $\mathbf{u}$ such that the following set of equations is satisfied:

$$A_{e=1,nelem}\left[\int_{\Omega e}\mathbf{B}^T\sigma(\epsilon(\mathbf{u}))d\Omega_e\right]-\mathbf{f}^{ext}=\mathbf{0} \tag{6}$$

where :

$A_{e=1,nelem}$= Assembly operator to obtain global variables from element variables.

$\mathbf{f}^{ext}$ = Global external force vector

$\int_{\Omega e}\mathbf{B}^T\sigma(\epsilon(\mathbf{u}))d\Omega_e$ =Element internal force vector

$\mathbf{B}$ = Matrix relating deformations and displacement in the following way: $\epsilon=\mathbf{Bu}$

$\sigma$ = Stress vector. The relationship between $\sigma$ and $\epsilon$ is non linear

Equation (20) may be also expressed in a more compact form as :

$$\mathbf{\Psi}(\mathbf{u})=\mathbf{0}$$

with $\mathbf{\Psi}(\mathbf{u})$ = Residual force vector = $A_{e=1,nelem}\left[\int_{\Omega e}\mathbf{B}^T\sigma(\epsilon(\mathbf{u}))d\Omega_e\right]-\mathbf{f}^{ext}$

The problem, as stated is a non linear problem in $\mathbf{u}$ and as such, an iterative procedure should be used to solve it. Using the Newton-Raphson to do so a displacement state $\mathbf{u}_o$ may be assumed to exist such that $\mathbf{\Psi}_o=\mathbf{\Psi}(\mathbf{u}_o)\neq\mathbf{0}$. A new displacement state $\mathbf{u}_n=\mathbf{u}_o+\delta\mathbf{u}$ is looked for such that $\mathbf{\Psi}_n=\mathbf{\Psi}(\mathbf{u}_n)=\mathbf{0}$. Using a two term Taylor expansion as in the one dimensional case:

$$\mathbf{\Psi}_n=\mathbf{\Psi}_o+\frac{\partial\mathbf{\Psi}}{\partial\mathbf{u}}\delta\mathbf{u}=\mathbf{\Psi}_o+\mathbf{K}_t\delta\mathbf{u}=\mathbf{0} \tag{7}$$

where :

$$
\begin{aligned}
\frac{\partial \boldsymbol{\Psi}}{\partial \mathbf{u}} &= \frac{\partial (A_{e=1,nelem}\left[\int_{\Omega e}\mathbf{B}^T\sigma(\epsilon)d\Omega_e\right])}{\partial \mathbf{u}} = A_{e=1,nelem}\left[\int_{\Omega e}\mathbf{B}^T\frac{\partial\sigma(\epsilon)}{\partial \mathbf{u}}d\Omega e\right] \\
&= A_{e=1,nelem}\left[\int_{\Omega e}\mathbf{B}^T\frac{\partial\sigma(\varepsilon)}{\partial\varepsilon}\frac{\partial\varepsilon}{\partial \mathbf{u}}d\Omega e\right] = A_{e=1,nelem}\left[\int_{\Omega e}\mathbf{B}^T\mathbf{D}^{ep}\mathbf{B}d\Omega e\right] = \mathbf{K}_t
\end{aligned}
$$

Expression (21) would lead to the following linear equation set:

$$\mathbf{K}_t\delta\mathbf{u} = -\boldsymbol{\Psi}_o$$

Solving for $\delta\mathbf{u}$ :

$$\delta\mathbf{u} = -\mathbf{K}_t^{-1}\boldsymbol{\Psi}_o$$

In this format the problem is load controlled, that is external forces $\mathbf{f}^{ext}$ are fixed and displacements $\mathbf{u}$ are the unknowns. In this context, the Newton-Raphson procedure is usually applied in an incremental fashion, that is, the load vector $\mathbf{f}^{ext}$ is divided into increments. The iterative procedure sketched in figure 2, is applied at each increment, in the following way:

**1)** Let us assume a displacement state $\mathbf{u}_A$ exists such that $\boldsymbol{\Psi}_A$= $\mathbf{0}$.This means the internal forces $\mathbf{f}_A^{int}$are in equilibrium with external forces $\mathbf{f}_A^{ext}$. Pressumably this state was reached by a Newton-Raphson procedure which converged after several iterations. The convergence condition was:

$$\frac{\|\boldsymbol{\Psi}_A\|}{\|\mathbf{f}_A^{ext}\|} = \frac{\left\|\mathbf{f}_A^{ext} - \mathbf{f}_A^{int}\right\|}{\|\mathbf{f}_A^{ext}\|} < tol \tag{8}$$

where $\|\|$ is the vector norm operator and $tol$ is a tolerance factor previously defined by the analyst.

**2)** We are now looking for displacements $\mathbf{u}_B$ associated to a load vector $\mathbf{f}_B^{ext}$. The next load increment to be applied is therefore $\mathbf{f}_B^{ext} - \mathbf{f}_A^{ext}$. Since we know $\mathbf{u}_A$ we are really looking for a displacement corrector $\Delta\mathbf{u}$ such that:

$$\mathbf{u}_B = \mathbf{u}_A + \Delta\mathbf{u}$$

Our first predictor for $\Delta\mathbf{u}$ will come from solving the following linear equation system:

$$\mathbf{f}_B^{ext} - \mathbf{f}_A^{ext} = \mathbf{K}_1\delta\mathbf{u}_1$$

where $\mathbf{K}_1$ is the stiffness matrix at $A$, that is $\mathbf{K}_1 = \mathbf{K}_A$

**3)** Once $\delta\mathbf{u}_1$ has been obtained, and the displacement vector updated $\mathbf{u}_1 = \mathbf{u}_A + \delta\mathbf{u}_1$ the internal force vector $\mathbf{f}_1^{int}(\mathbf{u}_1)$ is evaluated. To do so stresses at all Gauss points will have to be integrated to compute the following integral:

$$\mathbf{f}_1^{int}(\mathbf{u}_1) = \int \mathbf{B}^T \sigma(\epsilon(\mathbf{u}_1)) d\Omega$$

**4)** The next step is to evaluate the residual force vector $\mathbf{G}_1$ to check whether the convergence test is satisfied or not:

$$\frac{\|\mathbf{\Psi}_1\|}{\|\mathbf{f}_B^{ext}\|} = \frac{\|\mathbf{f}_B^{ext} - \mathbf{f}_1^{int}\|}{\|\mathbf{f}_B^{ext}\|} < tol \tag{9}$$

If the previous condition is satisfied the iterative process ends at this point.

**5)** If the convergence test is not satisfied a new stiffness matrix $\mathbf{K}_2$ will have to be obtained based on the new displacement state $\mathbf{u}_1$ to solve the following system of equations:

$$-\mathbf{\Psi}_1 = \mathbf{K}_2 \delta\mathbf{u}_2$$

In general $\mathbf{K}_2$ will be different from $\mathbf{K}_1$ since most likely there will be Gauss points that were elastic for $\mathbf{u}_A$ but have become plastic for $\mathbf{u}_1$. The iterative procedure will continue until at the $i^{th}$ iteration $\mathbf{f}_i^{int}$ is evalueated and the convergence condition $\frac{\|\mathbf{f}_B^{ext} - \mathbf{f}_i^{int}\|}{\|\mathbf{f}_B^{ext}\|}$ satisfied.

Norm operators used to evaluate the convergence condition can be defined in different ways. Typically the square root of the sum of squares is used:

$$\|\mathbf{v}\| = \sqrt{v_i^2}$$

but other definitions may be used such as:

$$\|\mathbf{v}\| = Max(v_i)$$

In the present version of the Newton-Raphson algorithm the convergence criterion was based on the norm of the residual force vector, but other criterions may be used. One of the most frequently used criteria, apart from the residual force one, is the iterative displacement correction:

$$\frac{\|\delta\mathbf{u}_i\|}{\|\mathbf{\Delta u}_i\|} < tol$$

where $\delta\mathbf{u}_i$ is the iterative displacement correction and $\mathbf{\Delta u}_i$ is the incremental displacement correction, both corresponding to the $i^{th}$ iteration:

$$\begin{aligned} \mathbf{\Delta u}_i &= \mathbf{\Delta u}_{i-1} + \delta\mathbf{u}_i \\ \mathbf{u}_i &= \mathbf{u}_A + \Delta\mathbf{u}_i \end{aligned}$$

Finally, another frequently used convergence criterion is the residual energy:

$$\frac{\mathbf{\Delta u}_i \cdot \mathbf{\Psi}_i}{\mathbf{\Delta u}_i \cdot \mathbf{f}^{ext}} < tol$$

As in the previous criterion the $_i$ subindex refers to the $i^{th}$ iteration

Although the most usual criterion is the residual force, it is advisable to use an additional criterion, typically the iterative displacement, and stop the iterative process when both criteria are satisfied.

The version of the Newton-Raphson method presented here corresponded to a single field static problem. The method can of course be applied in a more general context to multifield or time dependent problems. For example non linear equations a general problem formulated in displacements and pressures would be stated as:

$$\mathbf{\Psi}(\mathbf{u},\mathbf{p}) = \left\{ \begin{array}{c} \mathbf{\Psi}_u(\mathbf{u},\mathbf{p}) \\ \mathbf{\Psi}_p(\mathbf{u},\mathbf{p}) \end{array} \right\} = 0$$

or in a more compact manner:

$$\mathbf{\Psi}(\mathbf{x}) = \mathbf{0}$$

with:

$$\mathbf{x} = \left\{ \begin{array}{c} \mathbf{u} \\ \mathbf{p} \end{array} \right\}$$

The general Newton-Raphson equation would now be:

$$\mathbf{\Psi}_n = \mathbf{\Psi}_o + \frac{\partial \mathbf{\Psi}}{\partial \mathbf{x}} \delta \mathbf{x} = \mathbf{\Psi}_o + \mathbf{J}\delta\mathbf{x} = \mathbf{0} \tag{10}$$

with:

$$\delta\mathbf{x} = \left\{ \begin{array}{c} \delta\mathbf{u} \\ \delta\mathbf{p} \end{array} \right\} \qquad \mathbf{J} = \left[ \begin{array}{cc} \frac{\partial \mathbf{\Psi}_u}{\partial \mathbf{u}} & \frac{\partial \mathbf{\Psi}_u}{\partial \mathbf{p}} \\ \frac{\partial \mathbf{\Psi}_p}{\partial \mathbf{u}} & \frac{\partial \mathbf{\Psi}_p}{\partial \mathbf{p}} \end{array} \right]$$

If the problem were both mixed and time dependent, with first time derivatives, we would have:

$$\mathbf{\Psi}_{n+1}(\mathbf{u}_{n+1}, \dot{\mathbf{u}}_{n+1}, \mathbf{p}_{n+1}, \dot{\mathbf{p}}_{n+1}) = \left\{ \begin{array}{c} \mathbf{\Psi}_u(\mathbf{u_{n+1}}, \dot{\mathbf{u}}_{\mathbf{n+1}}, \mathbf{p_{n+1}}, \dot{\mathbf{p}}_{\mathbf{n+1}}) \\ \mathbf{\Psi}_p(\mathbf{u_{n+1}}, \dot{\mathbf{u}}_{\mathbf{n+1}}, \mathbf{p_{n+1}}, \dot{\mathbf{p}}_{\mathbf{n+1}}) \end{array} \right\} = 0$$

Assuming a time integration scheme of the following type:

$$\begin{array}{l} \mathbf{u}_{n+1} = \mathbf{u}_{n+1}(\mathbf{u}_n, \dot{\mathbf{u}}_n, \Delta\dot{\mathbf{u}}_n) \\ \dot{\mathbf{u}}_{n+1} = \dot{\mathbf{u}}_{n+1}(\dot{\mathbf{u}}_n, \Delta\dot{\mathbf{u}}_n) \\ \mathbf{p}_{n+1} = \mathbf{p}_{n+1}(\mathbf{p}_n, \dot{\mathbf{p}}_n, \Delta\dot{\mathbf{p}}_n) \\ \dot{\mathbf{p}}_{n+1} = \dot{\mathbf{p}}_{n+1}(\dot{\mathbf{p}}_n, \Delta\dot{\mathbf{p}}_n) \end{array}$$

the Newton-Raphson equation would now be:

$$\mathbf{\Psi}_{n+1,n} = \mathbf{\Psi}_{n+1,o} + \frac{\partial \mathbf{\Psi}}{\partial \mathbf{x}} \delta \mathbf{x} = \mathbf{\Psi}_o + \mathbf{J} \delta \mathbf{x} = \mathbf{0} \tag{11}$$

with:

$$\delta \mathbf{x} = \left\{ \begin{array}{c} \delta(\Delta \dot{\mathbf{u}}_n) \\ \delta(\Delta \dot{\mathbf{p}}_n) \end{array} \right\} \qquad \mathbf{J} = \left[ \begin{array}{cc} \frac{\partial \mathbf{\Psi}_u}{\delta(\Delta \dot{\mathbf{u}}_n)} & \frac{\partial \mathbf{\Psi}_u}{\delta(\Delta \dot{\mathbf{p}}_n)} \\ \frac{\partial \mathbf{\Psi}_p}{\delta(\Delta \dot{\mathbf{u}}_n)} & \frac{\partial \mathbf{\Psi}_p}{\delta(\Delta \dot{\mathbf{p}}_n)} \end{array} \right]$$

## 2.2 Modified Newton-Raphson

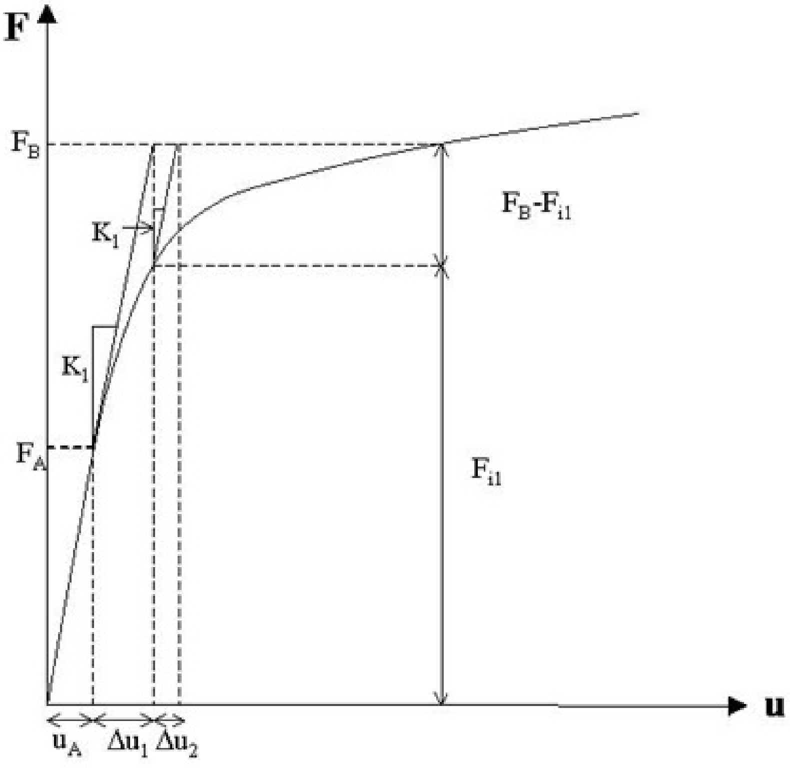


Figure 4: Modified Newton-Raphson method.

An alternative version of the method presented in the previous section is the method known as modified Newton-Raphson. The difference between this version and the previous one is that the stiffness matrix is only updated at the beginning of each increment or at least much less frequently than once every iteration. A single update per increment would imply, in graphical terms that tangents for consecutive iterations are paralel as seen in the sketch of the algorithm presented in figure 4. This updating strategy saves many numerical operations both at the element and global levels. Assuming a classical Gauss LU decomposition solver is being used a full Newton-Raphson would require a decomposition and a backsubstitution every iteration while the modified version would only require a single decomposition at the beginning of each increment and a back substitution at every iteration. This refers only to global level operations. If in addition to these, we consider the operations required to update the element stiffness matrices the computational cost per iteration is reduced very significantly.

On the other hand, assuming the tolerance factor is the same, the convergence speed would be slower, so the number of iterations required to converge for the modified Newton-Raphson is higher than for the full version.

## 2.3 Quasi Newton methods

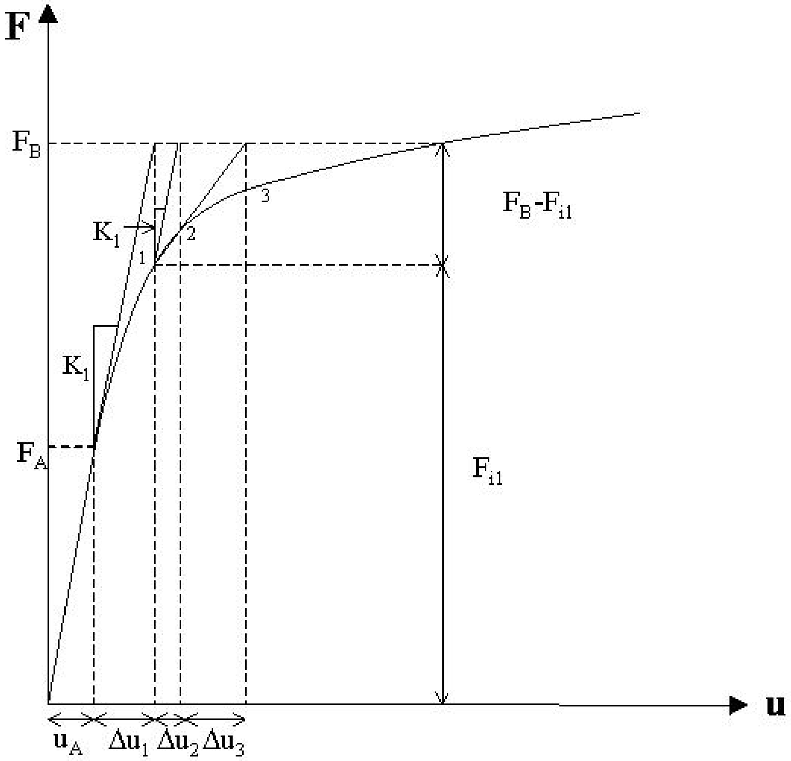


Figure 5: Quasi Newton methods.

Another very well known family of algorithms arising from the Newton-Raphson method is that of the secant or Quasi-Newton methods. In this type of algorithm the displacement of each iteration is obtained based on a secant approximation of the displacements from the two previous iterations as seen in figure 5. From the algorithmic point of view this implies that the new stiffness matrix should satisfy the so called secant condition:

$$\mathbf{\Psi}(\mathbf{u}_i) - \mathbf{\Psi}(\mathbf{u}_{i-1}) = \mathbf{K}_i(\mathbf{u}_i - \mathbf{u}_{i-1}) \qquad (12)$$

or in a more compact manner:

$$\mathbf{r}_i = \mathbf{K}_i \delta_i$$

where:

$$\mathbf{r}_i = \mathbf{\Psi}(\mathbf{u}_i) - \mathbf{\Psi}(\mathbf{u}_{i-1}) \qquad \delta_i = \mathbf{u}_i - \mathbf{u}_{i-1}$$

The stiffness matrix of each iteration is computed in a recursive fashion from the previous iteration's stiffness matrix. There are many ways to do this. The simplest way is through a rank one update such as:

$$\mathbf{K}_i = \mathbf{K}_{i-1} + \frac{(\mathbf{r}_i - \mathbf{K}_{i-1}\delta_i)\mathbf{v}^T}{\mathbf{v}^T \delta_i} \qquad (13)$$

where $\mathbf{v}^T \delta_i \neq 0$. It is straightforward to check that expression (13) satisfies the secant condition (12). A first version of this algorithm is due to Broyden [Bro65] taking $\mathbf{v} = \delta_i$ thus obtaining:

$$\mathbf{K}_i = \mathbf{K}_{i-1} + \frac{(\mathbf{r}_i - \mathbf{K}_{i-1}\delta_i)\delta_i^T}{\delta_i^T \delta_i}$$

A disadvantage of this method is that assuming $\mathbf{K}_{i-1}$ was symmetric, this type of update for $\mathbf{K}_i$ would not preserve $\mathbf{K}_{i-1}$'s symmetry. Davidon's version [GIH80] of the algorithm does preserve matrix symmetry by making $\mathbf{v} = \mathbf{r}_i - \mathbf{K}_{i-1}\delta_i$, thus obtaining:

$$\mathbf{K}_i = \mathbf{K}_{i-1} + \frac{(\mathbf{r}_i - \mathbf{K}_{i-1}\delta_i)(\mathbf{r}_i - \mathbf{K}_{i-1}\delta_i)^T}{(\mathbf{r}_i - \mathbf{K}_{i-1}\delta_i)^T \delta_i}$$

It is also possible to make Rank-two updates such as the one due to Davidon, Fletcher and Powell usually known as DFP [DM77] and the one due to Broyden, Fletcher, Goldfarb and Shanno usually known as BFGS [DM77], both of which preserve symmetry and positive definiteness.

An important advantage of quasi Newton methods is that apart from the computational cost saved on the stiffness matrix updating, as just shown, it is also posible to save on system solving through the so called inverse update. Through this scheme it is only necessary to perform matrix decomposition on the first iteration. The concept of inverse update is based on the classical matrix algebra formula known as Sherman and Morrison's:

$$(\mathbf{A} + \mathbf{b}\mathbf{c}^T)^{-1} = \mathbf{A} - \frac{\mathbf{A}^{-1}\mathbf{b}\mathbf{c}^T\mathbf{A}^{-1}}{1 + \mathbf{c}^T\mathbf{A}^{-1}\mathbf{b}}$$

where $\mathbf{b}$ and $\mathbf{c}$ are arbitrary vectors. The inverse update for Davidon's method based on Sherman and Morrison's formula is:

$$\mathbf{K}_i^{-1} = \mathbf{K}_{i-1}^{-1} + \frac{(\delta_i - \mathbf{K}_{i-1}^{-1}\mathbf{r}_i)(\delta_i - \mathbf{K}_{i-1}^{-1}\mathbf{r}_i)^T}{(\delta_i - \mathbf{K}_{i-1}^{-1}\mathbf{r}_i)^T\mathbf{r}_i}$$

or in more compact fashion:

$$\mathbf{K}_i^{-1} = \mathbf{K}_{i-1}^{-1} + a_i\mathbf{w}\mathbf{w}^T \tag{14}$$

where:

$$\mathbf{w} = \delta_i - \mathbf{K}_{i-1}^{-1}\mathbf{r}_i \qquad a_i = \tfrac{1}{\mathbf{w}^T\mathbf{r}_i}$$

Inverse updating does not preserve matrix sparsity. It is therefore necessary, if advantage is to be taken from this feature, to solve the equation system by matrix-vector multiplication using expression 14 in a recursive fashion down to the first iteration:

$$\mathbf{K}_i^{-1} = \mathbf{K}_0^{-1} + \sum_{k=1}^{i} a_k\mathbf{v}_k\mathbf{v}_k^T$$

where:

$$\mathbf{v}_i = \delta_i - \mathbf{K}_0^{-1}\mathbf{r}_i - \sum\nolimits_{k=1}^{i-1} a_k\mathbf{v}_k\mathbf{v}_k^T\mathbf{r}_i$$

A large number of quasi-Newton updates may cause an ill conditioned iteration matrix. It is therefore recommended to restart the iteration procedure either using the initial stiffness matrix $\mathbf{K}_0$ or obtaining a new stiffness matrix [GIH80]. Details on rank two inverse updates may be obtained in [GIH80] and [DM77].

## 2.4 Displacement control

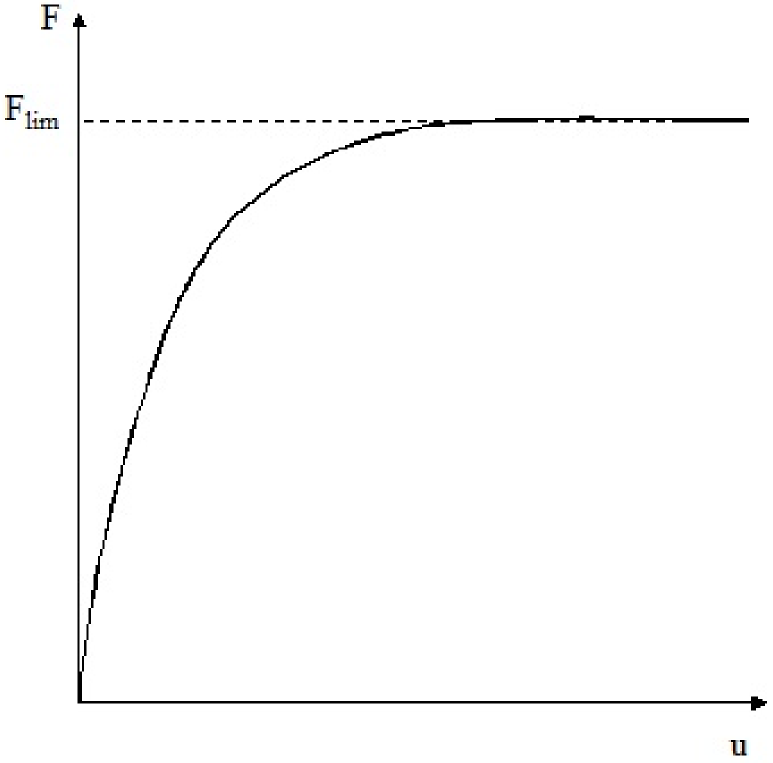


Figure 6: Limit load problem.

Non linear constitutive models such as those associated to elastic-perfectly plastic behaviour will frequently give rise to force-displacement diagrams such as the one in figure 6. In this case the load control presented at the beginning section 2.1 would obviously not converge. In such cases a displacement control would solve the problem. The first step would be to partition the problem into a first part including active degrees of freedom and a second part corresponding to restricted degrees of freedom. Thus, the global vectors would be of the following type:

$$\mathbf{\Psi}(\mathbf{u}^1,\mathbf{u}^2)=\left\{\begin{array}{c}\mathbf{\Psi}^1(\mathbf{u}^1,\mathbf{u}^2)\\ \mathbf{\Psi}^2(\mathbf{u}^1,\mathbf{u}^2)\end{array}\right\}\quad \mathbf{u}=\left\{\begin{array}{c}\mathbf{u}^1\\ \mathbf{u}^2\end{array}\right\}\quad \mathbf{f}^{ext}=\left\{\begin{array}{c}\mathbf{0}\\ \mathbf{f}^{ext,2}\end{array}\right\}$$

This means that $\mathbf{u}^1$ and $\mathbf{f}^{ext,2}$ would be the unknowns and $\mathbf{u}^2$ would be given.

$$A^1_{e=1,nelem}\left[\int \mathbf{B}^T\sigma(\epsilon(\mathbf{u}^1,\mathbf{u}^2))d\Omega_e\right] \qquad = \quad \mathbf{0} \tag{15}$$

$$A^2_{e=1,nelem}\left[\int \mathbf{B}^T\sigma(\epsilon(\mathbf{u}^1,\mathbf{u}^2))d\Omega_e\right] -\mathbf{f}^{ext,2} \quad = \quad \mathbf{0} \tag{16}$$

where $A^1_{e=1,nelem}$ is the assembly operator associated to the active degrees of freedom and $A^2_{e=1,nelem}$ is the one corresponding to the restricted degrees of freedom. The jacobian of this equation set would be:

$$\mathbf{J}=\begin{bmatrix} \frac{\partial\mathbf{\Psi}^1}{\partial\mathbf{u}^1} & \frac{\partial\mathbf{\Psi}^1}{\partial\mathbf{u}^2} \\ \frac{\partial\mathbf{\Psi}^2}{\partial\mathbf{u}^1} & \frac{\partial\mathbf{\Psi}^2}{\partial\mathbf{u}^2} \end{bmatrix} = \begin{bmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{bmatrix}$$

1) Let us assume a displacement state $\mathbf{u}_A$ exists such that $\mathbf{\Psi}_A$= $\mathbf{0}$.This means the internal forces $\mathbf{f}_A^{int}$are in equilibrium with external forces $\mathbf{f}_A^{ext}$. Pressumably this state was reached by a Newton-Raphson procedure which converged after several iterations.

2) We are now looking for displacements $\mathbf{u}^1_B = \mathbf{u}^1_A + \Delta\mathbf{u}^1$ associated to a given displacement vector $\mathbf{u}^2_B = \mathbf{u}^2_A + \Delta\mathbf{u}^2$. Displacement increment $\Delta\mathbf{u}^2$ is given, $\Delta\mathbf{u}^1$ is unknown and it should be such that:

$$\mathbf{\Psi}(\mathbf{u}^1_B,\mathbf{u}^2_B) = \mathbf{0}$$

Since $\mathbf{f}^{ext,2}$ are reactions, and are also unknowns, the solution strategy should be to solve the first $\mathbf{\Psi}^1(\mathbf{u}^1_B,\mathbf{u}^2_B) = \mathbf{0}$ obtaining $\Delta\mathbf{u}^1$ and substituting $\Delta\mathbf{u}^1$ and $\Delta\mathbf{u}^2$ into $\mathbf{\Psi}^2(\mathbf{u}^1_B,\mathbf{u}^2_B) = \mathbf{0}$ would automatically produce $\mathbf{f}^{ext,2}$.Our first predictor for $\Delta\mathbf{u}^1$shall be called $\Delta\mathbf{u}^1_1$ and will come from solving the following equation system:

$$\mathbf{\Psi}^1_1 = \mathbf{\Psi}^1_A + \begin{bmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \end{bmatrix} \begin{Bmatrix} \Delta\mathbf{u}^1_1 \\ \Delta\mathbf{u}^2_1 \end{Bmatrix} = \mathbf{0}$$

In other words:

$$\mathbf{K}^{11}_1\Delta\mathbf{u}^1_1 = -\mathbf{K}^{12}_1\Delta\mathbf{u}^2_1$$

where $\mathbf{K}^{11}_1$ and $\mathbf{K}^{12}_1$ are stiffness matrices at stress state $A$

3) Once $\Delta\mathbf{u}^1_1$ has been obtained, and displacement vectors updated through $\mathbf{u}^1_1 = \mathbf{u}^1_A + \Delta\mathbf{u}^1_1$ and $\mathbf{u}^2_1 = \mathbf{u}^2_A + \Delta\mathbf{u}^2_1$, internal force vectors $\mathbf{f}_1^{int,1}(\mathbf{u}^1_1,\mathbf{u}^2_1)$ and $\mathbf{f}_1^{int,2}(\mathbf{u}^1_1,\mathbf{u}^2_1)$ are evaluated. To do so stresses at all Gauss points will have to be integrated to compute the following integrals:

$$\mathbf{f}_1^{int,1}(\mathbf{u}^1_1,\mathbf{u}^2_1) = A^1_{e=1,nelem}\left[\int_{\Omega_e} \mathbf{B}^T\sigma(\epsilon(\mathbf{u}^1_1,\mathbf{u}^2_1))d\Omega_e\right]$$

$$\mathbf{f}_1^{int,2}(\mathbf{u}_1^1, \mathbf{u}_1^2) = A^2_{e=1,nelem}\left[\int_{\Omega_e} \mathbf{B}^T \sigma(\epsilon(\mathbf{u}_1^1, \mathbf{u}_1^2)) d\Omega_e\right]$$

Since $\mathbf{f}^{ext,2}$ are reactions it will also be necessary to update them:

$$\mathbf{f}_1^{ext,2} = \mathbf{f}_1^{int,2}(\mathbf{u}_1^1, \mathbf{u}_1^2) \tag{17}$$

4) The next step is check whether the convergence test is satisfied or not:

$$\frac{\|\boldsymbol{\Psi}_1\|}{\|\mathbf{f}_1^{ext}\|} = \frac{\|\mathbf{f}_1^{ext} - \mathbf{f}_1^{int}\|}{\|\mathbf{f}_1^{ext}\|} < tol \tag{18}$$

It is interesting to remark that only $\boldsymbol{\Psi}_1^1$ contributes to $\boldsymbol{\Psi}_1$ since $\boldsymbol{\Psi}_1^2$ is $\mathbf{0}$ due to equation (17) and only reactions stored in $\mathbf{f}_1^{ext,2}$ contribute to $\mathbf{f}_1^{ext}$.If the convergence condition is satisfied the iterative process ends at this point.

5) If the convergence test is not satisfied a new stiffness matrix $\mathbf{K}_2^{11}$ will have to be obtained based on the new displacement state $\mathbf{u}_1$ to solve the following system of equations:

$$-\boldsymbol{\Psi}_1^1 = \mathbf{K}_2^{11} \delta \mathbf{u}_2^1$$

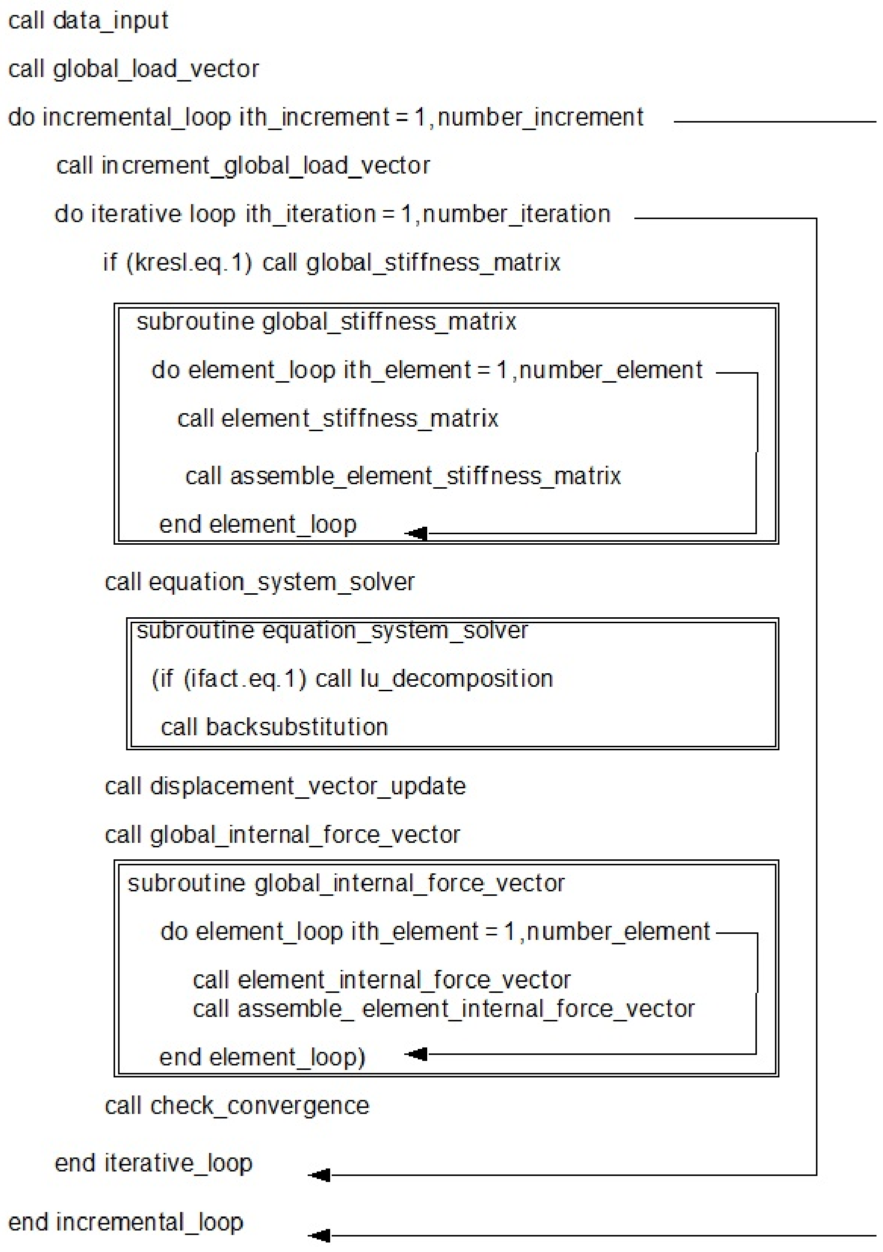
The iterative procedure will continue until at the $i^{th}$ iteration $\mathbf{f}_i^{int}$ is evaluated and the convergence condition satisfied

convergence is lost.

## 2.5 Implementation of the Newton-Raphson method in a FEM code

Implementation of the Newton-Raphson method in a simple finite element code would require an algorithmic program structure of the type shown in figure 7. Of course, there are different ways to do this, and the structure of a non linear finite element program does not have to reproduce exactly the one shown in figure 7. But the main ingredients are the ones shown there, and in any case the one presented there is a very typical non linear finite element program structure.

One of the main aspects of the program structure is a double loop consisting of an iterative loop nested inside an incremental loop. The incremental loop refers to loading increments in which the total external load is divided. The iterative loop is strictly speaking the Newton-Raphson procedure.



```
call data_input
call global_load_vector
do incremental_loop ith_increment = 1,number_increment
    call increment_global_load_vector
    do iterative loop ith_iteration = 1,number_iteration
        if (kresl.eq.1) call global_stiffness_matrix
          subroutine global_stiffness_matrix
            do element_loop ith_element = 1,number_element
              call element_stiffness_matrix
              call assemble_element_stiffness_matrix
            end element_loop
        call equation_system_solver
          subroutine equation_system_solver
            (if (ifact.eq.1) call lu_decomposition
            call backsubstitution
        call displacement_vector_update
        call global_internal_force_vector
          subroutine global_internal_force_vector
            do element_loop ith_element = 1,number_element
              call element_internal_force_vector
              call assemble_ element_internal_force_vector
            end element_loop)
        call check_convergence
    end iterative_loop
end incremental_loop
```

Figure 7: Non linear finite element program structure based on the Newton-Raphson method

It shall be assumed that the linear equation solving algorithm used is of the direct type, and more specifically a sparse Gauss LU decomposition, which is very frequently used in the finite element comunity. In this context, a complete equation solving procedure typically includes matrix decomposition and backsubstitution. As a consequence of this assumptions, the stiffness matrices and internal force vectors at the element level, have to be assembled into the corresponding global matrices and vectors. From the point of view of computational cost the iterative loop is divided into three main parts:

**1)** Computation of the global stiffness matrix. This task includes computation of tangent stiffness at each gauss point, element level integration of the stiffness matrix, and assemblage into global stiffness matrix

**2)** Linear equation system solving.

**3)** Computation of internal force vector. This task includes stress integration at each gauss point, element level integration of internal force vector and assemblage into global internal force vector.

Control variables are introduced in the iterative loop to decide whether a new stiffness matrix should be computed and assembled and to decide whether both matrix decomposidion and back substituion or just the latter of these tasks should be performed. These decisions depend on whether the present iteration is the first one or not, and on the algorithm used is a full or a modified Newton-Raphson. If the algorithm were a quasi Newton a few changes would have to be made in the stiffness matrix updating and linear equation solving tasks.

Finally, the iterative loop includes a convergence test to check whether convergence has beeen reached or on the contrary the iterative process should continue.

# 3 The Arc-length method

The simplest version of the Newton-Raphson method applied to the finite element method in the context of deformable solid mechanics is usually known as load control. These terms refer to the fact that at the beginning of each step the load level is fixed. Having fixed the value of the load vector an iterative process is carried out to compute the displacements at the nodes. This iterative process ends when equilibrium between internal and external forces is satisfied. This version of the Newton-Raphson method was presented in the previous chapter.

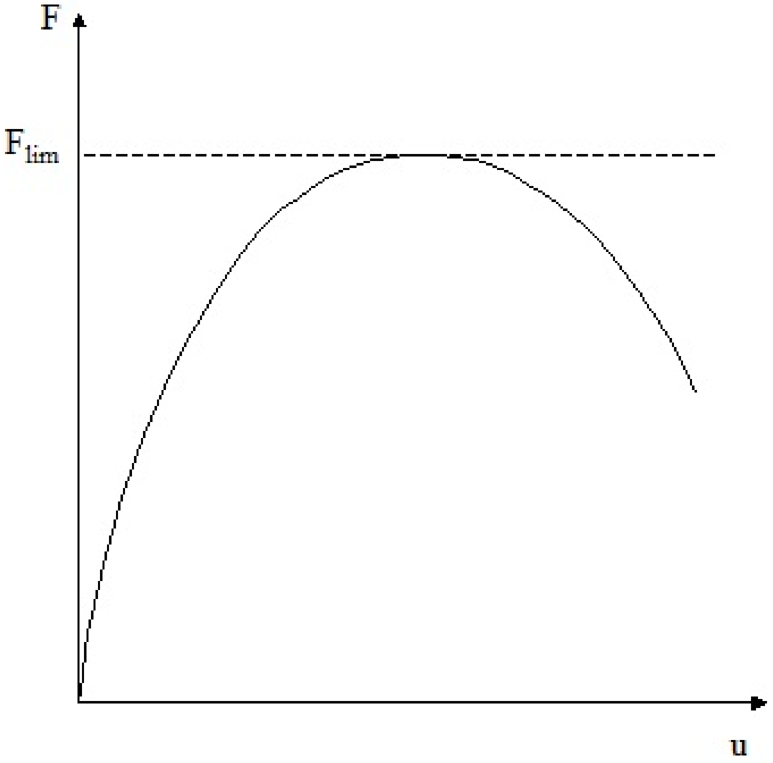


Figure 8: Limit load problem with descending post peak behaviour.

However, in problems that exhibit a force-displacement diagram such as the one of figure 8, this type of algorithm is not adequate since the limit load or upper bound of this curve is not known a priori. If the load control method were used to solve one of such problems, since the value of the limit load is not known a priori, and the load level at the beginning of each step is increased up to a certain level, a step would ultimately be reached where the value of the load vector would be higher than the limit load. At this point it would be impossible to make the iterative process converge because the norm of the residual forces would always have a finite value. In graphical terms, assuming that the solution at each step is obtained by intersection of the curve and a horizontal line drawn at the level of the fixed load value, if this value were higher than the limit load there would be no such intersection.

Following one step further this graphical type of reasoning, one would expect that fixing the displacement level instead of the force level would be a satisfactory solution for the figure 8 type of problems. However, this is only so for some limit load problems. In an elastoplastic material type context only in problems exhibitting a positive, zero or very slightly negative material hardening coefficient, would a displacement control approach be adequate. This is so because this type of graphical reasoning is only a 2D approximation of a problem which is really multidimensional and is therefore difficult to explain on a piece of paper.

The arc-length control is an interesting alternative method to overcome limit points in problems such as the one presented in figure 8.This method is based on a mixed type of control including forces and displacements.

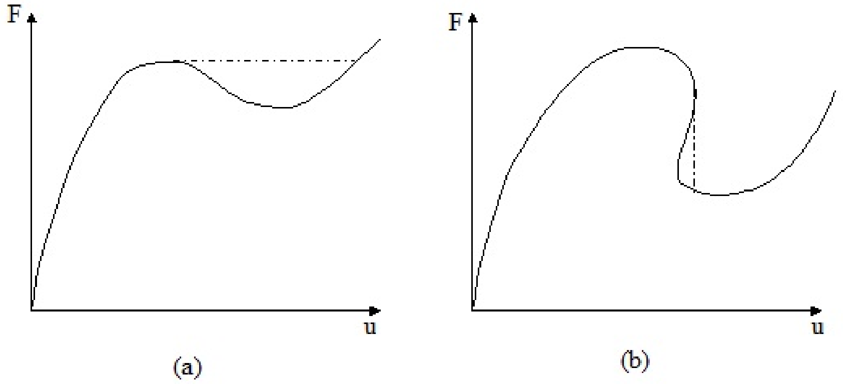


Figure 9: (a) Snap through problem (b) Snap back problem.

Until the arc-length control appeared, standard techniques based on load or displacement control, were able at best to overcome limit points with some difficulties and very frequently simply failed to do so and the process stalled slightly before the limit point. With arc-length control it is possible to overcome limit points in an automatic fashion both for snap-through (figure 9.a) and snap-back (figure 9.b) situations.

Very often, limit points constitute the beginning of structural collapse. One could, therefore, pose the following question: What is the use of obtaining the force-displacement diagram after the collapse of the structure has taken place?. Crisfield [Cri91] produces several answers to this question:

1) It is possible that the point identified as a limit point is not such, and is only a local maximum. The only way to confirm that a point is really a limit point is to overcome it. Frequently, when using standard techniques such as load and displacement control, as the process approaches the limit point serious convergence problems appear. Given impossibility to overcome such points with standard techniques, the assumption is frequently made that convergence problems are caused by the existence of a limit point.

2) It is possible the the point produces only the collapse of a substructure and not the global structural collapse. Some times it is also important to know whether the collapse is brittle or ductile or to investigate the stress state of different structural components.

The arc-length control consists basically of the following steps :

1) The starting point is the standard set of N equilibrium equations which may be expresssed as:

$$\mathbf{f}_{int}(\mathbf{u}) - \lambda \mathbf{f}_{ext} = 0$$

Where :

$N$ = Number of degrees of freedom of the system

$\mathbf{u}$ = N component vector representing displacements of the system.

$\lambda$ =Scalar representing the system load level.

$\mathbf{f}_{int}$ = N component vector containing internal forces of the system =

$$A_{e=1,nelem}\left[\int_{\Omega e} \mathbf{B}^T \sigma(\epsilon(\mathbf{u}))d\Omega_e\right]$$

$\mathbf{f}_{ext}$= N component vector containing external forces of the system .

2) A length restriction is enforced on a norm based on a combination of the displacement and external force incremental vectors.

$$\Delta \mathbf{u}^T \Delta \mathbf{u} + \boldsymbol{\Delta} \lambda^2 \varphi^2 \mathbf{f}_{ext}^T \mathbf{f}_{ext} = \Delta l^2 \tag{19}$$

where:

$\Delta \mathbf{u}$ = Present incremental displacement vector.

$\Delta \lambda$ = Present incremental load level factor

$\varphi$ = A constant with a value set at the beginning of the computations. This factor controls the relative importance of the displacement and external force vectors in the restriction. Further comments will be made on this value in the following sections

There are many versions of the arc-length method. Although the method was originally introduced by Riks[Rik79],[Rik72] and Wempner [Wem71], in the following sections the version due to Crisfield [Cri91] will be presented. One of the main differences between the original versions and Crisfield's is that in the former the restriction is enforced at the same time as the equations yielding a N+1 dimension unsymmetric equation system while in the latter the restriction is enforced after solving a symetric equation system with N unknowns. This, of course, is assuming the N dimension original equilibrium equation system is symetric. Symmetry is evidently a very important issue from the point of view of computational cost.The preservation of symmetry of the global equation system reduces significantly the number of operations and the CPU time.

## 3.1 Analytical description of the method

Before introducing the arc-length method we shall define the residual force vector $\boldsymbol{\Psi}$ as :

$$\boldsymbol{\Psi}(\mathbf{u},\lambda) = \mathbf{f}_{int}(\mathbf{u}) - \lambda \mathbf{f}_{ext} = A_{e=1,nelem} \left[ \int_{\Omega e} \mathbf{B}^T \sigma(\epsilon(\mathbf{u})) d\Omega_e \right] - \lambda \mathbf{f}_{ext} \tag{20}$$

This vector represents a measure of the degree of convergence of the iterative process. Given the scalar $\lambda$ representing the external force factor along with the displacements $\mathbf{u}$, and residual force $\boldsymbol{\Psi}(\mathbf{u},\lambda)$ vectors, the problem consists in finding the displacement and load factor corrections $\delta \mathbf{u}$ and $\delta \lambda$ such that the new residual force vector will be zero.

$$\boldsymbol{\Psi}_n = \boldsymbol{\Psi}_o + \frac{\partial \boldsymbol{\Psi}}{\partial \mathbf{u}} \delta \mathbf{u} + \frac{\partial \boldsymbol{\Psi}}{\partial \lambda} \delta \lambda = \boldsymbol{\Psi}_o + \mathbf{K}_t \delta \mathbf{u} - \mathbf{f}_{ext} \delta \lambda = \mathbf{0} \tag{21}$$

Subindices "o" and "n" refer to the old and new situation respectively, or in other words before and after the present iteration.. It is easy to prove from equation (20) that :

$$\frac{\partial \mathbf{\Psi}}{\partial \mathbf{u}} = \frac{\partial (A_{e=1,nelem}\left[\int_{\Omega e} \mathbf{B}^T \sigma(\epsilon) d\Omega_e\right])}{\partial \mathbf{u}} = A_{e=1,nelem}\left[\int_{\Omega e} \mathbf{B}^T \frac{\partial \sigma(\epsilon)}{\partial \mathbf{u}} d\Omega e\right]$$

$$= A_{e=1,nelem}\left[\int_{\Omega e} \mathbf{B}^T \frac{\partial \sigma(\varepsilon)}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial \mathbf{u}} d\Omega e\right] = A_{e=1,nelem}\left[\int_{\Omega e} \mathbf{B}^T \mathbf{D}^{ep} \mathbf{B} d\Omega e\right] = \mathbf{K}_t$$

$$\frac{\partial \mathbf{G}}{\partial \lambda} = \mathbf{f}_{ext}$$

One must keep in mind that $\delta\mathbf{u}$ and $\delta\lambda$ are iterative corrections while $\Delta\mathbf{u}$ and $\Delta\lambda$ that will appear in further developments along this section, are incremental corrections that come from the accumulation of iterative corrections starting at the beginning of the increment.

Solving for $\delta\mathbf{u}$ in equation (21):

$$\delta\mathbf{u} = -\mathbf{K}_{\mathbf{t}}^{\mathbf{-1}}(\mathbf{\Psi}_o - \delta\lambda\mathbf{f}_{ext})$$

$$\delta\mathbf{u} = -\mathbf{K}_{\mathbf{t}}^{\mathbf{-1}}\mathbf{\Psi}_o + \delta\lambda\mathbf{K}_t^{-1}\mathbf{f}_{ext} = \delta\overline{\mathbf{u}} + \delta\lambda\delta\mathbf{u}_t \tag{22}$$

As one may see in equation (22) $\delta\overline{\mathbf{u}}$ coincides with the classical displacement iterative correction while the term $\delta\lambda\delta\mathbf{u}_t$ allows the procedure to adapt to limit points through a small change in $\delta\lambda$.

It is now possible to update incremental displacements and load factors:

$$\Delta\mathbf{u}_n = \Delta\mathbf{u}_o + \delta\mathbf{u} = \Delta\mathbf{u}_o + \delta\overline{\mathbf{u}} + \delta\lambda\delta\mathbf{u}_t$$

$$\Delta\lambda_n = \Delta\lambda_o + \delta\lambda \tag{23}$$

Enforcing the restriction from equation (19) :

$$(\Delta\mathbf{u}_n^T\Delta\mathbf{u}_n + \boldsymbol{\Delta}\lambda_n^2\varphi^2\mathbf{f}_{ext}^T\mathbf{f}_{ext}) = \Delta l^2 \tag{24}$$

and substituting equation (23) into (24) :

$$a_1\delta\lambda^2 + a_2\delta\lambda + a_3 = 0 \tag{25}$$

where:

$$a_1 = \delta\mathbf{u}_t^T\delta\mathbf{u}_t + [\varphi^2\mathbf{f}_{ext}^T\mathbf{f}_{ext}] \tag{26}$$

$$a_2 = 2\delta\mathbf{u}_t^T(\Delta\mathbf{u}_o + \delta\overline{\mathbf{u}}) + [2\Delta\lambda_o\varphi^2\mathbf{f}_{ext}^T\mathbf{f}_{ext}] \tag{27}$$

$$a_3 = (\Delta\mathbf{u}_o + \delta\overline{\mathbf{u}})^T(\Delta\mathbf{u}_o + \delta\overline{\mathbf{u}}) - \Delta l^2 + [\Delta\lambda_o^2\varphi\mathbf{f}_{ext}^T\mathbf{f}_{ext}] \tag{28}$$

The next step now is to solve (25) for $\delta\lambda$ and choose one of the two possible solutions. The most usual criterion to do so is to choose the $\delta\lambda$ such that the resulting $\Delta\mathbf{u}_n$ is closest to $\Delta\mathbf{u}_o$, or in other words the angle between them is the smallest. It is therefore necessary to compute the cosine of this angle based on the scalar product. The procedure to make the choice can therefore be expressed as :

$$\Delta\mathbf{u}_n^1 = \Delta\mathbf{u}_o + \delta\mathbf{u}_1 {=} \Delta\mathbf{u}_o + \delta\overline{\mathbf{u}} + \delta\lambda_1\delta\mathbf{u}_t$$

$$\Delta\mathbf{u}_n^2 = \Delta\mathbf{u}_o + \delta\mathbf{u}_2 {=} \Delta\mathbf{u}_o + \delta\overline{\mathbf{u}} + \delta\lambda_2\delta\mathbf{u}_t$$

$$\cos\theta_1 = \frac{\Delta\mathbf{u}_o \cdot \Delta\mathbf{u}_n^1}{\|\Delta\mathbf{u}_o\| \, \|\Delta\mathbf{u}_n^1\|}$$

$$\cos\theta_2 = \frac{\Delta\mathbf{u}_o \cdot \Delta\mathbf{u}_n^2}{\|\Delta\mathbf{u}_o\| \, \|\Delta\mathbf{u}_n^2\|}$$

choosing $\delta\lambda_i$ such that $\cos\theta_i = \max(\cos\theta_1, \cos\theta_2)$.

The scheme of the procedure is presented in figure 10

An aspect of the procedure which has not been covered yet the computation of $\Delta\lambda$ y $\Delta l$ at the beginning of the increment. Assuming $\varphi = 0$, the usual way to procede is to fix the value of $\Delta\lambda$ at the beginning of the first increment and compute the equivalent value of $\Delta l$ based on the restriction:

$$\left.\begin{array}{r} \Delta\mathbf{u}_1^T\Delta\mathbf{u}_1 = \Delta l^2 \\ \Delta\mathbf{u}_1 = \mathbf{K_t^{-1}}\Delta\lambda_1\mathbf{f}_{ext} = \Delta\lambda_1\delta\mathbf{u}_t \end{array}\right\} \Longrightarrow \Delta l = \Delta\lambda_1\sqrt{\delta\mathbf{u}_t^T\delta\mathbf{u}_t}$$

The subindex of $\Delta\lambda_1$ and $\Delta\mathbf{u}_1$ refers to the first iteration of of the increment. For the first iteration of increments different from the first one, the order of the procedure would be the contrary, that is obtaining $\Delta\lambda_1$ from $\Delta l$:

$$\Delta\lambda_1 = sign\frac{\Delta l}{\sqrt{\delta\mathbf{u}_t^T\delta\mathbf{u}_t}} \tag{29}$$

The value of variable sign in equation 29 will be decided based on the following criterion:

a) If $\delta\mathbf{u}_t^T\mathbf{K}_t\delta\mathbf{u}_t > 0 \Rightarrow$ before limit point $\Rightarrow sign = +1$

b) If $\delta\mathbf{u}_t^T\mathbf{K}_t\delta\mathbf{u}_t < 0 \Rightarrow$ after limit point $\Rightarrow sign = -1$

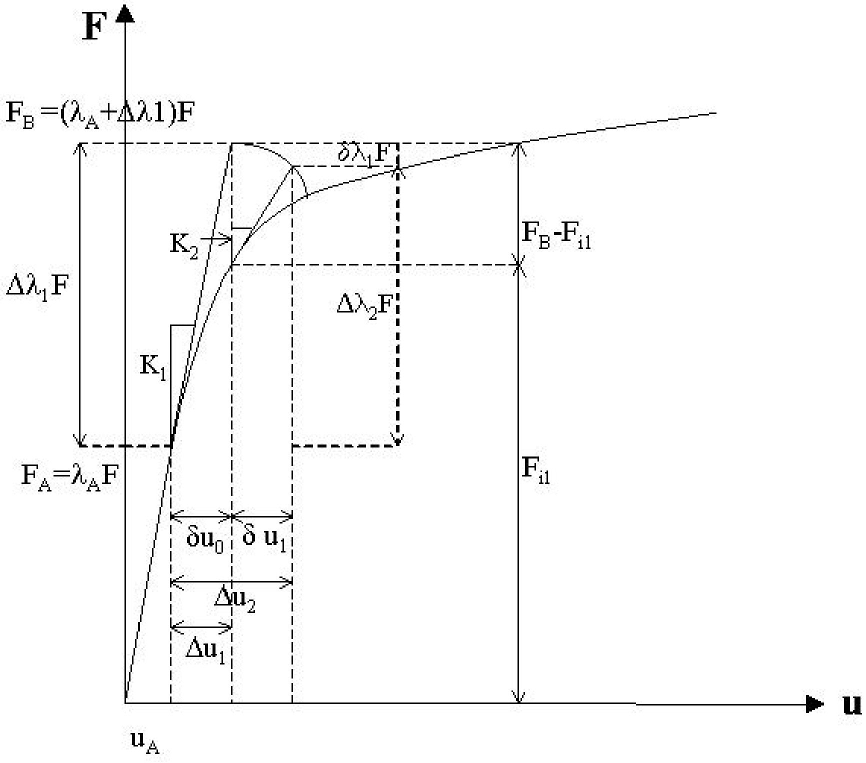


Figure 10: The arc-length method.

Fixing the value of $\Delta l$ in equation (29) is the same type of decision as fixing the value of load steps in a classical load control analysis. Although it is a decission which should be based on the experience of the analyst it is clear that the size of the step should diminish as the loading process advances, since the degree of non linearity usually increases with the loading factor. Although it requires additional programming, usually the best solution is to establish some sort of automatic time stepping algorithm, based on error estimation or some other criterion. Crisfield [Cri91] presents in his text a very simple time stepping algorithm which will in many cases solve the problem in a satisfactory manner. The algorithm is based on the concept of ”desirable number of iterations”. If in the previous increment the number of itereations required to converge has been ”too high” the length of the present increment should be smaller.If in the previous increment the number of itereations required to converge has been ”too small” the length of the present increment should be bigger. This, of course, requires for the analyst to establish what is a ”desirable number of iterations” based on his/her computational experience. Following these ideas, the expression controling the step length would be:

$$\Delta l_j = \Delta l_{j-1} \frac{I_d}{I_{j-1}}$$

where:

$\Delta l_j$ = length of present increment

$\Delta l_{j-1}$ = length of previous increment

$I_d$ = Desirable number of iterations, based on experience, usually around 5.

$I_{j-1}$ = Number of iterations of previous increment.

## 3.2 Alternative versions

### 3.2.1 Spherical versus cylindrical arc-length

The arc-length method as presented in the previous section is usually known as spherical arc-length. The term spherical refers to the fact that restriction (19) includes both forces and displacements. It is a restriction applied in a multidimensional space of forces and displacements, and therefore a spherical restriction. When $\varphi = 0$, the restriction includes only displacements, and is therefore referred to as cylindrical. In fact, the latter version is much more frequently used because it is simpler and the spherical arc-length does not have significant advantages. In equations 26,27 and 28, coefficients for the cylindrical versions may by obtained by omitting the terms in brackets.

Although they will not be covered in detail, one should mention that there are ”linear” versions for both cylindrical and spherical arc-length. These versions are linear in the sense that the arclength restriction is only enforced in an approximate fashion. Using a 2D graphical explanation, a linear approximatiom woul be to obtain the intersection between the tangent to a circumference arc and the force-displacement while the full arc-length restriction would be to obtain the intersection between the said circumference arc and the force-displacement curve.

### 3.2.2 Modified versus full Newton-Raphson

The arc-length method may be implemented following either a full Newton-Raphson algorithm where stiffness matrix is updated at every iteration or a modified Newton-Raphson where the stiffnes matrix is only updated at the first iteration of each increment. The first alternative is obviously more powerful and converges faster but each iteration is computationally more expensive. In spite of the latter disvantage it is usually more reliable than the modified alternative. The updating of vector $\delta\mathbf{u}_t$ is done at the same time as $\mathbf{K}_t$ is updated, at each iteration for the full Newton-Raphson and at the beginning of the increment for the modified version.

### 3.2.3 A displacement control type version of the arc-length method

One the possible versions of the arc-length method, due to Batoz [BD79], is similar although not equivalent to a displacement control. The method consists in applying the the length restriction to a specific degree of freedom, that is:

$$\Delta l^2 = \Delta u_i^2$$

where subindex $i$ refers to the degree of freedom which is being controlled.

This method has the limitation of requiring the certainty that the specific degree of freedom chosen for control varies monotonically during the loading process. This means that it will always have to increase or decrease during the the loading process. The possibility exists that no such degree of freedom exists in the model, that is, all the degrees of freedom in the model swith from an increasing to a decreasing pattern or viceversa. That would be the case in a ”snap-back” problem.

### 3.2.4 Bathe-Dvorkin version

An interesting version of the arc-length method was formulated by Bathe and Dvorkin [BD83] that consists in using a restriction based on a constant external work along the increment. This means that the external work is computed in the firs iteration of the increment according to the following expression:

$$\Delta l = \Delta W = (\lambda + \frac{1}{2}\Delta\lambda_p)\mathbf{f}_{ext}^T\Delta\mathbf{u}_p = \Delta\lambda_p(\lambda + \frac{1}{2}\Delta\lambda_p)\mathbf{f}_{ext}^T\delta\mathbf{u}_t$$

and is forced to remain constant during the subsequent iterations:

$$(\lambda_o + \frac{1}{2}\delta\lambda)\mathbf{f}_{ext}^T\delta\mathbf{u} = (\lambda_o + \frac{1}{2}\delta\lambda)(\mathbf{f}_{ext}^T\delta\overline{\mathbf{u}} + \delta\lambda\mathbf{f}_{ext}^T\delta\mathbf{u}_t) = 0$$

where $\lambda_o$ is the load factor in the previous iteration. This restriction gives rise to second degree equation in $\delta\lambda$, as in the standard arc-length. The difference between this version and the standard arc-length lies in the fact that in this case the roots of the second degree equation are guaranteed to be real. This is why some authors recommend to switch to the Bathe-Dvorkin version when the Crisfield version yields complex solutions.

# 4 Line search methods

As seen in previous sections the Newton-Raphson procedure is based on an iterative process in which the displacement vector correction is:

$$\delta\overline{\mathbf{u}} = -\mathbf{K}_t^{-1}\mathbf{\Psi}$$

where $\mathbf{\Psi}$ is the residual force vector. The line-search method is based on the idea that the optimum prediction for the new displacement vector $\mathbf{u}_n$ is not obtained by simply adding the displacement correction vector $\delta\overline{\mathbf{u}}$ to the old displacement vector $\mathbf{u}_o$ but according to the following expression:

$$\mathbf{u}_n = \mathbf{u}_o + \eta\delta\overline{\mathbf{u}} \tag{30}$$

where $\eta$ is a new parameter obtained by minimization of a certain potential function $\phi$ :

$$\phi_n(\eta + \delta\eta) = \phi_o(\eta) + \frac{\partial\phi}{\partial\eta}\delta\eta + ... = \phi_o + \frac{\partial\phi}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial\eta}\delta\eta + ...$$

The minimization condition for $\phi$ in terms of $\eta$ is:

$$\frac{\partial\phi}{\partial\eta} = 0$$

In the case of non linear elastic problems $\phi$ is taken as the total potential energy which may be precisely define from solid mechanics point of view. This choice is consistent with the classical variational approach to a non linear elastic problem as a total potential energy minimization problem. In this context it is possible to prove that the previous equation may be expressed as:

$$\frac{\partial\phi}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial\eta} = \mathbf{\Psi}(\eta)^T\delta\overline{\mathbf{u}} = 0$$

or similarly as:

$$\delta(\eta) = \frac{\partial \phi}{\partial \eta} = \delta \overline{\mathbf{u}}^T \boldsymbol{\Psi}(\eta) = 0 \tag{31}$$

Strictly speaking this approach can not be used with problems such as elastoplasticity in a materially non linear context. However, for this type of problems it is possible to use an algorithm with an energy approach such as line search, with a potential function of which it is known that the derivative with respect to the displacement vector $\frac{\partial \phi}{\partial \mathbf{u}}$ is the residual force vector $\boldsymbol{\Psi}$. Following this approach and according to equation (31) it is possible to define $\delta_0$ as $\delta(\eta = 0)$ according to the following expression:

$$\delta_0 = \delta(\eta = 0) = \delta \overline{\mathbf{u}}^T \boldsymbol{\Psi}(\eta = 0) = \delta \overline{\mathbf{u}}^T \boldsymbol{\Psi}_0 = -\boldsymbol{\Psi}_0^T \mathbf{K}_t^{-1} \boldsymbol{\Psi}_{\mathbf{0}} = -\delta \overline{\mathbf{u}}^T \mathbf{K}_t \delta \overline{\mathbf{u}}$$

Although the strict minimization condition is the one expressed in (31) from a numerical point of view it is more interesting to enforce a lax or non strict minimization condition such as :

$$|r(\eta)| = \left| \frac{\delta(\eta)}{\delta_0} \right| < \beta \tag{32}$$

where $\beta$ is a constant fixed by the analyst usually between 0.1 and 0.5. To solve equation (32) the following steps are necessary:

$1^o$) Obtain $\delta_0 = \delta \overline{\mathbf{u}}^T \boldsymbol{\Psi}_0$. where $\boldsymbol{\Psi}_0$ is available from the previous iteration.

$2^o$) Compute $\delta_1 = \delta(\eta = 1)$.

$3^o$) Obtain:

$$\eta_2 = \frac{-\eta_1 \delta_0}{\delta_1 - \delta_0} \tag{33}$$

Equation (33) comes from performing a linear interpolation of the value $\eta_2$ between $\eta_0$ and $\eta_1$ by enforcing $\delta(\eta_2) = 0$.

$4^o$) $\delta_2 = \delta(\eta_2)$ is obtained

$5^o$) Check if $\delta_2$ satisfies or not the non strict minimization condition $\left| \frac{\delta(\eta)}{\delta_0} \right| < \beta$

$6^o$) If the answer is :

a) **Yes** : the line search algorithm has ended and displacemnts are updated according to equation (30)

b) **No** : Return to step 3 obtaining $\eta_3 = \frac{-\eta_2 \delta_0}{\delta_2 - \delta_0}$ and continuing with the iterative process until convergence is reached.

As one may see from the described procedure, each iteration in a line search algorithm requires an additional update of the internal force vector, with computational cost that this implies. This is why a non strict line search with a few iterations is usually performed instead of strict minimation line search. The computaional cost of the latter would be prohibitive.

### 4.1 Combination of arc-length and line-search

The application of the line search method in combination with arc-length has also been treated by Crisfield [Cri97]. The combination of these two algorithms implies the application of equation minimization condition (31) to the total iterative displacement in equation (22), that is:

$$\delta = (\delta\overline{\mathbf{u}} + \delta\lambda\delta\mathbf{u}_t)^T\mathbf{\Psi}(\lambda,\eta) = 0 \tag{34}$$

The problem is that every time equation (33) is applied in search of a new value for $\eta$ that approximately satisfies equation (34) the arc-length condition ceases to be satisfied. This means that a new value for $\delta\lambda$ has to be obtained such that:

$$(\Delta\mathbf{u}_0 + \eta(\delta\overline{\mathbf{u}} + \delta\lambda\delta\mathbf{u}_t))^T(\Delta\mathbf{u}_0 + \eta(\delta\overline{\mathbf{u}} + \delta\lambda\delta\mathbf{u}_t)) = \Delta l^2 \tag{35}$$

This gives rise to an additional iterative procedure nested in the algorithm presented in the previous section. As expected equation (35) ends up in a second degree equation in $\delta\lambda$ with similar coefficients to (26),(27) and (28) but including $\eta$ . The additional iterative procedure consists in:

1) Application of equation (33).$\eta$ is obtained.

2) Solution of problem (35). A new value for $\delta\lambda_n$ is obtained.

If $(\frac{|\delta\lambda_n - \delta\lambda_o|}{|\delta\lambda_o|} < tolerancia) \Rightarrow$ End of the iterative process.

Else $\Rightarrow$ return to step 1

## References

[Bat96] KJ Bathe. *Finite element Procedures*. Prentice Hall, New Jersey, 1996.

[BD79] J. L. Batoz and G. Dhatt. Incremental displacement algorithms for non linear problems. *International Journal for Numerical and Analytical Methods in Engineering*, 14:1262–1266, 1979.

[BD83] KJ Bathe and EN Dvorkin. On the automatic solution of nonlinear finite element equations. *Computers and Structures*, 17:871–879, 1983.

[BLM00] T Belytschko, WK Liu, and B Moran. *Non-linear Finite Elements for Continua and Structures*. Wiley, 2000.

[Bro65] CG Broyden. *Mathematics of Computation*, volume 19. 1965.

[Cri91] MA Crisfield. *Non-linear Finite Element Analysis of Solids and Structures. Volume 1: Essentials*. John Wiley and Sons, 1991.

[Cri97] MA Crisfield. *Non-linear Finite Element Analysis of Solids and Structures. Volume 1: Advanced Topics*. John Wiley and Sons, 1997.

[DM77] J. E. Dennis and J. More. Quasi newton methods: motivation and theory. *SIAM Review*, 19:46–89, 1977.

[GIH80] M. Geradin, S. Idelsohn, and M. Hogge. Non linear structural dynamics via newton and quasi newton methods. *Nuclear Engineering Design*, 58:339–348, 1980.

[PTVF92] WH Press, SA Teukolsky, WT Vetterling, and BP Flannery. *Numerical recipes in FORTRAN*. Cambridge University Press,2nd ed, 1992.

[Rik72] E. Riks. The application of newton's method to the problem of elastic stability. *Journal of Applied Mechanics*, 39:1060–1069, 1972.

[Rik79] E. Riks. An incremental approach to the solution of snapping and buckling problems. *International Journal for Solids Structures*, 15:529–551, 1979.

[SH98] J. Simo and T. Hughes. *Computational inelasticity*. Springer, 1998.

[Wem71] G. A. Wempner. Discrete approximations related to non linear theories of solids. *International Journal for Solids Structures*, 7:1581–1599, 1971.

[ZT89] O. C. Zienkiewicz and R. L. Taylor. *The Finite Element Method*, volume 1 and 2. McGraw-Hill, 4th edition, 1989.

# Computational plasticity (II): numerical integration of elastoplastic constitutive equations

**Claudio Tamagnini**[a], **Kateryna Oliynyk**[a,b]

[a] *University of Perugia, Italy*
[b] *University of Dundee, UK*

*This chapter presents an overview of some of the most widely used numerical procedures for the integration of elastoplastic constitutive equations in the context of non–linear Finite Element analysis. The first part of the chapter is devoted to the formulation of the evolution equations and the discussion of the stress–point algorithms for infinitesimal plasticity. The second part focuses on the evolution equations of finite deformation multiplicative plasticity and the corresponding stress–point algorithms. Both the implicit Backward Euler method – based on a two–stage procedure with an elastic predictor problem and a plastic corrector problem – and explicit adaptive schemes with substepping and error control are covered for both infinitesimal and finite deformation plasticity models. This chapter was first published in the lecture notes of the 2021 ALERT School "Constitutive Modelling in Geomaterials".*

## 1 Introduction

In recent years the parallel development of: i) advanced constitutive theories for the mechanical behavior of geomaterials, ii) robust and accurate numerical methods for the solution of partial differential equations, and iii) powerful computer architectures, has led to a radical change in the analysis of geotechnical problems, notably in some areas such as the design of deep excavations or the analysis of complex soil–structure interaction problems where traditional design methods – based on the classical distinction between "failure" and "deformation" problems – are not able to capture the most relevant aspects of the soil–structure system behavior.

A common and almost universal feature of the constitutive models proposed for geomaterials – from those which have now became a standard design tool in geotechnical practice to the ones which were mainly developed for research purposes – is the fact that they are cast in *incremental* form. Rather than providing the state of stress associated to a specific state of strain, they define the *evolution laws* for the state variables. Therefore, the quantitative evaluation of the mechanical effects of a given "load", be it an imposed stress increment, strain increment or a combination of both, requires the solution of an initial value problem, consisting in the *integration* of the constitutive equation along the assigned loading path, with prescribed *initial conditions*. As this task cannot be performed analytically, except in very special cases, the development of a numerical algorithm for this purpose is a crucial part of any computational procedure for the solution of non–linear problems in geomechanics.

More specifically, in the application of numerical methods – such as the finite Element method – to the solution of a non–linear initial/boundary value problem, the following general strategy is usually adopted, see [SH98]:

1. from the original system of governing partial differential equations (PDEs), a non–linear system of algebraic *balance equations* is obtained by the introduction of appropriate space and time discretizations. Such a system is typically solved by adopting an incremental–iterative approach;

2. for any given *global* iteration, the discretized equilibrium equations generate incremental motions, which, in turn, are used to determine the incremental strain history by purely kinematic relationships;

3. for a given strain increment, updated values of the state variables are obtained by integrating numerically the constitutive equations at the *local* level, with given initial conditions; for their local scope, the procedures employed for this task are typically referred to as *stress–point algorithms*;

4. the discrete balance equations are then checked for convergence, and if the convergence criterion is not met, the iteration process is continued by returning to step (2).

As first pointed out by Hughes [Hug84], the integration of the constitutive equation at the local level – *i.e.*, step (3) – represents the central problem of computational plasticity, since it corresponds to the main role played by the constitutive equation in actual computations. There are of course many other important computational ingredients in the overall procedure, but they are particular to the type of solution strategy employed, and involve the constitutive theory only in a limited way, if at all. Moreover, the precision with which the constitutive equations are integrated has a direct impact on the overall accuracy of the analysis.

Since the early works on metal plasticity, summarized in [Hug84], a number of fundamental treatises have been published on this subject. Among them we cite the books of Simo and Hughes [SH98], de Souza Neto *et al.* [dPO11] and the chapter written by Simo [Sim98] for the Handbook of Numerical Analysis.

In this chapter, we present an overview of some of the most widely used stress–point algorithms for the integration of classical and advanced plasticity models for soils, reflecting our personal experience in this field. After a brief description of the notation (Sect. 2), in Sect. 3 we address the main problem of computational plasticity for the case of infinitesimal deformations, summarizing the evolution equations to be integrated and the different numerical procedures for their integration, separating explicit adaptive strategies with error control based on Runge–Kutta methods and the implicit Backward–Euler algorithm, which has now become a standard in computational plasticity. The evolution equations for finite deformation multiplicative plasticity and the corresponding explicit, semi–implicit and implicit integration algorithms are presented in Sect. 4. In both Sect. 3 and 4, particular attention is paid to the definition and the computation of the *consistent tangent stiffness matrix*, which guarantees the asymptotic quadratic convergence of the Newton–Raphson method when it is used for the iterative solution of the discrete equilibrium equations.

# 2 Notation

In the following, all stresses and stress–related quantities are effective, unless otherwise stated. The sign convention of continuum mechanics (traction and extension positive) is adopted throughout. Both direct and index notations will be used to represent vector and tensor quantities according to convenience. In direct notation, vectors and second–order tensors will be represented by boldface italic fonts. Boldface italic fonts and blackboard bold fonts – such as $\boldsymbol{c}^e$ and $\mathbb{C}^e$ – are used to represent fourth–order tensors, according to convenience. Following standard practice, for any two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3$, the dot product is defined as: $\boldsymbol{v} \cdot \boldsymbol{w} := v_i w_i$, and the dyadic product as: $[\boldsymbol{v} \otimes \boldsymbol{w}]_{ij} := v_i w_j$. Accordingly, for any two second–order tensors $\boldsymbol{X}, \boldsymbol{Y}$, $\boldsymbol{X} \cdot \boldsymbol{Y} := X_{ij} Y_{ij}$ and $[\boldsymbol{X} \otimes \boldsymbol{Y}]_{ijkl} := X_{ij} Y_{kl}$. The quantity $\|\boldsymbol{X}\| := \sqrt{\boldsymbol{X} \cdot \boldsymbol{X}}$ denotes the Euclidean norm of the second order tensor $\boldsymbol{X}$, unless otherwise stated.

# 3 Stress–point algorithms for infinitesimal plasticity

## 3.1 Evolution equations

The evolution equations of the theory of infinitesimal plasticity are briefly summarized below. Let $\boldsymbol{\epsilon}$ be the strain tensor and $\boldsymbol{q}$ be the vector (of dimension $n_{\text{int}}$) of the internal state variables accounting for the effects of the previous loading history. Also, let:

$$\mathbb{E} := \left\{ (\boldsymbol{\sigma}, \boldsymbol{q}) \;\middle|\; f(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0 \right\} \tag{1}$$

be the elastic domain, defined through a suitable yield function $f(\boldsymbol{\sigma}, \boldsymbol{q}) = 0$. Taking into account the usual additive decomposition of the strain rate tensor, $\dot{\boldsymbol{\epsilon}}$, into an elastic

($\dot{\boldsymbol{\epsilon}}^e$) and a plastic ($\dot{\boldsymbol{\epsilon}}^p$) part, we have:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^e \left[ \dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p \right] \tag{2}$$

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\gamma} \, \frac{\partial g}{\partial \boldsymbol{\sigma}}(\boldsymbol{\sigma}, \boldsymbol{q}) \tag{3}$$

$$\dot{\boldsymbol{q}} = \dot{\gamma} \boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{q}) \tag{4}$$

subject to the following Kuhn–Tucker complementarity conditions:

$$\dot{\gamma} \geq 0 \, , \quad f(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0 \, , \quad \dot{\gamma} f(\boldsymbol{\sigma}, \boldsymbol{q}) = 0 \tag{5}$$

which state that plastic processes ($\dot{\gamma} > 0$) can occur only for states on the yield surface, and to the consistency condition:

$$\dot{\gamma} \dot{f} = \dot{\gamma} \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \dot{\boldsymbol{\sigma}} + \frac{\partial f}{\partial \boldsymbol{q}} \cdot \dot{\boldsymbol{q}} \right) = 0 \tag{6}$$

requiring that the state of the material remains on the yield surface ($f = 0$) whenever plastic loading occurs. Eq. (2) is the elastic constitutive equation of the material in incremental form. The fourth–order tensor $\boldsymbol{D}^e$ is the elastic tangent stiffness of the material. Eq. (3) provides the flow rule for the plastic strain rate, defined in terms of the plastic potential function $g = \widehat{g}(\boldsymbol{\sigma}, \boldsymbol{q})$. The non–negative scalar $\dot{\gamma}$ is the plastic multiplier. The evolution of the internal variables $\boldsymbol{q}$ is provided by the hardening law (4), in which $\boldsymbol{h}$ is a prescribed hardening function.

From the consistency condition (6) the following expression for the plastic multiplier is obtained:

$$\dot{\gamma} = \frac{1}{K_p} \left\langle \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \dot{\boldsymbol{\epsilon}} \right\rangle \tag{7}$$

in which:

$$K_p := \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \boldsymbol{D}^e \frac{\partial g}{\partial \boldsymbol{\sigma}} + H_p > 0 \qquad\qquad H_p := -\frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{h} \tag{8}$$

Substituting the expression (7) for the plastic multiplier in eqs. (3) and (4), we obtain:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{ep} \dot{\boldsymbol{\epsilon}} \qquad\qquad \dot{\boldsymbol{q}} = \boldsymbol{H}^p \dot{\boldsymbol{\epsilon}} \tag{9}$$

where:

$$\boldsymbol{D}^{ep} := \boldsymbol{D}^e - \frac{\mathscr{H}(\dot{\gamma})}{K_p} \left( \boldsymbol{D}^e \frac{\partial g}{\partial \boldsymbol{\sigma}} \right) \otimes \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \boldsymbol{D}^e \right) \tag{10a}$$

$$\boldsymbol{H}^p := \frac{\mathscr{H}(\dot{\gamma})}{K_p} \boldsymbol{h} \otimes \left( \frac{\partial f}{\partial \boldsymbol{\sigma}} \boldsymbol{D}^e \right) \tag{10b}$$

where $\mathscr{H}(x)$ is the Heaviside step function, equal to one if $x > 0$ and zero otherwise, and $K_p$ is provided by eq. $(8)_1$.

## 3.2 State update

Let $\mathbb{I} = \bigcup_{n=0}^{N} [t_n, t_{n+1}]$ be a partition of the time interval of interest into time steps. It is assumed that at time $t_n \in \mathbb{I}$ the state of the material $(\boldsymbol{\sigma}_n, \boldsymbol{q}_n)$ is known at any quadrature point in the adopted finite element discretization. Also, let:

$$\{\boldsymbol{\epsilon}_i : i = 0, 1, \ldots, n+1\}$$

be the prescribed history of $\boldsymbol{\epsilon}$ up to time $t_{n+1}$. The computational problem to be addressed is the update of the state variables:

$$\boldsymbol{\sigma}_{n+1}^{(k)} \to \widehat{\boldsymbol{\sigma}}\left(\boldsymbol{\epsilon}_{n+1}^{(k)}; \boldsymbol{\sigma}_n, \boldsymbol{q}_n\right) \tag{11}$$

$$\boldsymbol{q}_{n+1}^{(k)} \to \widehat{\boldsymbol{q}}\left(\boldsymbol{\epsilon}_{n+1}^{(k)}; \boldsymbol{\sigma}_n, \boldsymbol{q}_n\right) \tag{12}$$

for a *given* increment $\Delta\boldsymbol{\epsilon}_{n+1}^{(k)} := \boldsymbol{\epsilon}_{n+1}^{(k)} - \boldsymbol{\epsilon}_n$, relative to the global iteration $(k)$, through the integration of the system of ordinary differential equations (ODEs) (2)–(5) or (9) provided by the elastoplastic constitutive equations. Note that the evolution problem defined by eqs. (2)–(5) belongs to the category of the so–called *stiff differential–algebraic systems* – see [HW91] for details – for which *implicit* methods are ideally suited. In the evolution problem governed by eqs. (9), the algebraic constraint posed by eqs. (5) has been linearized by imposing the consistency condition and then removed. This format is therefore best suited for the application of *explicit* methods.

Whenever the existence of a free energy function $\psi = \psi(\boldsymbol{\epsilon}^e)$ can be postulated, the stress tensor is linked to the elastic strain tensor by the relation:

$$\boldsymbol{\sigma}(\boldsymbol{\epsilon}^e) = \frac{\partial \psi}{\partial \boldsymbol{\epsilon}^e}(\boldsymbol{\epsilon}^e) \tag{13}$$

and thus can be considered a *dependent* quantity. As such, $\boldsymbol{\sigma}$ can be replaced in the set of state variables by the elastic strain tensor $\boldsymbol{\epsilon}^e$. The evolution equations (11) and (12) can then be recast in the following format:

$$\boldsymbol{\epsilon}_{n+1}^{e(k)} \to \widehat{\boldsymbol{\epsilon}}^e\left(\boldsymbol{\epsilon}_{n+1}^{(k)}; \boldsymbol{\epsilon}_n^e, \boldsymbol{q}_n\right) \tag{14}$$

$$\boldsymbol{q}_{n+1}^{(k)} \to \widehat{\boldsymbol{q}}\left(\boldsymbol{\epsilon}_{n+1}^{(k)}; \boldsymbol{\epsilon}_n^e, \boldsymbol{q}_n\right) \tag{15}$$

## 3.3 Consistent linearization of the stress update algorithm

In a standard finite element context, the starting point for the solution of a static equilibrium problem is the weak form of the balance of momentum equation, which, for the problem at hand, is stated as follows. Find the unknown function $\boldsymbol{u}(\boldsymbol{x})$ such that, for any test function (variation) $\boldsymbol{\eta}(\boldsymbol{x})$ satisfying homogeneous boundary conditions on the appropriate part of the boundary, the following non–linear functional equation is satisfied:

$$\mathscr{G}(\boldsymbol{u}, \boldsymbol{\eta}) = \int_{\Omega} \nabla^s \boldsymbol{\eta} \cdot \boldsymbol{\sigma}(\boldsymbol{u})\, dV - \int_{\Omega} \rho\, \boldsymbol{\eta} \cdot \boldsymbol{b}\, dV - \int_{\Gamma_t} \boldsymbol{\eta} \cdot \boldsymbol{t}\, dA = 0 \tag{16}$$

In the above equation, non–linearity stems from the non–linear dependence of the stress tensor on $\boldsymbol{u}$ induced by the constitutive equation. The iterative solution via Newton's method of the non–linear algebraic problem resulting after the introduction of a standard finite element discretization, requires the linearization of the non–linear functional $\mathscr{G}$ with respect to the independent field $\boldsymbol{u}$:

$$D_{\boldsymbol{u}}\mathscr{G}\left(\boldsymbol{u}_{n+1}^{(k)},\boldsymbol{\eta}\right)\left[\delta\boldsymbol{u}_{n+1}^{(k)}\right]=\int_{\Omega}\left\{\nabla^s\boldsymbol{\eta}\cdot\left(\widetilde{\boldsymbol{D}}\right)_{n+1}^{(k)}\nabla^s\left(\delta\boldsymbol{u}\right)_{n+1}^{(k)}\right\}dV \tag{17}$$

in which:

$$\left(\widetilde{\boldsymbol{D}}\right)_{n+1}^{(k)}:=\frac{\partial\boldsymbol{\sigma}_{n+1}^{(k)}}{\partial\boldsymbol{\epsilon}_{n+1}^{(k)}} \tag{18}$$

The fourth–order tensor $\widetilde{\boldsymbol{D}}_{n+1}^{(k)}$ is the so–called *consistent tangent stiffness matrix* to the update procedure defined by eqs. (11) and (12) or (14) and (15), *i.e.*, by the stress–point algorithm, see [ST85]. This quantity heavily depends on the adopted integration algorithm, and its accurate evaluation is crucial to achieve the quadratic convergence when using Newton–Raphson method to solve iteratively the global discrete equilibrium equations.

## 3.4 Explicit adaptive methods

Starting from the pioneering work of Sloan [Slo87], a significant amount of work has been devoted to the development of explicit stress–point algorithms for infinitesimal plasticity, based on the use of Runge–Kutta methods of various order. The key point in the application of classical methods to the solution of the differential–algebraic evolution problem posed by eqs. (2)–(4) and (5) is the removal of the algebraic constraint by its linearization through the consistency condition (6), in order to obtain the system of ODEs of eqs. (9).

Due to their conditional stability, explicit integration methods have been developed in connection with adaptive time–stepping strategies employing variable substep sizes. Adaptive time–stepping is usually implemented in two possible ways, see [SB92b]:

a) by comparing the solutions obtained with the same time step size with two explicit methods of different order (*embedded Runge–Kutta methods*);

b) by comparing the solutions obtained with the same algorithm using different step sizes (typically, a single step of size $h$ and two consecutive steps of size $h/2$).

Methods of the first group have been used in the works of Sloan and coworkers [Slo87, SB92a, SAS01, PSS08] and Tamagnini *et al.* [TVCD00]. A method of the second group based on the repeated use of the simple Forward Euler algorithm has been adopted by Fellin, Ostermann and Mittendorfer [FO02, FMO09]. In the following, we will focus our attention on these last two works, which, differently from the others mentioned, address the point of computing the consistent tangent stiffness matrix as a part of the integration algorithm.

### 3.4.1 Substepping, time rescaling and consistent linearization

Again, let $\mathbb{I} = \bigcup_{n=0}^{N} [t_n, t_{n+1}]$ be a partition of the time interval of interest $[t_0, t_{\text{fin}}]$ into time steps of amplitude $\Delta t_{n+1} = t_{n+1} - t_n$. When the material behavior is rate–independent, it is possible to rescale the time axis by introducing the following non–dimensional time factor:

$$T = \frac{(t - t_n)}{(t_{n+1} - t_n)} = \frac{(t - t_n)}{\Delta t_{n+1}} \qquad T \in [0, 1] \tag{19}$$

The (unit) non–dimensional time increment can then be divided in $m$ substeps of size:

$$\Delta T_{k+1} = T_{k+1} - T_k = \frac{t_{k+1} - t_k}{\Delta t_{n+1}} \qquad \text{provided that:} \quad \sum_{k=1}^{m} \Delta T_k = 1 \tag{20}$$

Considering that, during the time step $[t_n, t_{n+1}]$ the strain rate is assumed constant, we can write:

$$\dot{\boldsymbol{\epsilon}} = \frac{\Delta \boldsymbol{\epsilon}_{n+1}}{\Delta t_{n+1}} \qquad \frac{d\boldsymbol{\epsilon}}{dT} = \dot{\boldsymbol{\epsilon}} \frac{dt}{dT} = \Delta \boldsymbol{\epsilon}_{n+1} \tag{21}$$

and thus rewrite the evolution equations (9) as:

$$\frac{d\boldsymbol{\sigma}}{dT} = \boldsymbol{D}^{ep}(\boldsymbol{\sigma}, \boldsymbol{q}) \Delta \boldsymbol{\epsilon}_{n+1} = \boldsymbol{\xi}(\boldsymbol{\sigma}, \boldsymbol{q}, \Delta \boldsymbol{\epsilon}_{n+1}) \qquad \boldsymbol{\sigma}\big|_{T=0} = \boldsymbol{\sigma}_n \tag{22a}$$

$$\frac{d\boldsymbol{q}}{dT} = \boldsymbol{H}^{p}(\boldsymbol{\sigma}, \boldsymbol{q}) \Delta \boldsymbol{\epsilon}_{n+1} = \boldsymbol{\eta}(\boldsymbol{\sigma}, \boldsymbol{q}, , \Delta \boldsymbol{\epsilon}_{n+1}) \qquad \boldsymbol{q}\big|_{T=0} = \boldsymbol{q}_n \tag{22b}$$

where the strain increment $\Delta \boldsymbol{\epsilon}_{n+1}$ is to be considered a given data. As indicated by eq. (18), the consistent tangent stiffness emerging from the linearization of the algorithm employed to integrate eqs. (22) in the interval $[0, 1]$, with the initial conditions given in eqs. (22a)$_2$ and (22b)$_2$, measures the changes in the updated value of $\boldsymbol{\sigma}$ (*i.e.*, $\boldsymbol{\sigma}_{n+1}$) for an infinitesimal change of the prescribed strain increment, that is:

$$\widetilde{\boldsymbol{D}}_{n+1} = \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \boldsymbol{\epsilon}_{n+1}} = \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} \tag{23}$$

where the superscript $(k)$ has been dropped to ease the notation. By deriving eqs. (22) with respect to $\Delta \boldsymbol{\epsilon}_{n+1}$ we obtain:

$$\frac{d}{dT}\left(\frac{\partial \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}}\right) = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\sigma}} \frac{\partial \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} + \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{q}} \frac{\partial \boldsymbol{q}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} + \frac{\partial \boldsymbol{\xi}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} \tag{24a}$$

$$\frac{d}{dT}\left(\frac{\partial \boldsymbol{q}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}}\right) = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\sigma}} \frac{\partial \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} + \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{q}} \frac{\partial \boldsymbol{q}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} + \frac{\partial \boldsymbol{\eta}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} \tag{24b}$$

By setting:

$$\widetilde{\boldsymbol{D}} = \frac{\partial \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} \qquad \widetilde{\boldsymbol{G}} = \frac{\partial \boldsymbol{q}}{\partial \Delta \boldsymbol{\epsilon}_{n+1}} \tag{25}$$

eqs. (24) provide the following evolution equations for $\widetilde{\boldsymbol{D}}$ and $\widetilde{\boldsymbol{G}}$:

$$\frac{d\widetilde{\boldsymbol{D}}}{dT} = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\sigma}} \widetilde{\boldsymbol{D}} + \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{q}} \widetilde{\boldsymbol{G}} + \boldsymbol{D}^{ep} \qquad \widetilde{\boldsymbol{D}}\big|_{T=0} = \boldsymbol{0} \tag{26a}$$

$$\frac{d\widetilde{\boldsymbol{G}}}{dT} = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\sigma}} \widetilde{\boldsymbol{D}} + \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{q}} \widetilde{\boldsymbol{G}} + \boldsymbol{H}^{p} \qquad \widetilde{\boldsymbol{G}}\big|_{T=0} = \boldsymbol{0} \tag{26b}$$

The ordinary differential equations (22) and (26), integrated over the dimensionless time interval $[0, 1]$ with the prescribed initial conditions, will yield, at the end of the integration process ($T = 1$), the updated values of the state variables $(\boldsymbol{\sigma}_{n+1}, \boldsymbol{q}_{n+1})$. The final integrated value of $\widetilde{\boldsymbol{D}}$ at $T = 1$ will be the tangent stiffness consistent with the numerical integration algorithm adopted to solve the evolution problem. This approach to the consistent linearization of the integration algorithm has been proposed by Fellin and Ostermann [FO02].

In view of the analytical difficulties in computing the derivatives of the functions $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with respect to $\boldsymbol{\sigma}$ and $\boldsymbol{q}$ for realistic constitutive models, Fellin and Ostermann suggest to replace the RHS of eqs. (26a) and (26b) with the following approximation, obtained by numerical differentiation:

$$\frac{d\widetilde{\boldsymbol{D}}_{kl}}{dT} \simeq \frac{1}{\vartheta} \left\{ \boldsymbol{\xi} \left( \boldsymbol{\sigma} + \vartheta \widetilde{\boldsymbol{D}}_{kl}, \boldsymbol{q} + \vartheta \widetilde{\boldsymbol{G}}_{kl}, \Delta\boldsymbol{\epsilon}_{n+1} + \vartheta \widetilde{\boldsymbol{I}}_{kl} \right) - \boldsymbol{\xi} \left( \boldsymbol{\sigma}, \boldsymbol{q}, \Delta\boldsymbol{\epsilon}_{n+1} \right) \right\} \tag{27a}$$

$$\frac{d\widetilde{\boldsymbol{G}}_{kl}}{dT} \simeq \frac{1}{\vartheta} \left\{ \boldsymbol{\eta} \left( \boldsymbol{\sigma} + \vartheta \widetilde{\boldsymbol{D}}_{kl}, \boldsymbol{q} + \vartheta \widetilde{\boldsymbol{G}}_{kl}, \Delta\boldsymbol{\epsilon}_{n+1} + \vartheta \widetilde{\boldsymbol{I}}_{kl} \right) - \boldsymbol{\eta} \left( \boldsymbol{\sigma}, \boldsymbol{q}, \Delta\boldsymbol{\epsilon}_{n+1} \right) \right\} \tag{27b}$$

for $k = 1, 2, 3$ and $l = 1, 2, 3$, with the initial conditions:

$$\widetilde{\boldsymbol{D}}_{kl}\big|_{T=0} = \boldsymbol{0} \qquad \widetilde{\boldsymbol{G}}_{kl}\big|_{T=0} = \boldsymbol{0} \qquad \forall\,(k, l) = 1, 2, 3 \tag{28}$$

In eqs. (27) and (28), the quantities $\widetilde{\boldsymbol{D}}_{kl}$, $\widetilde{\boldsymbol{G}}_{kl}$ and $\widetilde{\boldsymbol{I}}_{kl}$ are defined as:

$$\widetilde{\boldsymbol{D}}_{kl} := \frac{\partial \boldsymbol{\sigma}}{\partial \Delta\epsilon_{kl,n+1}} \qquad \widetilde{\boldsymbol{G}}_{kl} := \frac{\partial \boldsymbol{q}}{\partial \Delta\epsilon_{kl,n+1}} \qquad \widetilde{\boldsymbol{I}}_{kl} = (\delta_{ik}\delta_{jl}) \boldsymbol{e}_i \otimes \boldsymbol{e}_j \tag{29}$$

If Voigt notation is adopted to represents second–order and fourth–order tensors, with the following index mapping:

| $(ij)$/$(kl)$ | 11 | 22 | 33 | 12 | 23 | 31 |
|---|---|---|---|---|---|---|
| $\alpha$/$\beta$ | 1 | 2 | 3 | 4 | 5 | 6 |

then the quantities in eq. (29) can be interpreted as the $\beta$–th column vectors of the Voigt matrices $\widetilde{\boldsymbol{D}}$, $\widetilde{\boldsymbol{G}}$ and $\widetilde{\boldsymbol{I}}$, this last being the Voigt representation of the fourth–order identity tensor.

### 3.4.2 Adaptive time integration

Let the unknowns of the evolution problem – $\boldsymbol{\sigma}$, $\boldsymbol{q}$, $\widetilde{\boldsymbol{D}}$ and $\widetilde{\boldsymbol{G}}$ – be collected into a single vector:

$$\boldsymbol{y} = \left\{\boldsymbol{\sigma}^T, \boldsymbol{q}^T, \widetilde{\boldsymbol{D}}_{11}^T, \widetilde{\boldsymbol{D}}_{22}^T, \dots \widetilde{\boldsymbol{D}}_{31}^T, \widetilde{\boldsymbol{G}}_{11}^T, \widetilde{\boldsymbol{G}}_{22}^T, \dots, \widetilde{\boldsymbol{G}}_{31}^T\right\}^T \tag{30}$$

in which the stress $\boldsymbol{\sigma}$ and the eventual tensorial internal variables collected in $\boldsymbol{q}$ are represented in Voigt notation as 6–dimensional vectors and the matrices $\widetilde{\boldsymbol{D}}$ and $\widetilde{\boldsymbol{G}}$ are stored columnwise. Then, the ODEs of eqs. (22) and (27) can be recast in the following standard format:

$$\frac{d\boldsymbol{y}}{dT} = \boldsymbol{f}(\boldsymbol{y}) \qquad\qquad T \in [0,1] \qquad\qquad \boldsymbol{y}\big|_{T=0} = \boldsymbol{y}_0 \tag{31}$$

in which the vector $\boldsymbol{f}$ collects the RHSs of eqs. (22) and (27). Eq. (31) could be integrated by means of different adaptive explicit algorithms with error control, such as Forward Euler with Richardson extrapolation [FO02, FMO09], or various types of embedded Runge–Kutta schemes of different orders [Slo87, TVC00, SAS01]. Here, we discuss in detail the implementation of the second–order adaptive substepping scheme based on the simple Forward Euler method coupled with Richardson extrapolation, first proposed by Fellin and Ostermann [FO02], for its good properties of simplicity, robustness and accuracy.

Let $[T_k, T_{k+1}] \in [0,1]$ a generic substep of size $\Delta T_{k+1}$, and let $\boldsymbol{y}_k$ the known value of $\boldsymbol{y}$ at the beginning of the step. Using the Forward Euler method, the following first approximation to $\boldsymbol{y}_{k+1}$ is obtained:

$$\boldsymbol{v} = \boldsymbol{y}_k + \Delta T_{k+1} \boldsymbol{f}(\boldsymbol{y}_k) \tag{32}$$

A second approximation to $\boldsymbol{y}_{n+1}$ is obtained by applying the Forward Euler method to two steps of size $\Delta T_{k+1}/2$:

$$\boldsymbol{w} = \boldsymbol{y}_k + \frac{\Delta T_{k+1}}{2}\boldsymbol{f}(\boldsymbol{y}_k) + \frac{\Delta T_{k+1}}{2}\boldsymbol{f}\left\{\boldsymbol{y}_k + \frac{\Delta T_{k+1}}{2}\boldsymbol{f}(\boldsymbol{y}_k)\right\} \tag{33}$$

both $\boldsymbol{v}$ and $\boldsymbol{w}$ are first–order approximations to $\boldsymbol{y}_{k+1}$ but a straightforward Taylor expansion shows that:

$$\boldsymbol{y}_{n+1} = 2\boldsymbol{w} - \boldsymbol{v} + \mathcal{O}(\Delta T_{k+1}^2) \tag{34}$$

*i.e.*, the difference $2\boldsymbol{w} - \boldsymbol{v}$ is a second–order approximation to the local solution.

The norm:

$$EST := \|\boldsymbol{w} - \boldsymbol{v}\|_{\max} \qquad\qquad \|\boldsymbol{w} - \boldsymbol{v}\|_{\max} := \max_{i=1,\dots,n_y} \left|\frac{w_i - v_i}{s_i}\right| \tag{35}$$

with $s_i$ a suitable scaling factor, is an asymptotically correct estimate for the local integration error of $\boldsymbol{w}$. Setting the quantity $TOL$ as the user–supplied tolerance, the comparison between $TOL$ and $EST$ provides an indicator of the accuracy of the numerical integration procedure and an estimate of the optimal substep to be used. In particular:

a) If $EST < TOL$: the substep is accepted, with $\boldsymbol{y}_{k+1}$ given by eq. (34). The next substep size can be increased according to the relation:

$$\Delta T_{k+2} = \Delta T_{k+1} \min \left\{ r_I, \max \left( r_D, 0.9\sqrt{\frac{TOL}{EST}} \right) \right\} \tag{36}$$

b) If $EST \geq TOL$: the substep is rejected, and the integration step repeated with a smaller substep size given by:

$$\Delta T_{k+1} \leftarrow \Delta T_{k+1} \min \left\{ r_I, \max \left( r_D, 0.9\sqrt{\frac{TOL}{EST}} \right) \right\} \tag{37}$$

In eqs. (36) and (37), the coefficient 0.9 multiplying the square root of $TOL/EST$ is a "safety factor" accounting from the approximation introduced in the error estimation, while $r_I$ and $r_D$ represent the maximum increase and decrease in the step size allowed. Typically they are set to $r_I = 2.0$ and $r_D = 0.2$.

### 3.4.3 Drift correction and other computational aspects in explicit integration

When using explicit integration algorithms, the updated state variables $(\boldsymbol{\sigma}_{k+1}, \boldsymbol{q}_{k+1})$ may violate the consistency condition, so that:

$$f_{k+1} = f(\boldsymbol{\sigma}_{k+1}, \boldsymbol{q}_{k+1}) > FTOL$$

with $FTOL$ a prescribed error tolerance for the consistency condition. This situation, which corresponds to a stress state $\boldsymbol{\sigma}_{k+1}$ outside the final yield surface, is commonly known in computational plasticity as *yield surface drift*. The reason for this pathology is that, in explicit methods, the algebraic constraint imposed by eq. $(5)_2$ is linearized, and thus enforced in a weak form. The extent of this violation depends on the accuracy of the integration scheme, so it could be reduced by adopting stringent error tolerances on the adaptive substepping scheme. Nonetheless, in order to prevent error accumulation, it is highly recommendable to implement a drift correction algorithm at the end of each substep, particularly for complex constitutive models.

Different types of drift correction algorithms have been proposed in literature. A detailed discussion on the advantages and drawbacks of some of the more widely used strategies for drift correction, focusing on their application to plasticity models developed for soils, can be found in the works of [PG85, SAS01].

In addition to drift correction, the adoption of explicit integration methods in classical plasticity – where there is a non–smooth transition between elastic and plastic behavior along a predefined stress–path – requires particular attention for those time integration steps which:

a) start from an elastic state and – if elastic response is maintained for the entire step – end outside the current yield surface;

b) start from a plastic state (on the current yield surface) and crosses the yield surface once before ending up on a new plastic state;

c) start from an elastic state and end on another elastic state, crossing the yield surface twice during the path from the initial to the final state.

Situations of type (a) are quite common, particularly when relatively large integration steps are used. Situations of type (b) may occur in presence of relatively small elastic domains – *e.g.*, in models for sands with rotational hardening, where the yield surface is a cone with a small opening. Both these issues have been addressed in [SAS01]. Situations of type (c) may occur when the yield surface is non–convex. While the opportunity of adopting a non–convex yield surface is questionable on both theoretical and experimental grounds, the treatment of this case has been effectively addressed by Pedroso *et al.* [PSS08].

## 3.5 Implicit Generalized Backward Euler method

Implicit algorithms based on the concepts of operator split and closest point projection return mapping, as discussed for example in [SH87, SG91], have been applied to computational geomechanics in a number of works, among which we mention [BL90, Bor91, ARS92, MWA97, JS97].

The starting point for this approach is the exploitation of the additive structure of the governing equations of the differential–algebraic problem eqs. (2)–(5) to split the update processes into two consecutive steps, as detailed in the following section.

### 3.5.1 Operator split and product formula algorithm

The constitutive equation of infinitesimal plasticity are amenable to the *elastic–plastic operator split* of the original problem of evolution, into an *elastic predictor* problem and a *plastic corrector* problem, as shown in Tab. 1 [SH87, SH98]. Note that in Tab. 1, exploiting the existence of a free energy function and thus of the elastic constitutive equation (13), the elastic constitutive equation in rate–form has been replaced by the additive split of the strain rate: $\dot{\boldsymbol{\epsilon}}^e = \dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p$.

Starting from this operator split, a product formula algorithm is constructed as follows. First, the elastic predictor problem is solved and a so–called *trial elastic state* is obtained. Then, the constraints (5) are checked for the trial state, and if they are violated, the trial state is taken as the initial condition for the plastic corrector problem.

### 3.5.2 Problem 1: elastic predictor

From the physical point of view, the elastic predictor problem can be derived from the original problem of evolution by *freezing* the plastic flow (i.e., setting $\dot{\gamma} = 0$), and taking an incremental *elastic* step which ignores the constraints placed on the stress state by the yield function. The solution of the predictor stage (*trial state*) in terms of

| | *Global* | *Elastic predictor* | *Plastic corrector* |
|---|---|---|---|
| | $\dot{\boldsymbol{\epsilon}} = \nabla^s(\dot{\boldsymbol{u}})$ | $\dot{\boldsymbol{\epsilon}} = \nabla^s(\dot{\boldsymbol{u}})$ | $\dot{\boldsymbol{\epsilon}} = \mathbf{0}$ |
| Evolution eqs. | $\dot{\boldsymbol{\epsilon}}^e = \dot{\boldsymbol{\epsilon}} - \dot{\gamma}\dfrac{\partial g}{\partial \boldsymbol{\sigma}}$ | $\dot{\boldsymbol{\epsilon}}^e = \dot{\boldsymbol{\epsilon}}$ | $\dot{\boldsymbol{\epsilon}}^e = -\dot{\gamma}\dfrac{\partial g}{\partial \boldsymbol{\sigma}}$ |
| | $\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}$ | $\dot{\boldsymbol{q}} = \mathbf{0}$ | $\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}$ |
| Initial conds. | $\boldsymbol{\epsilon}^e(t_n) = \boldsymbol{\epsilon}^e_n$ | $\boldsymbol{\epsilon}^e(t_n) = \boldsymbol{\epsilon}^e_n$ | $\boldsymbol{\epsilon}^e\Big\vert_{(\dot{\gamma}=0)} = \boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1}$ |
| | $\boldsymbol{q}(t_n) = \boldsymbol{q}_n$ | $\boldsymbol{q}(t_n) = \boldsymbol{q}_n$ | $\boldsymbol{q}\Big\vert_{(\dot{\gamma}=0)} = \boldsymbol{q}^{\mathrm{tr}}_{n+1}$ |
| | $f(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0$ | | $f(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0$ |
| Constraints | $\dot{\gamma} \geq 0$ | none | $\dot{\gamma} \geq 0$ |
| | $f(\boldsymbol{\sigma}, \boldsymbol{q})\dot{\gamma} = 0$ | | $f(\boldsymbol{\sigma}, \boldsymbol{q})\dot{\gamma} = 0$ |

Table 1: Operator split of the evolution problem of infinitesimal plasticity, formulated in terms of strain rates.

elastic strains is given by the following geometric update:

$$\boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1} = \boldsymbol{\epsilon}^e_n + \boldsymbol{\epsilon}_{n+1} - \boldsymbol{\epsilon}_n \tag{38}$$

As for the internal variables, since they do not change during an elastic process, the trivial solution for their trial values is:

$$\boldsymbol{q}^{\mathrm{tr}}_{n+1} = \boldsymbol{q}_n \tag{39}$$

Finally, the trial state of stress is obtained from $\boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1}$ by a simple function evaluation:

$$\boldsymbol{\sigma}^{\mathrm{tr}}_{n+1} := \frac{\partial \psi}{\partial \boldsymbol{\epsilon}^e}\left(\boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1}\right) \tag{40}$$

At the end of the elastic predictor stage, the trial state is checked for consistency with the yield locus. If:

$$f^{\mathrm{tr}}_{n+1} := f\left(\boldsymbol{\sigma}^{\mathrm{tr}}_{n+1}, \boldsymbol{q}_n\right) \leq 0$$

the trial state satisfies the constraints imposed by the Kuhn–Tucker conditions. The process is then declared *elastic* and the trial state represents the actual final state of the material. If, on the contrary, $f^{\mathrm{tr}}_{n+1} > 0$, the process is declared *plastic*, and consistency is restored by solving the plastic corrector problem.

### 3.5.3 Problem 2: plastic corrector

If $f^{\mathrm{tr}}_{n+1} > 0$, the trial state lies outside the yield locus, and thus violates the constraints. Consistency is then restored by solving the plastic corrector problem, which takes place at *fixed total strain* ($\dot{\boldsymbol{\epsilon}} = \mathbf{0}$). Since the objective of the plastic corrector stage is to map the trial state back to the yield surface, the algorithms performing such task are commonly referred to as *return mapping algorithms*.

Typically, the plastic corrector problem is solved numerically by integrating the corresponding system of ODEs by an implicit *Backward Euler* scheme, taking the trial state as the new initial condition:

$$\boldsymbol{\epsilon}^{e}_{n+1} = \boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1} - \Delta\gamma_{n+1}\left(\frac{\partial g}{\partial \boldsymbol{\sigma}}\right)_{n+1} \tag{41}$$

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} \tag{42}$$

As $\Delta\gamma_{n+1} > 0$, the constraints of eq. (5) reduce to:

$$f_{n+1} = f(\boldsymbol{\sigma}_{n+1}, \boldsymbol{q}_{n+1}) = 0 \tag{43}$$

Equations (41)–(43) provide a system of $7 + n_{\mathrm{int}}$ non–linear algebraic equations in the $7 + n_{\mathrm{int}}$ unknowns $\boldsymbol{\epsilon}^{e}_{n+1}$, $\Delta\gamma_{n+1}$, and $\boldsymbol{q}_{n+1}$, which can be solved iteratively by Newton's method, at the Gauss point level.

Let:

$$\boldsymbol{x}_{n+1} := \left\{\boldsymbol{\epsilon}^{eT}_{n+1} \quad \boldsymbol{q}^{T}_{n+1} \quad \Delta\gamma_{n+1}\right\}^{T} \in \mathbb{R}^{7+n_{\mathrm{int}}} \tag{44}$$

be a vector containing the the unknowns of the problem and

$$\widetilde{\boldsymbol{x}}_{n+1} := \left\{\boldsymbol{\epsilon}^{eT}_{n+1} \quad \boldsymbol{q}^{T}_{n+1}\right\}^{T} \qquad \text{so that:} \qquad \boldsymbol{x}_{n+1} = \left\{\widetilde{\boldsymbol{x}}^{T}_{n+1} \quad \Delta\gamma_{n+1}\right\}^{T}$$

The return mapping equations (41)–(43) require the vanishing of the following *residual vector*:

$$\boldsymbol{R}_{n+1}\left(\boldsymbol{x}_{n+1}\right) := \begin{Bmatrix} \boldsymbol{r}^{\epsilon}_{n+1} \\ \boldsymbol{r}^{q}_{n+1} \\ f_{n+1} \end{Bmatrix} := \begin{Bmatrix} -\boldsymbol{\epsilon}^{e}_{n+1} + \boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1} - \Delta\gamma_{n+1}\boldsymbol{Q}_{n+1} \\ -\boldsymbol{q}_{n+1} + \boldsymbol{q}^{\mathrm{tr}}_{n+1} + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} \\ f_{n+1} \end{Bmatrix} = \mathbf{0} \tag{45}$$

where $\boldsymbol{Q}_{n+1} = (\partial g/\partial \boldsymbol{\sigma})_{n+1}$. The steps required for the iterative solution of eq. (45) via Newton's method are outlined in Tab. 2.

A first difficulty in applying the procedure outlined in Tab. 2 is that Step 3 requires the inversion of a $(7 + n_{\mathrm{int}}) \times (7 + n_{\mathrm{int}})$ square matrix. By observing that the last component of the residual vector $\boldsymbol{R}^{(j)}_{n+1}$ does not depend on $\Delta\gamma_{n+1}$, the resulting linearized system of equation can be reduced in size by one via static condensation. However, the inversion of the resulting tangent operator in closed form can still be

1. Initialize:

$$\boldsymbol{\epsilon}^{e}_{n+1} = \boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1} \qquad \boldsymbol{q}_{n+1} = \boldsymbol{q}^{\mathrm{tr}}_{n+1} \qquad \Delta\gamma_{n+1} = 0$$

2. Check for convergence:

$$\text{IF: } \left\{ \begin{array}{c} \left\| \boldsymbol{r}^{\epsilon(j)}_{n+1} \right\| < TOL_{\epsilon} \cdot \left\| \boldsymbol{\epsilon}^{e,\mathrm{tr}}_{n+1} \right\| \\ \left\| \boldsymbol{r}^{q(j)}_{n+1} \right\| < TOL_{q} \cdot \left\| \boldsymbol{q}^{\mathrm{tr}}_{n+1} \right\| \\ f^{(j)}_{n+1} < TOL_{f} \end{array} \right\} \text{ THEN exit, ELSE:}$$

3. Find update at local iteration $(j)$:

$$\delta \boldsymbol{x}^{(j)}_{n+1} = - \left[ \left( \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}} \right)^{(j)}_{n+1} \right]^{-1} \boldsymbol{R}^{(j)}_{n+1}$$

4. Update state variables and plastic multiplier:

$$\boldsymbol{x}^{(j+1)}_{n+1} = \boldsymbol{x}^{(j)}_{n+1} + \delta \boldsymbol{x}^{(j)}_{n+1}$$

5. Set: $j \leftarrow j + 1$, GO TO 2.

Table 2: Iterative solution of the plastic corrector problem.

very difficult, especially in presence of a large number of internal variables (*i.e.*, in the case of anisotropic hardening models). In the most difficult cases, this problem can be solved by resorting to symbolic computation tools (as, *e.g.*, MATHEMATICA) or by numerical methods, as in [TCN02].

Another classical problem in the application of the implicit Backward Euler algorithm to complex, three–invariants plasticity models lies in the need of computing the second gradients of the plastic potential function $\partial^2 g / \partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}$ and the derivatives with respect to $\boldsymbol{\sigma}$ and $\boldsymbol{q}$ of the hardening function $\boldsymbol{h}$. In the most complex situations, this task can be performed by resorting to numerical differentiation, as suggested, *e.g.*, in [PFRFH00].

### 3.5.4 Formulation of the corrector step in principal elastic strain space

A considerable simplification in the application of the implicit Backward Euler algorithm to complex plasticity models can be obtained in the case of isotropic–hardening plasticity, by formulating the return mapping stage in principal elastic strain space. By exploiting the spectral decomposition of the tensors $\boldsymbol{Q}_{n+1}$, $\boldsymbol{\epsilon}^e_{n+1}$ and $\boldsymbol{\epsilon}^{e,\text{tr}}_{n+1}$, eq. (41) transforms into:

$$\sum_{A=1}^{3} (\epsilon^e_A)_{n+1} \, \boldsymbol{n}^{(A)}_{n+1} \otimes \boldsymbol{n}^{(A)}_{n+1} = \sum_{A=1}^{3} \left(\epsilon^{e,\text{tr}}_A\right)_{n+1} \boldsymbol{n}^{(A),\text{tr}}_{n+1} \otimes \boldsymbol{n}^{(A),\text{tr}}_{n+1} - \Delta\gamma_{n+1} \sum_{A=1}^{3} \left(\frac{\partial g}{\partial \sigma_A}\right)_{n+1} \boldsymbol{n}^{(A)}_{n+1} \otimes \boldsymbol{n}^{(A)}_{n+1} \quad (46)$$

in which $\boldsymbol{n}^{(A)}_{n+1}$ and $\boldsymbol{n}^{(A),\text{tr}}_{n+1}$ are the $A$–th unit eigenvectors of $\boldsymbol{\epsilon}^e_{n+1}$ and $\boldsymbol{\epsilon}^{e,\text{tr}}_{n+1}$. Then, it follows at once that:

$$\boldsymbol{n}^{(A)}_{n+1} = \boldsymbol{n}^{(A),\text{tr}}_{n+1} \quad (47)$$

and:

$$(\epsilon^e_A)_{n+1} = \left(\epsilon^{e,\text{tr}}_A\right)_{n+1} - \Delta\gamma_{n+1} \left(\frac{\partial g}{\partial \sigma_A}\right)_{n+1} \quad (48)$$

for $A$ = 1, 2 or 3. Note that, as the trial elastic strain is known, so are its principal directions. Therefore, the only unknown quantities to be determined remain the three principal elastic strains $(\epsilon^e_A)_{n+1}$, the $n_{\text{int}}$ internal variables $\boldsymbol{q}_{n+1}$ and the plastic multiplier $\Delta\gamma_{n+1}$. Introducing for convenience the following vector notation:

$$\widehat{\boldsymbol{\epsilon}}^e := \begin{Bmatrix} \epsilon^e_1 \\ \epsilon^e_2 \\ \epsilon^e_3 \end{Bmatrix} \qquad \widehat{\boldsymbol{\epsilon}}^{e,\text{tr}} := \begin{Bmatrix} \epsilon^{e,\text{tr}}_1 \\ \epsilon^{e,\text{tr}}_2 \\ \epsilon^{e,\text{tr}}_3 \end{Bmatrix} \qquad \widehat{\boldsymbol{\sigma}} := \begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{Bmatrix} \qquad \widehat{\boldsymbol{Q}} := \begin{Bmatrix} \partial g / \partial \sigma_1 \\ \partial g / \partial \sigma_2 \\ \partial g / \partial \sigma_3 \end{Bmatrix} \quad (49)$$

the return mapping problem in principal elastic strain space can be recast as follows:

$$\widehat{\boldsymbol{\epsilon}}^{e}_{n+1} = \widehat{\boldsymbol{\epsilon}}^{e,\mathrm{tr}}_{n+1} - \Delta\gamma_{n+1}\widehat{\boldsymbol{Q}}_{n+1} \tag{50a}$$

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}^{\mathrm{tr}}_{n+1} + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} \tag{50b}$$

$$f_{n+1} := f(\widehat{\boldsymbol{\sigma}}_{n+1}, \boldsymbol{q}_{n+1}) = 0 \tag{50c}$$

The iterative solution of the return mapping equations (50) follows a scheme similar to that in Tab. 2. The number of equations to be solved is now reduced by 3. Moreover, only the evaluation of the $(3 \times 3)$ matrix:

$$\nabla\nabla g = \frac{\partial^2 g}{\partial\hat{\boldsymbol{\sigma}} \otimes \partial\hat{\boldsymbol{\sigma}}} \tag{51}$$

is now required to compute the tangent operator $\partial\boldsymbol{R}/\partial\boldsymbol{x}$.

### 3.5.5 Consistent tangent stiffness

One of the advantages of the proposed algorithm is the possibility of evaluating the consistent tangent operators *in closed form*, as shown in the following.

In the *global* iteration process, any (infinitesimal) variation in the total strain increment induces, by definition, an equal variation in the trial elastic strain:

$$d\boldsymbol{\epsilon} = d\boldsymbol{\epsilon}^{e,\mathrm{tr}} \tag{52}$$

where the subscript $n+1$ and the superscript $(k)$ have been omitted to ease the notation. Moreover, the return mapping equations associate to each trial elastic strain a well defined elastic strain tensor, obtained as a result of the local iteration process. Therefore, for an infinitesimal variation of $\boldsymbol{\epsilon}^{e,\mathrm{tr}(k)}_{n+1}$ one has:

$$d\boldsymbol{\epsilon}^{e} = \boldsymbol{L}\, d\boldsymbol{\epsilon}^{e,\mathrm{tr}} \tag{53}$$

On the other hand, from the hyperelastic constitutive equation, we have:

$$d\boldsymbol{\sigma} = \left(\frac{\partial^2\psi}{\partial\boldsymbol{\epsilon}^e \otimes \partial\boldsymbol{\epsilon}^e}\right) d\boldsymbol{\epsilon}^e = \boldsymbol{D}^e\, d\boldsymbol{\epsilon}^e = \boldsymbol{D}^e\boldsymbol{L}\, d\boldsymbol{\epsilon}^{e,\mathrm{tr}} = \boldsymbol{\Xi}\, d\boldsymbol{\epsilon}^{e,\mathrm{tr}} \tag{54}$$

By virtue of the definition (18) and of the identity (52), the tensor $\boldsymbol{\Xi}$ is the required consistent tangent stiffness tensor. Differentiation of the return mapping equations (45) yields:

$$\boldsymbol{A}\, d\widetilde{\boldsymbol{x}} = \boldsymbol{T}\, d\boldsymbol{\epsilon}^{e,\mathrm{tr}} - d(\Delta\gamma)\boldsymbol{U} \tag{55}$$

where:

$$d\widetilde{\boldsymbol{x}} := \left\{ d\boldsymbol{\epsilon}^{eT} \quad d\boldsymbol{q}^T \right\}^T \qquad \boldsymbol{A} := \begin{bmatrix} \boldsymbol{I} + \Delta\gamma\,\boldsymbol{A}_\sigma\boldsymbol{D}^e & +\Delta\gamma\,\boldsymbol{A}_q \\ -\Delta\gamma\boldsymbol{B}_\sigma\boldsymbol{D}^e & \boldsymbol{I}_q - \Delta\gamma\boldsymbol{B}_q \end{bmatrix} \tag{56}$$

$$\boldsymbol{T} := \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0}_q^T \end{bmatrix}^T \qquad \boldsymbol{U} := \left\{ \boldsymbol{Q}^T, -\boldsymbol{h}^T \right\}^T \tag{57}$$

$$\boldsymbol{A}_\sigma := \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\sigma}} = \frac{\partial^2 g}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}} \qquad \boldsymbol{A}_q := \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} = \frac{\partial^2 g}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{q}} \tag{58}$$

$$\boldsymbol{B}_\sigma := \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\sigma}} \qquad \boldsymbol{B}_q := \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{q}} \tag{59}$$

and $\boldsymbol{I}_q$ and $\boldsymbol{0}_q$ are the identity matrix in $\mathbb{R}^{n_{\text{int}}}$ and the zero $(n_{\text{int}} \times 6)$ matrix. The variation in the plastic multiplier $d(\Delta\gamma)$ can be evaluated by enforcing the consistency condition $df_{n+1}^{(k)} = 0$. Defining:

$$\boldsymbol{P} := \boldsymbol{D}^e \frac{\partial f}{\partial \boldsymbol{\sigma}} \qquad \boldsymbol{W} := \frac{\partial f}{\partial \boldsymbol{q}} \qquad \boldsymbol{V} := \left\{ \boldsymbol{P}^T, \boldsymbol{W}^T \right\} \tag{60}$$

the consistency condition reads:

$$\boldsymbol{V} d\widetilde{\boldsymbol{x}} = 0 \tag{61}$$

Solving eq. (55) for $d\widetilde{\boldsymbol{x}}$, substituting the result in eq. (61) and solving for $d(\Delta\gamma)$, the plastic multiplier increment is obtained as:

$$d(\Delta\gamma) = \frac{1}{\boldsymbol{V} \cdot \left[\boldsymbol{A}^{-1}\right] \boldsymbol{U}} \boldsymbol{V} \cdot \left[\boldsymbol{A}^{-1}\right] \boldsymbol{T} \, d\boldsymbol{\epsilon}^{e,\text{tr}} \tag{62}$$

From eqs. (55) and (62) we obtain:

$$\begin{aligned} d\boldsymbol{\sigma} = \boldsymbol{D}^e \, d\boldsymbol{\epsilon}^e = \boldsymbol{D}^e \boldsymbol{T}^T \, d\widetilde{\boldsymbol{x}} \\ = \left\{ \boldsymbol{D}^e \boldsymbol{T}^T \left[ \boldsymbol{A}^{-1} - \frac{(\boldsymbol{A}^{-1}\boldsymbol{U}) \otimes (\boldsymbol{V}\boldsymbol{A}^{-1})}{(\boldsymbol{V} \cdot \boldsymbol{A}^{-1}\boldsymbol{U})} \right] \boldsymbol{T} \right\} d\boldsymbol{\epsilon}^{e,\text{tr}} \end{aligned} \tag{63}$$

By comparing eq. (63) with (54) the expression for the consistent tangent stiffness tensor follows:

$$\widetilde{\boldsymbol{D}}_{n+1}^{(k)} = \boldsymbol{\Xi}_{n+1}^{(k)} = \boldsymbol{D}^e \boldsymbol{T}^T \left[ \boldsymbol{A}^{-1} - \frac{(\boldsymbol{A}^{-1}\boldsymbol{U}) \otimes (\boldsymbol{V}\boldsymbol{A}^{-1})}{(\boldsymbol{V} \cdot \boldsymbol{A}^{-1}\boldsymbol{U})} \right]_{n+1}^{(k)} \boldsymbol{T} \tag{64}$$

Note that the consistent tangent stiffness is, in general, *non symmetric*, even in the case of associative flow rule ($f \equiv g$).

# 4 Stress–point algorithms for finite deformation multiplicative plasticity

## 4.1 Evolution equations

The evolution equations of finite deformation multiplicative plasticity for isotropic materials are briefly summarized below, see [OT21] for details. Let the deformation gradient $\boldsymbol{F}$ be multiplicatively decomposed into an elastic part $\boldsymbol{F}^e$ and a plastic part $\boldsymbol{F}^p$:

$$\boldsymbol{F} = \boldsymbol{F}^e \boldsymbol{F}^p \tag{65}$$

Recalling the expression for the spatial velocity gradient $\boldsymbol{l} = \nabla \boldsymbol{v} = \dot{\boldsymbol{F}} \boldsymbol{F}^{-1}$ and defining accordingly the the *elastic* and *plastic velocity gradients* as:

$$\boldsymbol{l}^e := \dot{\boldsymbol{F}}^e \boldsymbol{F}^{e-1} \qquad \overline{\boldsymbol{L}}^p := \dot{\boldsymbol{F}}^p \boldsymbol{F}^{p-1} \qquad \boldsymbol{l}^p := \boldsymbol{F}^e \overline{\boldsymbol{L}}^p \boldsymbol{F}^{e-1} \tag{66}$$

it is easy to show that:

$$\boldsymbol{l} = \boldsymbol{l}^e + \boldsymbol{l}^p \tag{67}$$

*i.e.*, the multiplicative decomposition of the deformation gradient is consistent with an additive split of the spatial velocity gradient. From the spatial elastic and plastic velocity gradients, $\boldsymbol{l}^e$ and $\boldsymbol{l}^p$, the *elastic* and *plastic rates of deformation* and *spins* can be defined as:

$$\boldsymbol{d}^e := \mathrm{sym}\,(\boldsymbol{l}^e) \qquad \boldsymbol{w}^e := \mathrm{skw}\,(\boldsymbol{l}^e) \tag{68}$$

$$\boldsymbol{d}^p := \mathrm{sym}\,(\boldsymbol{l}^p) \qquad \boldsymbol{w}^p := \mathrm{skw}\,(\boldsymbol{l}^p) \tag{69}$$

In the following, consistently with the assumption of material isotropy, we will assume that the plastic spin $\boldsymbol{w}^p$ is always equal to zero, and $\boldsymbol{l}^p = \boldsymbol{d}^p$, as in [Sim98].

Then let us assume that the material possesses a *free energy function* per unit reference volume of the form:

$$\psi = \psi(\boldsymbol{b}^e) = \bar{\psi}(\overline{\boldsymbol{C}}^e) = \hat{\psi}(\lambda_1^e, \lambda_2^e, \lambda_3^e) \tag{70}$$

where $\boldsymbol{b}^e$ be the left elastic Cauchy–Green strain tensor, $\overline{\boldsymbol{C}}^e$ is the right elastic Cauchy–Green strain tensor, and $\lambda_A^e$, with $A = 1, 2, 3$ are the principal elastic stretches, eigenvalues of $\boldsymbol{F}^e$. The Kirchhoff stress tensor is linked to the elastic strains by the following alternative hyperelastic constitutive equations:

$$\boldsymbol{\tau} = 2 \frac{\partial \psi}{\partial \boldsymbol{b}^e} \boldsymbol{b}^e \qquad \boldsymbol{\tau} = 2\, \boldsymbol{F}^e \frac{\partial \bar{\psi}^e}{\partial \overline{\boldsymbol{C}}^e} \boldsymbol{F}^{eT} \tag{71}$$

To incorporate irreversible behavior, let us assume the existence of an elastic domain:

$$\mathbb{E} := \Big\{ (\boldsymbol{\tau}, \boldsymbol{q}) \;\Big|\; f\,(\boldsymbol{\tau}, \boldsymbol{q}) \leq 0 \Big\} \tag{72}$$

defined via a suitable yield function $f(\boldsymbol{\tau}, \boldsymbol{q})$ depending on Kirchhoff stress and a vector $\boldsymbol{q}$ (of dimension $n_{\mathrm{int}}$) of scalar internal variables, accounting for the effects of the previous loading history.

Adopting the left elastic Cauchy–Green tensor $\boldsymbol{b}^e$ and the internal variables $\boldsymbol{q}$ as the main state variables, the problem of evolution of non–associative multiplicative plasticity can be cast in the following form:

$$\dot{\boldsymbol{b}}^e = \boldsymbol{l}\boldsymbol{b}^e + \boldsymbol{b}^e\boldsymbol{l}^T + \mathscr{L}_v[\boldsymbol{b}^e] \tag{73}$$

$$\boldsymbol{d}^p = \dot{\gamma}\frac{\partial g}{\partial \boldsymbol{\tau}} \tag{74}$$

$$\mathscr{L}_v[\boldsymbol{b}^e] = -2\,\mathrm{sym}\,(\boldsymbol{d}^p\boldsymbol{b}^e) = -2\dot{\gamma}\,\frac{\partial g}{\partial \boldsymbol{\tau}}\boldsymbol{b}^e \tag{75}$$

$$\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}(\boldsymbol{\tau}, \boldsymbol{q}) \tag{76}$$

where $\mathscr{L}_v[\boldsymbol{b}^e]$ is the Lie derivative of $\boldsymbol{b}^e$, $\dot{\gamma} \geq 0$ is the plastic multiplier, $g(\boldsymbol{\tau}, \boldsymbol{q})$ is the plastic potential and $\boldsymbol{h}$ is the hardening function for the internal variables $\boldsymbol{q}$. Note that the Kirchhoff stress tensor, appearing as an argument of the yield function $f$ and of the plastic potential $g$, can be considered a derived quantity by virtue of the constitutive equation $(71)_1$.

The yield function $f$ and the plastic multiplier $\dot{\gamma}$ are subjected to the Kuhn–Tucker complementarity conditions:

$$\dot{\gamma} \geq 0 \qquad\qquad f(\boldsymbol{\tau}, \boldsymbol{q}) \leq 0 \qquad\qquad \dot{\gamma} f(\boldsymbol{\tau}, \boldsymbol{q}) = 0 \tag{77}$$

as well as to the consistency condition:

$$\dot{\gamma}\dot{f} = \dot{\gamma}\left(\frac{\partial f}{\partial \boldsymbol{\tau}} \cdot \dot{\boldsymbol{\tau}} + \frac{\partial f}{\partial \boldsymbol{q}} \cdot \dot{\boldsymbol{q}}\right) = 0 \tag{78}$$

requiring that the state of the material remains on the yield surface ($f = 0$) whenever plastic loading occurs ($\dot{\gamma} > 0$).

Eq. (73) shows that the rate of change of the elastic left Cauchy–green tensor is the sum of two contributions, the second of which – the Lie derivative of $\boldsymbol{b}^e$ – is associated to the development of plastic deformations. Eq. (74) is the non–associative flow rule for the plastic rate of deformation $\boldsymbol{d}^p$, while eq. (75) provides the link between $\boldsymbol{d}^e$ and the Lie derivative of $\boldsymbol{b}^e$. The evolution equation for the internal variables $\boldsymbol{q}$ is provided by the non–associative hardening law (76).

Differentiating the hyperelastic constitutive equation $(66)_2$ and taking into account that:

$$\boldsymbol{d} = \boldsymbol{d}^e + \boldsymbol{d}^p \qquad \boldsymbol{d}^e = \boldsymbol{F}^{e-T}\left(\frac{1}{2}\dot{\bar{\boldsymbol{C}}}^e\right)\boldsymbol{F}^{e-1} \qquad \boldsymbol{w}^e = \boldsymbol{w} - \boldsymbol{w}^p = \boldsymbol{w}$$

the following expression for the Jaumann rate of Kirchhoff stress is obtained:

$$\overset{\triangledown}{\boldsymbol{\tau}} = \mathbb{a}^e\,(\boldsymbol{d} - \boldsymbol{d}^p) = \mathbb{a}^e\left(\boldsymbol{d} - \dot{\gamma}\frac{\partial f}{\partial \boldsymbol{\tau}}\right) \tag{79}$$

where:

$$\mathbb{a}^e_{ijkl} = \mathbb{c}^e_{ijkl} + \tau_{ik}\delta_{jl} + \tau_{il}\delta_{jk} \tag{80}$$

and:

$$\mathbb{c}^e_{ijkl} = \mathbb{C}^e_{ABCD} F^e_{iA} F^e_{jB} F^e_{kC} F^e_{lD} \qquad \mathbb{C}^e_{ABCD} = 4\,\frac{\partial^2 \bar{\psi}^e}{\partial \bar{C}_{AB} \partial \bar{C}_{CD}} \tag{81}$$

Substituting eqs. (76) and (79) in eq. (78) – after noting that

$$\frac{\partial f}{\partial \boldsymbol{\tau}} \cdot \dot{\boldsymbol{\tau}} = \frac{\partial f}{\partial \boldsymbol{\tau}} \cdot \overset{\triangledown}{\boldsymbol{\tau}}$$

since, by isotropy, $\boldsymbol{\tau}$ and $\partial f / \partial \boldsymbol{\tau}$ commute – and solving for the plastic multiplier, the following expression for $\dot{\gamma}$ is obtained:

$$\dot{\gamma} = \frac{1}{\widehat{K}_p} \left\langle \frac{\partial f}{\partial \boldsymbol{\tau}} \cdot \mathbb{a}^e \boldsymbol{d} \right\rangle \tag{82}$$

in which:

$$\widehat{K}_p := \frac{\partial f}{\partial \boldsymbol{\tau}} \cdot \mathbb{a}^e \frac{\partial g}{\partial \boldsymbol{\tau}} - \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{h} > 0 \tag{83}$$

The elastoplastic constitutive equation in rate–form then reads

$$\overset{\triangledown}{\boldsymbol{\tau}} = \mathbb{a}^{ep} \boldsymbol{d} \qquad \mathbb{a}^{ep} = \mathbb{a}^e - \frac{\mathscr{H}(\dot{\gamma})}{\widehat{K}_p} \left( \mathbb{a}^e \frac{\partial g}{\partial \boldsymbol{\tau}} \right) \otimes \left( \mathbb{a}^e \frac{\partial f}{\partial \boldsymbol{\tau}} \right) \tag{84}$$

where $\mathbb{a}^{ep}$ is the elastoplastic continuum tangent stiffness of the material and $\mathscr{H}(x)$ is the Heaviside step function, equal to one if $x > 0$ and zero otherwise.

## 4.2 State update

Let $\mathbb{I} = \bigcup_{n=0}^{N} [t_n, t_{n+1}]$ be a partition of the time interval of interest into time steps. It is assumed that at time $t_n \in \mathbb{I}$ the state of the material $(\boldsymbol{b}^e_n, \boldsymbol{q}_n)$ is known at any quadrature point in the adopted finite element discretization. Also, let:

$$\{\boldsymbol{F}_i : i = 0, 1, \ldots, n+1\}$$

be the prescribed history of $\boldsymbol{F}$ up to time $t_{n+1}$. The computational problem to be addressed is the update of the state variables:

$$\boldsymbol{b}^{e(k)}_{n+1} \to \widehat{\boldsymbol{b}}^e \left( \boldsymbol{F}^{(k)}_{n+1}; \boldsymbol{b}^e_n, \boldsymbol{q}_n \right) \tag{85}$$

$$\boldsymbol{q}^{(k)}_{n+1} \to \widehat{\boldsymbol{q}} \left( \boldsymbol{F}^{(k)}_{n+1}; \boldsymbol{b}^e_n, \boldsymbol{q}_n \right) \tag{86}$$

for a *given* deformation gradient $\boldsymbol{F}^{(k)}_{n+1}$, through the integration of the system of ODEs (73)–(77) provided by the elastoplastic constitutive equations. Note that the evolution problem defined by (73)–(77) belongs to the category of the so–called *stiff differential–algebraic systems* due to the algebraic constraints of eqs. (77) – see [HW91] for details. At the end of the update process, the Kirchhoff stress tensor $\boldsymbol{\tau}_{n+1}$ at time $t_{n+1}$ can be evaluated from $\boldsymbol{b}^e_{n+1}$ by means of the hyperelastic constitutive equation $(71)_1$.

## 4.3 Consistent linearization of the stress update algorithm

In a standard finite element context, the starting point for the solution of a static equilibrium problem is the weak form of the balance of momentum equation, which, for the problem at hand, is stated as follows. Find the unknown deformation $\boldsymbol{\varphi}_{n+1} = \boldsymbol{X} + \boldsymbol{u}_{n+1}$ such that, for any test function (variation) $\boldsymbol{\eta}$ satisfying homogeneous boundary conditions on the appropriate part of the boundary, the following non–linear functional equation is satisfied:

$$\mathscr{G}(\boldsymbol{\varphi}_{n+1}, \boldsymbol{\eta}) = \mathscr{G}^{\text{int}}(\boldsymbol{\varphi}_{n+1}, \boldsymbol{\eta}) - \mathscr{G}^{\text{ext}}_{n+1} = \int_{\mathcal{B}} \boldsymbol{\tau}(\boldsymbol{\varphi}_{n+1}) \cdot (\nabla \boldsymbol{\eta})\, dV - \mathscr{G}^{\text{ext}}_{n+1} = 0 \quad (87)$$

The iterative solution via Newton's method of the non–linear algebraic problem resulting after the introduction of a standard finite element discretization, requires the linearization of the non–linear functional $\mathscr{G}$ with respect to the independent field $\boldsymbol{\varphi}_{n+1}$ in the direction $\delta\boldsymbol{u}$:

$$D_u \mathscr{G}^{\text{int}}(\boldsymbol{\varphi}^{(k)}_{n+1}, \boldsymbol{\eta})[\delta \boldsymbol{u}] = \int_{\mathcal{B}} \left\{ \nabla^s \boldsymbol{\eta} \cdot (\widetilde{\mathfrak{c}})^{(k)}_{n+1}\, \nabla^s \delta\boldsymbol{u} \right\} dV$$

$$+ \int_{\mathcal{B}} \left\{ \boldsymbol{\tau}^{(k)}_{n+1} \cdot (\nabla \delta \boldsymbol{u})^T (\nabla \boldsymbol{\eta}) \right\} dV \quad (88)$$

in which:

$$\widetilde{\mathfrak{c}}_{ijkl} = \widetilde{\mathbb{C}}_{ABCD} F^e_{iA} F^e_{jB} F^e_{kC} F^e_{lD} \quad (89)$$

and:

$$\widetilde{\mathbb{C}}^{(k)}_{n+1} = \left( 2 \frac{\partial \boldsymbol{S}}{\partial \boldsymbol{C}} \right)^{(k)}_{n+1} \qquad \text{with} \qquad \boldsymbol{S} := \boldsymbol{F}^{-1} \boldsymbol{\tau} \boldsymbol{F}^{-T} \qquad \boldsymbol{C} := \boldsymbol{F}^T \boldsymbol{F} \quad (90)$$

In eq. (88), the fourth–order tensor $\widetilde{\mathfrak{c}}$ is the so–called *spatial algorithmic tangent stiffness tensor*, obtained from the *material algorithmic tangent stiffness tensor* $\widetilde{\mathbb{C}}$ by the pull–back operation (89). This last quantity represents the variation of the updated second Piola–Kirchhoff stress tensor $\boldsymbol{S}^{(k)}_{n+1}$ associated to the infinitesimal change of the right Cauchy–Green deformation tensor $\boldsymbol{C}^{(k)}_{n+1}$ induced by the infinitesimal perturbation of the deformation field $\delta\boldsymbol{u}$. As such, the tensor $\widetilde{\mathbb{C}}^{(k)}_{n+1}$ is strongly dependent on the adopted integration algorithm [ST85]. Its accurate evaluation is crucial to achieve the quadratic convergence when using Newton–Raphson method to solve iteratively the global discrete equilibrium equations.

## 4.4 IMPLEX algorithm

In finite–deformation plasticity, explicit methods have not been so widely used as in infinitesimal plasticity. Notable exceptions are represented by the works of refs. [SO85, RFPH97, BRB16]. More recently, Monforte *et al.* [MCC$^+$19] extended the

IMPLicit–EXplicit (IMPLEX) algorithm proposed by Oliver *et al.* [OHC08] to increase the robustness and efficiency of classical fully–implicit return mapping algorithms to finite deformations. Applications of the IMPLEX method to computational geomechanics problems are reported in [MCC$^{+}$19, MGA$^{+}$21, OCT21, HS21].

The basic structure of the IMPLEX algorithm consists in a two–step solver:

1. *Extrapolation step*: the boundary–value problem is computed using an extrapolated value of the plastic multiplier increment.
2. *Correction step*: the final converged state is computed at each integration point using the displacement field obtained in Step 1. The resulting final plastic multiplier is then used for the next extrapolation step.

In a typical time step $[t_n, t_n + 1] \in \mathbb{I}$, the extrapolation step updates the state variables to their so–called IMPLEX values:

$$(\tilde{\boldsymbol{b}}^e_{n+1}, \tilde{\boldsymbol{q}}_{n+1})$$

obtained through explicit integration of the evolution equations by assuming a constant plastic multiplier increment:

$$\widetilde{\Delta\gamma}_{n+1} = \frac{\Delta t_{n+1}}{\Delta t_n} \Delta\gamma_n$$

To obtain an explicit update for $\boldsymbol{b}^e$, we observe that eq. $(66)_2$ provides an evolution equation for $\boldsymbol{F}^p$ in the form:

$$\dot{\boldsymbol{F}}^p = \overline{\boldsymbol{L}}^p \boldsymbol{F}^p = \left\{ \boldsymbol{F}^{e-1} \left( \dot{\gamma} \frac{\partial g}{\partial \boldsymbol{\tau}} \right) \boldsymbol{F}^e \right\} \boldsymbol{F}^p \tag{91}$$

By adopting an explicit exponential mapping to integrate the evolution equation (91) we have:

$$\boldsymbol{F}^p_{n+1} = \exp\left\{ \Delta\gamma_{n+1} \boldsymbol{F}^{e-1}_n \left( \frac{\partial g}{\partial \boldsymbol{\tau}} \right)_n \boldsymbol{F}^e_n \right\} \boldsymbol{F}^p_n \tag{92}$$

from which, replacing $\Delta\gamma_{n+1}$ with the *known* extrapolated plastic multiplier $\widetilde{\Delta\gamma}_{n+1}$, we finally obtain, after some algebra:

$$\tilde{\boldsymbol{b}}^e_{n+1} = \boldsymbol{f}_{n+1} \exp\left\{ -\widetilde{\Delta\gamma}_{n+1} \left( \frac{\partial g}{\partial \boldsymbol{\tau}} \right)_n \right\} \boldsymbol{b}^e_n \exp\left\{ -\widetilde{\Delta\gamma}_{n+1} \left( \frac{\partial g}{\partial \boldsymbol{\tau}} \right)_n \right\}^T \boldsymbol{f}^T_{n+1} \tag{93}$$

where $\boldsymbol{f}_{n+1} = \boldsymbol{F}_{n+1} \boldsymbol{F}^{-1}_n = \boldsymbol{1} + \nabla_n \boldsymbol{u}_{n+1}$ is the relative deformation gradient. The details of the derivation of eq. (93) are provided, for example, in [OCT21]. Using the elastic constitutive equation $(71)_1$, the derived Kirchhoff stress $\tilde{\boldsymbol{\tau}}_{n+1} = \boldsymbol{\tau}(\tilde{\boldsymbol{b}}^e_{n+1})$ is then obtained. Analogously, from the evolution equations (76), the following IMPLEX values for the internal variables are obtained:

$$\tilde{\boldsymbol{q}}_{n+1} = \boldsymbol{q}_n + \widetilde{\Delta\gamma}_{n+1} \boldsymbol{h}_n \tag{94}$$

According to eq. (94), the IMPLEX internal state variables depend only on known quantities, while $\tilde{\boldsymbol{b}}^e_{n+1}$ and $\tilde{\boldsymbol{\tau}}_{n+1}$ depend also on the unknown displacement field at the end of the step, $\boldsymbol{u}_{n+1}$. This field is determined by solving the global discretized equilibrium equations. In solving the global equilibrium problem, the global stiffness matrix coming from the linearization of the internal force vector can be computed using the elastic tangent stiffness tensor of eq. (81), since the plastic flow is independent of the displacement field.

Once the extrapolation step is completed, the correction step is performed at constant spatial configuration (*i.e.*, constant $\boldsymbol{u}_{n+1}$) to determine more accurate values of the state variables at the end of the step $(\boldsymbol{b}^e_{n+1}, \boldsymbol{q}_{n+1})$. In the original IMPLEX method [OCW$^+$07] this step is carried out by implicit numerical integration of the evolution equations. In the method proposed by [MCC$^+$19], an explicit adaptive scheme with substepping and error control is adopted to update the left elastic Cauchy–Green tensor and the internal variables. For a typical substep $[t_k, t_{k+1}] \in [t_n, t_{n+1}]$ we thus have:

$$\boldsymbol{b}^e_{k+1} = \boldsymbol{f}_{k+1} \exp\left\{-\Delta\gamma_{k+1}\left(\frac{\partial g}{\partial \boldsymbol{\tau}}\right)_k\right\} \boldsymbol{b}^e_k \exp\left\{-\Delta\gamma_{k+1}\left(\frac{\partial g}{\partial \boldsymbol{\tau}}\right)_k\right\}^T \boldsymbol{f}^T_{k+1} \tag{95}$$

and:

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k + \Delta\gamma_{k+1}\boldsymbol{h}_k \tag{96}$$

The plastic multiplier appearing in the above equations is provided by the explicit integration of eq. (82):

$$\Delta\gamma_{k+1} = \Delta t_{k+1}\dot{\gamma}_k = \frac{1}{(\widehat{K}_p)_k}\left(\frac{\partial f}{\partial \boldsymbol{\tau}}\right)_k \cdot \mathbf{a}^e_k \nabla^s_k\left(\Delta \boldsymbol{u}_{k+1}\right) \tag{97}$$

where $\nabla^s_k\left(\Delta \boldsymbol{u}_{k+1}\right)$ is the symmetric part of the spatial gradient of the displacement increment within the substep. The final value of the plastic multiplier at the end of the step $(t = t_{n+1})$ is then used for the extrapolation stage of the next computational step.

## 4.5 Implicit Generalized Backward Euler method

Until the beginning of the '80, computational methods for finite deformation elastoplasticity relied on models based on the additive decomposition of the rate of deformation tensor, see [OT21] in this volume. Therefore, they remained restricted to small elastic strains. Early works on computational applications of finite deformation plasticity models based on the multiplicative decomposition of the deformation gradient are presented, *e.g.*, in [AD79, SO85, Sim85]. For the case of isotropic plasticity, a very important contribution has been given by the work of Simo [Sim92] where he advocated the use of principal elastic logarithmic strains as primary state variables, in connection to an hyperelastic characterization of the elastic behavior of the material, to formulate an implicit Backward Euler elastic predictor–plastic corrector algorithm with the same structure of the corresponding integration scheme of infinitesimal plasticity. Applications of this approach to computational geomechanics are reported in

| | *Global* | *Elastic predictor* | *Plastic corrector* |
|---|---|---|---|
| Evol. eqs. | $\dot{\boldsymbol{f}} = \boldsymbol{l}\boldsymbol{f}$<br>$\dot{\boldsymbol{b}}^e = \boldsymbol{l}\boldsymbol{b}^e + \boldsymbol{b}^e\boldsymbol{l}^T - 2\dot{\gamma}\,\dfrac{\partial g}{\partial \boldsymbol{\tau}}\boldsymbol{b}^e$<br>$\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}$ | $\dot{\boldsymbol{f}} = \boldsymbol{l}\boldsymbol{f}$<br>$\dot{\boldsymbol{b}}^e = \boldsymbol{l}\boldsymbol{b}^e + \boldsymbol{b}^e\boldsymbol{l}^T$<br>$\dot{\boldsymbol{q}} = \boldsymbol{0}$ | $\dot{\boldsymbol{f}} = \boldsymbol{0}$<br>$\dot{\boldsymbol{b}}^e = -2\dot{\gamma}\,\dfrac{\partial g}{\partial \boldsymbol{\tau}}\boldsymbol{b}^e$<br>$\dot{\boldsymbol{q}} = \dot{\gamma}\boldsymbol{h}$ |
| Init. conds. | $\boldsymbol{b}^e(t_n) = \boldsymbol{b}^e_n$<br>$\boldsymbol{q}(t_n) = \boldsymbol{q}_n$ | $\boldsymbol{b}^e(t_n) = \boldsymbol{b}^e_n$<br>$\boldsymbol{q}(t_n) = \boldsymbol{q}_n$ | $\boldsymbol{b}^e\Big\rvert_{(\dot{\gamma}=0)} = \boldsymbol{b}^{e,\mathrm{tr}}_{n+1}$<br>$\boldsymbol{q}\Big\rvert_{(\dot{\gamma}=0)} = \boldsymbol{q}^{\mathrm{tr}}_{n+1}$ |
| Constr. | $f(\boldsymbol{\tau}, \boldsymbol{q}) \leq 0$<br>$\dot{\gamma} \geq 0$<br>$f(\boldsymbol{\tau}, \boldsymbol{q})\dot{\gamma} = 0$ | none | $f(\boldsymbol{\tau}, \boldsymbol{q}) \leq 0$<br>$\dot{\gamma} \geq 0$<br>$f(\boldsymbol{\tau}, \boldsymbol{q})\dot{\gamma} = 0$ |

Table 3: Operator split of the evolution problem of multiplicative plasticity, formulated in terms of elastic deformation rates.

the works of [SM93, BT98, CAS98, SSS02, OT20]. In the remainder of this section, we focus on this class of stress–point algorithms following closely the work of [OT20].

### 4.5.1 Operator split and product formula algorithm

For the implicit numerical integration of the evolution equations (73)–(77), we proceed as in the case of infinitesimal plasticity by adopting the *operator split* shown in Tab. 3, suggested by the additive structure of the evolution problem.

Again, computational strategy is to solve the elastic predictor problem first, with initial conditions provided by $(\boldsymbol{b}^e_n, \boldsymbol{q}_n)$, obtaining the so–called *trial solution* $(\boldsymbol{b}^{e,\mathrm{tr}}_{n+1}, \boldsymbol{q}^{\mathrm{tr}}_{n+1})$. Then, if the constraints posed by the complementarity conditions are violated, solve the plastic corrector problem using the trial solution as initial conditions. The attractiveness of this strategy stands in the geometric interpretation which can be given to each Problem, as detailed below.

### 4.5.2 Problem 1: elastic predictor

The evolution equations of the elastic predictor problem are obtained from the original problem by assuming that no dissipative processes take place ($\dot{\gamma} = 0$) and ignoring the constraint placed on the state variables by the yield function.

From a geometric point of view, during the elastic predictor stage, the update of the current configuration from $\mathcal{S}_n$ to $\mathcal{S}_{n+1}$ takes place at fixed intermediate configuration (modulo a rigid body rotation), with $\boldsymbol{F}^{p,\mathrm{tr}}_{n+1} = \boldsymbol{F}^p_n$. Thus we have:

$$\boldsymbol{F}_{n+1} = \boldsymbol{f}_{n+1}\boldsymbol{F}_n = \boldsymbol{F}^{e,\mathrm{tr}}_{n+1}\boldsymbol{F}^p_n \qquad \Rightarrow \qquad \boldsymbol{F}^{e,\mathrm{tr}}_{n+1} = \boldsymbol{f}_{n+1}\boldsymbol{F}^e_n \tag{98}$$

From this last result and the (trivial) evolution equation for $\boldsymbol{q}$ of the elastic predictor problem (see Tab. 3), the complete trial state is obtained:

$$\boldsymbol{b}^{e,\mathrm{tr}}_{n+1} = \boldsymbol{f}_{n+1}\boldsymbol{b}^e_n\boldsymbol{f}^T_{n+1} \qquad\qquad \boldsymbol{q}^{\mathrm{tr}}_{n+1} = \boldsymbol{q}_n \tag{99}$$

Then, the trial Kirchhoff stress is evaluated via the hyperelastic constitutive equation $(71)_1$ as $\boldsymbol{\tau}^{\mathrm{tr}}_{n+1} = \boldsymbol{\tau}(\boldsymbol{b}^{e,\mathrm{tr}}_{n+1})$.

It is worth noting that, due to its formulation in terms of kinematics, the elastic predictor problem can be solved exactly. The trial value of $\boldsymbol{b}^e$ at the end of the step is just the geometric update (actually, the push–forward) of $\boldsymbol{b}^e_n$ to the current configuration $\mathcal{S}_{n+1}$ via the relative deformation gradient.

### 4.5.3 Problem 2: plastic corrector

If the trial state satisfies the constraint posed by the Kuhn–Tucker conditions, *i.e.*:

$$f^{\mathrm{tr}}_{n+1} := f\left(\boldsymbol{b}^{e,\mathrm{tr}}_{n+1}, \boldsymbol{q}^{\mathrm{tr}}_{n+1}\right) \leq 0$$

then the trial state provides the exact update of the material state sought after. Otherwise, the intermediate configuration needs to be modified in order to restore the consistency with the yield surface:

$$f_{n+1} = f\left(\boldsymbol{b}^{e}_{n+1}, \boldsymbol{q}_{n+1}\right) = 0 \tag{100}$$

where $\boldsymbol{b}^e_{n+1}$ and $\boldsymbol{q}_{n+1}$ are the solution of the differential–algebraic plastic corrector problem. Since $\dot{\boldsymbol{f}} = \boldsymbol{0}$ in this case, the plastic corrector problem is formulated on a fixed current configuration $\mathcal{S}_{n+1}$.

The numerical solution of the plastic corrector problem is typically obtained by adopting an implicit strategy such as the Backward Euler method. In particular, the structure of the evolution equation for $\boldsymbol{b}^e$ suggest the use of the following exponential approximation, see [Sim92]:

$$\boldsymbol{b}^e_{n+1} = \exp\left\{-2\Delta\gamma_{n+1}\left(\frac{\partial g}{\partial \boldsymbol{\tau}}\right)_{n+1}\right\}\boldsymbol{b}^{e,\mathrm{tr}}_{n+1} \tag{101}$$

where $\Delta\gamma_{n+1}$ is the increment of the plastic multiplier associated to the plastic deformations, to be determined as part of the solution.

Finally, using the Backward Euler algorithm to integrate the evolution equation for $\boldsymbol{q}$, we obtain:

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}^{\mathrm{tr}}_{n+1} + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} = \boldsymbol{q}_n + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} \tag{102}$$

In principle, the system of (6 + $n_{\text{int}}$ + 1) non–linear algebraic equations (101) and (102) can be solved to provide the unknowns $\boldsymbol{b}^e_{n+1}$, $\boldsymbol{q}_{n+1}$ and $\Delta\gamma_{n+1}$. However, as first shown by Simo [Sim92], the solution of the plastic corrector problem can be significantly simplified by exploiting the isotropy of the material response, as shown in the following.

### 4.5.4 Plastic corrector problem in principal logarithmic elastic strains space

Due to the assumption of material isotropy, the tensor $(\partial g/\partial\boldsymbol{\tau})_{n+1}$ has the same principal directions of $\boldsymbol{\tau}$ and hence of $\boldsymbol{b}^e_{n+1}$, due to eqs. (70) and $(71)_1$. Therefore, the spectral decomposition of the tensors $\boldsymbol{b}^e_{n+1}$, $(\partial g/\partial\boldsymbol{\tau})_{n+1}$ and $\boldsymbol{b}^{e,\text{tr}}_{n+1}$ appearing in eq. (101) read:

$$\boldsymbol{b}^e_{n+1} = \sum_{A=1}^{3} \left(\lambda^e_{A,n+1}\right)^2 \boldsymbol{n}^{(A)}_{n+1} \otimes \boldsymbol{n}^{(A)}_{n+1} \tag{103a}$$

$$\left(\frac{\partial g}{\partial\boldsymbol{\tau}}\right)_{n+1} = \sum_{A=1}^{3} \left(\frac{\partial g}{\partial\tau_A}\right)_{n+1} \boldsymbol{n}^{(A)}_{n+1} \otimes \boldsymbol{n}^{(A)}_{n+1} \tag{103b}$$

$$\boldsymbol{b}^{e,\text{tr}}_{n+1} = \sum_{A=1}^{3} \left(\lambda^{e,\text{tr}}_{A,n+1}\right)^2 \boldsymbol{n}^{(A),\text{tr}}_{n+1} \otimes \boldsymbol{n}^{(A),\text{tr}}_{n+1} \tag{103c}$$

where the quantities $\lambda^{e,\text{tr}}_A$ and $\boldsymbol{n}^{(A),\text{tr}}$ denote the trial principal elastic stretches (eigenvalues of $\boldsymbol{F}^{e,\text{tr}}$) and the unit eigenvectors of $\boldsymbol{b}^{e,\text{tr}}$, respectively, while the scalars $\partial g/\partial\tau_A$ are the derivatives of the plastic potential functions with respect to the principal values of $\boldsymbol{\tau}$.

Rewriting eq. (101) as:

$$\exp\left\{2\Delta\gamma_{n+1}\left(\frac{\partial g}{\partial\boldsymbol{\tau}}\right)_{n+1}\right\}\boldsymbol{b}^e_{n+1} = \boldsymbol{b}^{e,\text{tr}}_{n+1} \tag{104}$$

and incorporating the spectral decompositions (103), it easy to show that:

a) the principal directions of $\boldsymbol{b}^e_{n+1}$ coincide with the (known) principal directions of $\boldsymbol{b}^{e,\text{tr}}_{n+1}$:

$$\boldsymbol{n}^{(A)}_{n+1} = \boldsymbol{n}^{(A),\text{tr}}_{n+1} \qquad (A = 1,2,3) \tag{105}$$

b) the principal values of the three tensors $\boldsymbol{b}^e_{n+1}$, $(\partial g/\partial\boldsymbol{\tau})_{n+1}$ and $\boldsymbol{b}^{e,\text{tr}}_{n+1}$ are related by the following equations:

$$\left(\lambda^e_{A,n+1}\right)^2 = \exp\left\{-2\Delta\gamma_{n+1}\left(\frac{\partial g}{\partial\tau_A}\right)_{n+1}\right\}\left(\lambda^{e,\text{tr}}_{A,n+1}\right)^2 \tag{106}$$

with $A$ = 1, 2, 3.

The result in eq. (106) is particularly relevant since, taking the natural logarithm of both sides, we obtain:

$$\varepsilon^e_{A,n+1} = \varepsilon^{e,\mathrm{tr}}_{A,n+1} - \Delta\gamma_{n+1} \left(\frac{\partial g}{\partial \tau_A}\right)_{n+1} \tag{107}$$

where:

$$\varepsilon^{e,\mathrm{tr}}_{A,n+1} := \ln(\lambda^{e,\mathrm{tr}}_{A,n+1}) \qquad \varepsilon^{e}_{A,n+1} := \ln(\lambda_{A,n+1})$$

Introducing the following vector notation:

$$\hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}} := \begin{Bmatrix} \varepsilon_1^{e,\mathrm{tr}} \\ \varepsilon_2^{e,\mathrm{tr}} \\ \varepsilon_3^{e,\mathrm{tr}} \end{Bmatrix} \qquad \hat{\boldsymbol{\varepsilon}}^{e} := \begin{Bmatrix} \varepsilon_1^{e} \\ \varepsilon_2^{e} \\ \varepsilon_3^{e} \end{Bmatrix} \qquad \hat{\boldsymbol{Q}} := \begin{Bmatrix} \partial g/\partial \tau_1 \\ \partial g/\partial \tau_2 \\ \partial g/\partial \tau_3 \end{Bmatrix}$$

The system of algebraic equations governing the return mapping problem formulated in principal logarithmic elastic strains space takes the following form:

$$\hat{\boldsymbol{\varepsilon}}^e_{n+1} = \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}_{n+1} - \Delta\gamma_{n+1}\hat{\boldsymbol{Q}}_{n+1} \tag{108a}$$

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} \tag{108b}$$

$$f_{n+1} = f\left(\boldsymbol{b}^e_{n+1}, \boldsymbol{q}_{n+1}\right) = 0 \tag{108c}$$

This set of (3+$n_{\mathrm{int}}$+1) non–linear algebraic equations can be solved using Newton's method to obtain the updated state at the end of the step and the plastic multiplier increment, as shown in Tab. 4.

As noted by [Sim92], the use of the exponential algorithm in connection with the choice of formulating the plastic corrector problem in principal logarithmic elastic strain space leads to an algebraic system of equations which are formally similar to the Generalized Backward Euler algorithm of infinitesimal plasticity, see eqs. (50).

### 4.5.5 Consistent tangent stiffness

By differentiating the expression for $\boldsymbol{\tau}_{n+1}$ provided by the spectral decomposition of Tab. 4, the following expression for the spatial consistent tangent stiffness tensor $\widetilde{\mathfrak{c}}$ of eq. (89) is obtained (see [Sim98]):

$$\widetilde{\mathfrak{c}} = \sum_{A=1}^{3}\sum_{B=1}^{3} \hat{d}_{AB}\, \boldsymbol{m}^A \otimes \boldsymbol{m}^B - \sum_{A=1}^{3} 2\tau_A\, \boldsymbol{m}^A \\ + \sum_{A\neq B} \left\{ \frac{\tau_A (\lambda_B^{e,\mathrm{tr}})^2 - \tau_B (\lambda_A^{e,\mathrm{tr}})^2}{(\lambda_A^{e,\mathrm{tr}})^2 - (\lambda_B^{e,\mathrm{tr}})^2} \right\} \boldsymbol{M}^{AB} \tag{109}$$

1. Determine the trial principal elastic stretches $\lambda_{A,n+1}^{e,\mathrm{tr}}$ and the principal eigenvectors $\boldsymbol{n}_{n+1}^{(A),\mathrm{tr}}$ via the spectral decomposition of $\boldsymbol{b}_{n+1}^{e,\mathrm{tr}}$.
2. Set:
$$\boldsymbol{n}_{n+1}^{(A)} = \boldsymbol{n}_{n+1}^{(A),\mathrm{tr}}$$
for $A$ = 1, 2, 3.
3. Solve the system of nonlinear algebraic equations:
$$\begin{aligned}
\boldsymbol{R}_\varepsilon &:= -\hat{\boldsymbol{\varepsilon}}_{n+1}^{e} + \hat{\boldsymbol{\varepsilon}}_{n+1}^{e,\mathrm{tr}} - \Delta\gamma_{n+1}\hat{\boldsymbol{Q}}_{n+1} = \boldsymbol{0}\\
\boldsymbol{R}_q &:= -\boldsymbol{q}_{n+1} + \boldsymbol{q}_n + \Delta\gamma_{n+1}\boldsymbol{h}_{n+1} = \boldsymbol{0}\\
R_f &:= -f\left(\boldsymbol{b}_{n+1}^{e}, \boldsymbol{q}_{n+1}\right) = 0
\end{aligned}$$
via Newton's method, to obtain the updated state variables at the end of the step.
4. Recover $\boldsymbol{b}_{n+1}^{e}$ and $\tau_{n+1}$ using the spectral decomposition and the hyperelastic constitutive equation:
$$b_{A,n+1}^{e} = \exp\left(2\varepsilon_{A,n+1}^{e}\right) \qquad \boldsymbol{b}_{n+1}^{e} = \sum_{A=1}^{3} b_{A,n+1}^{e}, \boldsymbol{n}_{n+1}^{(A)} \otimes \boldsymbol{n}_{n+1}^{(A)}$$
$$\tau_{A,n+1} = \left(\frac{\partial\psi}{\partial\varepsilon_{A}^{e}}\right)_{n+1} \qquad \boldsymbol{\tau}_{n+1} = \sum_{A=1}^{3} \tau_{A,n+1}, \boldsymbol{n}_{n+1}^{(A)} \otimes \boldsymbol{n}_{n+1}^{(A)}$$

Table 4: Solution strategy for the plastic corrector problem of isotropic multiplicative plasticity.

where:

$$\boldsymbol{m}^A := \boldsymbol{n}^{(A)} \otimes \boldsymbol{n}^{(A)} \quad \boldsymbol{m}^{AB} := \boldsymbol{n}^{(A)} \otimes \boldsymbol{n}^{(B)} \quad \boldsymbol{m}^{BA} := \boldsymbol{n}^{(B)} \otimes \boldsymbol{n}^{(A)}$$

$$\boldsymbol{M}^{(AB)} := \boldsymbol{m}^{AB} \otimes \boldsymbol{m}^{AB} + \boldsymbol{m}^{AB} \otimes \boldsymbol{m}^{BA}$$

The quantities $\hat{d}_{AB}$ in eq. (109), defined as:

$$\hat{d}_{AB} := \frac{\partial \tau_A}{\partial \varepsilon_B^{e,\mathrm{tr}}} \tag{110}$$

are the components of the $(3 \times 3)$ matrix $\hat{\boldsymbol{d}} := \partial \hat{\boldsymbol{\tau}} / \partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}$ of tangent moduli in principal strain space. In presence of repeated eigenvalues for $\boldsymbol{b}^{e,\mathrm{tr}}$, the third term on the RHS of eq. (109) becomes singular. The singularity can be easily eliminated as shown in [Ogd84], Ch. 6.

For the case at hand, the exact calculation of the matrix $\hat{\boldsymbol{d}}$ is possible only if, during the current time step, the loading process is elastic. When the plastic deformations occur, the Kirchhoff stress tensor is a function of $\boldsymbol{b}_{n+1}^e$ which is determined numerically via the algorithm of Tab. 4. In such conditions, the evaluation of $\hat{\boldsymbol{d}}$ requires the linearization of the integration algorithm and proceeds as follows.

In terms of principal values of Kirchhoff stresses and principal elastic logarithmic strains, the hyperelastic constitutive equation reads:

$$\tau_A = \frac{\partial \hat{\psi}}{\partial \varepsilon_{A,n+1}^e} \qquad \text{or, in vector format} \qquad \hat{\boldsymbol{\tau}} = \frac{\partial \hat{\psi}}{\partial \hat{\boldsymbol{\varepsilon}}^e} \tag{111}$$

where $\hat{\boldsymbol{\tau}} := \{\tau_1, \tau_2, \tau_3\}^T$. From this equation we have:

$$\hat{\boldsymbol{d}}_{n+1}^{(k)} = \left(\frac{\partial \hat{\boldsymbol{\tau}}}{\partial \hat{\boldsymbol{\varepsilon}}^e}\right)_{n+1}^{(k)} \left(\frac{\partial \hat{\boldsymbol{\varepsilon}}^e}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}}\right)_{n+1}^{(k)} = (\hat{\boldsymbol{D}}^e)_{n+1}^{(k)} \left(\frac{\partial \hat{\boldsymbol{\varepsilon}}^e}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}}\right)_{n+1}^{(k)} \tag{112}$$

where:

$$(\hat{\boldsymbol{D}}^e)_{n+1}^{(k)} := \left(\frac{\partial \hat{\boldsymbol{\tau}}}{\partial \hat{\boldsymbol{\varepsilon}}^e}\right)_{n+1}^{(k)} = \left(\frac{\partial^2 \hat{\psi}}{\partial \hat{\boldsymbol{\varepsilon}}^e \otimes \partial \hat{\boldsymbol{\varepsilon}}^e}\right)_{n+1}^{(k)} \tag{113}$$

is the $(3 \times 3)$ elastic tangent stiffness matrix in principal directions. Now, let us define:

$$\boldsymbol{x}_{n+1}^{(k)} := \begin{Bmatrix} (\hat{\boldsymbol{\varepsilon}}^e)_{n+1}^{(k)} \\ \boldsymbol{q}_{n+1}^{(k)} \\ \Delta\gamma_{n+1}^{(k)} \end{Bmatrix} \qquad \hat{\boldsymbol{K}}_{n+1}^{e(k)} := \begin{bmatrix} \hat{\boldsymbol{D}}^e & \boldsymbol{0}_{(3 \times n_{\mathrm{int}})} & \boldsymbol{0}_{(3 \times 1)} \end{bmatrix}_{n+1}^{(k)} \tag{114}$$

as the vector of unknown state variables and plastic multiplier increment, and a auxiliary matrix containing the elastic stiffness matrix, eq. $(112)_2$ can be rewritten in the following alternative form:

$$\hat{\boldsymbol{d}}^{(k)}_{n+1} = \hat{\boldsymbol{K}}^{e(k)}_{n+1} \left( \frac{\partial \boldsymbol{x}}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}} \right)^{(k)}_{n+1} \tag{115}$$

The derivative $\partial \boldsymbol{x} / \partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}$ measures the variation in the converged solution of the iterative algorithm used to solve the plastic corrector problem for an infinitesimal change in the relative displacement gradient $\boldsymbol{f}_{n+1}$, and thus in $\hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}_{n+1}$. This quantity can be obtained by linearizing the plastic corrector problem equations of Tab. 4.

Let:

$$\boldsymbol{x}^{(k),\mathrm{tr}}_{n+1} := \begin{Bmatrix} \left(\hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}\right)^{(k)}_{n+1} \\ \boldsymbol{q}_n \\ 0 \end{Bmatrix} \quad \boldsymbol{R}^{(k)}_{n+1} := \begin{Bmatrix} -\hat{\boldsymbol{\varepsilon}}^{e}_{n+1} + \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}_{n+1} - \Delta\gamma_{n+1} \hat{\boldsymbol{Q}}_{n+1} \\ -\boldsymbol{q}_{n+1} + \boldsymbol{q}_n + \Delta\gamma_{n+1} \boldsymbol{h}_{n+1} \\ -f\left(\boldsymbol{b}^{e}_{n+1}, \boldsymbol{q}_{n+1}\right) \end{Bmatrix} \tag{116}$$

be the vector of trial values for the problem unknowns and the residual vector of the plastic corrector problem. Then, let:

$$\boldsymbol{g}^{(k)}_{n+1} := \boldsymbol{x}^{(k),\mathrm{tr}}_{n+1} - \boldsymbol{R}^{(k)}_{n+1} = \begin{Bmatrix} \left(\hat{\boldsymbol{\varepsilon}}^{e}\right)^{(k)}_{n+1} + \Delta\gamma^{(k)}_{n+1} \left(\hat{\boldsymbol{Q}}^{*}\right)^{(k)}_{n+1} \\ \boldsymbol{q}^{(k)}_{n+1} - \Delta\gamma^{(k)}_{n+1} \boldsymbol{h}^{(k)}_{n+1} \\ f^{(k)}_{n+1} \end{Bmatrix} \tag{117}$$

be the difference between $\boldsymbol{x}^{(k),\mathrm{tr}}_{n+1}$ and the residual vector $\boldsymbol{R}^{(k)}_{n+1}$ of eq. (116), *i.e.*, the only part of the residual vector which actually depends on the problem unknowns. Then the governing equations of the plastic corrector problem in Tab. 4 can be recast as follows:

$$\boldsymbol{g}\left(\boldsymbol{x}^{(k)}_{n+1}\right) = \boldsymbol{x}^{(k),\mathrm{tr}}_{n+1} \tag{118}$$

Deriving both sides of eq. (118) with respect to $\hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}$, we have:

$$\left( \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \right)^{(k)}_{n+1} \left( \frac{\partial \boldsymbol{x}}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}} \right)^{(k)}_{n+1} = \left( \frac{\partial \boldsymbol{x}^{\mathrm{tr}}}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}} \right)^{(k)}_{n+1} \tag{119}$$

Noting that:

$$\left( \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \right)^{(k)}_{n+1} = - \left( \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}} \right)^{(k)}_{n+1} = -\boldsymbol{J}^{(k)}_{n+1} \tag{120}$$

$$\left( \frac{\partial \boldsymbol{x}^{\mathrm{tr}}}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}} \right)^{(k)}_{n+1} = \begin{Bmatrix} \boldsymbol{I}_3 \\ \boldsymbol{0}_{(n_{\mathrm{int}}+1 \times 3)} \end{Bmatrix} =: \boldsymbol{T} \tag{121}$$

and considering that the Jacobian matrix $\boldsymbol{J}^{(k)}_{n+1}$ is non–singular if the plastic corrector problem is well–posed, we obtain:

$$\left(\frac{\partial \boldsymbol{x}}{\partial \hat{\boldsymbol{\varepsilon}}^{e,\mathrm{tr}}}\right)^{(k)}_{n+1} = -\left(\boldsymbol{J}^{-1}\right)^{(k)}_{n+1} \boldsymbol{T} \tag{122}$$

and, finally:

$$\hat{\boldsymbol{d}}^{(k)}_{n+1} = -\hat{\boldsymbol{K}}^{e(k)}_{n+1} \left(\boldsymbol{J}^{-1}\right)^{(k)}_{n+1} \boldsymbol{T} \tag{123}$$

The evaluation of the RHS of eq. (123) is relatively easy as the inverse of the Jacobian matrix needs to be computed for the iterative solution of the local plastic corrector problem.

# References

[AD79] J. H. Argyris and J. S. Doltsinis. On the large strain inelastic analysis in natural formulation part I: Quasistatic problems. *Comp. Meth. Appl. Mech. Engng.*, 20(2):213–251, 1979.

[ARS92] H. Alawaji, K. Runesson, and S. Sture. Implicit integration in soil plasticity under mixed control for drained and undrained response. *Int. J. Num. Anal. Meth. Geomech.*, 16:737–756, 1992.

[BL90] R. I. Borja and S. R. Lee. Cam–clay plasticity, part I. implicit integration of elastoplastic constitutive relations. *Comp. Meth. Appl. Mech. Engng.*, 78:49–72, 1990.

[Bor91] R. I. Borja. Cam–clay plasticity, part II. implicit integration of constitutive equation based on a non–linear elastic stress predictor. *Comp. Meth. Appl. Mech. Engng.*, 88:225–240, 1991.

[BRB16] K. C. Bennett, R. A. Regueiro, and R. I. Borja. Finite strain elastoplasticity considering the eshelby stress for materials undergoing plastic volume change. *International Journal of Plasticity*, 77:214–245, 2016.

[BT98] R. I. Borja and C. Tamagnini. Cam–clay plasticity, part III: Extension of the infinitesimal model to include finite strains. *Comp. Meth. Appl. Mech. Engng.*, 155:73–95, 1998.

[CAS98] C. Callari, F. Auricchio, and E. Sacco. A finite-strain cam-clay model in the framework of multiplicative elasto-plasticity. *Int. J. of Plasticity*, 14(12):1155–1187, 1998.

[dPO11] E. A. de Souza Neto, D. Peric, and D. R. J. Owen. *Computational methods for plasticity: theory and applications*. John Wiley & Sons, 2011.

[FMO09] W. Fellin, M. Mittendorfer, and A. Ostermann. Adaptive integration of constitutive rate equations. *Comp. & Geotechnics*, 36(5):698–708, 2009.

[FO02] W. Fellin and A. Ostermann. Consistent tangent operators for constitutive rate equations. *Int. J. Num. Anal. Meth. Geomech.*, 26(12):1213–1233, 2002.

[HS21] L. Hauser and H. F. Schweiger. Numerical study on undrained cone penetration in structured soil using g-pfem. *Comp. & Geotechnics*, 133:104061, 2021.

[Hug84] T. J. R. Hughes. Numerical implementation of constitutive models: rate–independent deviatoric plasticity. In S. Nemat-Nasser, R. Asaro, and G. Hegemier, editors, *Theoretical Foundations for Large Scale computations of Non Linear Material Behavior*, pages 29–57, Horton, Greece, 1984. Martinus Nijhoff Publisher, Dordrecht.

[HW91] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential–Algebraic Problems, 2nd. Ed.* Springer Verlag, New York, 1991.

[JS97] B. Jeremić and S. Sture. Implicit integration in elastoplastic geotechnics. *Mech. Cohesive–Frictional Materials*, 2:165–183, 1997.

[MCC$^+$19] L. Monforte, M. O. Ciantia, J. M. Carbonell, M. Arroyo, and A. Gens. A stable mesh–independent approach for numerical modelling of structured soils at large strains. *Comp. & Geotechnics*, 116:103215, 2019.

[MGA$^+$21] L. Monforte, A. Gens, M. Arroyo, M. Mánica, and J. M. Carbonell. Analysis of cone penetration in brittle liquefiable soils. *Comp. & Geotechnics*, 134:104123, 2021.

[MWA97] E. J. Macari, S. Weihe, and P. Arduino. Implicit integration of elastoplastic constitutive models for frictional materials with highly non–linear hardening functions. *Mech. Cohesive–Frictional Materials*, 2:1–29, 1997.

[OCT21] K. Oliynyk, M. O. Ciantia, and C. Tamagnini. A finite deformation multiplicative plasticity model with non–local hardening for bonded geomaterials. *Comp. & Geotechnics*, 137, 2021.

[OCW$^+$07] J. Oliver, J. C. Cante, R. Weyler, C. González, and J. Hernández. Particle finite element methods in solid mechanics problems. *Computational Methods in Applied Sciences*, 7:87–103, 2007.

[Ogd84] R. Ogden. *Nonlinear Elastic Deformations*. Ellis Horwood, Chichester, 1984.

[OHC08] J. Oliver, A. E. Huespe, and J. C. Cante. An implicit/explicit integration scheme to increase computability of non-linear material and contact/friction problems. *Comp. Meth. Appl. Mech. Engng.*, 197(21-24):1865–1889, 2008.

[OT20] K. Oliynyk and C. Tamagnini. Finite deformation hyperplasticity theory for crushable, cemented granular materials. *Open Geomechanics*, 2:1–33, 2020.

[OT21] K. Oliynyk and C. Tamagnini. Finite deformation plasticity. In C. Tamagnini and D. Mašín, editors, *Constitutive modelling of soils*. ALERT Geomaterials, 2021. This volume.

[PFRFH00] A. Perez-Foguet, A. Rodriguez-Ferran, and A. Huerta. Numerical differentiation for non-trivial consistent tangent matrices: an application to the mrs-lade model. *Int. J. Num. Meth. Engng.*, 48:159–184, 2000.

[PG85] D. M. Potts and A. Gens. A critical assessment of methods of correcting for drift from the yield surface in elasto-plastic finite element analysis. *Int. J. Num. Anal. Meth. Geomech.*, 9(2):149–159, 1985.

[PSS08] D. M. Pedroso, D. Sheng, and S. W. Sloan. Stress update algorithm for elastoplastic models with nonconvex yield surfaces. *Int. J. Num. Meth. Engng.*, 76(13):2029–2062, 2008.

[RFPH97] A. Rodríguez-Ferran, P. Pegon, and A. Huerta. Two stress update algorithms for large strains: accuracy analysis and numerical implementation. *Int. J. Num. Meth. Engng.*, 40(23):4363–4404, 1997.

[SAS01] S. W. Sloan, A. J. Abbo, and D. Sheng. Refined explicit integration of elastoplastic model with automatic error control. *Engineering Computations*, 18(1):121–154, 2001.

[SB92a] S. W. Sloan and J. R. Booker. Integration of tresca and mohr–coulomb constitutive relations in plane strain elastoplasticity. *Int. J. Num. Meth. Engng.*, 33(1):163–196, 1992.

[SB92b] J. Stoer and R. Bulirsch. *Introduction to numerical analysis, 2nd ed.* Springer Verlag, New York, 1992.

[SG91] J. C. Simo and S. Govindjee. Non–linear B–stability and symmetry preserving return mapping algorithms for plasticity and viscoplasticity. *Int. J. Num. Meth. Engng.*, 31:151–176, 1991.

[SH87] J. C. Simo and T. J. R. Hughes. General return mapping algorithms for rate–independent plasticity. In C.S. Desai et al., editors, *Constitutive Laws for Engineering Materials*, Horton, Greece, 1987. Elsevier Science Publishing.

[SH98] J. C. Simo and T. J. R. Hughes. *Computational inelasticity*, volume 7. Springer Science & Business Media, 1998.

[Sim85] J. C. Simo. On the computational significance of the intermediate configuration and hyperelastic stress relations in finite deformation elastoplasticity. *Mechanics of Materials*, 4(3-4):439–451, 1985.

[Sim92] J. C. Simo. Algorithms for static and dynamic multiplicative plasticity that preserve the classical return mapping schemes of the infinitesimal theory. *Comp. Meth. Appl. Mech. Engng.*, 99(1):61–112, 1992.

[Sim98] J.C. Simo. Numerical analysis and simulation of plasticity. *Handbook of numerical analysis*, 6:183–499, 1998.

[Slo87] SW Sloan. Substepping schemes for the numerical integration of elastoplastic stress–strain relations. *Int. J. Num. Meth. Engng.*, 24(5):893–911, 1987.

[SM93] J. C. Simo and G. Meschke. A new class of algorithms for classical plasticity extended to finite strains. application to geomaterials. *Computational Mechanics*, 11(4):253–278, 1993.

[SO85] J. C. Simo and M. Ortiz. A unified approach to finite deformation elastoplastic analysis based on the use of hyperelastic constitutive equations. *Comp. Meth. Appl. Mech. Engng.*, 49(2):221–245, 1985.

[SSS02] L. Sanavia, B. A. Schrefler, and P. Steinmann. A formulation for an unsaturated porous medium undergoing large inelastic strains. *Computational Mechanics*, 28(2):137–151, 2002.

[ST85] J. C. Simo and R. L. Taylor. Consistent tangent operators for rate independent elasto–plasticity. *Comp. Meth. Appl. Mech. Engng.*, 48:101–118, 1985.

[TCN02] C. Tamagnini, R. Castellanza, and R. Nova. A Generalized Backward Euler algorithm for the numerical integration of an isotropic hardening elastoplastic model for mechanical and chemical degradation of bonded geomaterials. *Int. J. Num. Anal. Meth. Geomech.*, 26:963–1004, 2002.

[TVC00] C. Tamagnini, G. Viggiani, and R. Chambon. A review of two different approaches to hypoplasticity. In D. Kolymbas, editor, *Constitutive Modelling of Granular Materials*, pages 107–145. Springer, Berlin, 2000.

[TVCD00] C. Tamagnini, G. Viggiani, R. Chambon, and J. Desrues. Evaluation of different strategies for the integration of hypoplastic constitutive equations: Application to the CLoE model. *Mech. Cohesive–Frictional Materials*, 5:263–289, 2000.

# Finite element analysis of non-isothermal multiphase porous media in quasi-statics and dynamics

**Lorenzo Sanavia***, **Toan Duc Cao****, **Maria Lazari***

**Department of Civil, Environmental and Architectural Engineering, University of Padova, Italy*

***Department of Civil Engineering Technology, Environmental Management and Safety, Rochester Institute of Technology, Rochester, NY, USA*

*This chapter presents a mathematical and a numerical model for the analysis of the thermo-hydro-mechanical (THM) behavior of multiphase deformable porous materials in both quasi-statics and dynamics. The fully coupled governing equations are developed within the Hybrid Mixture Theory. To analyze the THM behavior of soil structures in the low frequency domain, e.g. under earthquake excitation, the u-p-T formulation is invoked by neglecting the relative acceleration of the fluids and their convective terms. The standard Bubnov-Galerkin method is applied to the governing equations for the spatial discretization, whereas the generalized Newmark scheme is used for the time discretization. The final non-linear and coupled system of algebraic equations is solved by the Newton method within the monolithic approach. The formulation and the implemented solution procedure are validated through the comparison with other finite element solutions. Moreover, the issue of spurious mesh sensitivity in strain localization analysis is addressed by employing visco-plastic and non-local theories and some numerical results where shear bands develop are presented.*

## 1 Introduction

The analysis of the dynamic response of multiphase porous media has many applications in civil engineering. Onset of landslides due to earthquakes and/or rainfall and the seismic behavior of dams are examples where inertial forces cannot be neglected. Moreover, there are situations where it is important to consider also the effect of temperature variation. It is the case of fast catastrophic landslides, where the mechanical energy dissipated in heat inside the slip zone may lead to vaporization of the pore water creating a cushion of zero friction, which may accelerate the movement of the landslides [Var02]. Another interesting case is the seismic analysis of deep nuclear waste disposal.

Many authors have developed models for the analysis of the dynamic behavior of multiphase porous media in isothermal conditions. A state of art can be found in Zienkiewicz et al. [Zie99] and Schanz [Sch09]. Recently, Nenning and Schanz [Nen10] presented an infinite element for wave propagation problems; Heider et al. [Hei11] analyzed a numerical solution of dynamic wave propagation problems in infinite half spaces with incompressible constituents and Albers [Alb10] analyzed wave propagation problems in saturated and partially saturated porous media.

This work presents a formulation of a fully coupled model for deformable multiphase geomaterials in dynamics including thermal effects. The model is derived introducing the u-p-T (displacements, pressures, temperature) formulation in the multiphase model developed in Lewis and Schrefler [Lew98], in which the relative acceleration of the fluids and their convective terms have been neglected following [Cha88], [Zie99], [Cha22]. This reduced model is valid for low frequency problems, as in earthquake engineering, [Cha88], [Zie99], [Cha22]. The standard Galerkin method is applied to the governing equations for the spatial discretization, while the generalized Newmark scheme is used for the time discretization. The final nonlinear set of equations is solved by the Newton method with a monolithic approach.

The model has been implemented in the finite element code COMES-GEO, [Gaw96], [Lew98], [San06], [San08], [San09], [Gaw09], [Gaw10], [San12] and has been validated through the comparison with analytical or finite element quasi-static or dynamic solutions. In addition, the present contribution deals with the elimination of spurious mesh sensitivity problems in strain localization simulation of multiphase geomaterials, under the scope of realistic modeling of the shear zone thickness in geotechnical applications. Visco-plasticity and non-local theories are employed and implemented in COMES-GEO code to eliminate mesh dependency in strain localization even in the case of weakly rate-sensitive materials (i.e. dense sand) [LSS15], [Laz16].

## 2 Macroscopic balance equations

The full mathematical model necessary to simulate the thermo-hydro-mechanical behavior of partially saturated porous media in dynamics was developed within the Hybrid Mixture Theory (HMT) by Lewis and Schrefler [Lew98], using averaging theories according to Hassanizadeh and Gray [Has79a], [Has79b], [Has80], [Gra91]. See also [Ocz99] or [Cha22]. This model can be derived from the more advanced averaging theory TCAT - Thermodynamically Constrained Averaging Theory [Gra14] and its references listing the journal papers on this topic, or, as an introduction, the chapter of the Alert Doctoral School 2015 by Gray and Miller [Gra15].

The variably saturated porous medium is treated as a multiphase system composed of solid skeleton (*s*) with open pores filled with liquid water (*w*) and gas (*g*). The latter, is assumed to behave as an ideal mixture of dry air (non-condensable gas, *ga*) and water vapor (condensable gas, *gw*). At the macroscopic level the porous material is modeled by a substitute continuum of volume $B$ with boundary $\partial B$ that simultaneously fills the entire domain, instead of the real fluids and the solid which fill only a part of it. In this substitute continuum each constituent $\pi$ has a reduced density which is obtained through the volume fraction $\eta^{\pi}(\mathbf{x},t) = dv^{\pi}(\mathbf{x},t) / dv(\mathbf{x},t)$, where $dv$ is the volume of the average volume element (representative elementary volume, REV) of the porous medium and $dv^{\pi}$ is the volume occupied by the constituent $\pi$ in $dv$. $\mathbf{x}$ is the vector of the spatial coordinates and $t$ the current time.

The solid is deformable and non-polar and the fluids, solid and thermal effects are coupled. All fluids are in contact with the solid phase. In the model, heat conduction and heat convection, vapor diffusion, (liquid) water flow due to pressure gradients or capillary effects and water phase change (evaporation and condensation) inside the pores are taken into account.

In the partially saturated zones the liquid water is separated from its vapor by a concave meniscus (capillary water). Due to the curvature of this meniscus, the sorption equilibrium equation [Gray91] gives the relationship $p^c=p^g-p^w$ between the capillary pressure $p^c(\mathbf{x},t)$ (also known as matrix suction), gas pressure $p^g(\mathbf{x},t)$ and liquid water pressure $p^w(\mathbf{x},t)$. This expression is approximated in dynamics; it is used here because of lack of experimental results. In the following, pore pressure is defined as compressive positive for the fluids, while stress is defined as tension positive for the solid phase.

The state of the medium is described by gas pressure $p^g(\mathbf{x},t)$, capillary pressure $p^c(\mathbf{x},t)$, temperature $T(\mathbf{x},t)$ and displacements of the solid matrix $\boldsymbol{u}(\mathbf{x},t)$ [San06]. The balance equations are developed in geometrically linear framework and are written here at the macroscopic level.

For sake of completeness the equations of the model are only summarized in this chapter; the interested reader is refereed to [Cao16] for more details regarding the development of the mathematical model and its finite element implementation. Direct notation is adopted. Boldface letters denote vector or tensors and lightface italic letters are used for scalar quantities.

After neglecting the relative velocity and acceleration of the fluids in the governing equations of Lewis and Schrefler [Lew98], a set of balance equations for the whole multiphase medium is obtained as follows.

The linear momentum balance equations of the mixture in term of the generalized effective Cauchy's stress $\boldsymbol{\sigma}'(\mathbf{x},t)$ [Lew98], [Nut08] takes the form

$$div\left(\boldsymbol{\sigma}' - \left[p^g - S_w p^c\right]\mathbf{1}\right) + \rho\mathbf{g} = \boxed{\rho\mathbf{a}^s} \tag{1}$$

where $\rho = \left[1-n\right]\rho^s + nS_w\rho^w + nS_g\rho^g$ is the mass density of the overall medium, $S_w(\mathbf{x},t)$ is the degree of saturation of the liquid water $n(\mathbf{x},t)$ is the porosity and $S_g(\mathbf{x},t)$

is the degree of saturation of the gas, with $S_w$+ $S_g$=1. $\rho^s(\mathbf{x},t)$ is the density of the solid grains, $\rho^w(\mathbf{x},t)$ is the density of liquid water and $\rho^g(\mathbf{x},t)$ is the density of the gas phase. $\mathbf{g}$ is the gravity acceleration vector, $\mathbf{1}$ is the second order identity tensor and $\mathbf{a}^s(\mathbf{x},t)$ the acceleration of the solid phase. The form of Eq. (1) assumes incompressible grains, which is common in soil mechanics. In order to consider compressible grains, the Biot coefficient should be set in front of the solid pressure (this becomes important when dealing with rock and concrete). The total stress of equation (1), using saturation as weighting functions for the partial pressures, was introduced in [Sch84] using volume averaging for the bulk materials and is thermodynamically consistent, e.g. [Gra91].

The mass balance equations for the dry air and the liquid water and its vapor are, respectively:

$$\begin{aligned}&\operatorname{div}\left(\rho^{ga}\frac{k^{rg}\boldsymbol{k}}{\mu^g}\left[-\operatorname{grad}p^g+\rho^g\boldsymbol{g}\right]\right)+\operatorname{div}\left(\rho^g\frac{M_aM_w}{M_g^2}\boldsymbol{D}_g^{ga}\operatorname{grad}\left(\frac{p^{gw}}{p^g}\right)\right)\\&+\rho^{ga}S_g\operatorname{div}\boldsymbol{v}^s+nS_g\dot{\rho}^{ga}-\rho^{ga}n\dot{S}_w-\rho^{ga}\beta_s\left(1-n\right)S_g\dot{T}=0\end{aligned}\tag{2}$$

and

$$\begin{aligned}&\operatorname{div}\left(\rho^w\frac{k^{rw}\boldsymbol{k}}{\mu^w}\left(-\operatorname{grad}p^w+\rho^w\boldsymbol{g}\right)\right)+\operatorname{div}\left(\rho^{gw}\frac{k^{rg}\boldsymbol{k}}{\mu^g}\left(-\operatorname{grad}p^{gw}+\rho^{gw}\boldsymbol{g}\right)\right)\\&-\operatorname{div}\left(\rho^g\frac{M_aM_w}{M_g^2}\boldsymbol{D}_g^{gw}\operatorname{grad}\left(\frac{p^{gw}}{p^g}\right)\right)+\left[\rho^wS_w+\rho^{gw}S_g\right]\operatorname{div}\boldsymbol{v}^s+\boxed{\rho^w\frac{nS_w}{K_w}\left[\dot{p}^g-\dot{p}^c\right]}\\&-\left[\rho^w\beta_{sw}+\rho^{gw}\beta_s\left(1-n\right)S_g\right]\dot{T}+\left[n\rho^w-n\rho^{gw}\right]\dot{S}_w+nS_g\dot{\rho}^{gw}=0\end{aligned}\tag{3}$$

where $\mathbf{k}$ ($\mathbf{x}$,$t$) = $k$($\mathbf{x}$,$t$)$\mathbf{1}$ is the intrinsic permeability tensor of the porous matrix in water saturated condition [m$^2$], which is assumed to be isotropic, $k^{r\pi}(\mathbf{x},t)$ is the fluid relative permeability parameter and $\mu^\pi(\mathbf{x}, t)$ is the dynamic viscosity of the fluid [Pa s], with $\pi = w, g$. $K_w$ is the bulk modulus of the liquid water. $\beta_{sw}$ = [1-n]$\beta_s$[$S_g\rho^{gw}$ + $\rho^w S_w$], with $\beta_s(\mathbf{x}, t)$ the cubic thermal expansion coefficient of the solid. $\mathbf{D}_g^{gw}(\mathbf{x})$ is the effective diffusivity tensor of water vapor in the gas phase contained within the pore space, and $M_a$, $M_w$ and $M_g(\mathbf{x},t)$ are the molar mass of dry air, liquid water and the gas mixture $M_g=\left[\frac{\rho^{gw}}{\rho^g}\frac{1}{M_w}+\frac{\rho^{ga}}{\rho^g}\frac{1}{M_a}\right]^{-1}$, respectively. These equations contain the mass balance equation of the solid phase,

$$\frac{\partial\left(1-n\right)\rho^s}{\partial t}+\rho^s\left(1-n\right)\operatorname{div}\boldsymbol{v}^s=0$$

which has been introduced to eliminate the time derivative of the porosity.

The enthalpy balance equation for the multiphase medium is:

$$\begin{aligned}&-\mathrm{div}\left(\rho^{w}\frac{k^{rw}\mathbf{k}}{\mu^{w}}\left[-\mathrm{grad}\left(p^{w}\right)+\rho^{w}\mathbf{g}\right]\right)\Delta H_{vap}-\mathrm{div}\left(\chi_{eff}\,\mathrm{grad}T\right)-\rho^{w}S_{w}\mathrm{div}\mathbf{v}^{s}\Delta H_{vap}\\&+\left[C_{p}^{w}\rho^{w}\frac{k^{rw}\mathbf{k}}{\mu^{w}}\left[-\mathrm{grad}\left(p^{w}\right)+\rho^{w}\mathbf{g}\right]+C_{p}^{g}\rho^{g}\frac{k^{rg}\mathbf{k}}{\mu^{g}}\left[-\mathrm{grad}p^{g}+\rho^{g}\mathbf{g}\right]\right]\cdot\mathrm{grad}T\\&+\left(\rho C_{p}\right)_{eff}\dot{T}-\boxed{\rho^{w}\frac{nS_{w}}{K_{w}}\dot{p}^{w}\Delta H_{vap}}+\beta_{sw}\dot{T}\Delta H_{vap}-n\left[\rho^{w}-\rho^{gw}\right]\dot{S}_{w}\Delta H_{vap}=0\end{aligned}\qquad(4)$$

where $\left(\rho C_{p}\right)_{eff}$ (**x**,*t*) is the effective thermal capacity of the porous medium, $C_{p}^{w}$ (**x**,*t*) and $C_{p}^{g}$ (**x**,*t*) are the specific heat of water and gas, respectively, and $\chi_{eff}$(**x**,*t*) is the effective thermal conductivity of the porous medium. The last term of Equation (4) considers the contribution of the evaporation and condensation.

In equations (2)-(4) the advective fluxes have been described using Darcy’s law for liquid water and gas, while the diffusion of vapor in the gas phase has been modeled with Fick’s law.

A recent development of a model which considers the air dissolved in the liquid water and its desorption at lower water pressures in quasi-statics loading conditions is presented in [Gaw09].

It should be noted that the quasi-static version of the model is obtained by neglecting the terms in boxes in the above equations [San06].

# 3 Constitutive relationships

For the gaseous mixture of dry air and water vapor, the ideal gas law is introduced. The equation of state of perfect gas (Clapeyron’s equation) and Dalton's law are applied to dry air (ga), water vapor (gw) and moist air (g)

$$p^{ga}=\rho^{ga}TR/M_{a},\quad p^{gw}=\rho^{gw}TR/M_{w},\quad p^{g}=p^{ga}+p^{gw},\quad \rho^{g}=\rho^{ga}+\rho^{gw}\qquad(5)$$

In the partially saturated zones, the equilibrium water vapor pressure $p^{gw}$(**x**,*t*) can be obtained from the Kelvin-Laplace equation, where the water vapor saturation pressure, $p^{gws}$(**x**,*t*), depending only upon the temperature, can be calculated from the Clausius-Clapeyron equation or from an empirical correlation.

The saturation degree $S_{\mathrm{w}}$(**x**,*t*) and the relative permeability $k^{\mathrm{r}\pi}$(**x**,*t*) are experimentally determined functions dependent on capillary pressure and temperature (e.g. [Fra08] for $S_{\mathrm{w}}$).

The bulk density of liquid water that is dependent on the temperature is modeled using the relationship proposed by Furbish [Fur97].
The liquid water viscosity, dry air and water vapor viscosity, and the latent heat of evaporation are also temperature dependent relationships [Gaw12].

The solid skeleton is assumed elastic, elasto-plastic or elasto-viscoplastic, homogeneous and isotropic in the numerical simulations described in Section 5. Its mechanical behavior is described within the classical rate-independent elasto-plasticity theory for geometrically linear problems or in the framework of elasto-viscoplasticity. The latter implies a time-delayed inelastic response of the material, which is accordingly referred to as rate-sensitive or time-dependent and is adopted herein to eliminate the spurious mesh dependency in strain localization simulations.
The yield function restricting the effective stress state $\boldsymbol{\sigma}'(\mathbf{x},t)$ is developed in the form of Drucker-Prager model for simplicity, with linear isotropic softening and non-associated plastic flow to take into account the post-peak and dilatant behavior of dense sands.

For the rate-independent elastoplastic model, the return mapping and the consistent tangent operator for the Jacobian matrix, are developed in [San06], where the singular behavior of the Drucker-Prager yield surface in the zone of the apex is solved by using the multi-surface plasticity theory (following the formulation developed in [San02] for isotropic linear hardening/softening and volumetric-deviatoric non-associative plasticity in case of large strain elasto-plasticity).

The Drucker-Prager yield function with linear isotropic hardening/softening has been used for both elasto-plastic and elasto-visco-plastic models, in the form

$$F\left(p,\mathbf{s},\xi\right)=3\alpha_F\,p+\left\|\mathbf{s}\right\|-\beta_F\sqrt{\tfrac{2}{3}}\left[c_0+h\xi\right] \tag{6}$$

in which $p=\frac{1}{3}\left[\boldsymbol{\sigma}':\mathbf{1}\right]$ is the mean effective Cauchy pressure, $\left\|\mathbf{s}\right\|$ is the $L_2$ norm of the deviator part of the effective Cauchy stress tensor $\boldsymbol{\sigma}'(\mathbf{x},t)$, $c_0(\mathbf{x})$ is the initial apparent cohesion, $\alpha_F(\mathbf{x})$ and $\beta_F(\mathbf{x})$ are two material parameters related to the friction angle $\phi\,(\mathbf{x})$ of the soil,

$$\alpha_F=2\frac{\sqrt{\tfrac{2}{3}}\sin\phi}{3-\sin\phi}\qquad\beta_F=\frac{6\cos\phi}{3-\sin\phi} \tag{7}$$

$h(\mathbf{x})$ the hardening/softening modulus and $\xi(\mathbf{x},t)$ the equivalent (visco-)plastic strain.

To take into account the effect of capillary pressure and temperature on the evolution of the yield surface, the interested reader can refer, for example, to the chapter by Manzanal et al. of this book and [Fra08], [Bol05] for capillary dependent constitutive relationships in isothermal or non-isothermal conditions, respectively.

For the elasto-viscoplastic model, the Perzyna model [Per63] is adopted and the return mapping scheme along with the consistent tangent operator is developed in [Laz16]. Moreover, the non-local elasto-viscoplastic model is used to obtain mesh insensitive results even in case of weakly rate-sensitive materials (i.e. dense sand), for which classical visco-plasticity cannot achieve a mesh independent solution [dPIA02]. The non-local model introduces a characteristic length directly related to the shear band width and when this internal length approaches zero, the (local) elasto-viscoplastic model is regained. More details for the analytical formulation and the numerical treatment can be found in [LSS15], [Laz16] and will briefly described below.

## 3.1 Elasto-plasticity

The mechanical behaviour of the solid skeleton is assumed to be governed by the Helmholtz free energy $\psi$ function in the form

$$\psi = \psi(\boldsymbol{\varepsilon}^e, \xi) \tag{8}$$

dependent on the small elastic strain tensor, $\boldsymbol{\varepsilon}^e(\mathbf{x},t)$, and the internal strain-like scalar hardening variable, $\xi(\mathbf{x},t)$, i.e., the equivalent plastic strain. The second law of thermodynamic yields, under the restriction of isotropy, the constitutive relations

$$\boldsymbol{\sigma}' = \frac{\partial\psi}{\partial\varepsilon^e}, \quad q = -\frac{\partial\psi}{\partial\xi} \tag{9}$$

and the remaining dissipation inequality

$$\boldsymbol{\sigma}' : \dot{\boldsymbol{\varepsilon}}^e - q\dot{\xi} \geq 0 \tag{10}$$

where $q(\mathbf{x},t)$ is the stress-like internal variable accounting for the evolution of the yield locus in the stress space. The evolution equations for the rate terms of the dissipation inequality (10) can be derived from the postulate of the maximum plastic dissipation in the case of associative flow rules [Sim98]

$$\dot{\boldsymbol{\varepsilon}}^e = \dot{\boldsymbol{\varepsilon}} - \dot{\gamma}\frac{\partial F}{\partial \boldsymbol{\sigma}'} \text{ and } \dot{\xi} = \dot{\gamma}\frac{\partial F}{\partial q} \tag{11}$$

subjected to the classical loading-unloading conditions in Kuhn-Tucker form

$$\dot{\gamma} \geq 0 \quad F(\boldsymbol{\sigma}',q) \leq 0 \quad \dot{\gamma}F = 0 \tag{12}$$

where $\dot{\gamma}(\mathbf{x},t)$ is the continuum consistency parameter and $F = F(\boldsymbol{\sigma}',q)$ the isotropic yield function (Equation 6).

### Algorithmic formulation for elasto-plasticity

The problem of the calculation of $\dot{\boldsymbol{\varepsilon}}^e$, $\xi$ and $\boldsymbol{\sigma}'$ is typically solved by an operator split into an elastic predictor and plastic corrector [SH98]. The calculation of the trial elastic state $(\bullet)^{tr}$ is based on freezing the plastic flow at time $t_{n+1}$. The $[\boldsymbol{\varepsilon}^e_{n+1}]^{tr}$ is hence obtained from the load step by means of $[\boldsymbol{\varepsilon}^e_{n+1}]^{tr} = \boldsymbol{\varepsilon}_{n+1}$. The corresponding trial elastic state is obtained from the hyperelastic free energy function as

$$\boldsymbol{\sigma}'^{tr}_{n+1} = \left[\frac{\partial\psi}{\partial\varepsilon^e}\right]_{\boldsymbol{\varepsilon}^e=\left[\boldsymbol{\varepsilon}^e_{n+1}\right]^{tr}}, \quad q^{tr}_{n+1} = \left[\frac{\partial\psi}{\partial\xi}\right]_{\xi=\xi^{tr}_{n+1}} \tag{13}$$

If this trial state is admissible, it does not violate the inequality $F'^{tr}_{n+1} = F(\boldsymbol{\sigma}'^{tr}_{n+1}, q^{tr}_{n+1}) \leq 0$ and the stress state is hence already computed. Otherwise the return mapping or plastic corrector algorithm is applied to compute $\Delta\gamma_{n+1}$ satisfying the consistency condition $F_{n+1} = 0$ .

From the knowledge of $\Delta\gamma_{n+1}$ the equivalent plastic strain is computed by the backward Euler integration of Equation (9)$_2$

$$\xi_{n+1} = \xi_n + \Delta\gamma_{n+1} \left.\frac{\partial F}{\partial q}\right|_{n+1} \tag{14}$$

The Cauchy stress components are then computed by the hyperelastic constitutive law Equation (9)$_1$ with the free energy $\psi = \hat{\psi}(\boldsymbol{\varepsilon}_e, t)$ written as function of the principal elastic strain components and the equivalent plastic strain (for isotropic linear hardening) is

$$\hat{\psi} = \frac{L}{2}\left[\varepsilon_{1\varepsilon} + \varepsilon_{2\varepsilon} + \varepsilon_{3\varepsilon}\right]^2 + G\left(\varepsilon^2_{1\varepsilon} + \varepsilon^2_{2\varepsilon} + \varepsilon^2_{3\varepsilon}\right) + \frac{1}{2}h\xi^2 \tag{15}$$

where *L* and *G* are the elastic Lame' constants and h the linear hardening modulus.

## 3.2 Local elasto-viscoplasticity

The total strain rate in an elasto-viscoplastic material is additively decomposed into an elastic and a viscoplastic strain rate:

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^{\mathrm{e}} + \dot{\boldsymbol{\varepsilon}}^{\mathrm{vp}} \tag{16}$$

where the superimposed dot denotes time derivative. Considering linear elasticity, the stress rate is related to the strain rate via the following constitutive relation:

$$\dot{\boldsymbol{\sigma}} = \mathbf{D}^{\mathrm{e}} : \left(\dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^{\mathrm{vp}}\right) \tag{17}$$

where $\mathbf{D}^{\mathrm{e}}(\mathbf{x})$ is the fourth-order elastic tensor and double dots ":" denote the doubly contracted tensor product.

In the viscoplastic model proposed by Perzyna [Per63] (which from this point on will be referred as local to distinguish from non-local), the viscoplastic strain rate is directly linked to the yield function through the viscous nucleus, $\Phi(\mathbf{x}, t)$. The time dependency is introduced by modifying the flow rule and by abolishing the consistency rule.

The plastic potential defines the direction of the viscoplastic strain rate tensor while the yield function influences its modulus by means of the viscous nucleus. The viscous nucleus quantifies the "overstress" (i.e. $f(\sigma') > 0$) and the choice of its form determines the regularizing effect of the viscoplastic model.

The choice of the viscous nucleus is fundamental in determining the temporal material mechanical response [Laz19], which is well described also in the chapter by di Prisco et al. of this book. The simplest choice is to assume that $\Phi$ is linearly dependent on $f$ as follows:

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{vp}} = \gamma \left\langle \Phi\left(\frac{f}{f_0}\right)^{\mathrm{N}} \right\rangle \frac{\partial \mathrm{g}}{\partial \boldsymbol{\sigma}'} \tag{18}$$

with $f(\mathbf{x}, t)$ being the yield function, $f_0$ introduced as a reference fixed value making the viscous nucleous dimensionless, $\gamma(\mathbf{x})$ is a fluidity parameter which depends on the viscosity $\eta(\mathbf{x})$ of the material ($\gamma = 1/\eta$), $\mathrm{N}(\mathbf{x})$ is a calibration parameter ($\mathrm{N} \geq 1$) and $\mathrm{g}(\mathbf{x}, t)$ is the viscoplastic potential function, "<•>" are the McCauley brackets. Associative flow is obtained by $\mathrm{g} = f$.

Invoking the viscoplastic model of Perzyna with the plastic potential function and applying the chain rule of partial differentiation, the flow rule specifies to:

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{vp}} = \gamma\Phi\left(f\right)\left\{\frac{\partial \mathrm{g}}{\partial \mathrm{p}'} : \frac{\partial \mathrm{p}'}{\partial \boldsymbol{\sigma}'} + \frac{\partial \mathrm{g}}{\partial \mathbf{s}} : \frac{\partial \mathbf{s}}{\partial \boldsymbol{\sigma}'}\right\} \tag{19}$$

and leads to:

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{vp}} = \lambda\left\{\alpha_{\mathrm{g}}\mathbf{1} + \mathbf{n}\right\} \tag{20}$$

where $\mathbf{n} = \mathbf{s} / \|\mathbf{s}\|$ is the unit normal field.

Accordingly, the equivalent viscoplastic strain rate is defined in terms of the viscoplastic strain rate, as follows:

$$\dot{\xi}^{vp} = \left\|\dot{\boldsymbol{\varepsilon}}^{vp}\right\| \tag{21}$$

and considered the definition of the Euclidean norm of a second order tensor, one obtains:

$$\dot{\xi}^{vp} = \gamma\Phi\left(f\right)\left\{\left(\alpha_g\mathbf{1}+\mathbf{n}\right):\left(\alpha_g\mathbf{1}+\mathbf{n}\right)\right\}^{1/2} = \gamma\Phi\left(f\right)\sqrt{3\alpha_g+1} \tag{22}$$

**Integration algorithm**

In displacement-based finite element method the update of stress takes place at Gauss points. Assuming that at time $t_n$ the value of total and viscoplastic strains is known, the stress state is known as well. Then, suppose that an increment in total strain ($\Delta\varepsilon$) is given which drives the state to time $t_{n+1} = t_n + \Delta t$. The incremental strain, $\Delta\varepsilon = \varepsilon_{n+1} - \varepsilon_n$, is used to update the stress at time $t_{n+1}$. A trial and error strain driven process is adopted, in which an elastic trial step is first assumed by freezing the viscoplastic flow to distinguish between elastic and viscoplastic loading

$$\boldsymbol{\varepsilon}_{n+1}^{vp,\,trial} = \boldsymbol{\varepsilon}_n^{vp} \qquad \xi_{n+1}^{vp,\,trial} = \xi_n^{vp} \tag{23}$$

Then, the trial stress is tested to see if it is inside or outside the yield surface:

$$\begin{gathered} p'^{tr}_{n+1} = p'_n + K\mathrm{tr}\Delta\boldsymbol{\varepsilon}^{el,\,tr}_{n+1} = K\mathrm{tr}(\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\varepsilon}_n^{vp}) \\ \mathbf{s}^{tr}_{n+1} = \mathbf{s}_n + 2G\Delta\mathbf{e}^{el,\,tr}_{n+1} = 2G(\mathbf{e}_{n+1} - \mathbf{e}_n^{vp}) \\ f^{tr}_{n+1} = 3\alpha_f p'^{tr}_{n+1} + \left\|\mathbf{s}^{tr}_{n+1}\right\| - \beta_f\sqrt{2/3}\left(c + h\xi_n^{vp}\right) \end{gathered} \tag{24}$$

If it falls within or is on the yield surface the process is elastic and the trial state $(\bullet)^{tr}$ represents the actual final state of the material. Otherwise, the process is viscoplastic and the viscoplastic strain increment is computed by integrating Equation (20) with the unconditionally stable Backward Euler scheme:

$$\boldsymbol{\varepsilon}^{vp}_{n+1} = \boldsymbol{\varepsilon}^{vp}_n + \Delta\lambda_{n+1}\left\{\alpha_g\mathbf{1} + \mathbf{n}_{n+1}\right\} \tag{25}$$

The unit normal $\mathbf{n}_{n+1}$ is determined exclusively in terms of the trial elastic stress [SH98]:

$$\mathbf{n}_{n+1} = \frac{\mathbf{s}_{n+1}}{\left\|\mathbf{s}_{n+1}\right\|} = \frac{\mathbf{s}^{tr}_{n+1}}{\left\|\mathbf{s}^{tr}_{n+1}\right\|} \tag{26}$$

In the absence of the consistency rule, the inelastic multiplier is directly computed as:

$$\Delta\lambda_{n+1} = \frac{f^{tr}_{n+1}}{\dfrac{\eta f_0}{\Delta t} + 9\alpha_f\alpha_g K + 2G + \sqrt{\dfrac{2}{3}}\beta_f h\sqrt{3\alpha_g^2+1}} \tag{27}$$

Subsequently, from the knowledge of Δλ$_{n+1}$, the components of the stress tensor and the equivalent viscoplastic strain can be updated:

$$p'_{n+1} = p'^{tr}_{n+1} - 3K\Delta\lambda_{n+1}\alpha_g\,, \quad \mathbf{s}_{n+1} = \mathbf{s}^{tr}_{n+1} - 2G\Delta\lambda_{n+1}\mathbf{n}_{n+1} \tag{28}$$

$$\boldsymbol{\sigma}'_{n+1} = p'_{n+1}\mathbf{1} + \mathbf{s}_{n+1} = \boldsymbol{\sigma}'^{tr}_{n+1} - 3K\Delta\lambda_{n+1}\alpha_g\mathbf{1} - 2G\Delta\lambda_{n+1}\mathbf{n}_{n+1}\,, \quad \xi^{vp}_{n+1} = \xi^{vp}_{n} + \Delta\lambda_{n+1}\sqrt{\left(3\alpha_g^2+1\right)}$$

Finally, the algorithmic procedure is completed with the derivation of the algorithmic (consistent) viscoplastic tangent moduli

$$\begin{aligned}\mathbf{C}^{vp}_{n+1} = \frac{\partial\boldsymbol{\sigma}_{n+1}}{\partial\boldsymbol{\varepsilon}^{el,trial}_{n+1}} &= \left(1-\frac{9\alpha_g\alpha_f K}{d}\right)K\mathbf{1}\otimes\mathbf{1} + 2G\mathbf{I}\left(1-\frac{2G\Delta\lambda_{n+1}}{\left\|\mathbf{s}^{trial}_{n+1}\right\|}\right) - \frac{6KG\alpha_g}{d}\mathbf{1}\otimes\frac{\mathbf{s}_{n+1}}{\left\|\mathbf{s}_{n+1}\right\|}\\ &-\frac{6KG\alpha_f}{d}\frac{\mathbf{s}_{n+1}}{\left\|\mathbf{s}_{n+1}\right\|}\otimes\mathbf{1} - 4G^2\left(\frac{1}{d}-\frac{\Delta\lambda_{n+1}}{\left\|\mathbf{s}^{trial}_{n+1}\right\|}\right)\frac{\mathbf{s}_{n+1}}{\left\|\mathbf{s}_{n+1}\right\|}\otimes\frac{\mathbf{s}_{n+1}}{\left\|\mathbf{s}_{n+1}\right\|}\end{aligned}$$

where

$$d = \frac{\eta f_0}{\Delta t} + 9\alpha_g\alpha_f K + 2G + \sqrt{\frac{2}{3}}\beta_f h\sqrt{3\alpha_g^2+1}$$

From the previous expression of the parameter d it can be inferred that the viscoplastic modulus tends to the elastoplastic limit [San06] as viscosity η tends to zero. More details can be found in [Laz15], [Laz16].

It should be noted that in viscoplasticity, '*the use of consistent tangent moduli is not only desirable but also necessary*' as remarked in [J90]. This necessity stems from the fact that in viscoplastic models a continuum tangent stiffness operator does not exist, as a result of abolishing the consistency condition which prevents a direct incremental relationship to be established between the stress and the total strain increments.

Table 1: Numerical algorithm for local Perzyna model [Laz15].

```
1. Compute trial elastic state
   $p'^{tr}_{n+1} = K\mathrm{tr}(\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\varepsilon}^{vp}_{n})$ ; $\mathbf{s}^{tr}_{n+1} = 2G(\mathbf{e}_{n+1} - \mathbf{e}^{vp}_{n})$
2. Check viscoplastic flow potential
   $f^{tr}_{n+1} = 3\alpha_f p'^{tr}_{n+1} + \left\|\mathbf{s}^{tr}_{n+1}\right\| - \beta_f \sqrt{2/3}\left(c + h\xi^{vp}_{n}\right)$
   IF: $f^{tr}_{n+1} \le 0$ → elastic step → Set $(\bullet)_{n+1} = (\bullet)^{tr}_{n+1}$ & EXIT
   else go to 3
3. Compute $\Delta\lambda_{n+1}$: $\Delta\lambda_{n+1} = \dfrac{f^{tr}_{n+1}}{\dfrac{\eta f_0}{\Delta t} + 9\alpha_f \alpha_g K + 2G + \sqrt{\dfrac{2}{3}}\beta_f h\sqrt{3\alpha_g^2 + 1}}$
4. Update viscoplastic strain and stress
   $\boldsymbol{\varepsilon}^{vp}_{n+1} = \boldsymbol{\varepsilon}^{vp}_{n} + \Delta\lambda_{n+1}\left(\alpha_g \mathbf{1} + \mathbf{n}_{n+1}\right)$
   $\xi^{vp}_{n+1} = \xi^{vp}_{n} + \Delta\lambda_{n+1}\sqrt{\left(3\alpha_g^2 + 1\right)}$
   $\boldsymbol{\sigma}'_{n+1} = p'_{n+1}\mathbf{1} + \mathbf{s}_{n+1}$
5. Compute consistent viscoplastic tangent moduli
```

## 3.3 Non-local elasto-viscoplasticity

Non-local approach is introduced next because in case of weakly rate-sensitive materials (such as dense sand) artificial viscosities have to be chosen to obtain objective finite element results. To ensure a regularized numerical solution physically based and following [dPI02], the local viscoplastic model of Perzyna presented in section 3.2 is expanded with respect to the non-local approach.

Following Jirásek [Jir02] in non-local approach, a certain variable is substituted with its non-local counterpart obtained by weighted averaging over a spatial neighborhood of each point under consideration. If $f(\mathbf{x})$ is a "local" field, the corresponding non-local field is defined as:

$$\hat{f}\left(\mathbf{x}\right) = \int_V \alpha(\mathbf{x}, \xi)\, d\xi \qquad (29)$$

where $\alpha(\mathbf{x}, \xi)$ is a given non-local weight function of the point under consideration located at **x** and the neighboring (or distributing) points located at $\xi$ and V is the volume of the entire body.

For an infinite body the weight function depends only on the distance $r = \|\mathbf{x} - \xi\|$ and can be expressed as $\alpha(\mathbf{x}, \xi) = \alpha_0(\|\mathbf{x} - \xi\|)$ where $\alpha_0$ is a function of r. For a finite body, the weight function is usually adjusted such that the non-local field corresponding to a constant local field remains constant even in the vicinity of a boundary. This is guaranteed if the weight function satisfies the normalizing condition:

$$\int_V \alpha(\mathbf{x},\xi)\, d\xi = 1 \,\forall\, \mathbf{x} \in V \tag{30}$$

This condition can be achieved by imposing that the weight function is expressed by:

$$\alpha(\mathbf{x},\xi) = \frac{\alpha_0\left(\|\mathbf{x} - \xi\|\right)}{\int_V \alpha_0\left(\|\mathbf{x} - \zeta\|\right) d\zeta} \tag{31}$$

where $\alpha_0$ is the basic weight function.

Key points for the formulation and implementation of an integral non-local approach are the shape function for the averaging, the non-local variable and its discretization. In the following, these factors are discussed in detail.

### Basic shape function

The weight function always contains at least one parameter with the dimension of length which incorporates, in the simplest possible way, information about the microstructure and controls the size of the localized plastic zone [Jir03]. Herein, a Gaussian weighting function with a bounded support is selected:

$$\alpha_0(\mathbf{x} - \xi) = \begin{cases} \exp\left(-\dfrac{2\cdot(\mathbf{x} - \xi)}{l}\right)^2 & \text{if } \|\mathbf{x}-\xi\| \le R \\ 0 & \text{if } \|\mathbf{x}-\xi\| > R \end{cases} \tag{32}$$

where R is the interaction radius representing a parameter linked to the internal length l.

### Non-local variable

In non-local theories, an internal length enters as a material parameter by allowing a dependency on the so-called non-local variables in the constitutive equations. A non-local variable is a weighted average of the local variable over all the material points in the body, and the length parameter determines how the value of the variable at a certain point is weighted. The way non-locality is introduced into the constitutive equations is dominant because an inappropriate treatment of the non-local variable, may lead to instability of the numerical analysis.

Herein, inspired by [dPI03], the yield function is chosen to be the non-local variable. Choosing yield function as the non-local variable $\left(\hat{f}\right)$, the viscous nucleus and consequently the viscoplastic flow rule are modified:

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{vp}} = \gamma\Phi\left(\hat{f}\right)\frac{\partial \mathrm{g}}{\partial \boldsymbol{\sigma}'} \tag{33}$$

**Discretization of the non-local variable**

The non-local variable of Equation 29 evaluated using Gauss quadrature. The Gaussian integration process allows the spatial integrals of a polynomial to be replaced with finite sums over a discrete set of integration points. In each element we perform Gauss integration

$$\hat{f}(\mathbf{x}_i) = \frac{\sum\limits_{e=1}^{el}\sum\limits_{j=1}^{n}\omega_j^e\alpha\left(\left\|\mathbf{x}_i - \xi_j^e\right\|\right)\det\mathbf{J}_j^e\, f(\xi_j^e)}{\sum\limits_{e=1}^{el}\sum\limits_{j=1}^{n}\omega_j^e\alpha\left(\left\|\mathbf{x}_i - \xi_j^e\right\|\right)\det\mathbf{J}_j^e} \tag{34}$$

in which *i* is the integration point under consideration, *j* is the *j*th Gauss point of element *e*; *el* is the total number of elements inside the interaction volume defined by a sphere centered at x with radius R, *n* is the number of Gauss points of this element inside the interaction volume; ω and **J** are, respectively, the weight and Jacobian matrix at Gauss point *j* of element *e*.

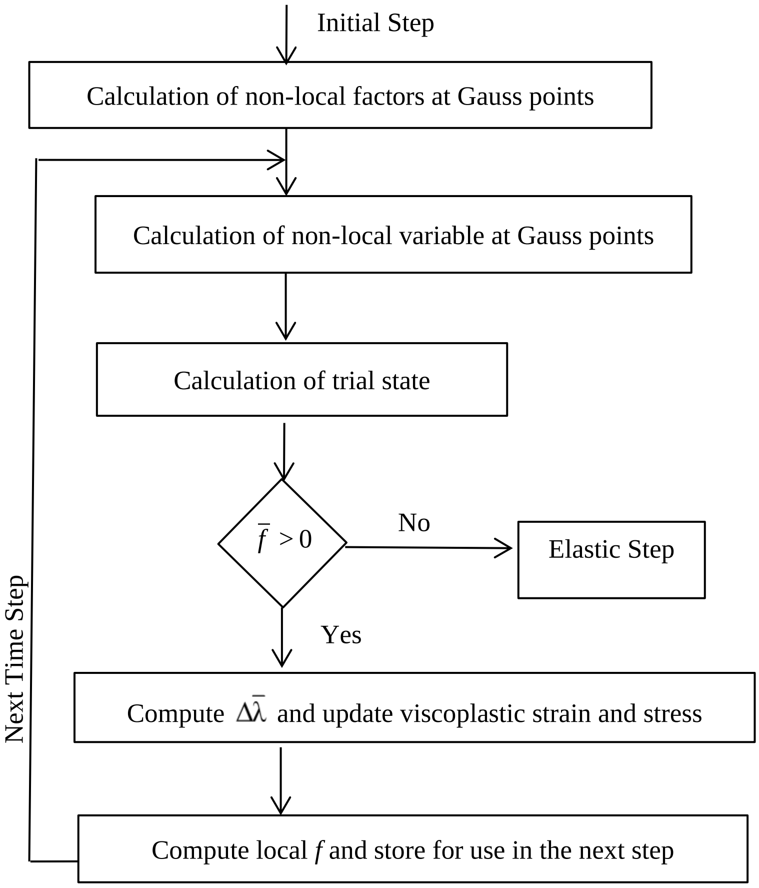
**Integration algorithm**

The algorithmic treatment of the non-local approach and its implementation can be divided in two main steps.

In the first step the factors $\omega_{\mathrm{j}}^{\mathrm{e}}$, $\mathbf{J}_{\mathrm{j}}^{\mathrm{e}}$ and $\alpha\left(\left\|\mathbf{x}_{\mathrm{i}} - \xi_{\mathrm{j}}^{\mathrm{e}}\right\|\right)$ of Equation (34) are computed.

These factors depend on the finite element mesh itself and not on the material model considered. Therefore, this step is applied only once at the beginning of the analysis and the values of the calculated factors can be reused in the subsequent iterations, whenever the non-local formulation is activated. This fact has twofold advantage as it allows for non-local extension to more sophisticated yield criteria and in addition reduces significantly the computational burden. Moreover, carrying out integration over the whole domain with a relatively small value for the internal length l, may lead to the summation of zero values. This is because in such a case, the sphere of influence of the weighting function is only in the closest neighborhood of the regarded integration point. To prevent such an inefficient strategy, at the beginning of the calculation, the set of elements, which have an influence on the non-local quantity, are determined by calculating the distance between the respective points and comparing the distance with some reference value, which depends on the internal length. In the second step the non-local quantity and the integration of the non-local elasto-viscoplastic constitutive equations are computed. In non-local context an implicit scheme is difficult to be applied, as the integration of the constitutive equations is no longer a local stage [Str96]. For this purpose an explicit integration scheme is adopted only for the integration of non-local viscoplastic constitutive equations (whereas the rest of the algorithm operates implicitly) and the non-local values at the current time step are calculated from the local values of the previous time step.

Table 2: Flowchart of the non-local implementation in COMES-GEO code [Laz16].


Initial Step
Calculation of non-local factors at Gauss points
Calculation of non-local variable at Gauss points
Calculation of trial state
$\bar{f} > 0$
No
Elastic Step
Yes
Compute $\Delta\bar{\lambda}$ and update viscoplastic strain and stress
Compute local $f$ and store for use in the next step
Next Time Step

Due to its explicit nature, the stability of the algorithm is maintained using a relatively small time step. Using forward Euler method for the integration of Equation (33) one obtains:

$$\boldsymbol{\varepsilon}_{\mathrm{n+1}}^{\mathrm{vp}} = \boldsymbol{\varepsilon}_{\mathrm{n}}^{\mathrm{vp}} + \Delta\bar{\lambda}_{\mathrm{n+1}} \left\{ \alpha_{\mathrm{g}} \mathbf{1} + \mathbf{n}_{\mathrm{n}} \right\} \tag{35}$$

in which the increment of the inelastic multiplier is obtained by integrating the non-local version of plastic multiplier:

$$\Delta\bar{\lambda}_{\mathrm{n+1}} = \frac{\Delta \mathrm{t}}{\eta\, f_0} \left( \bar{f}_{\mathrm{n}} \right) \tag{36}$$

Having computed $\Delta\bar{\lambda}$, strain and stresses can be updated. Since this procedure uses the local values of the previous time step for the calculation of non-local values at the current time step, the local variable has to be stored and ensure that this information will be available in the next time step. In fact the local yield function, *f*, is a column matrix that collects the local values at all Gauss points.

Table 2 summarizes the algorithm of non-local formulation, which extends the local stress update of the viscoplastic model of Perzyna.

# 4 Spatial and time discretization

The finite element model is derived by applying the Galerkin procedure for the spatial integration and the generalized Newmark method for the time integration of the weak form of the balance equations (1)-(4) [Lew98], [Zie99], [Zie00].

In particular, after spatial discretization within the isoparametric formulation, the following non-symmetric, non-linear and coupled system of equations is obtained:

$$\begin{cases} \boldsymbol{C}_{gg}\dot{\bar{\boldsymbol{p}}}^{g} + \boldsymbol{C}_{gc}\dot{\bar{\boldsymbol{p}}}^{c} - \boldsymbol{C}_{gT}\dot{\bar{\boldsymbol{T}}} + \boldsymbol{C}_{gu}\dot{\bar{\boldsymbol{u}}} + \boldsymbol{K}_{gg}\bar{\boldsymbol{p}}^{g} - \boldsymbol{K}_{gc}\bar{\boldsymbol{p}}^{c} - \boldsymbol{K}_{gT}\bar{\boldsymbol{T}} = \boldsymbol{f}_{g} \\ \boldsymbol{C}_{cg}\dot{\bar{\boldsymbol{p}}}^{g} + \boldsymbol{C}_{cc}\dot{\bar{\boldsymbol{p}}}^{c} + \boldsymbol{C}_{cT}\dot{\bar{\boldsymbol{T}}} + \boldsymbol{C}_{cu}\dot{\bar{\boldsymbol{u}}} - \boldsymbol{K}_{cg}\bar{\boldsymbol{p}}^{g} + \boldsymbol{K}_{cc}\bar{\boldsymbol{p}}^{c} + \boldsymbol{K}_{cT}\bar{\boldsymbol{T}} = \boldsymbol{f}_{c} \\ -\boldsymbol{C}_{Tg}\dot{\bar{\boldsymbol{p}}}^{g} - \boldsymbol{C}_{Tc}\dot{\bar{\boldsymbol{p}}}^{c} + \boldsymbol{C}_{TT}\dot{\bar{\boldsymbol{T}}} - \boldsymbol{C}_{Tu}\dot{\bar{\boldsymbol{u}}} - \boldsymbol{K}_{Tg}\bar{\boldsymbol{p}}^{g} + \boldsymbol{K}_{Tc}\bar{\boldsymbol{p}}^{c} + \boldsymbol{K}_{TT}\bar{\boldsymbol{T}} = \boldsymbol{f}_{T} \\ \boldsymbol{M}_{uu}\ddot{\bar{\boldsymbol{u}}} + \int \boldsymbol{B}^{T}\boldsymbol{\sigma}' dW - \boldsymbol{K}_{ug}\bar{\boldsymbol{p}}^{g} + \boldsymbol{K}_{uc}\bar{\boldsymbol{p}}^{c} = \boldsymbol{f}_{u} \end{cases} \tag{37}$$

where the displacements of the solid skeleton $\boldsymbol{u}(\mathbf{x},t)$, the capillary pressure $p^{c}(\boldsymbol{x},t)$, the gas pressure $p^{g}(\mathbf{x},t)$ and the temperature $T(\mathbf{x},t)$ are expressed in the whole domain by global shape function matrices $\mathbf{N}_{u}(\mathbf{x})$, $\mathbf{N}_{c}(\mathbf{x})$, $\mathbf{N}_{g}(\mathbf{x})$, $\mathbf{N}_{T}(\mathbf{x})$ and the nodal value vectors $\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{p}}^{c}(t), \bar{\boldsymbol{p}}^{g}(t), \bar{\boldsymbol{T}}(t)$.

Following the Generalized Newmark Method, equations (37) are rewritten at time $t_{(n+1)}$. The elements of the matrices $\boldsymbol{C}_{ij}$, $\boldsymbol{K}_{ij}$ and the vectors $\boldsymbol{f}_{i}$ are given in [San15].

In this study, the generalized Newmark time integration scheme [Zie00] is applied to the non-linear equation system (8) and a non-linear system of algebraic equations is obtained, in which the unknowns are $\mathbf{X} = \left[ \Delta\dot{\bar{p}}^g , \Delta\dot{\bar{p}}^c , \Delta\dot{\bar{T}} , \Delta\ddot{\bar{\boldsymbol{u}}} \right]$. The non-linear system is solved by Newton-Raphson method, thus obtaining the equation system that can be solved numerically (written below in a compact form) as:

$$\left.\frac{\partial \mathbf{G}}{\partial \mathbf{X}}\right|_{\mathbf{X}^i_{n+1}} \Delta\mathbf{X}^{i+1}_{n+1} \cong -\mathbf{G}\left(\mathbf{X}^i_{n+1}\right) \tag{38}$$

with the symbol $\left(\bullet\right)^{i+1}_{n+1}$ to indicate the current iteration ($i$+1) in the current time step ($n$+1) and where $\partial\mathbf{G}/\partial\mathbf{X}$ is the Jacobian matrix.
Owing to the strong coupling between the mechanical, thermal and the pore fluids fields, a monolithic solution of (38) is preferred.

# 5 Finite element simulations

This section addresses the numerical validation of the model previously derived and presents an application studying a biaxial strain localization test and a slope stability test.
Different tests have been simulated, aiming at validate:

a) the wave propagation in a solid material (Equation (1) restricted to single phase solid material),
b) the isothermal water saturated model (Equations (1) and (3) with $S_w$=1),
c) the isothermal variably saturated model (Equations (1), (2) and (3)) and
d) the non-isothermal water saturated model (Equations (1), (3) and (4) with $S_w$=1).

Analytical solutions are available in [Slu92] and [Boe93] for the first two tests respectively, while the numerical results from tests c) and d) have been compared with the numerical solution of the corresponding quasi-static models because of the lack of analytical solutions. Some representative results of tests c) and d) are illustrated here.

## 5.1 Drainage of liquid water from initially water saturated soil column

This numerical test is based on an experiment performed by Liakopoulos [Lia65] on a column 1 meter high (Figure 1) of Del Monte sand and instrumented to measure the moisture tension at several points along the column during its desaturation due to gravitational effects. Before the start of the experiment, water was continuously added from the top and was allowed to drain freely at the bottom through a filter, until uniform flow conditions were established. Then the water supply was ceased and the tensiometer readings were recorded. The finite element simulation is performed with the two-phase flow model in isothermal conditions. For the numerical calculation, a two-dimensional problem in plane strain conditions is solved; the spatial domain of the column is divided into 20 eight-node isoparametric finite elements of equal size. Furthermore, nine Gauss integration points were used. The material parameters are listed in [Gaw96], as well as the description of the boundary conditions and the equations for the saturation-capillary pressure and the relative permeability of water-capillary pressure relationships.

This problem has been solved considering single or two-phase flow mainly in quasi-static condition (e.g. [Gaw96]); a finite element solution in dynamics was presented in [Sch98]. The initial hydro-mechanical equilibrium state is obtained via a preliminary quasi-static solution.

The comparison between the dynamic and the quasi-static solution is plotted in Figures 2 to 4, where the profiles for liquid water pressure, liquid water saturation and vertical displacement along the column are plotted. Since the inertial loads are negligible in the experiment, the finite element solution in dynamics gives almost the same results of the quasi-static model [Gaw96], [Gaw09].

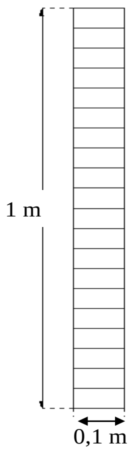


Figure 1: Geometry and finite element discretization of the sand column.

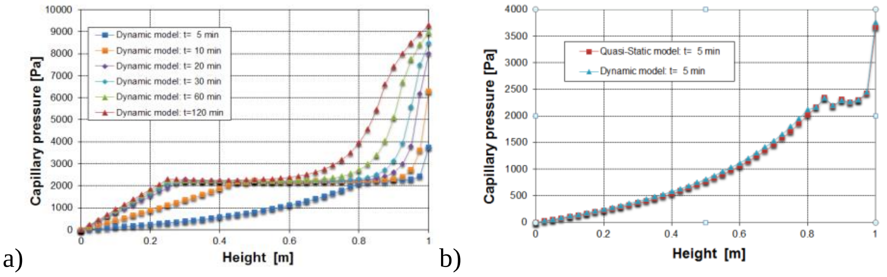


Figure 2: Profiles of capillary pressure versus height: a) dynamic solution; b) comparison between the quasi-static and the dynamic solution.

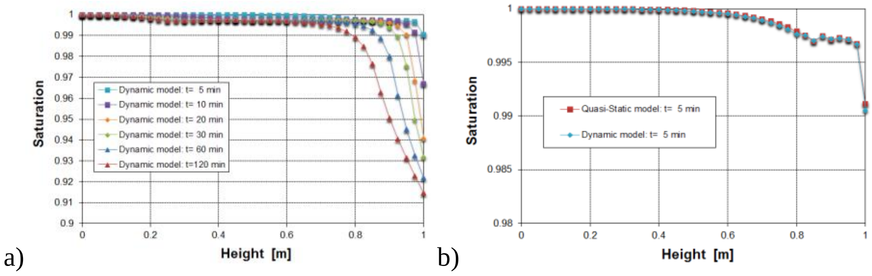


Figure 3: Profiles of liquid water saturation degree versus height: a) dynamic solution; b) comparison between the quasi-static and the dynamic solution.

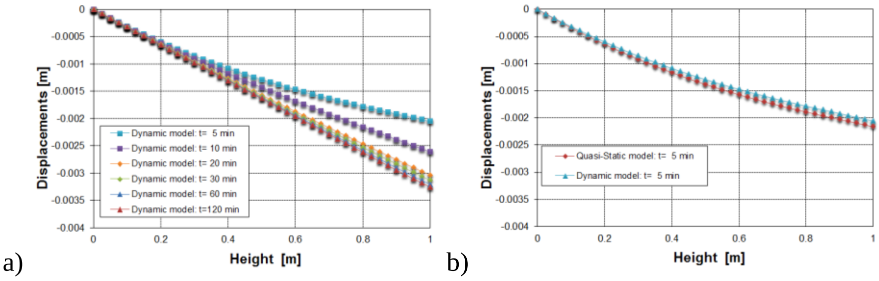


Figure 4: Profiles of vertical displacement versus height: a) dynamic solution; b) comparison between the quasi-static and the dynamic solution.

## 5.2 Numerical validation of the non-isothermal water saturated model

This problem deals with a water saturated thermo-elastic consolidation [Abo85], simulating a column, 7 m high and 2 m wide, of a linear elastic material subjected to an external surface load of 10 kPa and to a surface temperature jump of 50 K above the initial temperature of 293.15 K (Figure 5). The material parameters used in the computation are summarized in [San08]. The liquid water and the solid grain are assumed incompressible for the quasi-static analysis, whereas the compressibility of the liquid water is taken into account in the dynamic analysis. The initial and boundary conditions are described in [San08]. Plane strain condition is assumed. The spatial domain is discretized with eight-node isoparametric elements; nine Gauss points are used.

The solution of the finite element model presented in this work is compared with the quasi-static solution [San08] and is plotted in Figures 6 and 7. The results show that the dynamic solution is faster than the quasi-static one at the beginning of the analysis, and that the dynamic solution reaches the quasi-static one at the steady-state.

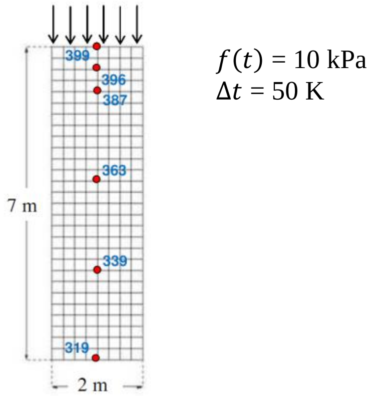


Figure 5: Description of the non-isothermal water saturated test.

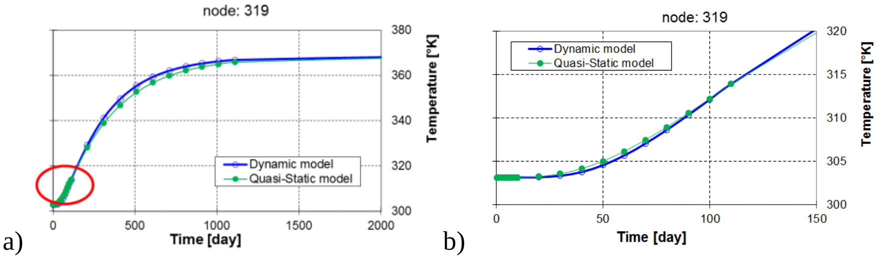


Figure 6: Temperature time history for node 319 up to the steady state solution (a) and in the first period (b) highlighted in a).

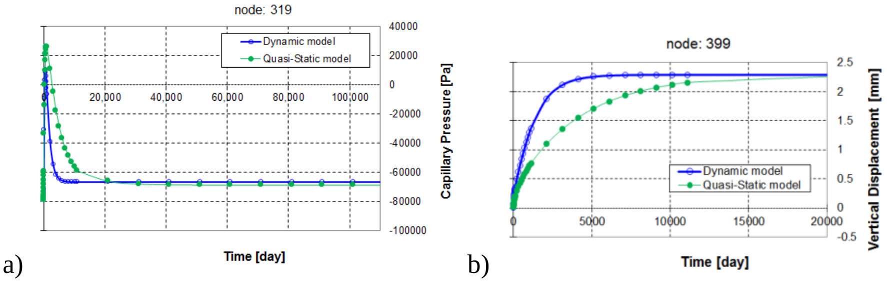


Figure 7: a) Capillary pressure time history for node 319 and b) vertical displacement time history for node 399.

## 5.3 Globally undrained biaxial compression test

A plane strain compression test of initially water saturated dense sand in globally undrained conditions is simulated here with the model developed in the previous sections. This case was solved in [San06] in quasi-static conditions and is inspired by the experimental work of Mokni and Desrues [Mok98], in which cavitation of the liquid water was experimentally observed at localization.

A sample of 34 cm height and 10 cm width is compressed with imposed vertical displacement applied to the top surface at a velocity of 3.6 mm/s (Figure 8). Vertical and horizontal displacements are constrained at the bottom surface; the boundary of the sample is impervious and adiabatic.

The mechanical behavior of the solid skeleton is simulated using the elasto-plastic Drucker-Prager constitutive model (with isotropic linear softening and non-associated plastic flow) summarized in Section 3. At time t= 0 seconds, the initial conditions for the domain are the hydrostatic water pressure, the gas pressure at atmospheric value and a temperature of 293.15 K. Gravity acceleration is taken into account; the initial stress state in equilibrium with the initial conditions and thermo-hydro boundary conditions is computed with the corresponding quasi-static model [San06]. The geomechanical characteristics of the dense sand are given in [San06].

Figures 9 and 10 show the contour plots at 13 seconds of the following thermo-hydro-mechanical variables: equivalent plastic strain, volumetric strain, capillary pressure, liquid water saturation and relative humidity. Positive volumetric strains are observed inside the dilatant shear bands (Figure 9b), inducing a liquid water pressure drop up to the development of capillary pressures (Figure 10a) desaturating the plastic zones (Figure 10b) because of the phase change of the liquid water into vapor due to cavitation (Figure 10c).

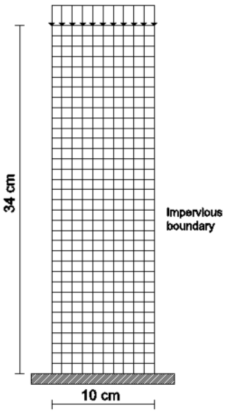


Figure 8: Finite element discretization and boundary conditions of the biaxial compression test.

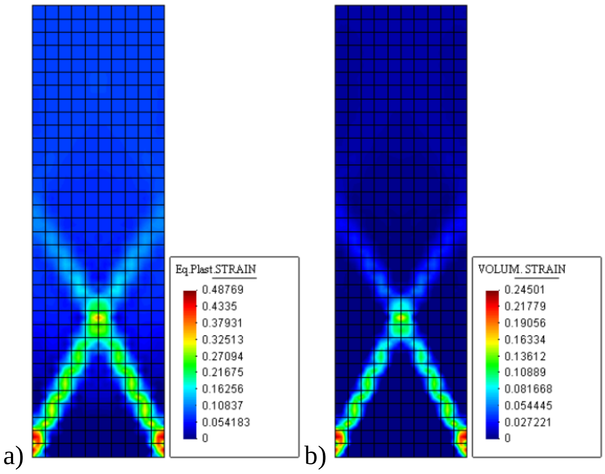


Figure 9: Numerical solution at 13 s: a) equivalent plastic strain, b) volumetric strain.

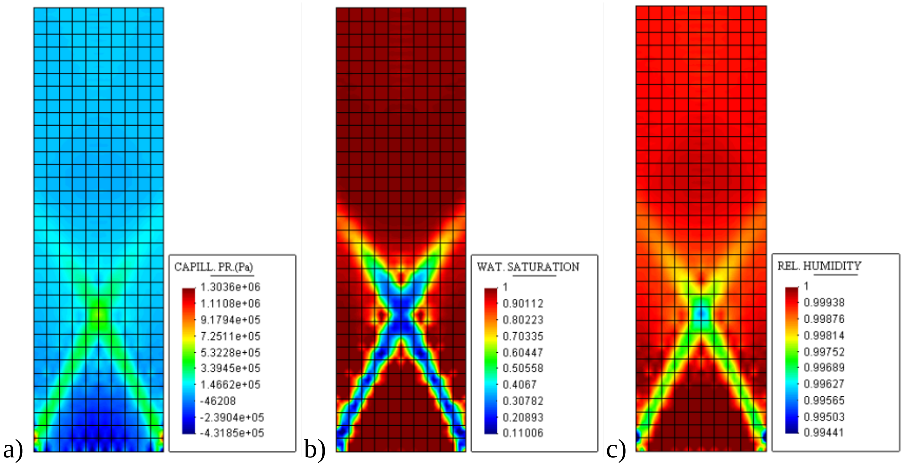


Figure 10: Numerical solution at 13 s: a) capillary pressure, b) liquid water saturation, c) relative humidity.

To study the independence of shear band width from the finite element size in dynamics, e.g. [Sch96], [Sch99], [Zha99] and [Sch06], test runs with meshes of 85, 340 and 1360 elements have been carried out. In this case, the analysis of the finite element results [Cao16] shows that the peak value of the equivalent plastic strain and, as a consequence, of the volumetric strain, the capillary pressure, the water vapor pressure and the relative humidity are sensitive to mesh refinement and a regularization scheme would be needed as expected, because the internal length scale given by the liquid water motion [Zha99] is not sufficient to regularize the numerical solution.

The effect of the local and non-local elasto-viscoplastic model of Perzyna in the regularization of this mesh dependency problem is illustrated in Figure 11.

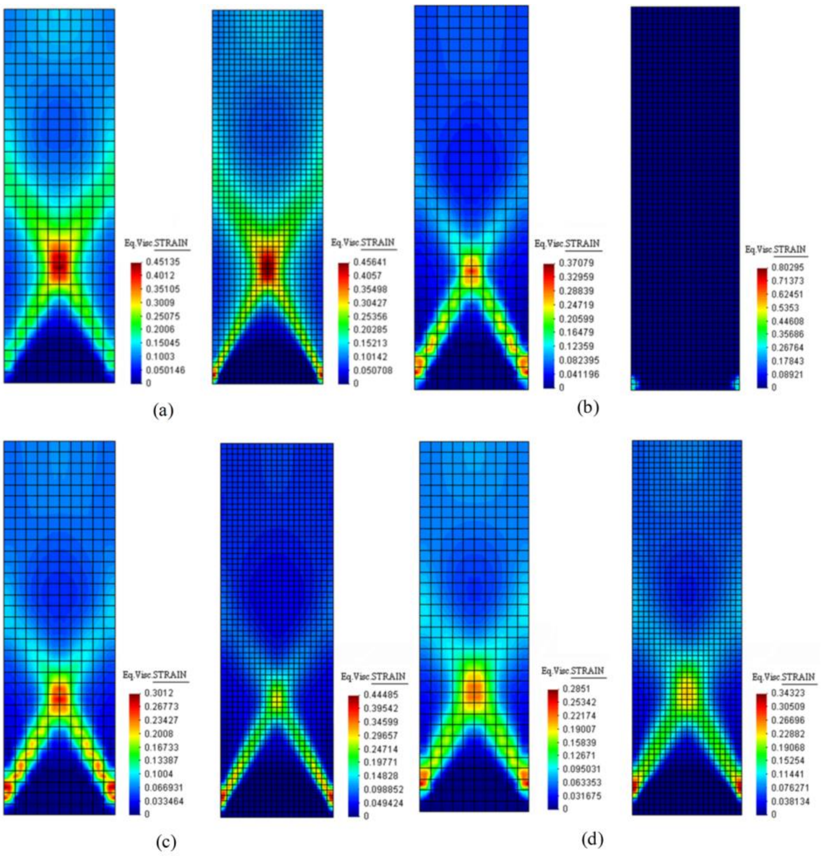


Figure 11: Numerical results for (a) local viscoplasticity with η=30s, (b) local viscoplasticity with η=10s, (c) non-local viscoplasticity with η=10s, l=0.01m and (d) non-local viscoplasticity with η=10s, l=0.02m, for two meshes (10x34 and 20x68 respectively).

The influence of the viscosity parameter, η, is clearly depicted for the local elasto-viscoplastic model. In the case of η=30s, a regularized solution is obtained for both meshes: the shear band width remains unaltered upon mesh refinement and the peak value of equivalent viscoplastic strain coincides for the two meshes. However, considering a less rate-sensitive material with η=10s (and as approaching the elasto-plastic solution), the regularizing effect of the local elasto-viscoplastic model is lost and the contour of the more refined mesh reveals strong mesh-sensitivity (Figure 11b). In this case, the non-local elasto-viscoplastic model proves to be sufficient to regularize the finite element solution for the given material parameters and the shear band propagates for internal length value, l=0.01m (Figure 11c). Moreover, in the case of internal length value l=0.02m (Figure 11d), the width of the shear band increases accordingly to the l value for both meshes adopted. In Figure 12 the numerical solution for the liquid phase is presented in terms of capillary pressure and water saturation. The results are presented for the case of non-local elasto-viscoplastic model, and mesh independency is apparent also for the fluid part. It is evident that pore water decreases up to the development of capillary pressure, accompanied by desaturation in the strain localization zones. It is noted that at the same time water pressure decreases below the vapor saturation pressure and the phase change of the liquid water to vapor occurs. A more detailed presentation and analysis of the influential parameters of the problem, such the loading velocity, the value of permeability and the interaction of the internal lengths (introduced by viscosity and non-locality) can be found in [Laz15].

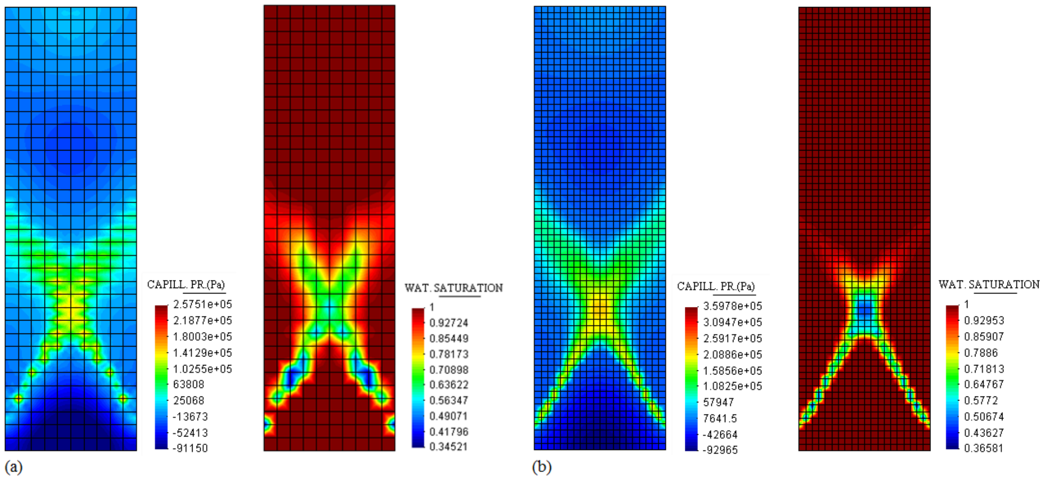


Figure 12: Non-local approach: Capillary pressure and water degree of saturation contours for (a) 10x34 mesh and (b) 20x68 mesh, in case of l=7.5mm and η=10s.

## 5.4 Slope stability test

A slope stability problem, inspired by [Re01], is presented to demonstrate the effectiveness of the regularization techniques presented in Section 3 towards strain localization simulation of geomaterials. The dimensions and boundary conditions of the problem are shown in Figure 13 whereas the soil parameters considered in the analysis can be found in [Laz16]. The initial stress field is given by geostatic stress state and drained conditions are imposed. Then, a downward displacement with a constant rate of $10^{-3}$ m/s is applied to a length of 4m on the top slope surface (Figure 13). Two meshes with 400 and 1600 eight node quadrilateral isoparametric elements are used to analyze the problem.

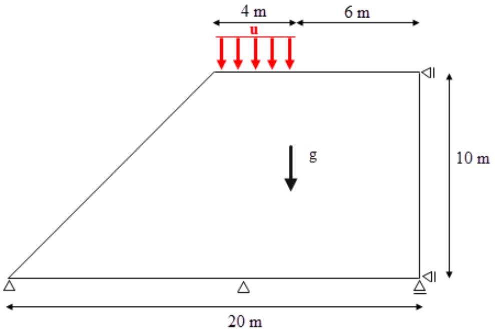


Figure 13: Slope stability problem. Geometry and boundary conditions.

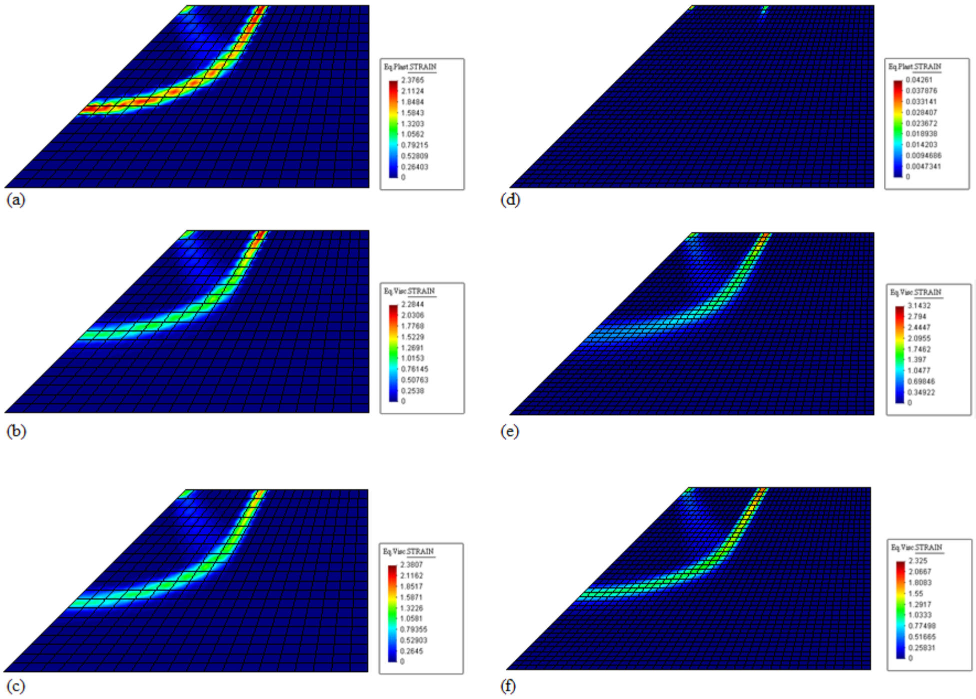


Figure 14: Equivalent (visco)plastic strain contours as calculated using the elastoplastic model (a, d), the local elasto-viscoplastic model (b, e) and the non-local elasto-viscoplastic model (c,f ) for a mesh of 400 elements (a), (b) and (c) and for a mesh of 1600 elements (d), (e) and (f), respectively.

The analyses are performed using: the elastoplastic model, the local elasto-viscoplastic model of Perzyna with viscosity η=100s and the non-local elasto-viscoplastic model with an internal length of l=0.8 m. The results from these models are shown in Figure 14 in terms of equivalent (visco-) plastic strain. The failure initiates in the element just to the right of the applied force and propagates in a manner dependent on the angle of friction. When the elastoplastic constitutive model is used to solve this initial boundary value problem, a classical mesh dependent numerical solution is observed with the model being unable to simulate the failure process of the slope when the mesh is refined (Figure 14d). As is shown in Figure 14b, e when the elasto-viscoplastic formulation of Perzyna is adopted, even if the number of elements is increased, the shear band formation is not affected by the element size. However, the peak value of the viscoplastic strain field depends on the element size of the mesh (Figure 14e). Finally, the non-local elasto-viscoplastic model is able to predict a clearly defined shear band independently of the mesh adopted (Figure 14f).

# 6 Conclusions

A model for the analysis of the thermo-hydro-mechanical behavior of porous media in dynamics was developed. Starting from the generalized mathematical model developed in [Lew98] for deforming porous media in non-isothermal conditions, the u-p-T formulation was derived following [Zie99] and [Cha22]. The validity of such an approximation is limited to low frequencies problems [Zie99] and [Cha22], as in earthquake engineering. In this formulation, the relative accelerations of the fluids and the convective terms related to these accelerations are neglected.

The numerical model was derived within the finite element method: the standard Bubnov-Galerkin procedure [Zie00] was adopted for the discretization in space, while the implicit and unconditionally stable generalized Newmark procedure was applied for the discretization in time [Zie00] or the chapter by Pastor in this volume.

The model was implemented in the finite element code Comes-Geo [Lew98], [Gaw96], [San06], [San08], [San09], [Gaw09], [Gaw10]. The formulation and the implemented solution procedure were validated through the comparison with literature benchmarks, finite element solutions or analytical solutions. In this work, comparison between the finite element solution in dynamics and the corresponding quasi-static solution is presented by studying the non-isothermal consolidation in a water saturated column and the drainage of liquid water in an initially water saturated soil column.

This work extends the model developed in [Sch98] to non-isothermal conditions and removes the passive air phase assumption of the multiphase porous media model in dynamics developed in [Zie99] and [Gaw98].

The efficiency of the soil constitutive models in terms of regularized performance has been illustrated using numerical examples of an undrained biaxial test and a slope failure problem. The numerical results indicate that only by using the applied regularization techniques the location and the propagation of the shear zone is reliably simulated in a mesh independent manner.

**Acknowledgements**. The authors would like to thank the 7th Framework Programme of the European Union (ITN MuMoLaDe project 289911) and University of Padova (BIRD208040) for the financial support.

## References

[Abo85] Aboustit, B. L., Advani, S. H. and Lee, J. K. Variational principles and finite element simulations for thermo-elastic consolidation. *Int. J. Numer. Anal. Methods Geomech.* 9: 49-69, 1985.

[Alb10] Albers, B. *Modeling and numerical analysis of wave propagation in saturated and partially saturated porous media*. Habilitation Thesis, Technischen Universität Berlin n.48, Shaker Verlag, 2010.

[Boe93] de Boer R, Ehlers W, Liu Z. One-dimensional transient wave propagation in fluid-saturated incompressible porous media. *Archive of Applied Mechanics*, 63(1): 59-72, 1993.

[Bol05] Bolzon, G., Schrefler, B.A. Thermal effects in partially saturated soils: a constitutive model, *Int. J. Numer. Anal. Methods Geomech*, 29(9), 861-877, 2005.

[Cao16] Cao, T.D, Sanavia L., Schrefler BA. A thermo-hydro-mechanical model for multiphase geomaterials in dynamics with application to strain localization simulation. *Int. J. Num. Meth. Engng., 107(4), 312-337*, 2016.

[Cha88] Chan, A.H.C. *A unified finite element solution to static and dynamic in geomechanics*. Ph.D. Thesis, University College of Swansea, 1988.

[Cha22] Chan, A.H.C., Pastor M., Schrefler B.A., Shiomi T. and O.C. Zienkiewicz. *Computational Geomechanics. Theory and Applications.* Wiley, 2022.

[dPI02] di Prisco, C., Imposimato, S. and Aifantis, E.C. A visco-plastic constitutive model for granular soils modified according to non-local and gradient approaches, *International Journal for Numerical and Analytical Methods in Geomechanics, 26, 121-138*, 2002.

[dPI03] di Prisco, C. and Imposimato, S. Nonlocal numerical analyses of strain localization in dense sand, *Mathematical and Computer Modelling, 37:497-506*, 2003.

[Fra08] François, B. and Laloui, L. ACMEG-TS: A constitutive model for unsaturated soils under non-isothermal conditions. *Int. J. Numer. Anal. Methods Geomech* 32:1955-1988, 2008.

[Fur97] Furbish DJ. *Fluid Physics in Geology: An Introduction to Fluid Motions on Earth's Surface and within Its Crust*. Oxford University Press, 1997.

[Gaw12] Gawin, D. and Pesavento F. *An overview of modeling cement based materials at elevated temperatures with mechanics of multi-phase porous media.* Fire Technology, 48, 753-793, 2012.

[Gaw09] Gawin, D., Sanavia, L. A unified approach to numerical modelling of fully and partially saturated porous materials by considering air dissolved in water. *CMES: Computer Modeling in Engineering & Sciences*, 53: 255-302, 2009.

[Gaw10] Gawin, D., Sanavia, L. Simulation of cavitation in water saturated porous media considering effects of dissolved air. *Transport in Porous Media*, 81: 141-160, 2010.

[Gaw96] Gawin, D., Schrefler, B.A. Thermo-hydro-mechanical analysis of partially saturated porous materials. *Engineering Computations*, 13: 113-143, 1996.

[Gaw98] Gawin, D., Sanavia, L., Schrefler, B.A. Cavitation modelling in saturated geomaterials with application to dynamic strain localisation, *International Journal for Numerical Methods in Fluids*, 27: 109-125, 1998.

[Gra13] Gray WG, Miller CT, Schrefler BA. Averaging theory for description of environmental problems: What have we learned, *Advances in Water Resources*; 51: 123–138, Doi.org/10.1016/j.advwatres.2011.12.005, 2013.

[Gra14] Gray WG, Miller CT. *Introduction to the Thermodynamically Constrained Averaging Theory for porous medium systems*, Springer, 2014.

[Gra15] Gray WG, Miller CT. *Thermodinamically Constrained Averaging Theory (TCAT) to model the coupled behaviour of multiphase porous system*, ALERT Doctoral School 2015 https://alertgeomaterials.eu/data/school/2015/2015_ALERT_schoolbook.pdf

[Gra91] Gray WG, Hassanizadeh M. Unsaturated flow theory including interfacial phenomena. *Water Resources Research*, 27: 1855-1863, 1991.

[Has79a] Hassanizadeh, M., Gray, W.G. General conservation equations for multiphase system: 1. Averaging technique. *Advances in Water Resources*, 2:131-144, 1979.

[Has79b] Hassanizadeh, M., Gray, W.G. General conservation equations for multiphase system: 2. mass, momenta, energy and entropy equations. *Advances in Water Resources*, 2: 191-201, 1979.

[Has80] Hassanizadeh M, Gray WG. General conservation equations for multiphase systems: 3. Constitutive theory for porous media flow. *Advances in Water Resources,* 3(1): 25-40, 1980.

[Hei11] Heider, Y., Markert, B., Ehlers, W. Dynamic wave propagation in infinite saturated porous media half spaces. *Computational Mechanics* 49: 319-336, 2011.

[Jir02] Jirásek, M. Objective modeling of strain localization. *Revue Française de Genie Civil, 6:1119-1132*, 2002

[Jir03] Jirásek, M. and Rolshoven, S. Comparison of integral-type nonlocal plasticity models for strain-softening materials, *International Journal of Engineering Science, 41:1553-1602*, 2003.

[Ju90] Ju, J. Consistent tangent moduli for a class of viscoplasticity, *Journal of Engineering Mechanics, 116: 1764-1779*, 1990.

[Laz15] Lazari, M., Sanavia, L., and Schrefler, B. A. Local and non-local elasto-viscoplasticity in strain localization analysis of multiphase geomaterials, *International Journal for Numerical and Analytical Methods in Geomechanics, 39: 1570–1592*, 2015.

[Laz16] Lazari, M. *Finite element regularization for post localized bifurcation in variably saturated media*. PhD thesis, University of Padova, Italy, 2016.

[Laz19] Lazari M, Sanavia L, di Prisco C, Pisanò F. Predictive potential of Perzyna viscoplastic modelling for granular geomaterials. International Journal for Numerical and Analytical Methods in Geomechanics. 43: 544–567, 2019.

[Lew98] Lewis, R.W. and Schrefler, B.A. *The finite element method in the static and dynamic deformation and consolidation of porous media*. Wiley, 1998.

[Lia65] Liakopoulos, A.C. *Transient flow through unsaturated porous media. PhD thesis*, University of California, Berkeley, USA, 1965.

[Mok98] Mokni, M., Desrues, J. Strain localisation measurements in undrained plane-strain biaxial tests on hostun RF sand. *Mechanics of Cohesive-frictional Materials*, 4, 419 – 441, 1998.

[Nen10] Nenning, M., Schanz, M. Infinite elements in a poroelastodynamic FEM. *Int. J. Numer. Anal. Methods Geomech.* 35: 1774-1800, 2010.

[Nut08] Nuth, M., Laloui, L. Effective stress concept in unsaturated soils: Clarification and validation of a unified approach. *Int. J. Numer. Anal. Methods Geomech.* 32:771-801, 2008.

[Per63] Perzyna, P. The constitutive equations for rate sensitive plastic materials, *Quarterly of applied mathematics, 20(4), 321-332*, 1963.

[Re01] Regueiro, R.A. and Borja, R.I.. Plane strain finite element analysis of pressure sensitive plasticity with strong discontinuity, *International Journal of Solids and Structures, 38, 3647-3672*, 2001.

[San02] Sanavia, L., B.A. Schrefler, and P. Steinmann, A formulation for an unsaturated porous medium undergoing large inelastic strains, *Computational Mechanics, 28*: 137-151, 2002

[San06] Sanavia, L., Pesavento, F., Schrefler, B.A. Finite element analysis of non-isothermal multiphase geomaterials with application to strain localization simulation. *Computational Mechanics*, 37: 331-348, 2006.

[San08] Sanavia, L., François, B., Bortolotto, R., Luison, L. and Laloui, L. Finite element modelling of thermo-elasto-plastic water saturated porous materials. *Journal of Theoretical and Applied Mechanics,* 38:7-24, 2008.

[San09] Sanavia, L. Numerical modelling of a slope stability test by means of porous media mechanics. *Engineering Computations,* 26: 245-266, 2009.

[Sch06] Schrefler, B.A., Zhang, H.W. and Sanavia, L. Interaction between different internal length scales in fully and partially saturated porous media – The 1-D case, *Int. J. Numer. Anal. Methods Geomech.,* 30: 45-70, 2006.

[Sch09] Schanz, M. Poroelastodynamics: linear models, analytical solutions, and numerical methods. *Applied Mechanics Reviews,* 62: 1-15, 2009.

[Sch84] Schrefler, B.A. *The Finite Element Method in Soil Consolidation (with applications to Surface Subsidence)*. PhD. Thesis, University College of Swansea, C/Ph/76/84, Swansea UK, 1984.

[Sch96] Schrefler BA, Sanavia L., Majorana CE. A multiphase medium model for localization and post localization simulation in geomaterials. *Mechanics of Cohesive-Frictional Materials*, 1:95-114, DOI: 10.1002/(SICI)1099-1484(199601)1:1<95::AID-CFM5>3.0.CO;2-D, 1996.

[Sch98] Schrefler, B.A., Scotta, R. A fully coupled dynamic model for two-phase fluid flow in deformable porous media. *Computer Methods in Applied Mechanics and Engineering*, 190: 3223-3246, 1998.

[Sch99] Schrefler, B.A., Zhang, H.W., Sanavia, L. Fluid-structure interaction in the localisation of saturated porous media. *ZAMM Zeitschrift für Ange-*

*wandte Mathematik und Mechanik (Journal of Applied Mathematics and Mechanics. Z. Angew. Math. Mech.)*, 79: 481-484, 1999.

[Sim98] Simo, J.C. *Numerical Analysis and Simulation of Plasticity*. In: Ciarlet PG, Lions JL (eds) Numerical Methods for Solids. Handbook of Numerical Analysis (Part3) vol.6, North-Holland, 1998.

[SH98] Simo, J.C., Hughes, T.J.R. *Computational inelasticity*. Springer, 1998.

[Slu92] Sluys, L.J. *Wave propagation, localization and dispersion in softening solids*, Ph.D. Dissertation, Delft University of Technology, 1992.

[Str96] Strömberg, L. and Ristinmaa, M. FE-formulation of a nonlocal plasticity theory, *Computer Methods in Applied Mechanics and Engineering, 136: 127-144*, 1996.

[Var02] Vardoulakis, I. Dynamic thermo-poro-mechanical analysis of catastrophic landslides. *Géotechnique*, 52: 157-171, 2002.

[Zha99] Zhang H.W., Sanavia L, Schrefler B.A. An interal length scale in dynamic strain localization of multiphase porous media. *Mechanics of Cohesive-frictional Materials*, 4(5): 443-460, 1999.

[Zie00] Zienkiewicz, O.C. and Taylor, R.L. *Finite element method* (5th edition) volume 1 - the basis. Elsevier, 2000.

[Zie99] Zienkiewicz, O.C., Chan, A.H., Pastor, M., Schrefler, B.A. and Shiomi, T. *Computational geomechanics with special reference to earthquake engineering*. Wiley, 1999.

©ALERT Geomaterials
Laboratoire 3SR / Bâtiment Galilée
CS 40700
38 058 Grenoble cedex 9
France

ISBN **978-2-9584769-3-9**

Fon: +33 (0) 456 528 621
Fax: +33 (0) 476 827 043
president@alertgeomaterials.eu
https://alertgeomaterials.eu

# ALERT Doctoral School 2024

*Numerical Methods in Geomechanics*

Editors: C. Tamagnini, L. Sanavia & M. Pastor

ISBN **978-2-9584769-3-9**