INTRODUCTION TO HYPOPLASTICITY (GeoMath1)

DIMITRIOS KOLYMBAS

Institute of Geotechnics and Tunnelling,
University of Innsbruck,
Techniker Str. 13, A-6020 Innsbruck, Austria
dimitrios.kolymbas@uibk.ac.at

(Received 19 April 2000)

Abstract: Rational mechanics offers a tool to objectively describe the mechanical behaviour of granular materials by means of non-linear tensorial functions. This new framework is called hypoplasticity and is characterized by the simplicity of the mathematical formulations. The stress dependency of incremental stiffness, yield, loading-unloading hysteresis and dilatancy-contractancy are included. One of the features of hypoplasticity is that yield is a natural outcome of the theory and needs not to be calibrated a priori. Also loss of uniqueness and localization of deformation are realistically predicted by hypoplasticity.

Keywords: granular materials, yield, hysteresis, dilatancy, constitutive equation

1. Fundamentals of continuum mechanics

1.1 Deformation

A motion consists of translation, rotation and deformation. A material point with the material (or initial or Lagrange) coordinates X_α ($\alpha = 1, 2, 3$) moves into a position with the spatial (or Euler) coordinates x_i ($i = 1, 2, 3$). Thus, the motion is described by the function $\mathbf{x} = \boldsymbol{\chi}(\mathbf{X}, t)$. Using a less exact notation we can write $\mathbf{x} = \mathbf{x}(\mathbf{X}, t)$. The deformation gradient is defined as

$$\mathbf{F} = F_{i\alpha} = x_{i,\alpha} = \frac{\partial x_i}{\partial X_\alpha} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}},$$

and can be decomposed into $\mathbf{F} = \mathbf{RU} = \mathbf{VR}$.

1.2 Stretching

Euler's stretching tensor **D** is obtained as the symmetric part of the velocity gradient $\mathbf{L} = \text{grad } \mathbf{v} = \upsilon_{i,j} = \dot{x}_{i,j}$. Thus we have:

$$\mathbf{D} = D_{ij} = \frac{1}{2}\left(\upsilon_{i,j} + \upsilon_{j,i}\right) = \frac{1}{2}\left(\dot{x}_{i,j} + \dot{x}_{j,i}\right) = \dot{x}_{(i,j)}. \tag{1}$$

Cauchy's spin tensor is obtained as the antimetric part of the velocity gradient:

$$\mathbf{W} = W_{ij} = \frac{1}{2}\left(\upsilon_{i,j} - \upsilon_{j,i}\right) = \frac{1}{2}\left(\dot{x}_{i,j} - \dot{x}_{j,i}\right) = \dot{x}_{[i,j]}.$$

We should not confuse **W** with the time rate of **R** ($\mathbf{W} \neq \dot{\mathbf{R}}$). The equality is only valid if the reference configuration is identical with the actual one, *i.e.* $\mathbf{W}(t) = \dot{\mathbf{R}}_{(t)}(t)$.

1.3 Cauchy stress

Cutting a body reveals the internal forces acting within it. Let us consider a particular point of the cutting surface with the unit normal **n** and the stress vector (*i.e.* areal density of force) **t**. Both vectors are connected by the linear transformation **T**:

$$\mathbf{t} = \mathbf{T}\mathbf{n}.$$

T is the Cauchy stress tensor. By lack of couple stresses the stress tensor **T** is symmetric. **T** can be decomposed in a deviatoric ($\mathbf{T}^*$) and a hydrostatic part ($\frac{1}{3}\,\text{tr}\,\mathbf{T}\mathbf{1}$):

$$\mathbf{T} = \mathbf{T}^* + \frac{1}{3}(\text{tr}\,\mathbf{T})\mathbf{1},$$

where tr **T** denotes the sum $T_1 + T_2 + T_3$.

1.4 Change in observer

Let **x**(**X**) be a motion. A so-called equivalent motion $\mathbf{x}^*$ is obtained from **x** by a *change in observer* if:

$$\mathbf{x}^*(\mathbf{X},t) = \mathbf{q}(t) + \mathbf{Q}(t)\left[\mathbf{x}(\mathbf{X},t) - \mathbf{o}\right]. \tag{2}$$

x and $\mathbf{x}^*$ differ by a rigid body motion, which consists of the translation $\mathbf{q}(t)$ and the rotation $\mathbf{Q}(t)$. With $\mathbf{F} = \partial\mathbf{x}/\partial\mathbf{X}$ and $\mathbf{F}^* = \partial\mathbf{x}^*/\partial\mathbf{X}$ it follows from (2):

$$\mathbf{F}^*(\mathbf{X},t) = \mathbf{Q}(t)\mathbf{F}(\mathbf{X},t). \tag{3}$$

The transformation rule for the Cauchy stress reads:

$$\mathbf{T}^* = \mathbf{Q}\mathbf{T}\mathbf{Q}^T. \tag{4}$$

1.5 Objectivity, objective time rates

The material behaviour is called *independent of the observer* if the stress is transformed according to (4). All tensors transformed according to (4) are called *independent of the observer* or *indifferent.*

A co-rotated observer registers the stress $\mathbf{T}^* = \mathbf{Q}\mathbf{T}\mathbf{Q}^T$. If the observer is at rest and the considered material is rotated by $\mathbf{R}$ (with $\mathbf{Q} = \mathbf{R}^T$), then the observer registers the stress $\mathbf{T}^* = \mathbf{R}^T\mathbf{T}\mathbf{R}$. Thus, $\mathbf{T}^*$ is the co-rotated stress. The observer registers the following time rate $\mathbf{T}^*$:

$$\dot{\mathbf{T}}^* = \dot{\mathbf{R}}^T\mathbf{T}\mathbf{R} + \mathbf{R}^T\dot{\mathbf{T}}\mathbf{R} + \mathbf{R}^T\mathbf{T}\dot{\mathbf{R}} .$$

Now we choose the actual configuration as our reference configuration. Then we have $\mathbf{R} = \mathbf{R}^T = \mathbf{1}$ and $\dot{\mathbf{R}} = -\dot{\mathbf{R}}^T = \mathbf{W}$, and we denote $\dot{\mathbf{T}}^*$ as $\mathring{\mathbf{T}}$:

$$\mathring{\mathbf{T}} = \dot{\mathbf{T}} - \mathbf{W}\mathbf{T} + \mathbf{T}\mathbf{W} .$$

$\mathring{\mathbf{T}}$ is the co-rotational or Zaremba[1] stress rate. $\mathring{\mathbf{T}}$ is the stress change that results solely from the deformation of the considered material, whereas any apparent parts (due to rotations of the observer or of the reference frame) are removed.

The *principle of material frame-indifference*, in short called *objectivity*, requires that a constitutive equation determines the stress $\mathbf{T}$ in such away, that an equivalent motion leads to $\mathbf{T}^*$, whereas $\mathbf{T}^*$ and $\mathbf{T}$ are related by $\mathbf{T}^* = \mathbf{Q}\mathbf{T}\mathbf{Q}^T$.

1.6 General constitutive equation

Following the *principle of determinism of the stress* and the *principle of local action* the stress within a non-polar material at a given time t, $\mathbf{T}(t)$, depends on the previous history of the motion χ of a neighbourhood of a material point $\mathbf{X}$. This history is represented as a functional:

$$\mathbf{T}(t) = \mathcal{F}_t(\chi) .$$

For an equivalent motion χ^* objectivity requires:

$$\mathbf{T}^*(t^*) = \mathcal{F}_{t^*}(\chi^*) .$$

For so-called *simple materials* the stress depends only on the history of the deformation gradient:

$$\mathbf{T}(t) = \mathcal{G}(\mathbf{F}^t) = \mathcal{G}_{s=0}^{\infty}(\mathbf{F}(t-s)) ,$$

whereas for so-called materials of the grade n also the n-th deformation gradient is important.

1.7 Principle of macrodeterminism

We consider two strain paths with identical initial and end points (see Figure 1). One path is smooth, whereas the other path results from the smooth one by superposition of small deviations. Let $\Delta\varepsilon$ be the maximum deviation between the two paths. These deviations may result *e.g.* from an automatic control obtained when trying to pursue the smooth path. The question is whether the corresponding maximum stress deviation $\Delta\sigma$ is large or not. More precisely, it is interesting to observe whether $\Delta\sigma \to 0$

[1] often attributed to Jaumann

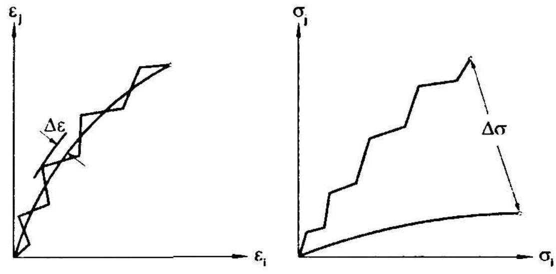

Figure 1. Smoooth and zig-zag-paths

implies $\Delta\varepsilon \to 0$ or not. The first case constitutes the so-called *principle of macrodeterminism*. By now there are no experimental results either to disprove or to corroborate this principle for real materials. Thus, it is still a (rather questionable) postulate than a principle. It is not fulfilled by hypoplastic equations.

1.8 Isotropy groups

There are particular deformations characterized by the deformation gradient **H**, that cannot be detected by subsequent investigation of the material behaviour:

$$\mathcal{G}^{\tau}_{s=0}\left(\mathbf{F}^{(t)}(s)\right)\mathbf{Q}_0^T = \mathcal{G}^{\tau}_{s=0}\left(\mathbf{F}^{(t)}(s)\mathbf{H}\right).$$

All deformation gradients **H** with this property constitute the so-called isotropy (or symmetry) group of a material. An isotropy group is defined with reference to a particular configuration of the body. If an orthogonal tensor **Q** belongs to the isotropy group, we infer from objectivity:

$$\mathbf{Q}_0\mathcal{G}^{\tau}_{s=0}\left(\mathbf{F}^{(t)}(s)\right)\mathbf{Q}_0^T = \mathcal{G}^{\tau}_{s=0}\left(\mathbf{Q}(s)\mathbf{F}^{(t)}(s)\right), \qquad \mathbf{Q}_0 := \mathbf{Q}(t=0).$$

Hence,

$$\mathbf{Q}\mathcal{G}^{\tau}_{s=0}\left(\mathbf{F}^{(t)}(s)\right)\mathbf{Q}^T = \mathcal{G}^{\tau}_{s=0}\left(\mathbf{Q}\mathbf{F}^{(t)}(s)\mathbf{Q}^T\right).$$

A material is called isotropic if there is at least one undistorted state such that its isotropy group is the full orthogonal group. For isotropic materials we cannot detect rotations by means of mechanical tests.

1.9 Rate dependence

Rate-independent materials are defined as materials without an internal time scale, *i.e.*, the rate of deformation is immaterial for the final stress. In other words, rate-independent materials are in variant with respect to changes of time scale. If we deform a rate-independent material twice as fast, then the stress rate will also be doubled. With respect to constitutive equations of the rate-type (*i.e.* constitutive equations of the type $\mathring{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$), rate-independence means that the stress rate $\mathring{\mathbf{T}}$ is positively homogeneous of the first degree with respect to **D**:

$$\mathbf{h}(\mathbf{T},\lambda\mathbf{D}) = \lambda\mathbf{h}(\mathbf{T},\mathbf{D}) \quad \text{for} \quad \lambda > 0.$$

Note that this homogeneity does by no means imply linearity (*cf.* the relation $y = |x|$, which is homogeneous in the above sense, but not linear). Soils are not exactly rate-independent. Clays are more pronouncedly rate dependent than sands, but also sands exhibit rate dependence. However, for a first approximation we can consider soils as rate-independent materials.

2. Hypoplasticity

2.1 Rate equations

A constitutive equation is expected to represent stress due to a strain (or deformation) history starting from some specified reference state. If we represent stress as a *function* of strain, this automatically means that the stress does not depend on the deformation *history*. This special case is called (by definition) elastic behaviour. Soil is not elastic, so we have to find another type of relation. How can we represent strain history? Some researchers introduced integral transformations using appropriate kernels. This approach is not useful for soils. A general way to introduce history (or path) dependence in physics is to use non-integrable differential forms (or Pfaffean forms), *i.e.* to represent y by the differential equation:

$$dy = a_1 dx_1 + a_2 dx_2 + \ldots a_n dx_n \ .$$

This equation connects increments dx_1, dx_2, ... with dy (or dy_1, dy_2, ..., if y is a vector) in such away that there is no closed-form representation of $y(x)$, *i.e.* the relation (which is called incremental, as it relates increments) $dy = f(dx_i)$ is not integrable. This is the way we proceed in soil mechanics when we represent the stress increment as a non-integrable function of the strain increment:

$$d\sigma = f(d\varepsilon) \ .$$

This approach is common to the theories of plasticity and hypoplasticity.

Now we can divide all increments by dt and obtain time rates:

$$\dot{\sigma} = d\sigma/dt \ , \qquad \dot{\varepsilon} = d\varepsilon/dt \ , \qquad etc.$$

Thus, an equation between increments is also representable as an equation between rates, as long as we refer to so-called rate-independent materials. An equation of the form $\dot{\sigma} = f(\dot{\varepsilon})$ is called a rate-equation. It does not imply the existence of an equation $\sigma = g(\varepsilon)$.

In tensor notation a constitutive equation of the rate type has the form $\mathring{\mathbf{T}} = \mathbf{f}(\mathbf{D})$. It is often reasonable to include $\mathbf{T}$ in the list of arguments, *i.e.* to write $\mathring{\mathbf{T}} = \mathbf{f}(\mathbf{T}, \mathbf{D})$. Note that, strictly speaking, $\mathbf{D}$ is not the time rate of any strain measure, and also $\mathring{\mathbf{T}} \neq \dot{\mathbf{T}}$. However, for the special case of rectilinear extensions ($\mathbf{W} = \mathbf{0}$) we have $\mathring{\mathbf{T}} \neq \dot{\mathbf{T}}$, and $\mathbf{D}$ is the time rate of logarithmic strain ε_{ij}.

2.2 Incremental non-linearity

$d\sigma/d\varepsilon = \dot{\sigma}/\dot{\varepsilon}$ represents the incremental stiffness of the material considered (see Figure 2). Since for anelastic (plastic) materials relation $|d\sigma|$, is much larger

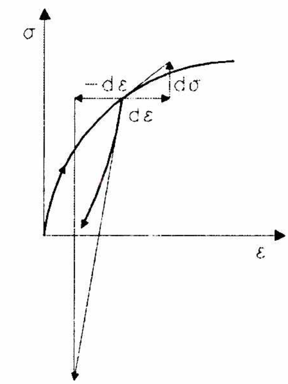

Figure 2. *Different stiffness at loading and unloading*

at unloading than at loading (*i.e.* the stiffness is much larger at unloading than at loading), we infer that for such a material the function $d\sigma = f(d\varepsilon)$ or $\dot{\sigma} = f(\dot{\varepsilon})$ must be nonlinear in $\dot{\varepsilon}$ (or $d\varepsilon$) This non-linearity remains, no matter how small $d\varepsilon$ is. Therefore it is called "non-linearity in the small" or "incremental non-linearity". Note that incremental non-linearity has nothing to do with the curved form of the stress-strain curve for loading. This curve can be, of course, linearized for small $|d\varepsilon|$, a fact which led many people to believe that in physics every relation can be linearized "in the small". Thus, all elastoplastic and hypoplastic relations are incrementally non-linear.

Incremental non-linearity is the seat of the hysteresis loop exhibited by stress-strain curves at cyclic stress. It also implies that constitutive equations of the form $\mathring{\mathbf{T}} = \mathbf{f}(\mathbf{T}, \mathbf{D})$ are non-linear in $\mathbf{D}$ and, also, non-differentiable at $\mathbf{D} = 0$. This fact imposes many mathematical difficulties.

2.3 Homogeneity in stress

On the base on tests obtained with a true triaxial apparatus Goldscheider formulated a principle according to which:

> proportional (*i.e.* straight) strain paths starting from a (nearly) stress free state are connected with proportional stress paths. If the initial state is not stress free, then the obtained stress path approaches asymptotically the path starting from the stress free state (see Figure 3).

This theorem has far-reaching consequences.

Assume that the relation $\mathring{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$ is homogeneous in $\mathbf{T}$, *i.e.*:

$$\mathbf{h}(\lambda\mathbf{T}, \mathbf{D}) = \lambda^n \mathbf{h}(\mathbf{T}, \mathbf{D}).$$

Let us investigate the consequences of this assumption. Consider a stress state $\mathbf{T}_1$. We now determine the stretching in such away that $\mathring{\mathbf{T}} = \mathbf{h}(\mathbf{T}_1, \mathbf{D}_1) = \lambda\mathbf{T}_1$. If we then continuously apply $\mathbf{D}_1$, then we shall obtain a stress path which is a straight line

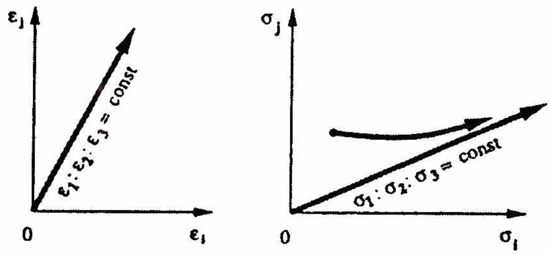

Figure 3. *Stress and strain paths referring to Goldscheider's principle*

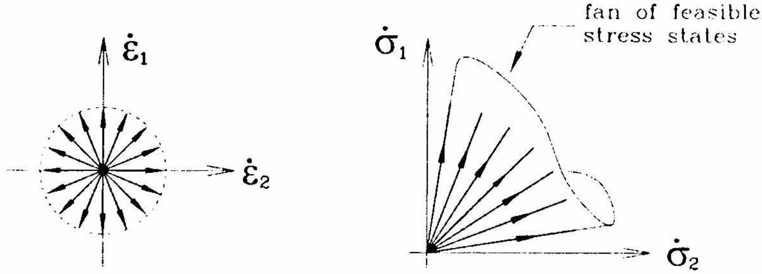

Figure 4. *Strain path directions (left) and correspond stress paths (right) obtained, each, by the constant application of the aforementioned strain directions and starting from the stress free state. Schematic representation showing only the components in 1- and 2-directions*

passing through the origin of stress space. This follows from our assumption, because:

$$\overset{\circ}{\mathbf{T}}(t+dt)=\mathbf{h}(\mathbf{T}_1+\lambda\mathbf{T}_1\,dt,\mathbf{D}_1)=(1+\lambda\,dt)^n\,\mathbf{h}(\mathbf{T}_1,\mathbf{D}_1)=(1+\lambda\,dt)^n\,\mathbf{T}(t)\,.$$

In other words, our assumption of homogeneity in **T** implies that proportional strain paths (*i.e.* paths with **D** = const) are connected with proportional stress paths (*i.e.* straight stress paths passing through the origin of the stress space) and conforms, thus, with Goldscheider's principle.

Note that proportional stress paths must be limited within a fan, because there are also inaccessible (infeasible) stress states (see Figure 4). *E.g.*, a stress state with tensile principal stresses is not feasible for cohesionless granulates. Referring to Figure 3 it is interesting to note that if we apply the proportional strain path shown in its left part, which starts *not* from the stress-free state, we obtain the curved stress path shown in this figure. Let us now consider the degree of homogeneity.

Knowing that $d\sigma/d\varepsilon=\dot{\sigma}/\dot{\varepsilon}$ or $\dot{\mathbf{T}}/\mathbf{D}$ is the stiffness, we infer that $(\dot{\mathbf{T}}/\mathbf{D})|_{\lambda\mathbf{T}} = \lambda^n(\dot{\mathbf{T}}/\mathbf{D})|_{\mathbf{T}}$. In other w ords, if we increase the stress by a factor λ, the stiffness is increased by the factor λ^n. Experimentalists in soil mechanics often remark that *normalized* stress-strain curves coincide (this is in particular the case with normally consolidated clays). The consequence is $n = 1$. Setting $n = 1$ would

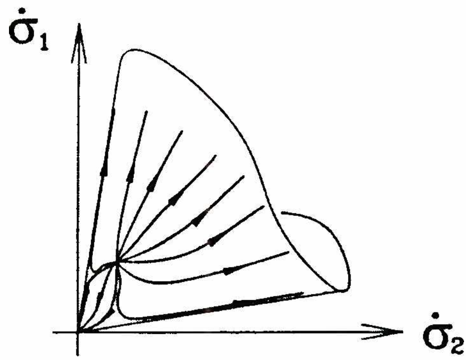

Figure 5. *Stress paths obtained with proportional strain paths starting not from the stress free state*

imply that the friction angle is invariant with respect to the stress level. This is an acceptable approximation to start with. If, however, the changes of stress level at a given void ratio are considerable, then the correspond variation of friction (and dilatancy) angles may not be neglected. Note that for the case $n = 1$ all material constants must be dimensionless.

2.4 *Hypoelasticity*

Truesdell introduced constitutive relations of the form $\overset{\circ}{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$. He required that the function **h**( ) be linear in **T** and in **D** and introduced the name hypoelasticity for such relations. Hypoelastic constitutive equations may produce curved stress-strain curves, and in some cases these stress-strain curves reach a horizontal plateau and can thus model yielding. However, the imposed incremental linearity implies equal stiffness for loading and unloading, and thus renders hypoelastic relations inappropriate to describe anelastic (plastic) materials. Despite this, some hypoelastic relations have been launched in soil mechanics (*e.g.* by Davis and Mullenger). They are (in most cases tacitly) endowed with additional stress-strain relations holding for (appropriately defined) unloading. Strictly speaking, such relations (regarded as a whole) are not linear any more, *i.e.* they are not hypoelastic.

2.5 *Hypoplasticity*

Elastoplastic and hypoplastic equations are both of the general form:

$$\overset{\circ}{\mathbf{T}} = \mathbf{h}(\mathbf{T},\mathbf{D}).$$

Starting from the fact that every function **h(T, D)** can be represented according to the general representation theorem,

$$\mathbf{h}(\mathbf{T},\mathbf{D}) = \psi_1 \mathbf{1} + \psi_2 \mathbf{T} + \psi_3 \mathbf{D} + \psi_4 \mathbf{T}^2 + \psi_5 \mathbf{D}^2 + \psi_6 (\mathbf{TD} + \mathbf{DT}) +$$
$$+ \psi_7 (\mathbf{TD}^2 + \mathbf{D}^2\mathbf{T}) + \psi_8 (\mathbf{T}^2\mathbf{D} + \mathbf{DT}^2) + \psi_9 (\mathbf{T}^2\mathbf{D}^2 + \mathbf{D}^2\mathbf{T}^2),$$

(ψ_i are scalar functions of invariants and joint invariants of **T** and **D**), the experiment was undertaken to find such a unique function which appropriately describes the mechanical properties of soils. In order to a void the shortcomings of hypoelasticity, this function has to be non-linear in **D**. On the other hand, it should be homogeneous of the first degree in **D** in order to describe rate-independent materials, and homogeneous in **T** in order to describe proportional stress-paths in case of proportional strain paths. Therefore, the design of such a function had to proceed along the above stated representation theorem and some general mathematical restrictions:

— non-linearity in **D**,

— homogeneity in **D** and **T**,

with avoidance of any recourse to notions from the theory of elastoplasticity such as yield functions, decomposition of strain *etc.*

This experiment (every theory is, virtually, an experiment) was more or less successful, as by trial anderror a function was found which was able to describe many aspects of soil behaviour. Thus, a new approach to constitutive modelling was created. The name "hypoplastic" equation fits very well, as the relation between hypoplasticity and (elasto)plasticity is the same as the one between hypoelasticity and elasticity: The theories with "hypo" do not use a potential. It should be mentioned that Dafalias formulated hypoplasticity earlier for something else, which can be considered as a general case of what we call hypoplasticity .

Let us now have a look at some hypoplastic equations. Most of them consist of 4 tensorial terms (so-called tensor generators) combined with 4 material parameters C_1, C_2, C_3 and C_4, *e.g.*:

$$\mathring{\mathbf{T}} = C_1 (\operatorname{tr}\mathbf{T})\mathbf{D} + C_2 \frac{(\operatorname{tr}\mathbf{T}\mathbf{D})}{\operatorname{tr}\mathbf{T}}\mathbf{T} + C_3 \frac{\mathbf{T}^2}{\operatorname{tr}\mathbf{T}}\sqrt{\operatorname{tr}\mathbf{D}^2} + C_4 \frac{\mathbf{T}^{*2}}{\operatorname{tr}\mathbf{T}}\sqrt{\operatorname{tr}\mathbf{D}^2} . \quad (5)$$

An alternative representation of hypoplastic constitutive equations is to summarize the linear terms by **LD**, with **L** being a linear operator applied to **D**, and the non-linear terms by $\mathbf{N}|\mathbf{D}|$ with $|\mathbf{D}| := \sqrt{\operatorname{tr}\mathbf{D}^2}$. Then, a hypoplastic equation assumes the general form

$$\mathring{\mathbf{T}} = \mathbf{L}\mathbf{D} + \mathbf{N}|\mathbf{D}| \quad (6)$$

or

$$\mathring{T}_{ij} = L_{ijkl} D_{kl} + N_{ij} |\mathbf{D}| .$$

The components L_{ijkl} and N_{ij} depend on the actual stress and can easily be numerically determined for a given constitutive relation $\mathring{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$: for any given values of the indices, say $k = k^*$ and $l = l^*$, we set $D_{k^*l^*} = 1$, else $D_{kl} = 0$, and obtain from the constitutive relation $\mathring{T}^+_{ij}$. Then we set $D_{k^*l^*} = -1$ and obtain $\mathring{T}^-_{ij}$. Subsequently we obtain:

$$L_{ijk^*l^*} = \frac{1}{2}\left(\overset{\circ}{T}{}^{+}_{ij} - \overset{\circ}{T}{}^{-}_{ij}\right),$$

$$N_{ij} = \frac{1}{2}\left(\overset{\circ}{T}{}^{+}_{ij} - \overset{\circ}{T}{}^{-}_{ij}\right).$$

Note that N_{ij} is independent of k^* and l^*, as it does not depend on $\mathbf{D}$. L_{ijkl} is not necessarily symmetric in the sense $L_{ijkl} \neq L_{klij}$. For the case of Equation (5), however, L_{ijkl} is symmetric in the aforementioned sense.

Another expression for constitutive relations of type (6) is:

$$\overset{\circ}{\mathbf{T}} = \mathbf{HD}\,,$$

with

$$\mathbf{H} := \mathbf{L} + \mathbf{N} \otimes \mathbf{D}^0\,.$$

Herein, $\mathbf{D}^0$ is the normalized stretching, *i.e.* $\mathbf{D}^0 := \mathbf{D}/|\mathbf{D}|$. In index notation the stiffness matrix $\mathbf{H}$ can be expressed as:

$$H_{ijkl} = L_{ijkl} + N_{ij}\, D^0_{kl}\,.$$

Several equations with only 4 material parameters C_1, C_2, C_3, C_4 could be found:

- the triaxial test as characterized by a stiffness decreasing down to zero at the limit state and a correspond volumetric strain curve exhibiting first contractancy and then dilatancy,
- incrementally non-linear behaviour, *i.e.* unloading stiffness much larger than at loading,
- realistic asymptotic properties (referring to proportional paths).

However, the void ratio was not taken into account, and, therefore, such simple hypoplastic constitutive models were not capable of describing the difference of friction angle and stiffness between dense and loose samples, or the decrease of the peak friction angle to the residual one with increasing strain (softening). But this was also not expected from such simple constitutive models. To achieve this, more recent versions have been elaborated in Karlsruhe scalar factors, which aim to model the influence of density and stress level as well as the transition to the so-called critical state. Of course, such factors increase the intricacy of the models.

Hypoplastic constitutive relations are directly presented without reference to any sort of surfaces in stress space. However, various surfaces can be derived from a hypoplastic equation, as will be explained in Section 5.

2.6 Second stretching tensor

Many of the present shortcomings of hypoplasticity appear upon changes of $\mathbf{D}$. It appears therefore interesting to introduce $\overset{\circ}{\mathbf{D}}$, the objective time rate of $\mathbf{D}$ (also called the second stretching tensor). In fact, consideration of $\overset{\circ}{\mathbf{D}}$ helps to describe the following effects:

— direction of undrained stress path immediately after isotropic consolidation,

— cyclic behaviour (in general),

— rate dependence, as manifested with tests with jump of deformation rate,

— creep.

There are effects due to change of the *direction* of **D** and also effects due to change of the *value* $|\mathbf{D}|$.

To start with, the following expression was investigated:

$$\overset{\circ}{\mathbf{T}} = \underbrace{\mathbf{h}(\mathbf{T},\mathbf{D},e)}_{\text{as already known}} + \int_{t'=-\infty}^{t} a\,\overset{\circ}{\mathbf{D}}_{t'} e^{-\beta\cdot g(t,t')} dt' \cdot |\mathbf{D}| \tag{7}$$

In Equation (7) all tensors **D** in the past are taken into account through integration over t'. The larger the time lapse $t = t'$, the lesser the influence of $\mathbf{D}(t')$ is. This is achieved by the fading factor $e^{-\beta g}$, where g is a measure of the time lapse. More precisely, g counts this time lapse in terms of deformation occuring in the time intervall from t' to t:

$$g(t,t') = \int_{t''=t'}^{t} |\mathbf{D}_{t''}| dt'' ,$$

Perhaps, it should be tried also with:

$$g(t,t') = \int_{t''=t'}^{t} \frac{\operatorname{tr}(\mathbf{T}_{t''}\mathbf{D}_{t''})}{\operatorname{tr}\mathbf{T}_{t''}} dt'' .$$

If the change of **D** occurs as a jump at time t_i, then we should write (instead of $\overset{\circ}{\mathbf{D}}$):

$$[\mathbf{D}_{t_i}] \cdot \delta(t - t_i) ,$$

where $[\mathbf{D}_{ti}] := \mathbf{D}_{ti+0} - \mathbf{D}_{ti-0}$ is the jump of **D** and δ is Dirac's function. In this case the constitutive equation reads:

$$\overset{\circ}{\mathbf{T}} = \mathbf{h}(\mathbf{T},\mathbf{D},e) + \left(\sum_{i=1}^{n} a[\mathbf{D}_{t_i}] e^{-\beta g(t,t_i)} \right) \cdot |\mathbf{D}| .$$

The material constant β defines how "fast" the obliviation sets on, whereas a determines the influence of this term. As a first approximation, a could be considered as a material constant. Later one can try to set a as a function of **T**, **D** and e.

If [**D**] is reduced by a factor of, say, 0.001 (*i.e.* if we reduce the values of **D** before and after the jump), then the contribution of the new term becomes negligible. To remove this shortcoming, a should depend on **D**. For example, in Equation (8) we should consider not the rate of **D** but the rate of $\mathbf{D}^0 := \mathbf{D}/|\mathbf{D}|$.

For the special case of rectilinear extensions we have $\overset{\circ}{\mathbf{D}} = \dot{\mathbf{D}}$. Then we have:

$$\dot{\mathbf{D}}^0 = \frac{\dot{\mathbf{D}}}{|\mathbf{D}|} - \frac{\operatorname{tr}(\mathbf{D}\dot{\mathbf{D}})}{|\mathbf{D}|^{3/2}} \mathbf{D} .$$

However, this expression does not allo w to describe the *rate dependence* after Leinenkugel. Perhaps, one should try with $a :=$ const + $|\mathbf{D}|$ or $a := \sqrt{\text{const} + \text{tr}\,\mathbf{D}^2}$.

2.7 Numerical sim ulation of element tests

How can we obtain simulations of lab oratory element tests by using an equation of the rate type? First, we have to start from a known stress state. If the test to be simulated has kinematical boundary conditions, then the stretching **D** is known, *e.g.* in case of the oedometer test all but one components of **D** are equal to zero and the only non-vanishing component corresponds to the rate of compression. With knowledge of **T** and **D** the constitutive equation $\overset{\circ}{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$ makes it possible to evaluate $\overset{\circ}{\mathbf{T}}$. Multiplying $\overset{\circ}{\mathbf{T}}$ with a sufficiently small time step Δt gives $\Delta\mathbf{T} \approx \overset{\circ}{\mathbf{T}}\,\Delta t$. The new stress state is then obtained to $\mathbf{T} + \Delta\mathbf{T}$. This process can be continued and corresponds to a numerical in tegration of the evolution equation $\overset{\circ}{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$ (so-called Euler forward integration). The procedure is a little more difficult if not all of the boundary conditions are of the kinematic type. In case of a static boundary condition (*e.g.* $\sigma_2 = \sigma_3 =$ const for triaxial test), the component D_2 of **D** must be determined by solving the algebraic equation $\dot{\sigma}_2\,(D_2) = 0$. A program to simulate element tests can be downloaded from *ftp://ftp.uibk.ac.at/pub/uni-innsbruck/igt/sources/*

2.8 Calibration

A constitutive relation is of no use if the involved material parameters cannot be adapted to a particular material. The values of these parameters constitute the identity card of this material with respect to a particular constitutive model. Moreover, a particular parameter is useless unless it is embedded within a constitutive model. *E.g.* the notion "viscosity" is unclear unless it is embedded within a Newton-type constitutive equation, say $\tau = \mu\dot{\varepsilon}$. The process of the determination of the values of the parameters of a constitutive model is called "calibration" or "parameter identification". In many publications on constitutive models the calibration is simply omitted as being not worth mentioning. In fact it is a task which can take up to several months of work! Considering hypoplastic constitutive equations, the calibration is straight forward by fitting the equation to the outcomes of one or several (say triaxial) tests values of the strain and stress increments at a particular stress state from experiments, the only remaining unknowns in the constitutive equation are the material constants. Thus, we have to solve a system of four linear equations.

2.9 Cyclic loading, ratcheting, shake-down

Cyclic loading is recognized as one of the most difficult fields in soil mechanics. Elastoplasticity and hypoplasticity bear some inherent deficiencies which become more important in the case of cyclic loading. In the realm of classical elastoplasticity all unloading — reloading cycles are completely elastic, a feature which is not realistic. On the other hand, in (the initial versions of) hypoplasticity the first and subsequent unloading-reloading cycles are identical to the virgin loading-unloading. This shortcoming is called ratcheting effect and is due to the fact that in hypoplasticity the stress is the only memory parameter.

In reality either a gradual transition from plastic to elastic behaviour (so-called shake-down) takes place or deformation increases unbounded with the number of cycles (so-called incremental collapse). Regarding shake-down, experiments show that the behaviour of soil never becomes completely elastic, as every cycle is connected with dissipation of energy, a fact which is modelled in soil dynamics by a fictitious viscous damping.

It turns out that the quality of the modelling of cyclic behaviour depends on whether the stress amplitudes are small or large. If the unloading is continued to the extension side (*i.e.* the stress deviator changes sign), then the hypoplastic models work satisfactorily. Furthermore, the proper incorporation of barotropy and pyknotropy by the advanced hypoplastic models enables that cyclic shearing produces gradually a high density (*i.e.* small void ratio) which cannot be exceeded by additional cycles.

A more general representation of the cyclic behaviour in hypoplasticity requires an additional state variable such as a structure tensor history. A "memory function" or an "intergranular strain" have been proposed for this purpose.

3. Uniqueness and limit loads

3.1 Limit states

A very important property of granular materials is their ability to flow (or yield), *i.e.* to undergo large deformations without stress change, as soon as the stresses and the void ratioobtain their critical values. This sort of flow should be attributed as "plastic" flow and distinguished from the flow of fluids. The latter has a pronounced viscous (rate-dependent) character.

Plastic flow occurs as soon as the stress state $\mathbf{T}$ and the strain rate $\mathbf{D}$ fulfil the condition $\mathbf{h}(\mathbf{T}, \mathbf{D}) = \mathbf{0}$. In the theory of elastoplasticity the condition in terms of $\mathbf{T}$ is called the yield (limit) surface, and the condition in terms of $\mathbf{D}$ is called the flow rule.

In elastoplasticity the yield function is the starting point and the mathematical relation connecting strain and stress increments at loading is based up on this yield function. In contrast, it can be shown that a yield function is contained in a hypoplastic formulation $\mathring{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$, *i.e.* the yield function $f(\mathbf{T})$ can be derived from the constitutive relation. To this purpose we rewrite (following a proposition of Desrues and Chambon):

$$\mathring{\mathbf{T}} = \mathbf{L}(\mathbf{T})[\mathbf{D}] + \mathbf{N}(\mathbf{T})|\mathbf{D}| = \mathbf{L}(\mathbf{T})\big[\mathbf{D} + \mathbf{B}|\mathbf{D}|\big],$$

with $\mathbf{L}(\mathbf{T})$ being a matrix operator applied to its tensorial argument. It is obvious that $\mathbf{h}(\mathbf{T}, \mathbf{D}) = \mathbf{0}$ occurs for:

$$\mathbf{D}^0 := \mathbf{D}/|\mathbf{D}| = -\mathbf{B}.$$

Consequently, the function $f(\mathbf{T})$ reads:

$$f(\mathbf{T}) = \operatorname{tr}\mathbf{B}^2 - 1\,,$$

with **B** being a function of **T**. In other words, the limit surface reads:

$$f(\mathbf{T}) = \operatorname{tr}\mathbf{B}^2 - 1 = 0\,,$$

For the constitutive equation (5) **B** reads as follows[2]:

$$\mathbf{B} = \mathbf{L}^{-1}\mathbf{N} = \frac{C_3\mathbf{T}^2}{C_1(\operatorname{tr}\mathbf{T})^2} + \frac{C_4\mathbf{T}^{*2}}{C_1(\operatorname{tr}\mathbf{T})^2} - \left(\frac{C_2C_3}{C_1}\frac{\operatorname{tr}(\mathbf{T}^3)}{(\operatorname{tr}\mathbf{T})^2} + \frac{C_2C_4}{C_1}\frac{\operatorname{tr}(\mathbf{T}\mathbf{T}^{*2})}{(\operatorname{tr}\mathbf{T})^2}\right)\times$$
$$\times\frac{\mathbf{T}}{C_1(\operatorname{tr}\mathbf{T})^2 + C_2\operatorname{tr}(\mathbf{T}^2)}\,.$$

Due to the homogeneity of $\mathbf{h}(\mathbf{T}, \mathbf{D})$ (and consequently also of **B**) in **T**, the surface $f(\mathbf{T}) = 0$ is a cone with apex at the origin $\mathbf{T} = \mathbf{0}$. The cross section of this cone with the deviatoric plane reveals the influence of the intermediate principal stress, *i.e.* the yield surface differs from the one determined by the Mohr-Coulomb criterion (where the in termediate principal stress does not play any role).

3.2 Invertibility and controllability

In kinematically controlled tests (such as o edometric test or undrained triaxial test) the stretching **D** is prescribed and the stress rate $\overset{\circ}{\mathbf{T}}$ can be uniquely determined by means of the hypoplastic constitutive equation. What about the unique determination of **D** when $\overset{\circ}{\mathbf{T}}$ is prescribed? To answer this question of unique in vertibility we[3] multiply the equation[4] $\overset{\circ}{\mathbf{T}} = \mathbf{L}\mathbf{D} + \mathbf{N}|\mathbf{D}|$ with the inverse operator $\mathbf{L}^{-1}$ and obtain:

$$\mathbf{A} := \mathbf{L}^{-1}\overset{\circ}{\mathbf{T}} = \mathbf{D} + \mathbf{L}^{-1}\mathbf{N}|\mathbf{D}|$$

or

$$\mathbf{D} = \mathbf{A} - \mathbf{B}|\mathbf{D}|\,, \tag{8}$$

with $\mathbf{B} := \mathbf{L}^{-1}\mathbf{N}$. With the notation $\mathbf{X}\cdot\mathbf{Y} := \operatorname{tr}(\mathbf{X}\mathbf{Y})$ we obtain from (8):

$$\mathbf{D}\cdot\mathbf{D} = (\mathbf{A} - \mathbf{B}|\mathbf{D}|)\cdot(\mathbf{A} - \mathbf{B}|\mathbf{D}|) = \mathbf{A}\cdot\mathbf{A} - 2\mathbf{A}\cdot\mathbf{B}|\mathbf{D}| + \mathbf{B}\cdot\mathbf{B}|\mathbf{D}|^2\,. \tag{9}$$

Noting that $\mathbf{D}\cdot\mathbf{D} \equiv |\mathbf{D}|^2$ we observe that (9) is a quadratic equation for $x := |\mathbf{D}|$. Its solution reads:

$$x_{1/2} = \frac{2\mathbf{A}\cdot\mathbf{B} \pm \sqrt{4(\mathbf{A}\cdot\mathbf{B})^2 - 4\mathbf{A}\cdot\mathbf{A}(\mathbf{B}\cdot\mathbf{B} - 1)}}{2(\mathbf{B}\cdot\mathbf{B} - 1)}\,.$$

---

[2] private communication by J. Nader

[3] the author is indebted to Dr. P. Wagner, Innsbruck, for many suggestions to this section

[4] for simplicity, the brackets in $\mathbf{L}[\mathbf{D}]$ are omitted

Since x is a modulus, only a solution $x > 0$ is meaningful. Moreover, in order to obtain a unique solution we have to require that only one solution is positive, *i.e.* $x_1 \cdot x_2 < 0$:

$$x_1 \cdot x_2 = \frac{4(\mathbf{A}\cdot\mathbf{B})^2 - 4(\mathbf{A}\cdot\mathbf{B})^2 + 4\mathbf{A}\cdot\mathbf{A}(\mathbf{B}\cdot\mathbf{B}-1)}{4(\mathbf{B}\cdot\mathbf{B}-1)^2} = \frac{\mathbf{A}\cdot\mathbf{A}}{\mathbf{B}\cdot\mathbf{B}-1} < 0 .$$

Since $\mathbf{A}\cdot\mathbf{A} > 0$ we infer that in vertibility is given for $\mathbf{B}\cdot\mathbf{B} - 1 < 0$, *i.e.* for all stress states inside the limit surface $\mathbf{B}\cdot\mathbf{B} - 1 = 0$, as already pointed by Chambon.

A more subtle question on unique solutions of the constitutive equation arises if some (say k) components of $\mathbf{D}$ and $6 - k$ components of $\mathring{\mathbf{T}}$ are prescribed, and the remaining components have to be determined[5]. The existence of a unique solution of this problem (which corresponds *e.g.* to the conve tional triaxial test with the mixed conditions $D_{11} = 1$ in axial direction and $\mathring{T}_{22} = \mathring{T}_{33} = 0$ in lateral directions) is called controllability $\mathbf{D}$ and $\mathring{\mathbf{T}}$ as column or row vectors $\mathbf{x}$ and $\mathbf{y}$, *i.e.* we take $x_1 := D_{11}$, $x_2 := D_{22}$, ... and similarly $y_1 := \mathring{T}_{11}$, $y_2 := \mathring{T}_{12}$, ...

The selection of the independent and dependent components can be accomplished by the partition matrices $\mathbf{P}$ and $\mathbf{Q}$, the components of which vanish for $i \neq j$. Their diagonal components are either 1 or 0. $\mathbf{P}$ and $\mathbf{Q}$ are related by $\mathbf{P} + \mathbf{Q} = \mathbf{1}$, with $\mathbf{1}$ being the unit matrix. For example:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can now obtain the independent (or controlling) variable $\mathbf{X}$ of a problem with mixed conditions as:

$$\mathbf{X} = \mathbf{Q}\mathbf{x} + \mathbf{P}\mathbf{y}, \tag{10}$$

and, likewise, the dependent variable $\mathbf{Y}$ as:

$$\mathbf{Y} = \mathbf{P}\mathbf{x} + \mathbf{Q}\mathbf{y}. \tag{11}$$

For example:

$$\mathbf{X} = (X_1, X_2, \ldots)^T = (D_{11}, \mathring{T}_{12}, \mathring{T}_{13}, D_{22}, \ldots)^T,$$

$$\mathbf{Y} = (Y_1, Y_2, \ldots)^T = (\mathring{T}_{11}, D_{12}, D_{13}, \mathring{T}_{22}, \ldots)^T.$$

[5] as Nova points out, the most general case of test control (*e.g.* a test with $T_1 + T_2 + T_3 = \text{const.}$, $D_2 = D_3$; $D_1 = 1$) is obtained if we replace $\dot{\mathbf{T}}$ by $\dot{\mathbf{T}}' := \mathbf{S}\dot{\mathbf{T}}$ and $\mathbf{D}$ by $\mathbf{D}' := \mathbf{S}^{-1}\mathbf{D}$ with some appropriately chosen non-singular matrix $\mathbf{S}$. Obviously, $\mathbf{T}'$ and $\mathbf{D}'$ are energy-conjugated in the sense that $\mathrm{tr}(\mathbf{TD}) = \mathrm{tr}(\mathbf{T}'\mathbf{D}')$ or $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D}) = \mathrm{tr}(\mathring{\mathbf{T}}'\mathbf{D}')$

Inverting the system of equations (10) and (11) and using $(\mathbf{1}-2\mathbf{P})^{-1}=(\mathbf{1}-2\mathbf{P})$ we obtain:

$$\mathbf{x}=\mathbf{QX}+\mathbf{PY}\,, \tag{12}$$

$$\mathbf{y}=\mathbf{PX}+\mathbf{QY}\,. \tag{13}$$

Inserting (12) into the constitutive equation $\mathbf{y}=\mathbf{h}(\mathbf{x})$ or $\mathbf{y}-\mathbf{h}(\mathbf{x})=\mathbf{0}$ we obtain an implicit relation between $\mathbf{X}$ and $\mathbf{Y}$:

$$\mathbf{F}(\mathbf{X},\mathbf{Y}):=\mathbf{PX}+\mathbf{QY}-\mathbf{h}(\mathbf{QX}+\mathbf{PY})=\mathbf{0} \tag{14}$$

A unique determination of $\mathbf{Y}$ from (14) (*i.e.* controllability) is possible if $\det(\partial\mathbf{F}/\partial\mathbf{Y})=\det(\partial F_i/\partial Y_j)\neq 0$. This means that:

$$\det\left(\mathbf{Q}-\mathbf{P}\frac{d\mathbf{h}}{d\mathbf{x}}\right)\neq 0\,.$$

$\partial\mathbf{h}/\partial\mathbf{x}$ the stiffness matrix. For a hypoplastic constitutive equation $\mathbf{h}(\mathbf{x}):=\mathbf{L}\mathbf{x}+\mathbf{N}|\mathbf{x}|$ the stiffness matrix reads:

$$\mathbf{H}(\mathbf{x}):=\frac{d\mathbf{h}}{d\mathbf{x}}=\mathbf{L}+\mathbf{N}\otimes\frac{\mathbf{x}}{|\mathbf{x}|}=\mathbf{L}+\mathbf{N}\otimes\mathbf{x}^0\,, \tag{15}$$

i.e. the stiffness matrix depends on the direction of $\mathbf{x}$. This is to be contrasted with elastoplastic formulations where:

$$\mathbf{H}=\begin{cases}\mathbf{L}_{elastic} & \text{for unloading or inside the yield surface,}\\ \mathbf{L}_{plastic} & \text{for loading.}\end{cases}$$

The application of the operator $\mathbf{P}$ to $\mathbf{H}(\mathbf{x})$ selects from $\mathbf{H}(\mathbf{x})$ only those rows which have a non-vanishing $\mathbf{P}$-component. *E.g.* for[6]

$$\mathbf{P}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}, \qquad \mathbf{Q}=\mathbf{1}-\mathbf{P}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix}$$

we obtain:

$$\mathbf{PH}(\mathbf{x})=\begin{pmatrix}H_{11}&H_{12}&H_{13}\\H_{21}&H_{22}&H_{23}\\0&0&0\end{pmatrix}.$$

Subtracting this from $\mathbf{Q}$ we obtain:

$$\mathbf{Q}-\mathbf{PH}(\mathbf{x})=-\begin{pmatrix}H_{11}&H_{12}&H_{13}\\H_{21}&H_{22}&H_{23}\\0&0&-1\end{pmatrix},$$

[6] for simplicity only 3 × 3 matrices are considered here

such that

$$\det\left(\mathbf{Q}-\mathbf{PH}(\mathbf{x})\right)=\det\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}.$$

We thus see that controllability is given if the determinants of all conceivable symmetric minors of $\mathbf{H}(\mathbf{x})$ are positive. This is the case if $\mathbf{H}(\mathbf{x})$ fulfils the condition[7]:

$$\mathbf{x}\cdot\mathbf{H}(\mathbf{x})\mathbf{x}>0 \qquad \text{for} \qquad \forall\mathbf{x}\neq 0\,. \tag{16}$$

Let $\hat{\mathbf{H}}$ be a symmetric minor of $\mathbf{H}$, and let $\mathbf{H}^s$ be the symmetric part of $\mathbf{H}$. According to the theorem of Ostrowsky and Taussky det $\hat{\mathbf{H}} \geq \det \hat{\mathbf{H}}^s > 0$. Note[8] that $\mathbf{x}\cdot\mathbf{H}(\mathbf{x})\mathbf{x}$ represents the so-called second order work $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})$. Thus positive second order work implies controllability. In other words, positive second order work is sufficient (but not necessary) condition for controllability:

$\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})>0$	$\rightarrow$	controllability is guaranted
lack of controllability	$\rightarrow$	$\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})<0$.

It is interesting to note that, with hypoplastic constitutive equations, the condition $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})=0$ is in fact encountered *before* the peak. More specifically, the condition $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})=0$ (*i.e.* vanishing second order work) constitutes a surface in the stress space. Since this surface is connected with possible loss of uniqueness, we call it "bifurcation surface". It can be easily determined if we insert the hypoplastic equation in to the equation $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})=0$. For simplicity we consider only rectilinear extensions, *i.e.* we restrict the dimensions of column vectors $\mathbf{x}$ and $\mathbf{y}$ to 3. We then obtain:

$$F(\mathbf{x})=\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})=L_{ij}x_ix_j+N_ix_i\cdot|\mathbf{x}|=0 \tag{17}$$

The bifurcation surface is defined as a surface in the stress space. It consists of stress states for which the equation $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D})=0$ possesses only one solution. With $\mathbf{x}^0 := \mathbf{x}/|\mathbf{x}|$ Equation (17) can be written as:

$$\mathbf{x}^0\cdot\left(\mathbf{L}\mathbf{x}^0+\mathbf{N}\right)=0\,. \tag{18}$$

This equation is fulfilled in two cases: first for $\mathbf{L}\mathbf{x}^0+\mathbf{N}=0$, or $\mathbf{x}^0=-\mathbf{L}^{-1}\mathbf{N}$, which occurs on the limit state. As an alternative (so-called Fredholm's alternative), Equation (18) is fulfilled if $\mathbf{x}^0$ is orthogonal to $\mathbf{L}\mathbf{x}^0+\mathbf{N}$. All stress states for which only one $\mathbf{x}$ fulfils this condition constitute the so-called bifurcation surface. To determine the bifurcation surface we require:

$$\nabla F-\lambda\nabla G=0\,, \tag{19}$$

with

$$G(x)=|x|^2-1=x_1^2+x_2^2+x_3^2-1=0\,. \tag{20}$$

[7] the more general case $\mathbf{y}\cdot\mathbf{H}(\mathbf{x})\mathbf{y}>0$ for $\forall$ $\mathbf{x}$, $\mathbf{y}$ is not considered here

[8] a constant matrix $\mathbf{H}$ fulfilling the condition $\mathbf{x}\cdot\mathbf{H}\mathbf{x}\equiv\mathbf{x}^T\mathbf{H}\mathbf{x}>0$ is called positive definite. All the eigenvalues of $\mathbf{H}^s := (\mathbf{H}+\mathbf{H}^T)/2$ are positive

Introducing (20) into (19) we obtain:

$$2L_{(ij)}x_j + N_i + N_k x_k x_i - 2\lambda x_i = 0 . \tag{21}$$

For $i = 1, 2, 3$ the vectorial equation (21) corresponds to three scalar equations. We can search for a point of the bifurcation surface on the deviatoric plane:

$$T_1 + T_2 + T_3 = \text{const} . \tag{22}$$

and on the ray:

$$\frac{\xi}{\eta} = \text{const} . \tag{23}$$

The system of 7 algebraic equations (17), (20), (21), (22) and (23) makes it finally possible to determine numerically the 7 unknowns $T_1, T_2, T_3, D_1, D_2, D_3, \lambda$. The analytical representation of the bifurcation surface in terms of **L** and **N** is too complex.

Corollary 1: As already noted, the theory of elastoplasticity uses (at least) two constant stiffness matrices, one for unloading or inside the yield surface and one (or more, depending on the direction of stress increment) for loading. A result in elastoplasticity is that the limit states are defined by $\det(\mathbf{L}_{plastic}) = 0$, whereas a bifurcation (*i.e.* a non-unique solution of an element test with mixed boundary conditions) may set on if $\det(\mathbf{L}^s_{plastic}) = 0$, where $\mathbf{L}^s_{plastic}$ is the symmetric part[9] of $\mathbf{L}_{plastic}$. We can easily see, that these results can also be obtained from the above derivations if we set $\mathbf{N} = \mathbf{0}$ and $\mathbf{L} = \mathbf{L}_{plastic}$. From the definition of limit state, $\overset{\circ}{\mathbf{T}} = \mathbf{LD} + \mathbf{N}\,|\mathbf{D}| = \mathbf{0}$ the limit state condition follows immediately, $\det(\mathbf{L}) = 0$ for $\mathbf{N} = \mathbf{0}$. With $\mathbf{N} = \mathbf{0}$ and with $\mathbf{x}^0 \cdot \mathbf{L}\mathbf{x}^0 = \mathbf{x}^0 \cdot \mathbf{L}^s\mathbf{x}^0$ we obtain from (18) the known equation of the bifurcation surface in elastoplasticity, $\det(\mathbf{L}^s) = 0$.

Corollary 2: As can be inferred from Figure 6, negative second order work means negative stiffness. Of course, we should keep in mind that stiffness is

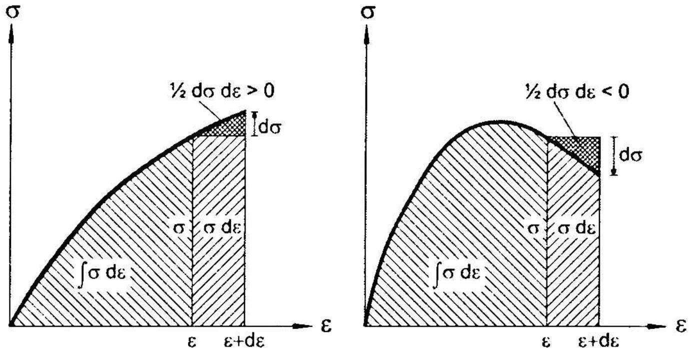

Figure 6. *Positive (left) and negative (right) second order work*

[9] Note that the eigenvalues of **L** and of $\mathbf{L}^s$ are not identical

a fourth order tensor, $d\sigma_{ij}/d\varepsilon_{kl}$, so that the notion "negative stiffness" is virtually meaningless. That means that we obtain negative slope of a graph representing a particular combination of stress components plotted over a particular combination of strain components. If the equation $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D}) = 0$ possesses only one solution, this means that there is only one $\mathbf{D} = \mathbf{D}_1$, for which $\mathrm{tr}(\mathring{\mathbf{T}}(\mathbf{D}_1)\mathbf{D}_1) = 0$. Similarly, for $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D}) < 0$ several tensors **D** can be found are connected with negative stiffness. If the condition $\mathrm{tr}(\mathring{\mathbf{T}}\mathbf{D}) = 0$ is encountered in the ascending part of the stress-strain curve of a conventional triaxial test, this means that there is a deformation modus (*i.e.* a specific **D**) — different than the one correspond to the homogeneous triaxial deformation — connected with vanishing stiffness.

3.3 Softening

It has been often discussed in soil mechanics, whether softening (see Figure 7) is a material property or not. In this context (and in contrast to the theory of elastoplasticity) softening is understood as negative stiffness. Traditionally, softening was considered as a principal part of soil behaviour. Later on it became fashionable to deny softening, as being only an apparent effect due to the inhomogeneous deformation of the sample. The view in hypoplasticity is that a large amount of the registered softening is due to the inhomogeneous sample deformation. However, the "material" softening, *i.e.* the softening which would be exhibited by a fictitious sample undergoing homogeneous deformation, is also there. We have to admit that the onset of inhomogeneous deformation makes the experimental approach infeasible. We can, however, proceed by reasoning: It is a matter of fact that dense samples have a higher strength (*i.e.* peak stress deviator) than loose ones. In the course of deformation, dilatancy transforms a sample from dense to loose. Consequently, its strength must decrease and this is material softening.

3.4 Shear Bands

A typical pattern of inhomogeneous deformation is the localization of deformation within a narrow zone called shear band (see Figures 8 and 9).

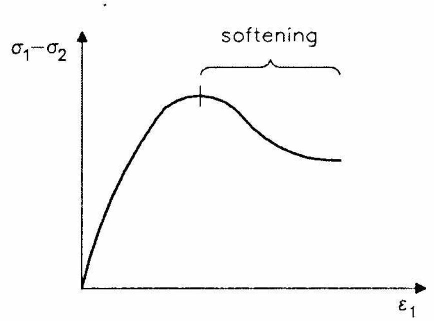

Figure 7. *Softening in triaxial test deformation*

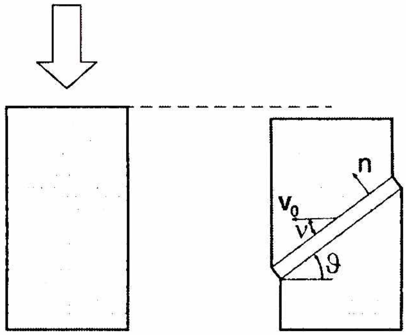

Figure 8. Soil sample before and after shear-banding in plane (biaxial) deformation

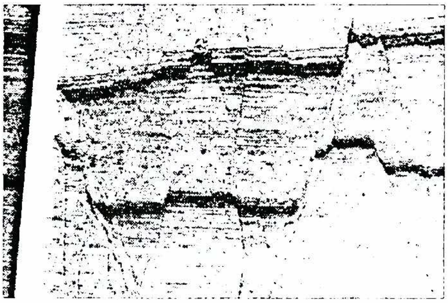

Figure 9. Examples of shear bands in nature

Such shear bands constitute one of the most fascinating phenomena in geomechanics. Due to the work of Desrues based on tomography, we know now that also apparently non-localized inhomogeneous deformation modes are actually localized. The transition to localized deformation may occur either gradually or suddenly. In the latter case it consists in a drastic change of the deformation direction, as experiments by Vardoulakis show. If a constitutive model is capable of realistically describing the material behaviour (*i.e.* the stiffness) also in this new direction, then it will be possible to predict when and under which inclination a shear band can occur. This ability is not self-evident since many constitutive models (*e.g.* the elastoplastic ones) are suggested or tested only for some particular fans of deformation directions. It is therefore a good check of constitutive relations to predict the formation of shear bands. This test has been passed by several versions of the hypoplastic relations.

The shear band divides the initially homogeneous sample into three parts (see Figure 8), the upper and the lower parts separated by a thin shear zone whose thickness is undetermined. It is realistic to assume that the upper and lower parts do not deform after the spontaneous formation of the shear band, *i.e.* they behave as rigid bodies (*i.e.* $\mathbf{D} = \mathbf{0}$, and consequently $\overset{\circ}{\mathbf{T}} = \mathbf{0}$), whereas inside the shear band the motion is described by the velocity gradient $\mathbf{v}_0 \otimes \mathbf{n}$ with $\mathbf{v}_0 := |\mathbf{v}_0|\mathbf{m}$. Since in the rigid parts $\overset{\circ}{\mathbf{T}} = \dot{\mathbf{T}} = \mathbf{0}$ holds true, it follows that the rate of the traction acting up on the discontinuity separating these parts from the shear zone must vanish:

$$\dot{\mathbf{T}}\mathbf{n} = \mathbf{0}\,. \tag{24}$$

Equation (24) is the condition for the spontaneous appearance of a shear band, *i.e.* a shear band can only appear if this equation possesses a solution. Clearly, $\dot{\mathbf{T}} = \overset{\circ}{\mathbf{T}} + \mathbf{WT} - \mathbf{TW}$ depends via the constitutive equation on the motion within the shear zone. This motion is described by $\mathbf{D} = \frac{1}{2}(\mathbf{m} \otimes \mathbf{n} + \mathbf{n} \otimes \mathbf{m})$ and $\mathbf{W} = \frac{1}{2}(\mathbf{m} \otimes \mathbf{n} - \mathbf{n} \otimes \mathbf{m})$ or $D_{ij} = m_{(i}n_{j)} := \frac{1}{2}(m_i n_j + n_i m_j)$, $W_{ij} = m_{[i}n_{j]} := \frac{1}{2}(m_i n_j - n_i m_j)$, where $\mathbf{m}$ and $\mathbf{n}$ are unit vectors. Introducing the hypoplastic constitutive equation (6) into (24) yields:

$$\left(L_{ijkl}m_{(k}n_{l)} + N_{ij}\sqrt{(n_k m_l)(n_k m_l)} + m_{[i}n_{k]}\sigma_{kj} - \sigma_{ik}m_{[k}n_{j]}\right)n_j = 0\,. \tag{25}$$

For plane deformation the unit vectors $\mathbf{m}$ and $\mathbf{n}$ can be expressed by means of the angles ϑ and ν. It turns out that the two scalar equations correspond to (25) for $i = 1, 2$ possess a solution for the unknowns ϑ and ν if a fourth degree polynomial in $\sin\nu$ possesses a solution. This can be easily checked by means of Sturm chains. Thus, hypoplasticity yields realistic and "class A" predictions of stress states with the earliest possible onset of shear bands as w ell as realistic predictions of ϑ and ν. With renamed indices (25) transforms to:

$$A_{ij}m_j = 0 \tag{26}$$

with

$$A_{ij} := n_l\left(L_{iljk} + L_{ikjl}\right)n_k + N_{ij}\left|\mathbf{D}\right| + n_k\sigma_{ki}n_i - n_i\sigma_{jk}n_k - \sigma_{ij} + \sigma_{ik}n_k n_j\,.$$

For the transition to elastoplasticity ($N_{ij} = 0$) the expression in brackets is called the "acoustic tensor" $\mathbf{A}$. Solubility of (26) requires that the determinant of A_{ij} vanishes. Thus, $\det(\mathbf{A}) = 0$ is the criterion for shear band formation in elastoplasticity. The condition $\det(\mathbf{A}) = 0$ is called "loss of strong ellipticity".

The name "acoustic tensor" originates from the theory of propagation of elastic waves in anisotropic elastic media. With u_i being the displacement and $\varepsilon_{ij} = u_{(i,j)}$ the strain, we can obtain from momentum balance $\rho\ddot{u}_i = \sigma_{ik,k}$ and constitutive equation $\sigma_{ik} = L_{iklm}\varepsilon_{lm}$ the wave equation:

$$\rho\ddot{u}_i = L_{iklm}\frac{\partial^2 u_m}{\partial x_k \partial x_l}\,.$$

To obtain solutions of the type $u_i = u_{oi}e^{i(\mathbf{kr}-\omega t)}$, the following condition must be fulfilled:

$$\left|L_{iklm}k_k k_l - \rho\omega^2\delta_{im}\right| = 0\,.$$

3.5 Bifurcation modes for 2D and 3D problems

3.5.1 Formulation with finite elements

Considering hypoplastic or elastoplastic materials, initial-boundary-value-problems can be numerically solved with the method of finite elements. Doing so we consider equilibrium of m nodal points and obtain $n = n_d \cdot m$ equations, where n_d is the number of spatial dimensions (*i.e.* 1 or 2 or 3). Usually we are interested in the time evolution of an equilibrium state as it changes due to *e.g.* an external loading process. We then consider equations that express the continuation of equilibrium and have the form:

$$K_{ij}\dot{x}_j = \dot{y}_i \,, \qquad i,j = 1,\dots n \,. \tag{27}$$

Herein K_{ij} is the incremental (or tangential) stiffness matrix of the considered body, $\dot{x}_j$ are the nodal velocities and $\dot{y}_i$ are the rates of nodal forces due to external volume or surface forces (tractions). Usually the loading process will be controlled by the prescription of the displacement and traction rates for some boundary nodes, whereas the remaining boundary nodes will be free of tractions or displacements. Thus, we re-define n in Equation (27) as the number of *unknown* nodal velocities. The vector $\dot{y}_i$ results from quantities controlling the loading process (even if control is purely kinematical, *i.e.* description of boundary displacement, $\dot{y}_i$ contains non--zero components). The global stiffness matrix K_{ij} is a constant matrix only for linear-elastic materials. For plastic materials, K_{ij} depends on the solution of Equation (27), *i.e.* $K_{ij} = K_{ij}(\dot{\mathbf{x}})$. Therefore, Equation (27) can only be solved iteratively, say by means of the Newton method. The dependency of K_{ij} on $\dot{\mathbf{x}}$ for elastoplastic materials is intricate, since the material stiffness is piecemeal linear and the distinction between loading and unloading is based on a series of criteria. In the analysis of bifurcation solutions of (27) for elastoplastic materials, the dependence of K_{ij} on $\dot{\mathbf{x}}$ is tacitly suppressed and K_{ij} is considered as a constant matrix. In hypoplasticity, the dependence of K_{ij} on $\dot{\mathbf{x}}$ is simpler:

$$K_{ij}\dot{x}_j = L_{ij}\dot{x}_j + N_{ip}\sqrt{a_{pkl}\dot{x}_k\dot{x}_l} \,.$$

The constants L_{ij}, N_{ip}, a_{pkl} depend on the hypoplastic constitutive equation and the discretization operations.

3.5.2 Bifurcation modes

In elastoplasticity the non-linearity of (27) is simply neglected and K_{ij} is considered as a constant matrix. As K_{ij} is non-symmetric, we have to distinguish between left and right eigenvectors. The right eigenvectors $\mathbf{v}$ are solutions to the problem $\mathbf{K}\mathbf{v} = \lambda\mathbf{v}$, and the left eigenvectors $\mathbf{w}$ are solutions to the problem $\mathbf{K}^T\mathbf{w} = \lambda\mathbf{w}$. Both problems have the same eigenvalues λ_i, but different eigenvectors. Eigenvectors belonging to different eigenvalues are orthogonal, *i.e.* $\mathbf{v}_i \cdot \mathbf{w}_j = 0$ if $i \neq j$. If we normalize the eigenvectors, we obtain $\mathbf{v}_i \cdot \mathbf{w}_j = \delta_{ij}$. Thus the vectors $\mathbf{v}_i$

(or $\mathbf{w}_i$) can serve as basis to represent any vector $\mathbf{x}$: $\mathbf{x} = \alpha_i \mathbf{v}_i$. Multiplying this equation with $\mathbf{w}_i$ and using the aforementioned orthogonality we obtain $\alpha_i = \mathbf{x} \cdot \mathbf{w}_i$. Thus we have $\mathbf{x} = (\mathbf{x} \cdot \mathbf{w}_i)\mathbf{v}_i$.

Let now $\dot{\mathbf{x}}_0$ be a solution of $\mathbf{K}\dot{\mathbf{x}} = \dot{\mathbf{y}}$. If this is not the unique solution, there must exist $\dot{\mathbf{x}}_1 \neq \mathbf{x}_0$ such that:

$$\mathbf{K}\dot{\mathbf{x}}_0 = \dot{\mathbf{y}} \,,$$
$$\mathbf{K}\dot{\mathbf{x}}_1 = \dot{\mathbf{y}} \,,$$

hence

$$\mathbf{K}\left(\dot{\mathbf{x}}_1 - \dot{\mathbf{x}}_0\right) = \mathbf{0} \,.$$

It follows that $\mathbf{K}$ must be singular (*i.e.* det($\mathbf{K}$) = 0), which means that at least one of its eigenvalues, say λ_1, must vanish: $\lambda_1 = 0$. If we represent $\dot{\mathbf{x}}$ by means of the right eigen vectors of $\mathbf{K}$ we obtain:

$$\mathbf{K}\left(\mathbf{w}_i \cdot \dot{\mathbf{x}}\right)\mathbf{v}_i = \lambda_i\left(\mathbf{w}_i \cdot \dot{\mathbf{x}}\right)\mathbf{v}_i = \dot{\mathbf{y}}, \qquad i = 2,3,\ldots n,$$

with $\lambda_2 \cdot \lambda_3 \ldots \lambda_n \neq 0$. Hence:

$$\lambda_i\left[\mathbf{w}_i \cdot \left(\dot{\mathbf{x}}_1 - \dot{\mathbf{x}}_0\right)\right]\mathbf{v}_1 = \mathbf{0} \,.$$

It then follows $\mathbf{w}_i \cdot (\dot{\mathbf{x}}_1 - \dot{\mathbf{x}}_0) = 0$ for $i = 2, 3, \ldots n$. This means that $\dot{\mathbf{x}}_1 - \dot{\mathbf{x}}_0$ has the direction of $\mathbf{v}_1$:

$$\dot{\mathbf{x}}_1 - \dot{\mathbf{x}}_0 = \alpha\, \mathbf{v}_1 \,,$$
$$\dot{\mathbf{x}}_1 = \dot{\mathbf{x}}_0 + \alpha\, \mathbf{v}_1 \,. \qquad (28)$$

For any $\alpha = 0$ we obtain with (28) a bifurcated solution. It is reported homogeneous deformation solution $\dot{\mathbf{x}}_0$, *i.e.* $\dot{\mathbf{x}}_1 \cdot \dot{\mathbf{x}}_0 = 0$. This can be obtained with $\alpha = -(\dot{\mathbf{x}}_0 \cdot \mathbf{v}_1)/(\mathbf{v}_1 \cdot \mathbf{v}_1)$. Some numerical methods to find the eigenvectors of $\mathbf{K}$ for zero (or even negative) eigen values are discussed in "eigenvector perturbation".

Another method which circumvents the search for eigenvectors is the so-called material perturbation: the material properties are assumed to scatter over their mean values. Then, the solution of (27) traces automatically the bifurcated solution. This is probably due to the onset of ill-posedness, according to which small perturbations grow exponentially.

For hypoplastic materials (and also for elastoplastic materials, a fact which is often overlooked) the stiffness matrix K_{ij} is not constant, therefore the eigenvector perturbation makes no sense. Material perturbation is, however, still applicable. Another possible procedure is to search the vector $\mathbf{x}$ that minimizes the second order work:

$$f(\mathbf{x}) = L_{ij}\dot{x}_i\dot{x}_j + N_{ip}\sqrt{a_{pkl}\dot{x}_k\dot{x}_l}\,\dot{x}_i \,,$$

and then to check whether this minimum is equal to zero.

References

[1] Bauer E. and Wu W., *A hypoplastic model for granular soils under cyclic loading*, in: D. Kolymbas (ed.), Modern Approaches to Plasticity, 247, Elsevier, 1993

[2] Kolymbas D., *Eine konstitutive Theorie für Böden und andere körnige Stoffe*, Publ. Series of Institut für Bodenmechanik und Felsmechanik der Universität Fridericiana in Karlsruhe, Vol. 109, 1988

[3] Kolymbas D., *Generalized hypoelastic constitutive equation*, in: Saada and Bianchini (eds), Constitutive Equations for Granular Non-Cohesive Soils, 349, Balkema, 1988

[4] Niemunis A., *Hypoplasticity vs. elastoplasticity, selected topics*, in: D. Kolymbas (ed.), Modern Approaches to Plasticity, 278, Elsevier, 1993

[5] Niemunis A. and Herle I., *Hypoplastic model for cohesionless soils with elastic strain range*, Mechanics of Cohesive-Frictional Materials **2**, 279, 1997