NUMERICAL METHODS IN GEOMECHANICS

Antonio Bobet

School of Civil Engineering, Purdue University, West Lafayette, IN, USA

الخلاصة:
تقدم هذه الورقة وصفًا للطرق العددية الأكثر استخداماً في الميكانيكا الجيولوجية. وهي أربعة طرق: (1) طريقة العنصر المميز (2) طريقة تحليل النشوة المتقطع (3) طريقة التحام الجسيمات (4) طريقة الشبكة العصبية الصناعية.
وتضمنت الورقة أيضًا وصفًا موجزاً لتطبيق المبادئ الخوارزمية لكل طريقة إضافةً إلى حالة بسيطة لتوضيح استخدامها.

*Corresponding Author:
E-mail: bobet@purdue.edu

Paper Received November 7, 2009; Paper Revised January 17, 2009; Paper Accepted February 3, 2010
ABSTRACT

The paper presents a description of the numerical methods most used in geomechanics. The following methods are included: (1) The Distinct Element Method; (2) The Discontinuous Deformation Analysis Method; (3) The Bonded Particle Method; and (4) The Artificial Neural Network Method. A brief description of the fundamental algorithms that apply to each method is included, as well as a simple case to illustrate their use.

Key words: numerical methods, geomechanics, continuum, discontinuum, finite difference, finite element, discrete element, discontinuous deformation analysis, bonded particle, artificial neural network
NUMERICAL METHODS IN GEOMECHANICS

1. INTRODUCTION

Analytical methods are very useful in geomechanics because they provide results with very limited effort and highlight the most important variables that determine the solution of a problem. Analytical solutions, however, have often a limited application since they must be used within the range of assumptions made for their development. Such assumptions usually include elastic behavior, homogeneous, isotropic material, time independent behavior, quasi-static loading, etc. Geomaterials such as soils and rock masses display non-linear behavior, either because this is inherent to the material or because it has been externally induced (e.g., past stress history). Rocks and soils may not be isotropic or homogeneous, and the loading may not be static, or the geometry of the problem may be complex. In these cases, solutions can only be obtained numerically.

Numerical methods give only approximations to the correct or exact mathematical solution. This is so because some simplifications are made to solve the system of differential equations either inside the continuum or at the boundaries of the discretization. It has to be mentioned also that the problem that is solved is the conceptualization that is done of the actual physical problem. The conceptualization applies to the geometry of the problem, the loading process or history, and the response of the geomaterials to loading. The better the approximation to the field problem through this conceptualization process, the more accurate the solution will be with respect to the response observed in the field.

Numerical methods have been extensively used in the past several decades due to advances in computing power. In a broad sense, numerical methods can be classified into continuum and discontinuum methods [1,2]. Continuum methods may incorporate the discontinuities in the medium, if present, explicit or implicitly, while in discontinuum methods, discontinuities are incorporated explicitly. The need to use, for a particular problem, continuum or discontinuum methods depends on the size, or scale, of the discontinuities with respect to the size, or scale, of the problem that needs to be solved. There are no quantitative guidelines to determine when one method should be used instead of the other one. Figure 1 (following Brady [3]) provides some qualitative guidance. For example, Figure 1(a) illustrates an opening in a medium without discontinuities; in this case the displacement field is continuous and thus continuum numerical methods are appropriate. Figure 1(b) shows a tunnel excavated in a medium with a small number of discontinuities which divide the medium into a small number of continuous regions. The displacement field will be continuous inside each region but may be discontinuous across the discontinuities. If a continuum model is used, the model should be able to consider the specific discontinuities. The medium depicted in Figure 1(c) is determined by a number of discontinuities with spacing and continuity such that the blocks defined are within the scale of the opening. In this case, displacements may be determined by the slip along the discontinuities and rotation of the blocks. Thus, a discontinuum numerical method seems appropriate. If the medium is heavily jointed such that the blocks defined by the discontinuities have a size much smaller than the opening, e.g., Figure 1(d), a pseudo-continuous displacement field is produced and the use of a continuum model seems reasonable.

There is quite a large number of numerical methods that have been used in the literature to estimate the behavior of geomaterials. The most important, or at least the most used methods are: Continuum, Finite Difference Method (FDM), Finite Element Method (FEM) and Boundary Element Method (BEM); Discontinuum, Distinct Element Method (DEM), Discontinuous Deformation Analysis (DDA), and Bonded Particle Model (BPM). There are two other methods which do not follow this classification: Meshless Methods (MM) and Artificial Neural Networks (ANN). While all methods are relevant, the paper focuses on DEM, DDA, BPM, and ANN, which have recently seen a significant use growth. The following sections provide the fundamental assumptions and the mathematical framework for each method and an overview of the range of problems where each method has been successfully used. A brief description of continuum methods is also included.

2. CONTINUUM METHODS: FINITE DIFFERENCE METHOD, FINITE ELEMENT METHOD, AND BOUNDARY ELEMENT METHOD

The Finite Element Method (FEM), the Finite Difference Method (FDM), and the Boundary Element Method (BEM) are the continuum methods most used in geomechanics [4,5]. In these methods, either the medium and the boundaries (FEM and FDM) or only the boundaries (BEM) are discretized.

The Finite Difference Method (FDM) is based on the premise that governing differential equations can be adequately represented by finite differences. The method is the oldest among the numerical methods in geomechanics and was used even before the arrival of computers. Timoshenko and Goodier [6] attribute the first application of the method to Runge, who in 1908 used it for the solution of torsion problems. With the FDM, the set of differential equations is reduced to a system of linear equations, which can be solved by any of the classical methods. Southwell [7] developed the relaxation method, which provides a fast solution of the system of equations; this promoted a much wider use of the FDM. The method really took off with the advent of computers.
With the method a grid is superimposed to the domain, as shown in Figure 2. In the figure, the sub-indices represent the position of the point in the grid; for example, $i,j$ represents a point with coordinates $(x_i, x_j)$, $i+1,j$ the point $(x_i+h, x_j)$, and so on. The method relies on the approximation of the field equations, i.e. equilibrium, strain compatibility, etc. by finite difference formulas. Discontinuities can be incorporated in the model by using grid points on each side of the discontinuity. The relative displacement between corresponding grid points determines the slip along the discontinuity, and frictional laws (e.g., Coulomb) can be enforced by adding new equations to the system of equations that relate shear stress with normal stress. Normal and shear displacements can also be related to the shear and normal stiffness of the discontinuity.

![Figure 2: Finite Difference Grid in 2D](image-url)
The method can also be readily used to solve dynamic problems, where displacements are a function of position and time. Explicit time integration techniques are often used to provide solutions using small time increment steps. Dynamic problems require a maximum time step to ensure stability of the solution, which is given by

$$\Delta t = \min \left( \frac{\Delta x}{C_p} \right)$$

$$C_p = \sqrt{\frac{K + \frac{4}{3}G}{\rho}}$$

Equation (1) indicates that the maximum time step is controlled by the stiffer material in the medium. It is not unusual to run tens of thousands of steps to complete a numerical analysis. While the number of steps is quite large, the time required to complete each step and the memory required to store the solution is small, and, thus, complex dynamic problems can be analyzed in a reasonable period of time.

The finite difference approach is very well-suited to incorporate non-linear behavior. The solution is then obtained on a stepwise process involving sufficiently small loading increments until the desired final state is reached. At the end of each loading step, displacements at the grid points are obtained; stresses are then updated based on the non-linear behavior of the material, and another small loading increment is added. The new increment starts with the updated stress field from the previous increment. This is a forward scheme that does not require iteration, unlike other techniques such as Finite Element Methods that use implicit solution methods.

The Finite Element Method (FEM) is by far the method used the most for the analysis of continuous or quasi-continuous media. The term “Finite Element”, according to Bathe [8], was first introduced by Clough [9]. The method consists of discretization of the continuum into small elements that intersect at their nodes (Figure 3). The method relies on the assumption that, through appropriately chosen interpolation functions, displacements at any point within the element can be accurately obtained from the displacements of the nodes. The method is based on the principle of virtual displacements, which states that, for a body in equilibrium, any compatible (i.e., satisfies boundary conditions) small virtual displacements applied to the body, the total internal work associated with the virtual displacement field must be equal to the total virtual external work.

With the Boundary Element Method (BEM), only the boundaries of the continuum need to be discretized. See Figure 4. This is in contrast to the other two continuum methods, the Finite Difference and the Finite Element methods, where the entire medium has to be discretized. Also, if the medium extends to infinity, which is common in problems in geomechanics, no artificial boundaries such as those needed in FDM and FEM are required. The BEM automatically satisfies far-field conditions. In the BEM, the solution is approximated at the boundaries while equilibrium and compatibility are exactly satisfied in the interior of the medium. In FDM and FEM, the approximations are made inside the medium. The advantage of limiting the discretization to the boundaries is that the problem is reduced by one order: from 3D to a 2D surface problem at the boundary, and from 2D to a line problem. Thus the method is very attractive for those problems where the volume to boundary surface ratio is large.
The technique used in BEM consists in essence of transforming the governing differential equations, which apply to the entire medium, to integral equations which only consider boundary values [10–12]. In a boundary value problem, some parameters such as stresses and displacements are known while others are not, which then are part of the solution. There are two approaches to solve for the unknown parameters. In the first approach (Direct BEM), the unknowns are solved directly, and once they are obtained, stresses and displacements at any point in the continuum can be obtained directly from the solution. In the second approach (Indirect BEM), the solution is found in terms of some “fictitious” quantities, typically stresses or displacements. The fictitious quantities are obtained first and the stresses and displacements at any point in the medium are expressed in terms of these fictitious quantities.

Boundary Element Methods are particularly well-suited to address static continuum problems with small boundary to volume ratios, with elastic behavior, and with stresses or displacements applied to the boundaries. Actual problems may not always conform to these limitations. For example, rocks and soil deposits may undergo significant yielding under moderate stresses, gravity forces may be significant for shallow geostuctures, and inertia may play an important role with dynamic loading (e.g., blasting, earthquake). Dynamic and body forces require integration over the entire volume domain which leads to the need for discretization of the entire continuum. The plasticity algorithms require integration at least over the volume of the material that undergoes yielding and convergence of the solution, as with FEM, is attained through iteration. With plastic deformations and with cases where integration needs to be extended over part or the entire volume, the advantage that the BEM offers regarding limited discretization of the continuum may be lost. Efficient hybrid BEM-FEM solutions are possible, where a FEM discretization is used for those parts of the continuum where plastic deformations occur, while Boundary Elements are used in elastic regions. The advantage of the coupled FEM-BEM is reduced discretization and automatic satisfaction of boundary conditions at infinity. The challenge of the hybrid approach is the generation of nodal forces and displacements from the BEM that are consistent with those of FEM, and that the resulting stiffness matrix is non-symmetric (in contrast with FEM where the stiffness matrix is generally symmetric). Figure 5 shows an example of a hybrid discretization of a tunnel, where the tunnel liner and a volume of the ground next to the tunnel where plastic deformations occur, are discretized with Finite Elements. Far from the tunnel and where the deformations are elastic, Boundary Elements are used.

3. DISCONTINUUM: THE DISTINCT ELEMENT METHOD

The Distinct Element Method (DEM) was introduced by Cundall [13] as a model to simulate large movements in blocky rock masses, and then used for soils which were modeled as discs [14]. Later on, the method has been applied to spherical and polyhedral blocks [4,15–19] for both soils and rocks.

The DEM belongs to the family of Discrete Element Methods, which Cundall and Hart [18] define as those that: (1) allow finite displacements and rotations of discrete bodies, including detachment; and (2) automatically recognize new contacts between bodies during calculations. Discrete Element Methods need to address three key issues: (1) representation of contacts; (2) representation of solid material; and (3) detection and revision of contacts during execution. An in-depth discussion of these issues is provided by Cundall and Hart [18].
In the DEM, it is assumed that the medium is divided by fully persistent discontinuities which delimit through their intersections a finite number of blocks, which in turn are interconnected through the discontinuities. Figure 6 provides an idealization of a discretization with DEM of a medium with two sets of discontinuities. The following provides key concepts for the formulation of the DEM. For clarity, the discussion is restricted to two-dimensional discretizations with rigid bodies.

\[ m \ddot{u}_i + c \dot{u}_i = F_i \]
\[ \dot{u}_i + \omega_i = M_i \]

(2)
where \( t \) is time, \( m \) is the mass, \( I \) is the moment of inertia of the element, \( u_i \) is the displacement of the gravity center of the element in the direction \( i \), \( \dot{u}_i \) and \( \ddot{u}_i \) are the acceleration and the velocity of the gravity center, \( \omega \) and \( \dot{\omega} \) are the angular rotation and angular velocity of the element, \( c \) is the viscous damping, and \( F_i \) and \( M \) are the resultant force and moment applied at the center of gravity. In the DEM Equation (2) is solved in the time domain using an explicit finite difference method. Using the central finite difference, approximation velocities and displacements are given by

\[
\begin{align*}
\ddot{u}_i^{t+\Delta t/2} &= \left[ D_1 \ddot{u}_i^{t-\Delta t/2} + \frac{F_i}{m} \right] / \Delta t \\
\dot{\omega}^{t+\Delta t/2} &= \left[ D_2 \dot{\omega}^{t-\Delta t/2} + \frac{M}{m} \right] / \Delta t \\
D_1 &= 1 - \frac{c \Delta t}{m} \\
D_2 &= \frac{1}{1 + \frac{c \Delta t}{m}}
\end{align*}
\]  
(3)

The forces acting at the boundaries are originated by the interaction of the element with the surrounding elements. At each boundary, a normal and a shear force appear as the result of the relative movements between the two elements that share the discontinuity. The forces at the interface may be obtained using a penalty method where the magnitude of the forces is related to the relative movements between the two elements and the stiffness of the discontinuity. Figure 7(a) shows the positive forces at the top of the element, and Figure 7(b) shows an idealization of the contact between blocks. The normal force is proportional to the relative movement of the two blocks across the contact and along the normal direction. The shear force is proportional to the relative movement along the direction of the contact. Expressions for the forces are

\[
\begin{align*}
F_n^{t+\Delta t} &= F_n^{t} - K_n \Delta u_n^{t+\Delta t} A_C - \beta K_n \Delta u_n^{t+\Delta t} A_C \\
F_s^{t+\Delta t} &= F_s^{t} - K_s \Delta u_s^{t+\Delta t} A_C - \beta K_s \Delta u_s^{t+\Delta t} A_C
\end{align*}
\]  
(4)

\( K_n \) and \( K_s \) are the normal and shear stiffness of the contact (subscripts \( n \) and \( s \) refer to the directions normal and parallel to the discontinuity, respectively); \( \Delta u_n \) and \( \Delta u_s \) are relative displacements between the two elements, and \( A_C \) is the contact area. A damping factor, the third term on the right-hand side of the equation, is normally included to attenuate or prevent “rattling” of the contact between blocks. Damping (\( C_n \) and \( C_s \) in Figure 7) is often expressed as proportional to the normal and shear stiffness (\( \beta K_n \) and \( \beta K_s \) in (4)), but other expressions for damping have been proposed (e.g., damping proportional to the rate of change of the kinetic energy of the element [15]).

Figure 7. Forces at the Boundary of DEM Elements
The magnitude of the shear force is limited by the constitutive relation used for the contact surface. For a Coulomb-type friction law,

\[ F_s^{t+\Delta t} \leq c A + F_n^{t+\Delta t} \tan \phi \]  

(5)

where \( c \) and \( \phi \) are the cohesion and friction angles of the contact surface. If the shear force obtained from (4) is larger that that from (5), it is reduced to the limiting magnitude given by (5).

The calculations are performed from one state, where the solution is fully known, to another state in small time increments. The procedure is as follows [17]: The law of motion is applied through Equation (3) with current forces to update the position of each element. As a result, the relative displacements and velocities at the contacts between elements are obtained. From the relative displacements, contact forces are updated using Equation (4) and new resultant forces and moments at the center of gravity of each element are computed. The cycle is repeated with small increments until the final solution is obtained. In the formulation, time can represent actual time when performing a dynamic analysis, or a fictitious parameter to represent loading increment from one loading stage to the next.

As with the Finite Difference Method, numerical stability requires a time increment smaller than the critical time step, which is given by [17]:

\[ \Delta t_{\text{crit}} = \kappa \frac{m_{\text{min}}}{2K_{\text{max}}} \]  

(6)

where \( m_{\text{min}} \) is the smallest element mass, \( K_{\text{max}} \) is the largest normal or shear stiffness in the discretization, and \( \kappa \) is a factor that takes into account the fact that an element may be in contact with more than one element. A value for \( \kappa \) equal to 0.1 has been suggested [17].

Typical runs are completed with thousands of cycles involving very small time increments. The solution of equations (3) and (4) is a forward process, and, thus, the computer time required in each cycle is very small; also, the storage information needed for each element is small. Therefore, the process discussed so far does not require intensive computation power or large storage capabilities. Where such requirements become significant is for the algorithm to recognize and keep track of all the contacts between elements during execution. A very simple procedure would be to compare the position of each element with the rest of the elements at the end of each cycle. For a discretization with \( n \) elements, this would require of the order of \( n^2 \) operations in each cycle, which would make the entire method impractical. Considerable effort has been done to develop efficient algorithms, which on the one hand need to accurately describe the interaction between elements, and on the other hand are not computationally intensive. The problem is complex as the algorithms need to identify not only what elements are in contact but also the type of contact: corner to corner, corner to edge, or edge to edge, since the magnitude and direction of the contact forces depend on the type of contact. A number of approaches has been proposed to identify contacts, such as global searching algorithms, buffer zone definition, contact or field zone, binary tree structures, space decomposition and alternating digital tree [16,20–23]. A comprehensive review of these methods is provided in [24].

The Distinct Element Method is nowadays a very versatile and extensively validated procedure. It has been developed for full three-dimensional problems, and by discretizing the elements with Finite Difference or Finite Element meshes, can be applied to deformable bodies [4,18,25] and to fragmentation of discontinua [20, 25]. It can be used for static and for dynamic calculations [26,27]. Heuze and Morris [28] provide an extensive overview of the DEM as applied to jointed rock masses. One fundamental advantage of the DEM is that pre-existing joints in rock can be incorporated into a DEM model directly, and the joints are allowed to undergo large deformations. Detailed joint constitutive models (see [29] for a review) can also be used to combine experimentally observed fracture properties (such as joint dilation, friction angle, and cohesion) with the DEM approach.

Figure 8(a) shows the discretization used to investigate the response of a tunnel in a discontinuous rock mass subjected to blast loading [30,31]. Figures 8(b) and (c) show the response of the tunnel immediately after detonation and 30 ms later. The simulations were run using parallel processing and the Livermore Distinct Element Code (LDEC), and consisted of 8 million blocks with approximately 100 million contacts, with typical block size of 30 cm, making these the largest simulations of this type performed to date.
4. DISCONTINUUM: DISCONTINUOUS DEFORMATION ANALYSIS

The Discontinuous Deformation Analysis (DDA) is a Discrete Element Method following the definition by Cundall and Hart [18], as outlined in the preceding section. The method started with the work of Shi and Goodman [32,33], and since then it has received considerable attention by the geoengineering community.

The method is fully described in [34,35]. In essence, the medium is discretized into elements or blocks which are in contact with each other only through the discontinuities. The discretization used in Figure 6 to illustrate the DEM could perfectly apply to the DDA. There are fundamental differences between the DEM and DDA. In the DEM each block is treated separately, while in the DDA, the total potential energy of the system is minimized to find the solution. In the DEM, stresses and forces are unknowns while displacements are computed from stresses; in DDA the displacements are the unknowns. In the DEM, the contacts are resolved using a penalty method which results in the definition of the contact forces, while in the DDA, interpenetration of blocks is prevented by adding springs to the contacts. The DEM uses an explicit procedure to solve the equilibrium equations and the DDA is an implicit method. While the DDA is a fully discontinuous analysis, it resembles and follows the procedures developed for FEM.

The DDA, similar to the DEM, needs to address three key issues: (1) representation of contacts; (2) representation of solid material; and (3) detection and revision of contacts during execution. The elements can be convex or non-convex, and their shapes are determined by the location of their contacts with the neighboring elements. Thus, blocks are represented by polyhedra, with the contacts between blocks consisting of edge to face, edge to edge, or face to face.

It is assumed that any large displacements or deformations are the result of the accumulation of small displacements and deformations after a sufficiently large number of steps. Within each step, the displacements of any block are small and, thus, they can be given, in 2-D, by a first order approximation of the form

\[ u = u_0 + (x-x_0) a_1 + (y-y_0) a_2 \]
\[ v = v_0 + (x-x_0) b_1 + (y-y_0) b_2 \]

where \( u \) and \( v \) are the x- and y-axis displacements of a point with coordinates \( x \) and \( y \); \( u_0 \) and \( v_0 \) are the rigid body motions at point \( x_0 \), \( y_0 \), and \( a_i \) and \( b_i \), \( i=1,2 \) are constants. Strains can be computed from (7). In turn displacements can be expressed as a function of strains as follows:
\[
\begin{align*}
    u &= u_0 + (x-x_0) \varepsilon_{xx} + (y-y_0) \left( \frac{1}{2} \varepsilon_{xy} - \gamma_0 \right) \\
    v &= v_0 + (y-y_0) \varepsilon_{yy} + (x-x_0) \left( \frac{1}{2} \varepsilon_{xy} + \gamma_0 \right)
\end{align*}
\]

where \( \varepsilon_{xx}, \varepsilon_{yy} \) and \( \gamma_{xy} \) are the axial strains and the shear strains in the x and y axis, respectively, and \( r_0 \) is the rigid block rotation, in radians, about point \( x_0, y_0 \). Equations (8) are expressed in matrix notation

\[
U = T D
\]

where \( U = (u, v) \), \( D^T = (u_0, v_0, r_0, \varepsilon_{xx}, \varepsilon_{yy}, \gamma_{xy}) \) and \( T \) are the appropriate coefficients from (8). The matrix \( D \) represents the unknowns for each element; thus, there are a total of 6 degrees of freedom or unknowns. Note that strains in each element are constant. For a system of \( N \) elements or blocks, the total number of unknowns is \( 6N \). Minimization of the potential energy of the system of blocks, following FEM convention, is expressed as

\[
D = \sum F_i K_{ij} D_j
\]

\( D_j \) is made of 6x1 sub-matrices that contain the 6 unknowns of each element; \( K_{ij} \) is composed of 6x6 stiffness sub-matrices associated with the corresponding degrees of freedom of element \( j \), and \( F_i \) is a set of 6x1 force sub-matrices of element \( i \). \( K_{ij} \) depends on the material properties of element \( i \) and \( K_{ij} \) (i.e.,) on the contacts between elements. The sub-matrices \( K_{ij} \) are obtained by minimizing the potential energy associated with strain energy, initial stresses, concentrated and distributed loads, body forces, inertia forces, viscosity, displacement constraints at the element contacts, etc. Full derivation of the equations are provided in [35].

In the DDA, no tension and no penetration between blocks are allowed. The kinematics of the block system are incorporated into the equations of equilibrium (10) by adding very stiff springs between appropriate elements to lock the movement in the corresponding direction. Tension between two elements can be modeled by applying a lock in the direction where tension is permitted; once the lock is removed (i.e., a critical tensile threshold is reached) the elements can separate. Hence, by adding or removing locks along appropriate directions, movements between blocks can be avoided, thus preventing penetration. Within a certain loading step (load increment), an iteration process is applied where locks are added or removed as appropriate until all kinematic constrains (e.g., no penetration) are satisfied. To impose the kinematics of the problem requires addressing two issues: (1) determine contacts between blocks, and (2) add to the global equilibrium equations the appropriate stiff springs.

The contact identification process starts after definition of the elements where some threshold distance is established such that only elements within the threshold distance are checked for contact. As the simulation proceeds, potential contacts between elements are updated. If within a single step, the relative displacement between two elements is smaller than their initial distance, no contact check is performed. If interpenetration between two elements is detected, then stiff springs are placed between the two elements and the system is recalculated.

The procedure of solving the equilibrium equations, determining interpenetration, and adding stiff springs is repeated until no interpenetration occurs. At the end of each iteration, the spring force is calculated. If the component of the force normal to the contact is tensile, the normal spring is removed. If the component of the force parallel to the contact is larger than the maximum allowed by the constitutive model (e.g., \( F \), \( \mu F_n \), Coulomb), a spring normal to the contact is placed to allow for sliding and prevent penetration in the normal direction; if smaller than the maximum allowed, springs both in the normal and parallel directions are placed to prevent any relative movement at the contact.

The method, which was originally developed for 2D problems [32,33,35], has been expanded to 3D [36,37]. The limitation that the original DDA had that the blocks could not break has been overcome by new developments in modeling, where blocks are divided into sub-blocks when tensile or shear stresses reach the strength of the material; thus, the DDA has been extended to fragmentation and fracture propagation problems [38,39]. Validation of the Displacement Discontinuity Analysis has been done extensively by comparing predictions from the method with analytical solutions, with other numerical methods, with laboratory and field measurements (e.g., [40–45; an extensive review can be found in [46]).

Figure 9 illustrates an example application of the DDA method [47]. In the figure, a shallow rectangular tunnel in a rock mass medium with two joint sets is subjected to a vertical load on the surface. The figure shows the different stages of the failure, from initial conditions, Figure 9(a), to final failure, Figure 9(f).
5. DISCONTINUUM: BONDED PARTICLE METHOD

The Bonded Particle Method [19] originates from the application of the DEM to a discontinuous medium modeled as discs in two dimensions or spheres in three dimensions. The key idea of the method is that the geomaterial can be approximated by an agglomerate of cemented grains; see Figure 10(a). The grains or particles are assumed rigid with circular or spherical shape with a non-uniform distribution. The particles interact with each other through their contacts such that deformation is produced at the particle contacts or by relative displacements between particles; see Figure 10(b). Tensile and shear cracks between particles occur when the tensile or shear strength of the contact is reached.

As with the DEM, Newton’s second law of motion is solved through a central finite difference algorithm to determine the displacements and velocities of each particle due to the forces acting on the particle. The forces arise from the weight of the particle and from the contact forces between particles. Equations (3) and (4) are used to determine the motions of any particle. The solution of a problem with static or dynamic loading is done incrementally with very small time steps (for static loading, time is an auxiliary variable related to the load increment during each step). The procedure follows that of the Distinct Element Method, displacements and velocities of each grain are computed using Equation (3) with the magnitude of the forces equal to those at the end of the previous step. From absolute displacements, the relative motions between particles in contact is obtained, which in turn are used to determine the magnitude of the forces and moments acting between particles. The updated loads
are then used to compute motions for the next time increment. The process is repeated until the complete solution of the problem is obtained. During the process, contact between particles is reviewed and updated as new contacts may be formed or old ones are destroyed, as bonds between particles break. Inter-particle forces and moments are obtained based on the relative motions between particles and on the properties of the particles and bond. The magnitude of the forces and moments, Figure 10(b), is given by

$$
F_i = F_i^n \bar{n} + F_i^s \bar{s}
$$

$$
\vec{M}_i = \vec{M}_i^n \bar{n} + \vec{M}_i^s \bar{s}
$$

where $F_i$ is the inter-particle force between particle A and particle B (Figure 10(b)), with components $F_i^n$ and $F_i^s$ in the directions normal and parallel, respectively, to the contact between the two particles; $\vec{F}_i$ and $\vec{M}_i$ are the force and moment carried by the bond between the two particles. The magnitude of the loads is given by [19]:

$$
\Delta F_i^n = \frac{k_A^A k_B^B}{k_A^n + k_B^n} \Delta U^n
$$

$$
\Delta F_i^s = \frac{k_A^B k_B^A}{k_A^s + k_B^s} \Delta U^s
$$

$$
\Delta F_i^n = \bar{k}_n A \Delta U^n
$$

$$
\Delta F_i^s = \bar{k}_s A \Delta U^s
$$

$$
\Delta M^n = \bar{k}_n I \Delta \theta^n
$$

$$
\Delta M^s = \bar{k}_s J \Delta \theta^s
$$

$k_A^A, k_A^n, k_B^B$ and $k_B^s$ are the normal and shear stiffnesses of particles A and B, and $\bar{k}_n$ and $\bar{k}_s$ are the normal and shear stiffness of the bond between particles; $\Delta U^n$ and $\Delta U^s$ are the incremental normal and shear displacements between particles, and $\Delta \theta^n$ and $\Delta \theta^s$ are the incremental rotational angles also in the normal and shear directions; A, I and J are the area, moment of inertia, and polar moment of inertia of the bond between the two particles, and are given by:
$A = \begin{cases} \frac{2 R}{\pi R^2} & \text{in 2D} \\ \frac{2}{3} \frac{R^3}{\pi R^2} & \text{in 3D} \end{cases}$

$J = \begin{cases} \frac{1}{4} \frac{\pi R^4}{\pi R^2} & \text{in 3D} \\ \text{n/a} & \text{in 2D} \end{cases}$

$J = \frac{1}{2} \frac{\pi R^4}{\pi R^2}$

$R$ is the bond radius, and $R_A$ and $R_B$ are the radius of particles A and B, respectively, as shown in Figure 10(b).

The maximum tensile and shear stresses acting on the bond are calculated as

$$\sigma_{\max} = \frac{p_i}{A} \frac{R_i^b}{R} \frac{\bar{F}_i}{J}$$

$$\tau_{\max} = \frac{p_i}{A} \frac{K_i^b}{J}$$

When the maximum tensile or shear stress reaches the tensile strength of the bond, $\sigma_{\max}$, or shear strength, $\tau_{\max}$, the bond breaks and it is removed from the model.

The shear force $F_i$ in (12) is limited by the constitutive law used for inter-particle friction (e.g., Coulomb with $F_i \leq \mu F_i$; $\mu$ is the coefficient of friction between particles). If the relative displacement between two particles is negative, there is a gap between the two particles and the normal and shear forces are set to zero; if it is positive, the two particles overlap and, thus, there are normal and shear forces between the particles.

Thus, the following microproperties are needed for the model: $k_n$, $k_s$, and $\mu$, which are associated with the grains, and $\bar{R}$, $\bar{R}_N$, $\bar{K}_S$, $\bar{\sigma}_C$ and $\bar{\tau}_C$, which depend on the bond.

Even though the Bonded Particle Model is relatively new, it has been already used for a wide range of applications within geotechnical engineering. The model has been applied to investigate the strength of soils and rock materials [19,48–50], slope stability [51], damage to rock mass during tunnel excavation and tunnel support [19,52–56], fracture mechanics [19,57], blasting and dynamic analysis [58–60], and the behavior of granular materials and powders [61–63].

The list of applications of the method is not exhaustive, and it is intended to provide a measure of the wide range of fields where the method is used. The method has been the focus of recent conferences where a large number of cases and applications has been presented, even in fields beyond civil engineering, e.g. [64].

Figure 11 illustrates the use of the model [53] to determine the damage zone around a circular opening, in the form of tensile and shear cracks. The model reproduces the experimental results conducted on Berea sandstone where an opening of 14 mm diameter was placed into a prismatic block which was loaded in plane strain with 7.5 MPa confinement. A uniform particle size distribution was used to model the rock, with average particle size 0.2 mm, similar to the actual size of the Berea sandstone grains. Figure 11(b) shows the final stage of failure of the opening with significant cracks and notches between grains.

6. OTHER METHODS: ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks (ANN) are based on a paradigm completely different than the other numerical methods visited. The methods discussed so far all reach a solution addressing the mechanics of the problem where equilibrium, constitutive model, strain compatibility, and boundary conditions are rigorously satisfied. What distinguishes one method from the other is how mathematically this is accomplished. ANNs are based on biological models such as the human brain and rely on information processing techniques based on establishing associations between parameters. As with the human brain, ANNs are composed of a number of interconnected units called neurons. Each neuron receives information, processes the information, and sends the results to other neurons. The characteristics of ANNs are that the information is stored over the entire network, are massively parallel processing systems, are fault-tolerant and can reach a solution with ill-defined or imprecise information, and can learn and adapt. The disadvantages are that ANN systems operate as “black boxes” in that there is no possibility of assessing how they work internally, their design guidelines and operation are somewhat arbitrary, training may be difficult or impossible, and their performance may not be easily predicted [65]. They may be perceived as highly sophisticated
curve fitting techniques but they have proven to provide reasonable solutions to imprecisely formulated problems or to phenomena only described through observations [1].

Figure 11. Example of BMP Model.Circular tunnel subjected to biaxial compression with $\sigma_3 = 7.5$ MPa. From Fakhimi et al. (2002).

The first generation of artificial Neural Networks started in the 1940s to 1960s, but it was not until the 1980s when the introduction of new architectures and learning processes made ANNs useful and practical tools. There are several types of Artificial Neural Networks depending on the characteristics of each neuron, the learning or training scheme, network topology, and network function, e.g., [65,66]. The Feedforward Network is still the preferred type in geotechnology, and is based on a series of two or more layers of neurons (Figure 12(a)). The first layer receives the input applied to the network and the last layer contains the output. The units or neurons in each layer are forward connected only to the units or neurons in the next layer. There is no connection between neurons in the same layer. Thus, ANNs are connected to the exterior by the input and output layers only. The layers between the input and output are called hidden layers. As shown in Figure 12(a), the input consists of $n$ units, each corresponding to an input parameter, and $m$ output units, each corresponding to a requested result parameter. There can be any number of hidden layers and each layer can have any number of units.

The information stored in each neuron, often called the state of the neuron, is passed forward to its connected neuron in the next layer and modified by a connection weight and a bias or threshold value. The resulting value is further modified in the receiving neuron by a function called the activation or transformation function (Figure 12(b)). For example, neuron $j$ in layer $L_k$ receives input from the neurons in layer $L_{k-1}$. If the state of neuron $j$ is denoted by $i_j$,

$$i_j = f\left(\sum_{h \in L_{k-1}} \left(w_{jh} i_h + \theta_j\right)\right) + \theta_j,$$  

where $f$ is the activation function; $w_{jh}$ is the weight associated with the connection between neuron $h$ in layer $L_{k-1}$ and neuron $j$ in layer $L_k$ (note that $w_{jh}$ does not exist since there is no connection back from neuron $h$ to neuron $j$); $\theta_j$ is the bias or threshold value associated with neuron $j$; and $o_j$ is the argument of the function.

The process in the network works as follows: an array of input values is defined as the state of the neurons in the input layer. These values are transmitted to the second, hidden layer, following the protocol defined in Equation (15); the state of the neurons in the second layer is transmitted to the third layer where new calculations are performed to obtain the state of the neurons in this layer. The process is repeated until the output layer is reached. The state of the neurons in the output layer constitutes the output of the system. The weights and biases are not known, which requires training of the ANN; the activation function, however, is defined within the code. Several
functions are possible (e.g., linear, multiplicative, etc.). The function used the most is the sigmoidal function, which has an expression

\[ f(o_j) = \frac{1}{1+e^{-o_j}} \]  

(16)

The sigmoidal function is part of the family of squashing functions which constrain the output to values in the range 0 to 1. It is a continuous function and its first derivative exists, which is necessary for the training of the ANN.

Training of the network (i.e., to obtain the values of weights and biases) is done by comparing the output provided by the ANN with actual results, \( t_m \), associated with a given input. The strategy normally used is to minimize the difference between actual and predicted results using the error norm

\[ E = \sum_{m \in L_N} [t_m - f(o_m)]^2 \]  

(17)
There are different strategies to minimize the error \( E \) in (17) by changing the values of the weights and the biases. The most common strategy is the backpropagation algorithm or delta rule \([65,66]\), where the derivatives of the error function \( E \) with respect to the weights or biases are set to zero; i.e., the error norm in (17) is minimized.

There are no rules to design ANNs. The input and output neurons, in terms of numbers and characteristics, are defined by the user. Thus, the user needs to decide what are the variables that may affect the results and what are the results needed. The number of hidden layers and the number of neurons per layer is problem-dependent. Increasing the number of neurons and/or hidden layers does not necessarily result in better predictions. In fact, overfitting the ANN is a real danger which may induce erroneous results. The strategy often followed consists of dividing the available data in two sets: one for training and the other for validation. A number of strategies can be tested with a different number of hidden layers, different number of neurons per hidden layer, and number of passes (epochs) for training. Each trained ANN is then tested against the validation data, selecting the ANN with the smallest differences \([67]\). There is no guarantee, however, that the process described will result in at least one of the ANNs providing satisfactory results. Once the ANN is trained and selected, it can be used for predictive purposes. It is very important to realize that the ANN should not be used to make predictions outside the range of cases within which it has been trained.

Despite the shortcomings of the ANNs, they have been successful in giving accurate predictions to problems that cannot be solved following the mechanics approach because some of the inputs or conditions needed are not well defined or the input data may be not completely reliable. ANNs are being used in many fields of geoengineering. For example, ANNs have been applied to obtain soil and rock properties \([68–73]\) including soil liquefaction \([74,75]\), slope stability \([76,77]\), deep excavation deformations \([78]\), mining and tunneling support \([79–83]\), and tunneling \([67,84–86]\). ANNs have also been coupled with FEM, where the Finite Element Method is used to solve the mechanics of the problem or to produce the data for training the ANN, or the ANN is used to obtain input parameters for the FEM from back-calculation or to make predictions based on input data from the FEM \([87]\).

7. DISCUSSION

Numerical methods are tools that the engineer has to evaluate qualitatively and quantitatively the effects of geology on the design and the consequences of the design on geology. The methods can be used both in a forward analysis where, given geometry and properties, results are obtained (e.g., stresses, displacements), or on a backward analysis where, given results or measurements, ground properties or ground behavior are approximated.

In any analysis, the following needs to be determined: geometry of the problem, including the geologic geometry in terms of layers, depth, extent, etc.; appropriate boundary conditions; actual material behavior such as elastic, plastic, visco-elastic, etc.; and construction process. Without exception, all the details and complexities of the problem cannot be introduced into the numerical model. This is so because in many cases the geology and material behavior are not fully known, the actual construction process cannot be predicted, or the numerical model is necessarily applied to a limited volume of the entire domain. In any case, assumptions and decisions need to be made. The goal is to create a model that is simple enough such that it can be implemented and interpreted within a reasonable amount of time, and yet it is accurate enough that the results sufficiently approximate the performance of the design.

All numerical models visited in this chapter are capable of providing reasonable results when sound engineering judgment is employed with their use. A word of caution needs to be added for Artificial Neural Networks since their use should be confined within the range of the database employed for their training.

The largest portion of time spent in modeling is during pre-processing or discretization and post-processing or results analysis. It is perhaps for this reason that the most used numerical methods in practice are those that include user-friendly pre- and post-processing capabilities. These are almost exclusively commercial codes. The following is a list of the codes most referenced in the literature: Finite Difference Method: FLAC and FLAC3D (ITASCA Consulting Group, Inc.); Finite Element Method: ABAQUS (Hibbit, Karlson and Sorensen, Inc.), PENTAGON-2D and -3D (Emerald Soft), PHASE2 (Rockscience), PLAXIS (Plaxis BV); Boundary Element Method: BEFE (coupled BEM-FEM, Computer Software and Services (CSS)), EXAMINE2D and EXAMINE3D (Rockscience); Distinct Element Method: EDEM (DEM Solutions), UDEC, 3DEC (ITASCA Consulting Group, Inc.); and Bonded Particle Method: PFC2D and PFC3D (ITASCA Consulting Group, Inc.). All codes are based on the principles of mechanics and they rigorously solve (in the context of numerical solutions) equilibrium equations, boundary conditions, strain compatibility, and the constitutive material model. The choice between one code or another, within the realm of continuum or discontinuum, is often based on personal or company preferences. All codes have a very steep learning curve and it may take significant time and effort for a company to train engineers in any one particular code. Thus there is a tendency to keep the expertise within a very reduced number of numerical codes. The codes listed can be divided into Continuum (FLAC, ABAQUS, PENTAGON, PHASE, PLAXIS, EXAMINE, BEFE) and Discontinuum (EDEM, UDEC, 3DEC, PFC).
For soils it is often assumed that a continuum approach is appropriate. For rocks, however, there are no guidelines to decide when a continuum or a discontinuum model should be used. If very few discontinuities are present in the medium, a continuum model can still be efficient; with a large number of discontinuities (e.g., the size of the blocks determined by the discontinuities is much smaller than the characteristic size of the geostructure) a pseudo-continuum model can still be applied. Otherwise, a discontinuum model seems more reasonable. This issue is still under debate; on the one hand, there is a large experience-based on continuum models successfully used in rock masses, but on the other hand there is mounting evidence that in discontinuous media the stress field obtained with a continuous model does not compare well with the stress jumps across discontinuities predicted by discontinuous models [15,88].

REFERENCES


