Thirty Years of Developments in Contact Modelling of Non-spherical Particles in DEM: A Selective Review

Y. T. Feng*
Zienkiewicz Centre for Computational Engineering
Faculty of Science and Engineering, Swansea University, UK

Abstract

The discrete element method has advanced significantly since it originated in 1979. This paper aims to provide a selective overview of some major developments on contact modelling methodologies for non-spherical particles in DEM over the last three decades. More attention is specifically paid towards developments in the early years. This review mainly focuses on the geometric aspects of contact modelling methods without touching upon material related physical issues. Various shape representation schemes for non-spherical particles are first presented using a classification system, followed by the critical review of contact modelling approaches for almost all contact types of non-spherical shapes. In addition to outlining key ideas, concepts and novel computational procedures that are proposed in these methodologies in a systematic and logically constructed manner, their evolutions and inter-relations in particular are also discussed in detail. Moreover, possible problems and unresolved issues for each method reviewed are highlighted, and possible solutions are pointed out where applicable.

KEYWORDS: Non-spherical shapes, Shape representation, Contact modelling methodology, Energy-conserving contact theory

1 General Introduction

The *discrete element method* (DEM) originated from a series of pioneering works by Cundall on rock mechanics in the 1970s in the form of scientific reports [1, 2, 3] and the first computer program BALL [4]. The method is now generally associated with the publication of the seminal paper by Cundall and Stack [5] in 1979, which details the computational framework of DEM that has been followed and used without much alteration until today. Historically, the method was initially termed the *distinct element method* in [5], while the discrete element method was specially reserved for problems where elements are discretised from finite element approaches and are often associated with applications in rock or geo-mechanics. However, mainly because the difference between these two terms is very subtle, and their abbreviations are the same, the use of the distinct element method gradually faded away in the 1990s. Now, the discrete element method has been universally accepted as the official name of the method.

Another version of DEM also exists and was mainly promoted by a research group in Swansea, UK in the 1990s, where discrete elements are defined as solids that are physically separated but deformable. The deformation of each discrete object can be accounted for, by finite

*e-mail: y.feng@swansea.ac.uk

elements for instance, while the contact between discrete objects mainly follows the procedure in DEM. This approach has been developed into the so-called combined finite-discrete element method [6, 7, 8], and has now been widely employed for problems with discontinuous nature, particularly for applications involving transition from continuum to dis-continuum due to progressive material damage and fracturing failure. This version of DEM will not be covered in this review.

The discrete element method was originally proposed to seek a better understanding of the mechanical behaviour of granular matter, such as sand, that largely differ from solid continua. The overall behaviour of such systems is mainly determined by the collective motion of individual particles mutually interacting through adhesive/frictional contact along their boundaries.

Granular systems often consist of an excessively large number of individual particles. In order to make DEM computationally feasible to model such systems, rather than using finite elements which is technically possible but computationally prohibited, the particles are assumed *rigid* and physically-based interaction laws describing their interactions or forces are *explicitly* specified based on their contact geometric information. The motion of each particle is then governed by Newton's second law where all forces acting on the particle are accounted for. The resulting governing equations are solved numerically mainly using explicit time integration schemes due to the highly nonlinear nature of the problem. The dominant choice is the central difference scheme for its simplicity (explicit) and relatively high accuracy (second order).

Initially, discrete or distinct particles are represented as discs for 2D problems in [5]. The interaction of two dry discs in contact is described by a linear force-displacement law which relates the displacement (or more precisely 'overlap' of the two discs) with the corresponding normal contact force between them. Frictional contact can also be considered by incorporating a frictional force along the tangential direction of the contact using a similar linear force-displacement relation. However, the maximum magnitude of the force is limited by the Coulomb friction law. Various damping forces can also be included, acting as the main mechanism to account for energy dissipation during contact.

Extension of DEM to 3D problems with spherical particles is straightforward. In addition to the linear contact law, the nonlinear Hertz contact law can also be employed for particles. This is a *physically correct* interaction for spherical particles exhibiting a linearly elastic material behaviour and a small deformation due to the contact.

Due to its three decades long developments, the discrete element method has been firmly established as one of the most powerful and widely used numerical techniques for modelling scientific and engineering problems exhibiting a discrete or particulate nature. This can be attributed to 1) the discrete and rigid nature of particles; 2) a simple computational framework with discs/spheres as the principle geometric primitives; and 3) the existence of a set of explicitly defined contact laws replacing complicated contact interactions between particles. Main application areas of DEM include rock and geo-mechanics, chemical, civil, mechanical, mining and processing engineering, defence and pharmaceuticals, to name a few.

The growing popularity of DEM, particularly in industrial applications, is also promoted by the availability of commercial software, such as PFC [9], EDEM [10], StarCCM+ [11], Rockey, and the availability of more accessible and affordable parallel computing platforms, particularly the rapid advance of inexpensive but powerful GPU or GPGPU computing in recent years. The availability of open-source code, such as LIGGGTHS [12], Yade [13] and MatDEM [14], has also facilitated and sped-up research in DEM particularly in academics.

Spherical particles have been the dominant particle shape for representing almost all discrete particles in DEM. Even for non-spherical particles, clumped particles where multiple spheres are rigidly bonded together to approximate a complex shape, have been recognised and used as an universal approach since the 1990s to take full advantage that spherical particles offer in DEM simulations. However, actual particle shapes in nature are far from spherical. A distinct feature that a non-spherical shape possesses but a sphere lacks is angularity. The angularity is a geometric property of particles that can resist relative rotational motion of particles in an assembly due to their inter-locking action. This geometric feature can significantly affect the macro-behaviour of granular systems. The geometric shape of particles can also have other profound influences on the physical behaviour of particle systems in many different aspects.

The importance of incorporating non-spherical particles into the DEM framework has long been realised since the late 1980s and a first period of rapid development occurred in the early 1990s when ellipses, ellipsoids and super-quadrics were introduced as additional primary geometric entities in DEM. Since then, developing contact models for non-spherical particles has been one of the most important research themes in the DEM community. Over the last 30 years, a wide range of diverse particle shapes and corresponding contact models have been proposed.

The primary aim of this paper is to critically reviewing the major developments in DEM for modelling non-spherical particles over the last three decades. However, this review is not intended to give a comprehensive coverage of all aspects of developments in DEM, but the focus will be on original and novel developments for contact interaction modelling of non-spherical particles from a methodology and numerical technique point of view. Thus the broad phase contact detection, time integration schemes, and parallel computing will not be covered. Also applications of DEM will not be covered. More comprehensive or in-depth review of the whole DEM, particularly the aspects that will not be covered here may be found in monographs [15, 16, 17] and the references therein, specially in a more recent book [18], and also in related review articles [19, 20, 21, 22], for instance. A literature survey of publications in DEM before 1990 was conducted in [23].

Developments of contact models for non-spherical particles will consider both geometric and physical aspects. The geometric aspect concerns with: 1) how to represent the shape of a particle, and 2) how to resolve the contact status of two particles in potential contact, and if they are indeed in contact, what quantities are associated with the contact geometry, such as contact overlap, direction, and contact point/centre. The physical aspect of the contact model deals with how the contact forces are related to the contact geometric features found and the material properties of the particles. These two aspects will be covered, but the magnitude of the contact force with the material properties or behaviour will not be touched upon. A comprehensive coverage of force-overlap laws used in DEM can be found, for instance, in [24]. Note that special efforts will be put to highlight the contributions of researchers in the early stages of developments, as these contributions have largely affected and also shaped the further development of DEM until now.

Another aim that this paper intends to achieve is not only to individually present these major developments, including original concepts introduced and novel computational procedures, but also to discuss how these key developments have been evolved, improvised, and also affected and shaped the subsequent developments. In particular, some important interconnections and relationships between different modelling methodologies will be discussed in detail. Moreover, existing problems and unresolved issues for some methods will be highlighted, and possible solutions will be highlighted where applicable.

The remainder of this review is structured as follows. The next section will focus on shape

representations where most commonly used shapes will be presented in a classification system adopted here. Subsequent sections will focus on contact models for various non-spherical particles. After covering some general issues related to contact modelling in Section 3, Section 4 is dedicated to reviewing major contact modelling methods for convex shapes. Existing contact modelling approaches for non-convex shapes will be discussed in Section 5. The energy-conserving contact theory and related contact models recently established will be presented in Section 6. Some concluding remarks will be made in Section 7.

2 Shape Representation and Reconstruction

Modelling the contact between a pair of particles largely depends on the shapes of the two particles. Thus it is not surprising that the development of contact models for non-spherical particles has been accompanied by gradually introducing various geometric entities or shape representation schemes into DEM. Different representations have also led to the proposal of various contact modelling schemes.

The shape representation is a purely geometric problem. The same problem also appears in many other scientific and application fields. Most closed related fields include computer graphics, computational geometry and computer aided design (CAD), from which many representation schemes used in DEM have been borrowed.

There exist many schemes to represent particle shapes, and these representations can also be classified in many different but often overlapping ways, mainly depending on the purpose or perspective of a particular classification. Some of possible classifications of shape representation schemes that have been used in the DEM literature include: 1) analytical vs topological, 2) implicit vs explicit, 3) parametric vs non-parametric, 4) continuous vs discrete, 5) simple vs compound, 6) single patch vs multiple-patch, 7) surface/boundary vs solid/volume construction, 8) convex vs non-convex (or concave); 9) any others. In addition, representation schemes can also be classified based on some unique features, such as a coarse-to-fine hierarchy, or discretisation. Note that terms 'non-convex' and 'concave' are often used interchangeable below to describe a shape that is non-convex.

In this review, most commonly used shape representation schemes are classified into the following four large categories: 1) *Continuous function representation*, 2) *Patched continuous function representation*, 2) *Topological representation*, 3) *Dilated or Minkowski-sum shapes*, and 4) *Level-set representation*. Shapes in each category share some common features, but the categories are not mutually exclusive, so one particular representation may belong to multiple categories. Also, each shape will be described in its local system where its mathematical expression may be the simplest, if possible. Transformation from the local to global system will not be mentioned, but can be done following standard procedures.

As any interior/internal structure of a particle, such as enclosed cavities, will not affect DEM modelling, all particles are assumed solid domains, and their shapes are solely determined by their boundaries or surfaces which are orientable. Note, however, that a particle domain cannot always be assumed simply connected, but can be multi-connected for some applications. To make this review more concise, a 2D case will be considered as the special case of a 3D case where the $z-$coordinate is zero, if appropriate.

The geometric properties of each shape, including volume, three principal moments of inertia (and their orientations for a non-symmetric shape), need to be evaluated, but may not be trivially done for some complex shapes. The related computational detail will not be given.

Shape representation and reconstruction are two closely related aspects of shapes, and may often be used interchangeably. In this review, shape representation refers to as how mathematical expressions or numerical procedures are used to represent a given shape, while shape reconstruction involves choosing a particular representation scheme and determining values of the required parameters for a single or group of realistic shapes when some macro or geometric properties, such as sphericity, are given statistically. Hence, shape reconstruction can be viewed as the reverse of shape characterisation which extracts some key macro indicators from real particles. This section will mainly focus on shape representation, while shape reconstruction will only be mentioned briefly when Fourier series or spherical harmonics are used to represent realistic particle shapes below.

2.1 Continuous function representations (CFR)

The continuous function representation is probably the most popular scheme used in DEM. The shape or boundary of a particle is represented by a single analytical function

$$f(\mathbf{x}) = 0 \qquad \text{or} \qquad f(x, y, z) = 0 \tag{1}$$

where $\mathbf{x} = (x, y, z)$ are the coordinates of a 3D point on the boundary of the shape. We restrict here that the function f is required be to an analytical mathematical expression with at least C^1 continuous, i.e. f defines the complete bounding surface of a smooth shape. Cases where f is the combination of multiple piecewise continuous functions are called patched continuous function representations (PCFR) and will be covered in the next subsection.

A single CFR is the most compact and thus effective representation. CFRs also provide some desirable analytical properties, including first and higher order derivatives, that can be utilised and lead to some unique contact modelling schemes. This aspect will be covered in later sections.

A CFR can either be *implicit* or *explicit*. The representation is implicit if the coordinates of boundary points $\mathbf{x}$ cannot be explicitly obtained from the function itself, but some forms of equation solving may have to be involved. The representation is explicit if the coordinates of points on the shape can be explicitly obtained in a general form

$$x = x(u, v), y = y(u, v), z = z(u, v), \qquad (u, v) \in D_{uv} \tag{2}$$

where $x(u, v), y = y(u, v), z = z(u, v)$ are all explicit mathematical functions of two parameters u, v defined in a 2D parametric domain D_{uv}. For 2D cases, only the parameter u in a 1D interval is required. Thus x, y and z can be simply obtained by function evaluations without involving any equation solving. It is clear that an explicit CFR is equivalent to a *parametric representation* of the shape. Some shapes may, however, not have an implicit, but only an explicit or parametric representation. One such example is the Fourier series or spherical harmonic representation which will be discussed later.

2.1.1 Ellipses/ellipsoids, and super-ellipses/ellipsoids, and super-quadrics

Most basic CFR shapes include ellipses/ellipsoids, super-ellipses/ellipsoids, and super-quadrics. These were all introduced into DEM in the early 1990s [25, 26, 27, 28, 29, 30, 31], and have since been the most important non-spherical shapes used [32, 33, 34, 35, 36]. Instead of presenting them individually, these shapes can all be expressed by the following CFR form as a super-quadric or its degenerate case:

$$\left|\frac{x}{a}\right|^r + \left|\frac{y}{b}\right|^s + \left|\frac{z}{c}\right|^t - 1 = 0 \tag{3}$$

where the exponents r, s and t are positive numbers, and a, b and c have the same geometric definitions as the semi-axes of an ellipsoid which is recovered when $r = s = t = 2$, and further reduces to a spheroid (axi-symmetrical ellipsoid) if any two semi-axes are equal. When $r = s = t \neq 2$, the super-quadric reduces to a super-ellipsoid. The 2D counterparts can also be obtained by setting $z = 0$.

The parametric equations of (3) are expressed in two parameters $(u, v) \in [-\pi, \pi] \times [-\pi/2, \pi/2]$ as

$$x(u,v) = a\, g\left(v, \frac{2}{r}\right) g\left(u, \frac{2}{r}\right), \quad y(u,v) = b\, g\left(v, \frac{2}{s}\right) f\left(u, \frac{2}{s}\right), \quad z(u,v) = c\, f\left(v, \frac{2}{t}\right) \tag{4}$$

where the two auxiliary functions $f(u, m)$ and $g(u, m)$ are

$$f(u,m) = \mathrm{sgn}(\sin u)\,|\sin u|^m, \quad g(u,m) = \mathrm{sgn}(\cos u)\,|\cos u|^m \tag{5}$$

in which $\mathrm{sgn}(x)$ is the sign function

$$\mathrm{sgn}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ +1, & x > 0. \end{cases}$$

The three coordinate planes, $x - y, y - z$ and $z - x$, are the three symmetric planes of super-quadrics. Thus, when constructing a superquadric, only the portion in the first quadrant needs to be generated from (4). Then the other portions can be easily obtained by symmetry to save the computational cost, as evaluations of functions of a power with a real valued exponent and trigonometry are costly.

The three exponents r, s, t can be varied independently to obtain a wide range of shapes. Super-quadrics are convex if $r, s, t \geq 1$. Otherwise, the shape becomes a concave octahedron-like object with concave faces, pointy vertices and sharp edges. Many shapes that resemble cubes, octahedra, cylinders, lozenges and spindles, with rounded or sharp corners, can be generated from (3) by the combination of different exponents and axis dimensions. Because of the flexibility and relative simplicity, super-quadrics have become popular geometric entities for modelling 3D non-spherical convex particles in DEM.

Although super-quadrics can be convex or non-convex, convex super-quadrics have been almost exclusively used in DEM, and non-convex super-quadrics have not been used until very recently [37]. It is mainly because no sound contact model has been available to handle the contact between non-convex super-quadrics. The details of contact modelling of non-convex or concave shapes will be covered in Section 5.

In the DEM literature, a particular super-quadric proposed in [38], which is referred to as Barr's form here, is often used:

$$\left(\left|\frac{x}{a}\right|^{n_2} + \left|\frac{y}{b}\right|^{n_2}\right)^{n_1/n_2} + \left|\frac{z}{c}\right|^{n_1} - 1 = 0 \tag{6}$$

which can also be parametrised in the following form

$$x(u,v) = a\, g\left(u, \frac{2}{n_2}\right) f\left(v, \frac{2}{n_1}\right), \quad y(u,v) = b\, f\left(u, \frac{2}{n_2}\right) f\left(v, \frac{2}{n_1}\right), \quad z(u,v) = c\, g\left(v, \frac{2}{n_1}\right) \tag{7}$$

Comparing the two forms of superquadrics (3) and (6) shows that neither form is contained in the other. In particular, Barr's form (6) has five parameters to define the shape, while the form (3) has six. Also, each cross-section along the $z-$axis is a super-ellipse in (6), but a 2D super-quadric in (3).

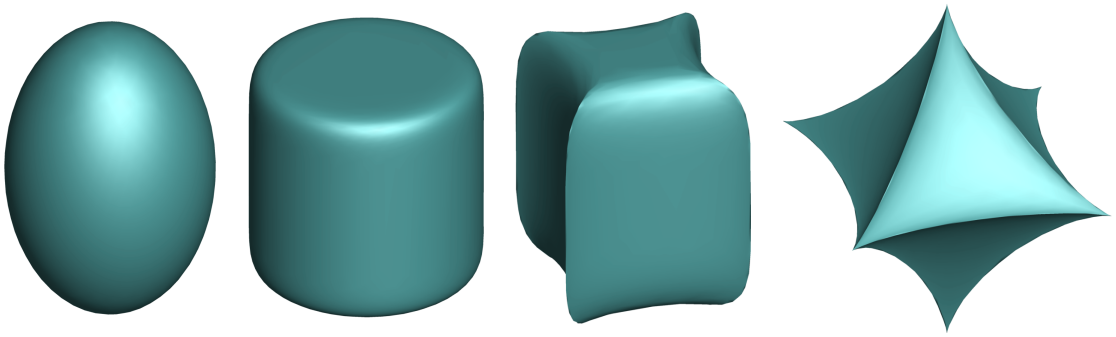


Figure 1: Four superquadric shapes represented in Barr's form (6).

Figure 1 shows a set of superquadric shapes with different parameters defined by Barr's form above, demonstrating the high representation capacity of this particular expression.

The common feature of all the above superquadric shapes is that the boundary/surface of such a shape is C^∞ smooth, in the sense that no shape variation occurs locally, except for concave cases ($r, s, t \leq 1$) with sharp vertices and edges. It is clearly a significant simplification if superquadrics are used to represent real particle shapes, such as rock, sand and soil particles. For such shapes with rich local textures, Fourier series and spherical harmonics are two continuous function representations for 2D and 3D cases respectively.

2.1.2 Fourier Series and Spherical Harmonic Representation

Most Fourier series and spherical harmonic representations are for a *star-shaped* object Ω where there exists an (interior) point $\mathbf{c} \in \Omega$ such that every (line) ray emitting from $\mathbf{c}$ in any direction intersects the boundary of Ω at exactly one point. This point is taken as the origin of the local coordinate system for the shape. The radial distance r of any point on the boundary of a 2D shape to the origin can be expressed by a Fourier series with the parameter $\theta \in [-\pi, \pi]$ as

$$r(\theta) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} [a_n \cos(n\theta) + b_n \sin(n\theta)] \tag{8}$$

where a_0, a_n and b_n are the Fourier coefficients. While for a 3D shape, the radial distance r is expressed by spherical harmonics with two parameters $(\theta, \phi) \in [0, \pi] \times [0, 2\pi]$ as

$$r(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{n,m} Y_n^m(\theta, \phi) \tag{9}$$

where $Y_n^m(\theta, \phi)$ is the spherical harmonic of degree n and order m:

$$Y_n^m(\theta, \phi) = c_{n,m} P_n^m(\cos\theta)\, e^{im\phi} \quad (n \geq 0; -n \leq m \leq n) \tag{10}$$

in which $c_{n,m}$ is a normalisation constant

$$c_{n,m} = \sqrt{\frac{(n+1)(n-m)!}{4\pi(n+m)!}} \tag{11}$$

and $P_n^m(x)$ is the associated Legendre polynomial.

In practice, the above expansion (8) or (9) is always truncated to include only limited terms, i.e. $n \leq N$ (a given number). This Fourier/spherical harmonic approximation provides a

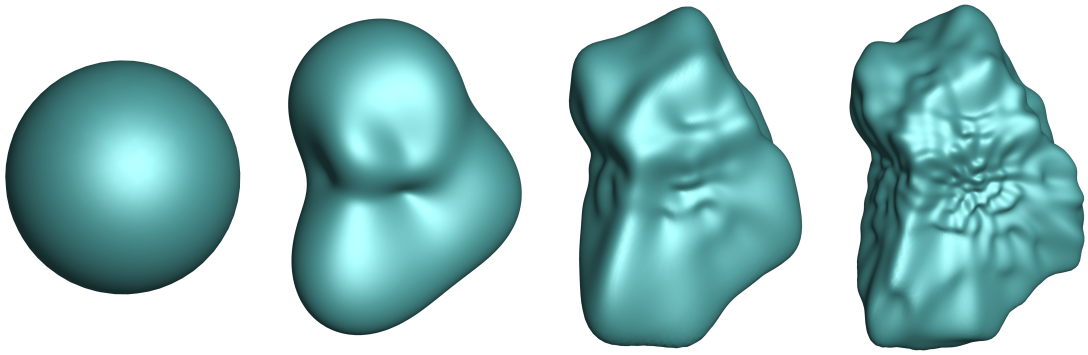


Figure 2: A star-shaped surface approximated by spherical harmonics with an increasing maximum degree $N = 0, 5, 15$ and 30 (from left to right), showing a coarse-to-fine representation hierarchy.

coarse-to-fine hierarchy representation in which higher frequency variations, normally with decreasing magnitude, are gradually added to enrich the shape representation starting from a perfect circle/sphere. Figure 2 shows how a 3D surface is approximated by spherical harmonics with different maximum degrees $N = 0, 5, 15, 30$, where fine features of the surface gradually emerge from a sphere as the basis shape when N increases.

The first few lower frequency components in the expansion of (8) or (9) are more important to determine the overall features of the particle morphology, such as sphericity, elongation, triangularity, squareness or roundness, asymmetry, or irregularity, while higher order frequency components provide the details of local roughness features or particle texture. Detailed discussions of particle shape characterisation can be found in [39].

The representation (8) or (9) mathematically defines a one-to-one mapping between each point on the boundary and one point on the unit circle/sphere. When the shape is not star-shaped, the representation is no longer valid. This is also called the re-entrant angle problem in 2D particle characterisation [141]. A solution to the problem, as described also in [141], is to use the discretised points along the boundary in the complex form of Fourier series. For DEM modelling of 2D irregular particles, the r-θ Fourier representation is often sufficient and easier to use, as demonstrated in [141]. Solutions to non-star-shaped 3D particles also exist.

Spherical harmonics provide the most generic CFR for arbitrarily shapes in 3D. In practice, the spheric harmonics for a 3D particle, such as a rock sample, are obtained in conjunct with particle shape acquisition techniques, such as 3D laser scanning, as the post-process step after a digital image of the particle has bee acquired. More details can be found in [90, 96, 91, 123], while the discussion on the related shape reconstruction can be found in [124, 125], for instance. Note that directly using digitalised images, i.e. pixels or voxels, to represent shapes in DEM for simulating packing of non-spherical shapes can be found in [94, 95] although no physical contact forces are considered.

2.1.3 Cylinders

The cylinder is a primitive geometric entity and has been used in DEM [143, 144, 145, 104] to model long and slender objects. Strictly speaking, a cylinder is not a CFR by our definition as it cannot be described by a single analytical function. Nevertheless it is covered under this CFR category because its analytical expression is very simple and particularly in the cylindrical coordinates. A cylinder with a radius of r and a length of l along the z-direction

and centred at the origin can be expressed as

$$x^2 + y^2 - r^2 = 0, \quad |z| \leq l/2 \tag{12}$$

or in the cylindrical coordinate system with the angle $\theta \in [0, 2\pi]$:

$$x = r\cos(\theta), \quad y = r\sin(\theta), \quad |z| \leq l/2 \tag{13}$$

The main difference between cylinders and other convex CFR shapes described above is the present of two end discs with two sharp edges. This seemly insignificant difference has caused many difficulties in their contact modelling as will be briefly mentioned in Section 4.3.

2.2 Patched Continuous Function Representation (PCFR)

Although super-qaudrics (9) provide a general CFR scheme that can not only include those basic geometric entities as the degenerate cases, but also create much more diverse shapes, a super-quadric is always symmetric about the three coordinate planes. This restricts the applicability of super-quadrics from representing more realistic particles where no symmetry is present. On the other hand, this inadequacy of representing unsymmetric shapes is totally eliminated by spherical harmonics (9), but with expenses of introducing a much higher degree of complexity in the representation.

Thus a question arises if there exist other approaches that may keep some advantages that CFR shapes can offer but are more effective and/or more capable of representing realistic particle shapes. Over the course of the DEM development, many approaches have been proposed, often in a compromised manner, to balance the compactness, efficiency and flexibility of various shape representations.

Some widely used techniques will be presented in a unified framework below, and all will be called the patched continuous function representation or PCFR. Mathematically, each continuous function defines the entire boundary/surface of a shape as one whole 'patch'. More complex shapes can be formed by putting together different shapes or their partial parts or patches. Different approaches are distinguished from each other mainly by the conditions under which shapes and/or 'patches' are 'glued' together.

The special names given below to describe some commonly used representing schemes may be different from what may be referred to in the literature. This is mainly because many terms have been used in the DEM community and currently there is no universally agreed terminology.

2.2.1 Clumped or Composite shapes

The first approach concerned is to rigidly bond together as many discs/spheres as required to represent a shape, either convex or concave. The resulting shape formed by the cluster of discs/spheres is often called a clumped or composite shape. In addition to the apparent diversity of shapes that a clumped/composite shape can represent, another substantial benefit is that the contact between these clumped particles can simply be reduced to the contact between different component discs/spheres, which can be trivially done, thereby inheriting the computational effectiveness of discs/spheres. Thus, not surprisingly clumped spheres started to be used in the early 1990s for modelling tetrahedral and cubic particles [51] and later for axi-symmetric particles [52]. Since then using clumped spheres has become the most commonly used approach to representing non-spherical particles.

i) Automatic cluster generation. Although with those desired features outlined above, this approach of clustering spheres has unfortunately achieved less success than would have hoped for, mainly because of practical difficulties in creating realistic representations of natural particles. Thus, there has been a practical requirement of developing an automatic procedure, instead of using a manual exercise/approach, that can produce a cluster of spheres to realistically represent real particles, but also minimise the number of spheres needed in order to reduce the DEM simulation cost. These two requirements are clearly contradictory, and thus need to be balanced in practice. To develop such a procedure, the problem may need to be formulated into an equivalent mathematical problem from which some numerical or approximate methods may emerge.

ii) Equivalent mathematical problem. A 2D shape can be fully covered by overlapping discs if its skeleton can be constructed first. By definition [42] the skeleton is the axis of symmetry of a shape, or equivalently it is the locus of centres of maximal disks contained in the shape. Thus using all (infinite number of) maximal discs will cover the whole shape. A limited number of discs can be obtained if the level of coverage can be relaxed. A computational algorithm was developed in [42] to generate the skeleton of a shape in 2D image based on the discrete Euclidean distance map [41]. The concept of skeleton can be extended to 3D shapes, but effective algorithms are hard to be developed.

iii) Heuristic numerical methods. Using a cluster of spheres with the smallest size (i.e the number of spheres) to represent a shape may be converted as a set-covering problem in combinatorics, as stated in [45]. The problem concerns finding the optimal number of subsets (spheres) whose union covers a set (shape) while minimising the total cost of the covering. The cost function here should account for both the level (accuracy) of covering and also the number of subsets used. Since this is a NP-hard problem, only heuristic methods can be expected.

However, there are issues with defining the target or cost function. First of all, for a given cluster of spheres, it is not easy to precisely calculate either the surface area or volume occupied, due to the curved and overlapping nature of sphere clusters. Secondly, there also is a choice of approximating either the surface or volume of the shape. As pointed out in [45], for some applications, the former may be more important than the latter. The final issue is to what quantitative level that the coverage accuracy and the number of spheres should be balanced.

There are three heuristic procedures that have been proposed within the set-covering framework over the years, including the burning algorithm [43], the clustered overlapping sphere (COS) algorithm [44], and the modified greedy heuristic (MGH) algorithm with three slightly different schemes [45]. Their numerical procedures are similar in some aspects but different in others. No details of these numerical procedures will be discussed here. Figure (3) shows the sphere cluster representations of two shapes using the MGH algorithm [45].

vi) Limitations. The surface of a clumped shape is the union of all the spheres in the cluster, and thus consists of spherical patches from different spheres. In this sense, a clumped sphere can be classified as a PCFR - patched continuous function representation, but the spherical patches concerned are connected with C^0 continuity. This C^0 continuity leads to an issue when a clumped particle has a surface which is larger and often much rougher than the original smooth shape. This may increase the locking effect/friction between clumped particles, resulting in some unsatisfactory physical behaviour in the system. The behaviour of clumped-sphere particles is investigated in [46, 47] to reveal some limitations.

v) Other possible clusters. Instead of using discs/spheres, any other geometric entity can also

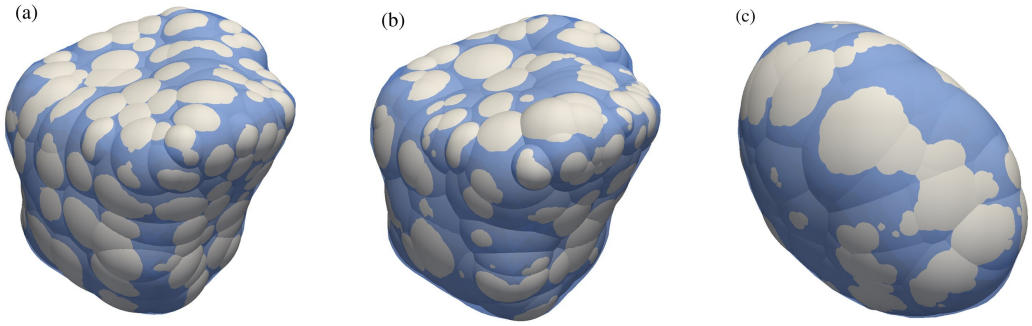


Figure 3: Clumped sphere representations of shapes using the MGH algorithm with three schemes: a) SC; b) TSC; and 3) BC, where the grey parts are from the sphere clusters, and the lighter blue from the original shapes [45].

Figure 4: Composite shapes formed by joining two (equal) spherical caps with a cylinder: (a) table; (b) capsule

be used, in principle, as a constituent component to form a composite shape, thus offering a further added flexibility to represent non-spherical particles. Nevertheless, because the contact modelling between non-spherical particles is much harder than that of discs/spheres, clumping non-spherical shapes have rarely been used in practice.

One exception is a *'tablet'* or *'capsule'* shape that can be constructed by joining two spherical caps to the two ends of a cylinder with the radius equal to that of the bases of the two caps. Note that in the special case when the two caps becomes semi-spherical, the connections become C^1 continuous, and the contact between two such capsules becomes easier than that of two cylinders. See Figure 4 for an illustration

2.2.2 Compound Shapes

Compound shapes, as another form of PCFR, differ from the above composite shape mainly by C^1, instead of C^0, continuity that different patches are connected. There are many different shapes developed in DEM that can be classified as compound shapes.

i) *4-arc approximated ellipse and ellipsoid.* In the first compound shape proposed in [126], each elliptical particle is approximated by four arcs connected in C^1 continuity. The centre and radius of each arc and co-ordinates of the four connecting points can be determined from the major and minor axes. By rotating this 4-arc shape around the major axis, an approximate ellipsoid is obtained [131]. Now the contact between two ellipses becomes the contact of arcs which can be done trivially, while the contact between two ellipsoids reduces to contact between axi-symmetric shapes with arcs as their profiles but with limited reduction

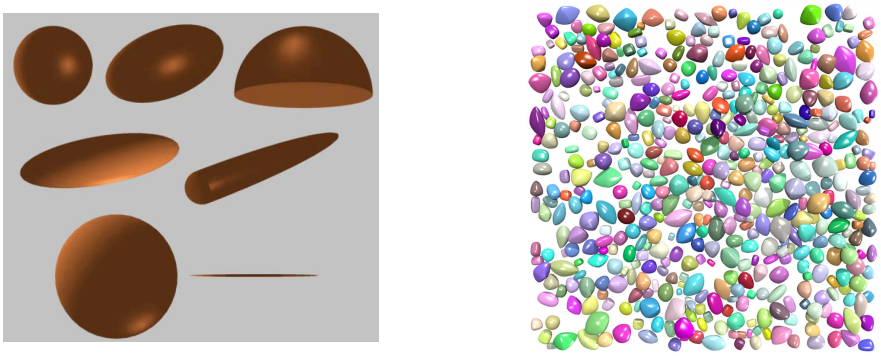


Figure 5: Compound shapes: (a) poly-ellipsoids [64]; (b) poly-super-ellipsoids [110].

in computational complexity though. The same approximation of ellipses using four-arcs is also adopted in [127] where using circular segments of different radii to represent other non-rounded shapes is also proposed.

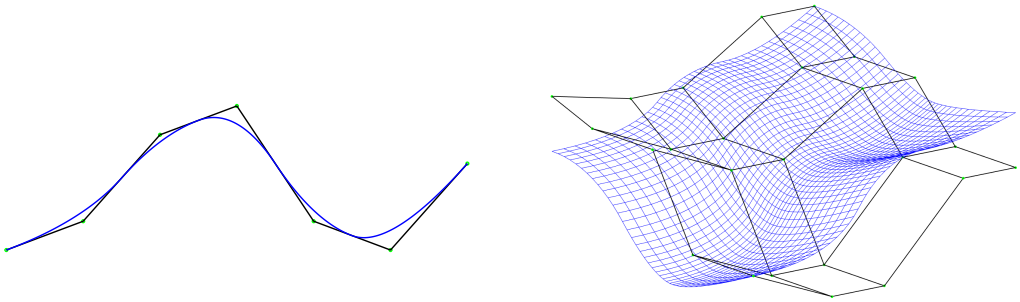
ii) *Poly-ellipsoid.* A poly-ellipsoidal shape is the first non-symmetric compound shape proposed in [64]. A poly-ellipsoid is formed by joining octants of eight ellipsoids, each with different aspect ratios, across their respective principal planes to ensure that the resulting compound surface is C^1 continuous. Thus only six parameters are required to determine a poly-ellipsoid, instead of the 24 parameters required to define the eight component ellipsoids. The geometric properties of this shape can also be easily computed using these six parameters in closed form. By varying the six parameters a wide range of *pebble-like* shapes can be obtained. The contact between poly-ellipsoids can be handled similar to that of ellipsoids. Several poly-ellipsoids from [64] are shown in Figure 5(a). Construction of poly-ellipsoidal grain shapes from SMT imaging on sand is reported in [65].

iii) *Poly-super-ellipsoid.* A further extension of the main idea of the above poly-ellipsoid to super-ellipsoids of Barr's form (6) leads to poly-super-ellipsoids proposed in [110]. A poly-super-ellipsoid is constructed by connecting eight patches of eight super-ellipsoid with different shape parameters from different octants in a C^1 continuous manner. Again, instead of 40 parameters required to define the eight component super-ellipsoids, only eight parameters are required to define the poly-super-ellipsoid, and the geometric properties of this shape can also be evaluated in an analytical form, but slightly harder [110]. A collection of various poly-super-ellipsoids from [110] is shown in Figure 5(b).

vi) *NURBS and Iso-geomrtric representations.* Non-uniform rational basis splines (NURBS) are very commonly used to represent curves and surfaces in computer graphics, computer-aided design (CAD), manufacturing (CAM), and engineering (CAE). A NURBS curve is defined parametrically as

$$\mathbf{C}(u) = \left(\sum_{i=0}^{n} N_{i,p}(u) w_i \mathbf{P}_i\right) / \left(\sum_{i=0}^{n} N_{i,p}(u) w_i\right) \tag{14}$$

where $u \in [0,1]$ is the curve parameter, $N_{i,p}(u)$ is the i-th B-Spline basis function of degree p which can be obtained recursively from $N_{i,0}$ defined in $[u_i, u_{i+1})$, a sub-interval of $[0,1]$; $w_i \geq 0$ are the weights, and $\mathbf{P}_i$ are control points. All subintervals $[u_i, u_{i+1})$ form a *knot* vector $U = [0 = u_0, u_1, \ldots, u_n, u_{n+1} = 1]$, and each knot value u_i can be repeated. The number of times a knot value repeats itself is called multiplicity k with the limit that $k \leq p$. The curve is infinitely continuously differentiable within each subinterval but is $(p-k)$-times continuously differentiable at a knot. Thus a NURBS is a piecewise (rational) function with C^{p-k} continuity.



(a) (b)

Figure 6: NURBS: (a) a curve determined by seven control points; (b) a 4-sided surface patch defined by a 5×5 grid of control points.

A NURBS surface or patch is obtained as the tensor product of two NURBS curves, using two independent parameters (u, v) defined over a square $[0, 1] \times [0, 1]$ as

$$S(u, v) = \sum_{i=1}^{k} \sum_{j=1}^{l} R_{i,j}(u, v)\mathbf{P}_{i,j} \tag{15}$$

with rational basis functions

$$R_{i,j}(u, v) = \frac{N_{i,n}(u)N_{j,m}(v)w_{i,j}}{\sum_{p=1}^{k} \sum_{q=1}^{l} N_{p,n}(u)N_{q,m}(v)w_{p,q}}$$

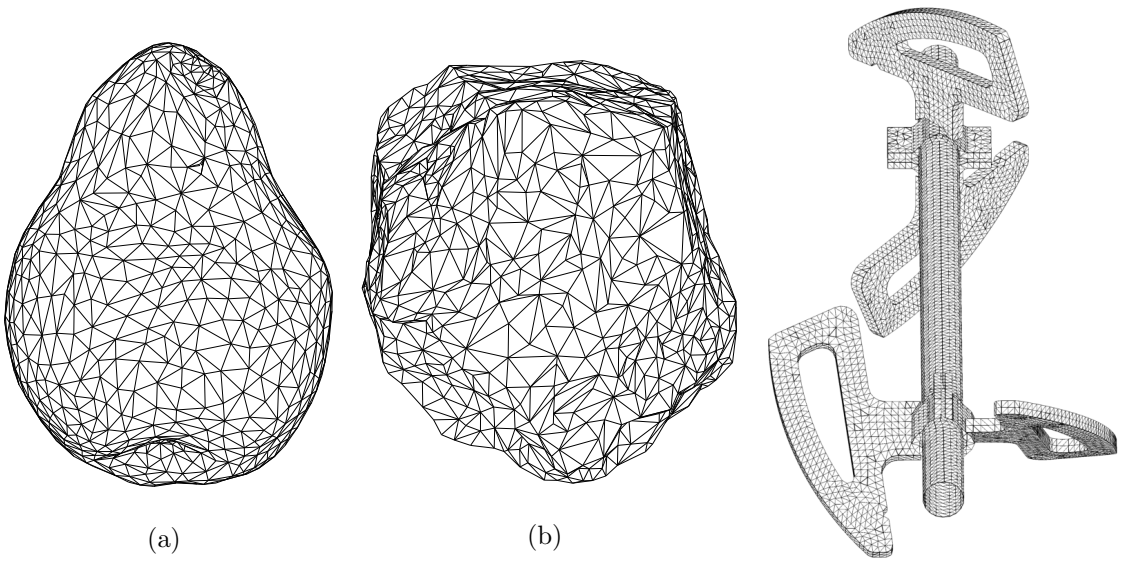
NURBS provide flexibility in representing arbitrary and complex but smooth geometries more effectively than conventional polygonal counterparts, and are part of numerous industry-wide standards in CAD/CAE. NURBS curves are first introduced to DEM in [84] for representing simple 2D shapes, while NURBS representations of 3D shapes are considered in [85].

However, as control points $\mathbf{P}_{i,j}$ must be defined in a rectangular grid, a NURBS patch is always 4-sided, and thus cannot represent a single closed surface without introducing singularities. Hence multiple NURBS patches have to be used to represent a 3D particle in DEM. Also, as control points are not on the curve/surface, it is not very easy to construct any particle shape in a simple DEM setting, unless the shape is already represented by NURBS for a special application. In addition, due to the use of rational basis functions $R_{i,j}(u, v)$, the computations involved in modelling the contact between two NUBRS shapes are substantially more complicated and algebraically demanding than other surfaces presented earlier.

The above difficulties have limited the use of NURBS particles in DEM. Nevertheless, the ability of NURBS surfaces to effectively represent 2D and 3D shapes in CAE has prompted the full development of the iso-geometric analysis (IGS) of solids and structures over the last 20 years [86], where the same rational basis functions $R_{i,j}(u, v)$ are also used for the finite element interpolation. Thus, combining NURBS based IGS for structures with DEM for discrete particles provides a natural and effective approach for modelling particle-structural interaction problems. This combined strategy is exploited in [87, 88].

2.3 Topological Representation

Instead of using continuous functions, topological representation (TR) uses vertices and their topological relationships to define shapes. Polygons and polyhedra are examples in 2D and 3D respectively. In our classification system, a polygon/polyhedron can also be classified as



(a) (b) (c)

Figure 7: Surface triangulated polyhedral representations of various 3D shapes.

a composite shape which is the union of line segments/(3D planar) polygons in a C^0 fashion, but using the coordinates of the vertices and their topological connectivity is a much more effective representation. The connectivity is mainly for defining side/polygonal faces, from which edges can also be defined if needed. In the above representation, a polygon/polyhedron is represented by its boundary as a polyline/polygonal mesh. Alternatively, a polygon/polyhedron as a solid region can also be represented by a triangular/tetrahedral mesh.

Polygons/polyhedra in fact were the first non-circular/spherical particle introduced to DEM in 1988 [56], even earlier than ellipses [25, 27]. Sharp corners/edges of polygons/polyhedra are the features that are lacking in CFR particles, but are key to problems in rock mechanics, for instance. Polygonal/polyhedral meshes are the most commonly used format in computer graphics to represent shapes obtained by digital technologies, such as computer tomography (CT). Polygons/polyhedra can be convex or non-convex. A few surface triangulated shapes as polyhedra are shown in Figures 7.

Furthermore, all CFR and PCFR shapes can be discretised as polygons/polyhedra. Thus, polygons/polyhedra provide the most flexible and versatile representation for any shape, convex or non-convex, that has been used in DEM [61, 62]. As a result, polygonal/polyhedral representation can be employed to compute geometric properties of any shape. In implementation, except for vertices and face connectivity which are essential, a number of slightly different schemes exist, mainly differing from how edges and other auxiliary information are presented, if included at all.

For convex polygons/polyhedra, a third format of representation exists: the i^{th} side/polygonal face is on the half space

$$\mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) \leq 0 \quad \text{or } a_i x + b_i y + c_i z + d_i \leq 0 \tag{16}$$

where $\mathbf{n}_i = (a_i, b_i, c_i)$ is the outer normal of the side/face, and $\mathbf{p}_i$ is any point on the side/face. Then the whole polygon/polyhedron as a solid region can be represented by the union of all the half-spaces corresponding to the sides/polygonal faces. This format of representation can be readily utilised to check if a point $\mathbf{x}$ is on/inside/outside the polygon/polyhedron: if any of (16) is not satisfied, the point is outside the shape; otherwise it is inside the shape but if one of the equality holds, the point is on the side/face. A more significant use of this format

will be discussed in Sections 4.4 and 4.2.3.

Nevertheless, the C^1 singularity at vertices/edges also imposes computational difficulties in contact modelling of polygons/polyhedra where those approaches developed for CFR or PCFR are no longer applicable. Thus new contact modelling approaches have to be developed to effectively utilise the unique features of polygons/polyhedra.

An immediate solution is to remove the C^1 singularity of polygons/polyhedra by rounding vertices/edges using arcs/spheres. This was widely used in DEM practice in the 1990s. The first rounded polygonal shape, called Polyarc, is introduced in [128], and the main idea is further refined and extended recently in [129] to general 2D shapes, obtained from digital images.

When the topological information between vertices is not provided, polygons/polyhedra reduce to *point clouds*. Point clouds are often obtained from digital acquisition of actual shapes. Because the topological information is essential in discrete contact modelling of shapes, point clouds alone are not sufficient to represent shapes. Hence, the triangular mesh needs to be generated from the point cloud to characterise the particle surface as a polygon/polyhedron. However, point clouds can be used as discretised or auxiliary points of a shape represented in other forms in certain contact models, as will be described in the next section.

2.4 Dilated or Minkowski-Sum Shapes

Dilated shapes are first introduced to DEM in [146, 106]. The main idea of the dilation process is to place the centre of a *dilating sphere* at every point on the surface of a *basis shape*, and then the envelope of the resulting shape forms a dilated shape. This dilation process can generate complicated shapes depending on the basis shapes used, but many primitive shapes can also be constructed in this way. For instance, dilating a point results in a sphere with the same diameter of the dilating sphere; dilating a sphere results in a larger sphere; dilating a line segment results in a capsule (cylinder with hemispherical ends). Most dilated shapes used in DEM are obtained by dilating (convex) polyhedra [147, 148], resulting in dilated polyhedra or *sphero-polyhedra* as termed in [122]. This essentially eliminates the singularity problem caused by vertices and edges in contact modelling of polyhedra. Hence dilating using spheres is more effective than artificially rounding vertices or edges as mentioned previously. Dilated non-spherical shapes can also be used to represent particles with equal-thickness interphase layers, see [109] for instance.

A dilated shape can be viewed as the union of the basic shape and a cluster of an infinite number of equal-sized spheres. Instead of using a sphere, other shapes can also be chosen to be the dilating shape. For instance, dilating a line segment by a two-dimensional flat disk results in a cylinder.

A more elegant and powerful approach than dilating basis shapes with spheres is by using the *Minkowski sum* operation to generate complex shapes. Let A and B be two shapes. The Minkowski sum of A and B is a new geometric shape, denoted as $A \oplus B$, defined by adding every point in A to every point in B:

$$A \oplus B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\} \tag{17}$$

The above algebraic definition of Minkowski sum $A \oplus B$ is rather abstract. A simpler geometric approach can be used to construct the Minkowski sum of two shapes A and B: Fix A and slide B in translation only around A to make sure that A and B are always kept in a touch contact condition. The profile of B formed while moving around A for a complete cycle is the

(a) (b)

Figure 8: (a) Minkowski-sum of a polyhedron and a sphere to form a sphero-polyhedron [108]; (b) Sphero-polyhedra with different sized spheres [18].

Minkowski sum $A \oplus B$. Note that the Minkowski sum of two convex shapes is also convex. The explicit construction of the Minkowski sum of two convex polyhedra is given in [79].

It is easy to see that a dilated shape using a dilating sphere mentioned above can be alternatively formed by the Minkowski sum of the basis shape and a sphere but with half the size of the dilating sphere [149]. Figure 8(b) shows a sphero-polyhedron constructed as the Minkowski sum of a polyhedron and a sphere, while Figure 8(b) shows several sphero-polyhedra with different sized spheres.

Using a Minkowski sum is a much more powerful and flexible way to form more complex shapes due to the following: A and B can be any shape, convex and no-convex included; A and/or B can also be a Minkowski sum, i.e. the Minkowski sum can be applied in a *hierarchical* manner. Because of these unique features, Minkowski sum has been widely utilised to construct various shapes/scenes in many fields [132], including computer graphics/image processing, robotics, computer-aided design, mathematical morphology and spatial planning. The key issue associated with using more complex Minkowski-sum shapes is how to conduct contact modelling between them. This issue will be briefly reviewed in Section 4.3.

2.5 Level-Set Representation

A particle occupies a solid region. All the previously mentioned shape representations describe the boundary of a particle shape, as the boundary is sufficient to define the shape. The level-set representation aims for defining a shape as a solid.

The concept of a level set is originally developed in [135] from implicitly defining a curve/-surface by the iso-contour/surface of a 2D/3D function

$$f(\mathbf{x}) = c \tag{18}$$

where c is a valid value within the range of f. Clearly the CFR form of a shape defined in (1) is simply a particular iso-contour/surface when $c = 0$, i.e. the zero-level set.

However, the level-set representation differs in the sense that it can define a field over the domain of the level-set function $f(\mathbf{x})$. A solid region bounded by the zero-level set is naturally defined when c takes all positive or negative values, depending on the choice of sign convection when defining the level-set function. The immediate benefit of using a level-set to represent a shape is that it is now trivial to determine if a given point is inside, on or outside the shape.

It is straightforward to convert CFR shapes, as those discussed in Section 2.1, into level-set representations, except for spherical harmonic shapes. However, as also commented earlier, CFR shapes have a very limited capacity of representing real particles encountered in engineering applications. Thus it is a substantial challenge to represent an arbitrarily shaped particle by defining a simple mathematical expression as the level-set function.

A particular choice of the level-set function is the signed distance function (SDF) which is defined as the minimum distance d of a point $\mathbf{x}$ to the boundary of a shape Ω:

$$d(\mathbf{x}) = \pm \min_{\mathbf{y}\in\partial\Omega} \left\{ \|\mathbf{y} - \mathbf{x}\| \right\} = \pm \|\mathbf{x}^* - \mathbf{x}\| \tag{19}$$

where the + or - sign is chosen based on the convention (if the point is inside or outside the shape or the other way around); $\mathbf{x}^*$ is the point on the boundary having the shortest distance to $\mathbf{x}$. $\mathbf{x}^* - \mathbf{x}$ is the corresponding vector distance, which gives not only the minimum distance but also the direction. The point $\mathbf{x}^*$, if also obtained, can be very valuable in contact modelling of two level-set represented shapes. This will be discussed in Section 5.3.

Clearly the signed distance function is very generic and is applicable to any shape and thus can be used as a universal level-set function to represent arbitrary shapes. However, except for a very few simple shapes, the signed distance function cannot be explicitly defined for a general shape. In addition, it is normally a very costly operation to compute the signed distance of a point to a shape. Consequently an effective way has to be found to define the signed distance function.

The most commonly used scheme is to use a regular grid to represent a shape, where the signed distance of each grid node is first constructed. Then the boundary of the shape can be implicitly but approximately extracted, if needed, from the grid cells where the associated grid nodes have both positive and negative distance values. The above procedure is equivalent to a discretised level-set representation of a shape, and appears to be the most widely used scheme in many fields including computer graphics/image processing, computational geometry and CFD, to name a few.

A regular Cartesian background grid is simple to use but is not very effective. The accuracy of using such a level-set grid to represent a shape depends on the grid spacing and is of first order which is not high. Often millions of grid cells/nodes are needed to achieve a moderate representation, and also generating all the SDF values of such a grid can be computationally substantial. A large number of algorithms have been developed, mainly in the computer graphics community, to effectively compute the SDF of a grid for complex shapes or scenes.

There are cases where level-set values of a full grid is not necessary but only the grid nodes sufficiently close to the boundary of the shape are required, for example in the DEM where the penetration of two contact particles are assumed small. In these cases, a better scheme exists, in which the grid is formed in a hierarchical manner using a quad-tree or octree [150], for instance, where the deepest level is controlled by the accuracy of approximating the surface.

The first use of a level set on a regular grid to represent a CFR shape in DEM is introduced in [133], and the resulting shape is called the *discrete function representation* (DFR). The boundary of the shape can be discretely sampled using a set of points independent of the grid. A simple contact modelling scheme is also proposed [133, 134]. The extension of the DFR to three-dimensional particles is reported in [19] using a 3D polar descriptor to generate surface nodes. Later on, level-sets are used to re-construct particle shapes that are digitally acquired. See [138] for more detail. The first sound contact model for two discrete level-set particles are reported in [139] and followed recently in [37]. The related contact modelling issues will be reviewed in Section 5.3.

3 Contact Modelling - Some General Issues

After having reviewed the major shape representation schemes for non-spherical particles, we now focus on the most widely used and influential contact modelling approaches in DEM

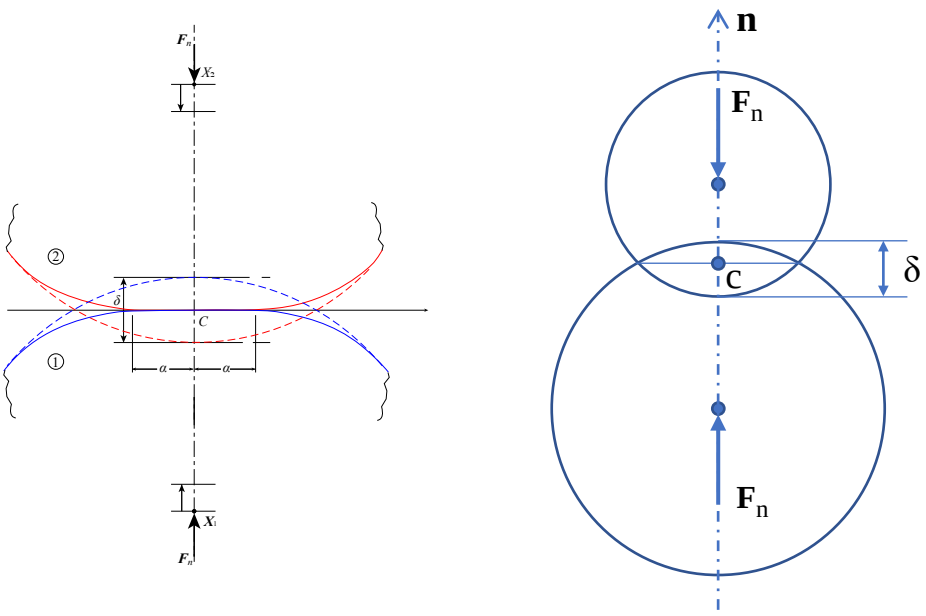


Figure 9: Hertz contact: (a) Deformation of two elastic spheres under compression; (b) DEM simplification

proposed over the last three decades. These approaches will be introduced similar to the categories used to classify shape representations in the previous section.

To ensure a coherent review of these major contact modelling approaches, we need to highlight some key features that a valid contact model should possess, and also a crucial issue that contact modelling of non-convex shapes will face - the possibility of multiple contact situations.

We will first recall the Hertz contact model for spheres as this is the most important interaction law in DEM, and in particular its main features involved have largely dominated the development of contact modelling for non-spherical particles until now. We will also briefly consider the issues with single and multiple contact situations. The actual review of current contact modelling approaches will be conducted in the next three sections.

3.1 Hertz Contact Model for Spheres

The Hertz contact theory established by Heinrich Hertz in 1882 was one of the most important developments in continuous mechanics in the 19th century, and in particular provides a theoretical foundation for modern problems in contact mechanics. The classic Hertz contact solution for two linearly elastic spheres describes the *magnitude* of *the total normal contact force* ($\mathbf{F}_n$) between the two spheres as [48]

$$F_n = \frac{4\pi E^*}{3} R^* \delta^{3/2} \tag{20}$$

where E^* is the equivalent Young's modulus of the two spheres, R^* is their equivalent radius, and δ is the *overlap*, as shown in Figure 9(a).

This analytical expression indicates that, in addition to the material properties and radii,

the total contact force F_n between two elastic spheres only depends on the overlap δ which can be easily determined by the position and size of the two spheres without knowing the very complex contact deformation in their contact region. In other words, the spheres can be nominally considered as *rigid* during the contact, thus validating the most essential assumption made in the original DEM [5], while the overlap can be viewed as an aggregated representation of the actual deformation.

The normal Hertz contact force as a vector can be expressed as

$$\mathbf{F}_n = F_n \mathbf{n}, \qquad \mathbf{n} = (\mathbf{x}_2 - \mathbf{x}_1)/\|\mathbf{x}_2 - \mathbf{x}_1\| \tag{21}$$

where $\mathbf{n}$ is the normal contact direction along the line linking the two sphere centres, $\mathbf{x}_1$ and $\mathbf{x}_2$. In this review, the positive normal contact direction between two shapes is assumed from the second shape towards the first one. In other words, the normal contact force $\mathbf{F}_n$ as a vector is the contact force exerted by the second shape on the first one. The reaction normal force from the first one to the second one is $\mathbf{F}'_n = -\mathbf{F}_n$.

The (common) contact point at which both forces $(\mathbf{F}_n, \mathbf{F}'_n)$ act is not explicitly specific in the Hertz contact theory, because the Hertz solution is mainly focused on the displacement and stress distributions around the contact region under elastic deformation. In DEM practice, the central point of the line joining the intersection of the two circles, the point $\mathbf{c}$ shown in Figure 9(b), has almost universally used as the contact point.

However, a recent revisit to the Hertz theory reveals [97] that a more appropriate contact point can be deduced, depending on the material properties of the two particles. If both spheres have the same property, the contact point should be at the position of the half overlap.

In general, a complete normal contact model should include three features: 1) *normal contact direction*, 2) *contact point*, and 3) *magnitude of the force*, which are termed the three *contact features*.

Thus, the Hertz contact model perfectly meets all the requirements of DEM for an interact law between two elastic spheres in contact, and is also physically correct under the assumptions made in the theory [48]. Furthermore, the Hertz contact law is the first physically correct contact law and hence plays a very important role in DEM. For more complex shapes with sharp edges and corners, such as polygons and polyhedra, no physically based theoretical model is available to serve as an interaction law. In this sense, existing contact models for non-spherical particles are of a heuristic nature and can all be classified as numerically-based.

The magnitude of the Hertz contact force (20) may admit a general multiply expression of two factors

$$F_n = k(E^*)g(\mathbf{x}_1, \mathbf{x}_2, R_1, R_2) \tag{22}$$

where the first term k denotes a material property dependant factor or *stiffness* - accounting for the material effect on the force; and the second term g denotes a *contact geometric* related factor - depending on the positions and sizes of the shapes concerned. For non-spherical cases, shape orientations will also be involved. In other words, the geometric and physical aspects of a contact model can essentially be decoupled.

This decoupled contact modelling framework has greatly facilitated the development for non-spherical particles. In particular, we can focus on the contact geometric aspect first for any shape. After the contact geometric features are obtained, the magnitude of the contact force can be specified by various constitutive laws involving both the material properties and the contact geometric features obtained. For instance, for spheres, after the overlap has been

computed, the force magnitude can be a linear or non-linear function of the overlap and multiplied by a material related stiffness k.

So in this review, contact modelling of two particles will concern all geometric aspects of their contact, while the contact interaction or institutive law specifies how the magnitude of the contact force is determined by one or several contact geometric qualities (such as overlap) and also the material properties concerned.

Different contact modelling approaches for a given two shapes mainly differ by how the contact geometric features are defined and which interaction laws are used to compute the magnitude of the contact force. *The focus of this review is on the aspect of modelling contact geometric features without referring to contact interaction laws.*

3.2 Contact Geometry Features and Single/Multiple Contacts

Obtaining the contact geometry and related features for two shapes in contact is a purely geometric problem, and usually treated as a contact detection or local resolution issue in DEM. The contact geometry is fully determined by the intersection region of the two shapes considered, from which various features can be attracted in general. However, in some situations, as will be shown later in Section 4.5.2, not just the vicinity of the contact, but the entirety of each shape may need to be considered.

In any case, the minimum set of contact geometry features must include the normal contact direction, contact point and at least one characteristic quality. The commonly used characteristic qualities include: overlap or penetration depth, (contact) area or volume. Due to Cundall's original work [5] and also the central role played in the Hertz contact, the overlap has almost universally become the default contact geometric quantity in DEM, particularly for convex shapes.

A distinct contact region between two shapes in contact is regarded as a single contact. The contact between any two convex shapes always results in a single contact, from which one set of contact geometric features can be evaluated.

When at least one shape is non-convex or concave, there may be more than one contact region that can be formed, resulting in a case of multiple contacts. Multiple contacts can occur in two situations. In the first situation, called the *distinct* multiple contact, multiple contact regions can formed from the very beginning of the contact, and evolve or disappear during the course of contact but cannot merge/separate into fewer/more contact regions. In the second case, called the *associated* multiple contacts [67], an initially single contact region develops into multiple contact regions as the contact progresses, but can merge back to a single region at the bounce back stage.

For a distinct multiple-contact case, each individual contact region can be processed to find the required contact geometric features. However if one of the contacting shapes is non-convex, it is not always known *in priori* whether their contact will be a distinct multi-contact or not. It is this problem and, more severely, the existence of associated multi-contact case that make the contact modelling of non-convex particles in DEM challenging and no significant progress has been made until very recently.

Contact modelling strategies that have been developed for convex shapes will be reviewed in the next section, while the review for non-convex shapes will be done in Section 5.

4 Contact Modelling for Convex Shapes

In this section, we will review major developments in the contact modelling of convex shapes in DEM. As stated in the previous section, the contact of convex shapes will always be a single contact from which the three elements of the contact geometric features need to be obtained: normal contact direction, contact point, and at least one characteristic quality. By default the characteristic quality will be overlap, unless stated otherwise. Thus, for each contact modelling method, we will focus on: 1) how these contact features are defined; 2) which computational procedures are available to evaluate these features; and 3) computational issues, if any.

Following the classification of shape representation schemes in Section 2, we will concentrate on major contact modelling methods for these classes of shapes and their combinations in the following subsections: (P)CFR shapes, polygons/polyhedra, dilated and Minkowski-sum shapes, and mixed shapes. Within each class, different contact detection or modelling methods will be reviewed.

4.1 Contact Modelling for (P)CFR Shapes

In this subsection for (P)CFR shapes, compound shapes (i.e. with C^0 continuity) are excluded until in Section 5.2, while clumped sphere particles will be covered in Section 5.1.

Assume that there are two contacting convex (P)CFR shapes, Ω_1 and Ω_2 respectively, whose boundaries $\partial\Omega_1$ and $\partial\Omega_2$ can be represented by two (piecewise) differentiable functions $f_1(\mathbf{x}) = 0$ and $f_2(\mathbf{x}) = 0$.

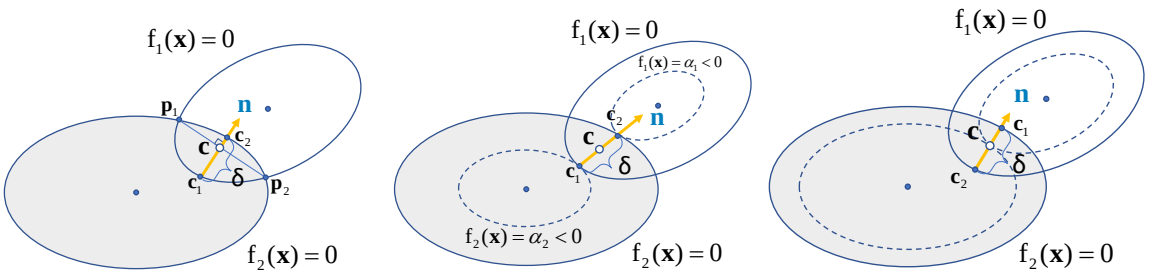
First consider the contact of two ellipses. Naturally this is the first non-spherical CFR contact case tackled in DEM. More importantly, the contact detection methods developed for this case and also the extensions to the contact of ellipsoids have subsequently dominated the contact modelling for almost all non-spherical (P)CFR contacts. For this reason, we will not review different contact models for different shapes, but group our discussion around a few key contact methods, and focus on how each method has evolved and how different variants have been developed. Any important theoretical and/or computational issues will also be discussed.

4.1.1 Contact Intersection Method

The first contact detection algorithm for two ellipses is proposed in [25, 27] and is called the *intersection* method [30], which determines the contact geometric features in the following order:

- Normal direction $\mathbf{n}$: find the two intersection points $\mathbf{p}_1$ and $\mathbf{p}_2$ by simultaneously solving two equations
$$f_1(\mathbf{x}) = 0, \quad f_2(\mathbf{x}) = 0 \tag{23}$$
Then the tangential direction of the contact is defined as along the segment p_1p_2. Now the normal direction $\mathbf{n}$ is just perpendicular to the segment.

- Contact point $\mathbf{c}$: it is defined as the midpoint of the segment p_1p_2, i.e. $\mathbf{c} = (\mathbf{p}_1 + \mathbf{p}_2)/2$. Both the normal direction and contact point define *the normal contact line* which is the line along the normal direction but passes the point $\mathbf{c}$:
$$\mathbf{x}(\lambda) = \mathbf{c} + \lambda\mathbf{n} \tag{24}$$



(a) (b) (c)

Figure 10: Contact modelling of ellipses: (a) Intersection method; (b) Geometric potential method; (c) Midway method

where $\lambda \in (-\infty, +\infty)$ is a free parameter.

- Contact overlap δ: the normal contact line will intersect the ellipses $f_1(\mathbf{x}) = 0$ and $f_2(\mathbf{x}) = 0$ at two points $\mathbf{c}_1$ and $\mathbf{c}_2$ respectively:

$$f_1(\mathbf{c}_1(\lambda_1)) = 0; \quad f_2(\mathbf{c}_2(\lambda_2)) = 0 \tag{25}$$

Then the overlap is the distance between $\mathbf{c}_1$ and $\mathbf{c}_2$

$$\delta = \|\mathbf{c}_1 - \mathbf{c}_2\| \tag{26}$$

This contact modelling method for two ellipses is illustrated in Figure 10(a).

Note that in the original development [25, 27], δ is calculated based on the radii of curvature of the two ellipses at the vicinity of contact as the contact force is also assumed to be related to the equivalent radius of curvature following the extension of the Hertz contact model to ellipses (see [48] for details). The above procedure is an equivalent way to compute the overlap. A similar scheme is also adopted in [28] to use the intersection points to determine the normal direction and contact point, but no explicit formula is given to compute the overlap. Instead an implicit procedure is used through an incremental update of the normal contact force.

The key step of this intersection method is to effectively compute the two intersections. Simultaneously solving two ellipse equations can be reduced to find two correct roots of a quartic equation. Although an analytical form for the roots is available and thus the problem is mathematically not difficult, a numerical problem [27] may occur when: 1) the overlap is small, as it may lead to two numerically close roots which may degrade computational accuracy when the relative orientation of the two ellipses is small; or 2) the eccentricity of an ellipse is very small. It was the attempt to overcome the aforementioned numerical difficulty that had led to the development of the next contact modelling method.

The main limitation of this intersection based method is that it cannot be extended to 3D ellipsoids, as now the intersection of two ellipsoids is a closed 3D curve which does not lie on a 3D plan, and thus cannot define the normal direction. An intersection curve based general contact modelling method has only been developed recently within the contact volume based energy-conserving contact model [68], which will be covered in Section 6.2.

4.1.2 Geometric Potential Method

The geometric potential based contact method is first proposed in [29, 30], and also derived in a slightly different manner in [28]. It attempts to overcome the computational accuracy

degradation issue due to solving the ill-conditioned quartic equation to find the intersections of two ellipses in the above method.

The crucial idea is to view each ellipse in the local form $f(\mathbf{x}) = (x/a)^2 + (y/b)^2 - 1 = 0$ as a domain with a potential field $c = f(\mathbf{x})$, which is negative inside the ellipse, positive outside and zero on the boundary. This concept is presented in Section 2.5 under the level-set representation of shapes. As this potential f is purely geometric and has no physical meaning attached to it, it is first coined as the *geometric potential* in [30].

The main difference from the previous intersection method is, instead of finding the intersection points, 'two special points' $\mathbf{c}_1$ and $\mathbf{c}_2$ are determined in the following way: $\mathbf{c}_1$ (or $\mathbf{c}_2$) is a point at the boundary of the first (or second) ellipse but has the minimal geometric potential in the field of the second (or first) ellipse:

$$\begin{cases} \mathbf{c}_1 = \underset{\forall \mathbf{x}: f_1(\mathbf{x})=0}{\operatorname{argmin}} f_2(\mathbf{x}) \\ \mathbf{c}_2 = \underset{\forall \mathbf{x}: f_1(\mathbf{x})=0}{\operatorname{argmin}} f_1(\mathbf{x}) \end{cases} \tag{27}$$

The contact point $\mathbf{c}$ is taken as the middle of $\mathbf{c}_1$ and $\mathbf{c}_2$: $\mathbf{c} = (\mathbf{c}_1 + \mathbf{c}_2)/2$. The normal direction $\mathbf{n}$ can be defined in two slightly different ways [29]:

$$\mathbf{n} = (\mathbf{c}_2 - \mathbf{c}_1)/\|\mathbf{c}_2 - \mathbf{c}_1\| \tag{28}$$

or

$$\mathbf{n} = (\mathbf{n}_2 - \mathbf{n}_1)/2 \tag{29}$$

where $\mathbf{n}_1 = \nabla f_1(\mathbf{c}_1)/\|\nabla f_1(\mathbf{c})\|$ and $\mathbf{n}_2 = \nabla f_2(\mathbf{c}_2)/\|\nabla f_2(\mathbf{c})\|$ are the unit outer normals of the two ellipses at $\mathbf{c}_1$ and $\mathbf{c}_2$ respectively. The overlap is defined in [29] as the distance between $\mathbf{c}_1$ and $\mathbf{c}_2$: $\delta = \|\mathbf{c}_1 - \mathbf{c}_2\|$.

This geometric potential based modelling method for two ellipses is illustrated in Figure 10(b) where the contact normal is defined in the first way (28) above.

As the gradients of the two ellipses (or their geometric potential fields) share the same (or opposite) direction at $\mathbf{c}_1$ (or $\mathbf{c}_2$), the problem (27) of finding $\mathbf{c}_1$ (or $\mathbf{c}_2$) can be re-written as solving the following two equations where λ_1 and λ_2 are two unknown scalars:

$$\begin{cases} \nabla f_1(\mathbf{x}) = \lambda_1 \nabla f_2(\mathbf{x}) \\ f_1(\mathbf{x}) = 0 \end{cases} \quad \text{or} \quad \begin{cases} \nabla f_1(\mathbf{x}) = \lambda_2 \nabla f_2(\mathbf{x}) \\ f_2(\mathbf{x}) = 0 \end{cases} \tag{30}$$

As noted in [30], solving (30) can again be reduced to finding the roots of a quartic equation which is more reliable than solving (23), and thus better accuracy can be obtained. However, as two sets of equations need to be solved, the total computational cost now doubles.

The main advantage of this geometric potential method is that it can be extended to 3D ellipsoidal contacts in a straightforward manner as stated in [30]. This method is also applied to the contact between egg-shapes in [40]. More importantly, this method has introduced the important concept of 'geometric potential' which in principle can apply to any shape, including ellipsoids and super-quadrics, and thus f_1 and f_2 denote the functions of two CFR shapes in either 2D or 3D cases. An equivalent concept called a 'potential particle' is introduced in [116] where potentials for polygons/polyhedra and more complicated shapes are defined. The details will be covered in Sections 4.2.3 and 4.4.

A slightly different solution scheme, called the 'midway method', is proposed in [116] where the contact point $\mathbf{c}$ is first determined as the point that has the equal potential of $f_1(\mathbf{x})$ and

$f_2(\mathbf{x})$ but the minimal total potential of $f_1(\mathbf{x}) + f_2(\mathbf{x})$. This can be formulated as solving a constrained optimisation problem:

$$\begin{cases} \min f_1(\mathbf{x}) + f_2(\mathbf{x}) \\ \text{s.t. } f_1(\mathbf{x}) - f_2(\mathbf{x}) = 0 \end{cases} \tag{31}$$

Introducing a Lagrange multiplier λ to convert the above problem into an unconstrained optimisation problem leads to the condition

$$(1+\lambda)\nabla f_1 + (1-\lambda)\nabla f_2 = 0 \tag{32}$$

Further introducing a parameter $\mu^2 = (1-\lambda)/(1+\lambda)$ gives

$$\begin{cases} \nabla f_1(\mathbf{x}) + \mu^2 \nabla f_2(\mathbf{x}) = 0 \\ f_1(\mathbf{x}) - f_2(\mathbf{x}) = 0 \end{cases} \tag{33}$$

This is a system of equations to be solved for the contact of super-quadrics adopted in [35], and also used in [36]. The introduction of μ^2, first in [63], is to ensure that the normals of f_1 and f_2 at the point $\mathbf{c}$ are in opposite directions. After $\mathbf{c}$ is obtained, the unit normal direction is simply $\mathbf{n} = \nabla f_2(\mathbf{c})/\|\nabla f_2(\mathbf{c})\|$. The normal contact line can be formed as in (24). Then the overlap can be evaluated exactly following the same procedure as in the intersection method above to determine the overlap, i.e. the distance of two points $\mathbf{c}_1$ and $\mathbf{c}_2$, which are the intersection points of the normal contact line with the two ellipses respectively.

This 'midway' method for two ellipses is illustrated in Figure 10(c). Although the method appears to perform well in practice, it has not been theoretically proved that the common point $\mathbf{c}$, satisfying (30) or (33), is guaranteed to be located within the contact region of the two shapes. In fact, an earlier but similar method, proposed in [105] in the field of molecular dynamics, introduces a so-called *contact potential function* for two ellipsoids in contact.

In this contact potential function method, the contact point is obtained as the common touch point when the two shapes are scaled equally. Mathematically, it has been proved [105] that the contact potential function is maximal at this point and that both functions f_1 and f_2 at the point are equal. This proves that the common point $\mathbf{c}$ is the solution of problem (30) or (33) must be within the contact area of the two shapes, thereby providing the theoretical support to the correctness of the problem (30) or (33). Because the point $\mathbf{c}$ is not exactly located the midway between $\mathbf{c}_1$ and $\mathbf{c}_2$, but rather has an equal potential, this particular variation of the geometric potential method may be more appropriately called the *equal potential method.*

4.1.3 Common Normal Method

The *common normal* method is first proposed also in [29], motivated by overcoming a deficiency of the original geometric potential method mentioned above where the normal vectors at the two special points, $\mathbf{n}_1$ and $\mathbf{n}_2$, to define the overall normal direction (29) may not be parallel to each other. The solution is to find two new special points $\mathbf{c}_1$ and $\mathbf{c}_2$ such that the normal directions at these points are parallel to the line that passes through the two points. This common line is thus also the contact normal - hence the resulting method is called the common normal method, and is illustrated in Figure 11.

Mathematically, $\mathbf{c}_1$ and $\mathbf{c}_2$ are the solution to the following problem:

$$\begin{cases} \nabla f_1(\mathbf{c}_1) + \mu^2(\mathbf{c}_2 - \mathbf{c}_1) = 0; \quad \nabla f_1(\mathbf{c}_1) + \nu^2(\mathbf{c}_1 - \mathbf{c}_2) = 0 \\ f_1(\mathbf{c}_1) = 0; \quad f_2(\mathbf{c}_2) = 0 \end{cases} \tag{34}$$

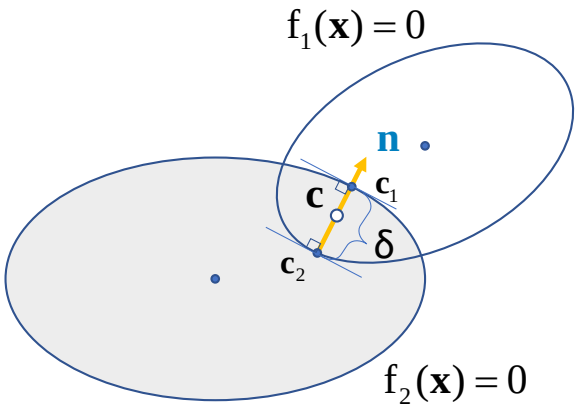


Figure 11: The common-normal method for ellipses

After having obtained $\mathbf{c}_1$ and $\mathbf{c}_2$, the contact point, contact normal and overlap can be determined as the same in the original geometric potential method: $\mathbf{c} = (\mathbf{c}_1 + \mathbf{c}_2)/2$, $\mathbf{n} = -\nabla f_1(\mathbf{c}_1)/\|\nabla f_2(\mathbf{c}_2)\| = \nabla f_2(\mathbf{c}_2)/\|\nabla f_2(\mathbf{c}_2)\|$, and $\delta = \|\mathbf{c}_1 - \mathbf{c}_2\|$.

The only issue with the presentation of this method in [29] is that no rigorous argument is provided to prove that there do exist two such points $\mathbf{c}_1$ and $\mathbf{c}_2$ at which the normal directions of the ellipses are parallel to the line that passes through them. Indeed, this is a well known fact that is related to the definition of a gap/overlap function between curves/surfaces in many fields. In contact mechanics [49], for instance, the gap or overlap of two smooth convex curves/surfaces, denoted as Γ_1 and Γ_2, is their *minimal distance* d defined as

$$d(\mathbf{x}_1, \mathbf{x}_2) = \max_{\mathbf{x}_1 \in \Gamma_1} \left\{ \min_{\mathbf{x}_2 \in \Gamma_2} \|\mathbf{x}_2 - \mathbf{x}_1\| \right\} \tag{35}$$

i.e. there exist two points, $\mathbf{x}_1$ on Γ_1 and $\mathbf{x}_2$ on Γ_2, such that each is an orthogonal projection on the other curve/surface. This essentially proves that the common normal method is theoretically correct.

Note that a similar common normal concept is also adopted in [63], but it is applied to find the time instance in the past that the two particles are just in contact at one common contact point where the two normals of the particles at the point are parallel. As the motion history of the two particles is involved, this approach is not ideal for contact modelling and coding. In addition, theoretically the mechanical energy may not be conserved when no damping or energy dissipation mechanism is considered - see Section 6 for an in-depth discussion.

Finally, the common normal approach is the most general approach and applicable to any convex CFR shape, but most importantly, it can be proved that it is also energy-conserving within the energy-conserving contact theory [67], as will be highlighted in Section 6.3.

4.1.4 Computational Issues

There are a number of computational issues that affect the efficiency of each contact method mentioned above and therefore are worth considering below. Note that except for the intersection method, the other methods are applicable to 3D cases and thus the related discussions below are for 3D in general.

i) Solution procedure. In the three methods, the key computational issue is how to solve the system of non-linear equations of either (23), (33), or (34). In general, they can be

solved in many different ways, but except for simple cases, such as ellipses, the most effective solution approach is using the Newton-Raphson (NR) method to find the solution.

In the original solution procedure proposed in [27] for ellipses in the intersection method, the problem (23) is reduced to finding the roots of a quartic equation. Of course, the problem can also be solved numerically, but similar numerical difficulties mentioned before will remain.

For the geometric potential method, there are *four unknown* variables (the contact point $\mathbf{c}$ and the parameter μ) involved and four equations available in (33). Thus the problem can be solved using any available solution procedure, including the NR method. While for the common normal method, there are *eight unknown* variables (the two points $\mathbf{c}_1$ and $\mathbf{c}_2$, and the parameters μ and ν) involved, which could be found by solving the eight equations of (34). It seems, therefore, that the geometric potential method is numerically more efficient than the common normal method. However, a more effective solution or contact detection algorithm exists for the common normal method, as outlined below.

ii) Alternative contact detection algorithm for the common normal method. This algorithm is proposed in [83], when applying the common normal concept to super-quadrics contact problems. It utilises the fact mentioned above that the correct points $\mathbf{c}_1$ and $\mathbf{c}_2$ minimises the overlap distance $\|\mathbf{x}_2 - \mathbf{x}_1\|$, and thus converts the original problem into a global optimisation problem. The proposed solution procedure consists of the following key steps:

1) instead of finding $\mathbf{c}_1$ and $\mathbf{c}_2$ directly, it seeks to obtain the unit normal direction $\mathbf{n}$ as the primary variables;

2) introduce two angles α_1 and α_2 as parameters to define the unit normal direction $\mathbf{n}(\alpha_1, \alpha_2)$;

3) utilising the fact that the point on the boundary of an super-quadric which has the outer normal equal to $\mathbf{n}$ can be explicitly expressed in terms of $\mathbf{n}$, two points $\mathbf{x}_1(\alpha_1, \alpha_2)$ and $\mathbf{x}_2(\alpha_1, \alpha_2)$ on the two super-quadrics respectively can be obtained, thereby automatically satisfying the equations $f_1(\mathbf{x}_1) = 0$ and $f_2(\mathbf{x}_2) = 0$; and

4) the true values α_1 and α_2 are found by minimising the distance $d(\alpha_1, \alpha_2) = \|\mathbf{x}_1(\alpha_1, \alpha_2) - \mathbf{x}_2(\alpha_1, \alpha_2)\|$ using a combined NR and Levenberg-Marquardt method to achieve good convergence and robustness.

Using this contact detection algorithm, the problem (33) is reduced to a global optimisation problem with only *two variables* instead of the original eight unknowns. The critical step is that there exists a smooth invertible mapping between a unit normal vector and the corresponding boundary point on the surface. Thus, this contact detection algorithm can also be applicable to any shape that has such an invertible mapping. For instance, the algorithm is applied to poly-super-ellipsoid in [110] with a minor extension that the corresponding surface point for a given unit normal vector is obtained using the so-called support point function which will be covered in Section 4.5.2.

iii) Initial guess. Iteratively solving (23), (33), or (34) may suffer from divergence if the initial guess is not properly chosen. A poor initial guess can also lead to an incorrect solution: for the common normal method, there exists another pair of points $\mathbf{c}_1$ and $\mathbf{c}_2$ that satisfy the equations (34) but maximise the distance or overlap instead and hence should be excluded. Thus, various techniques are needed to provide a better initial guess to increase the robustness of the contact detection approach concerned. A detailed procedure is proposed in [35] to deal with the initial guess issue for contact modelling of super-quadrics. Another scheme that can also provide good initial guesses will be mentioned in Section 4.5.2.

iv) Temporal coherence. It will be numerically inefficient if the same measure has to be taken to ensure a good initial guess at each time instance for each contact pair. This problem can be largely eliminated if the solution at the previous time step is used as the initial guess for the current step. This essentially utilises the *temporal coherence* feature of contacting pairs in DEM: since the time step used in DEM is normally sufficiently small to ensure the stability of time integration, two particles in contact would be only subjected to a relatively small spatial change during one time step. This temporal coherence feature has been utilised not only in the local contact resolution but also in the global search phase. The drawback of considering temporal coherence in the DEM code implementation is that some intermediate results have to be stored as additional history information, which inevitably increases the memory requirement. Additionally, different contact types of shapes may need to store different sets of data, thereby increasing the complexity of the coding.

v) No contact cases. Ideally a no-contact case should be identified as soon as possible in the contact detection procedure to reduce computational costs. However, most equation solving based procedures can only detect a no-contact case at a much later stage. For instance, in solving the intersection points of two ellipses from finding the roots of the quartic equation, a no-contact case can only be identified when no roots can be found, and thus the total computational operation involved is almost the same as for a contact case. Consequently an earlier detection mechanism embedded in contact detection for no-contact cases would be desirable. The alternative contact detection algorithm proposed in [83] has such a stopping criterion as soon as a no-contact case is detected. A more generic earlier termination approach for convex shapes will be introduced in Section 4.5.2.

4.2 Contact Modelling for Polygons/Polyhedra

Polygons/polyhedra are introduced into DEM to represent particles with sharp corners/edges and can also be used as generic representations for non-spherical and complex shapes including wall boundaries, as mentioned in Section 2. However, the derivative discontinuity at vertices/edges imposes considerable numerical difficulties for contact detection algorithms where derivatives are required, such as those for CFR shapes just discussed. Thus the contact features defined and determined in the preceding subsection are no longer applicable to polygons/polyhedra.

Consequently, the fundamental challenge to develop contact detection algorithms for polygons/polyhedra is how to define and then determine the contact (geometric) features appropriately. For instance, how to extend the definition of the overlap from discs/spheres to polygons/polyhedra, and which direction should be defined as the contact normal?

In this subsection below, three overlap based contact modelling methods will be briefly reviewed. A more rigorous and probably more effective approach will be covered in Section 4.5.1. More importantly this fourth method is applicable not only to convex polygons/polyhedra, but also to all convex shapes.

4.2.1 The Common-Plane Method

The first contact detection approach for polygons/polyhedra was proposed in [56, 57] in 1988, even earlier than the intersection method for ellipses which was proposed in [25, 27] in the early 1990s. This approach has since been referred to as the *common-plane* (CP) method. The central idea is to locate a common plane of two polygons/polyhedra, from which the required contact geometric features can be obtained. The method is applicable to both

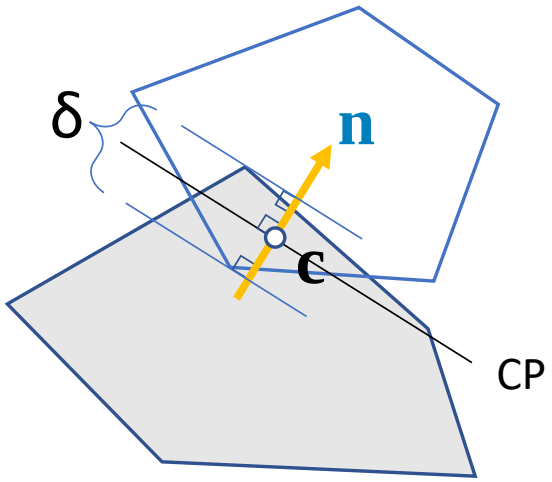


Figure 12: The common-plane method for two convex polygons

convex and concave cases, provided that a concave polygon/polyhedron can be decomposed into convex sub-polygons/polyhedra.

It is stated in [56] that '*a common plane is a plane that, in some sense, bisects the space between the two contacting particles*'. When the CP is determined, the normal of the CP will be the contact normal and the CP itself will be the tangent plane. The contact point is taken as the reference point on the CP at which the contact forces/moment will apply. The overlap is the sum of the distances of the two shapes to the CP. Figure 12 illustrates this common-plane concept.

A numerical procedure is proposed in [56] to iteratively determine the CP and the reference point. Before outlining the procedure, the distance of a polygon/polyhedron to a (common) plane is re-defined here which is different from the original version in order to simplify the description below without distinguishing between contact and separation cases. This definition also makes it more consistent with what will be used in later sections.

We state that the distance of a polygon/polyhedron (or shape) to a plane is *the minimum distance that the shape can move only in translation to just touch the plane*. The distance is taken as positive if the plane is in contact with the shape and as negative otherwise. For polygons/polyhedra, the distance can be readily obtained by evaluating the distances of all vertices to the plane (with a properly defined outer normal) via dot products.

The key steps of the original CP algorithm [56] are: choose an initial CP with a reference point; and then subsequently update the position (i.e the reference point) and orientation of the CP to reduce the overlap until a prescribed tolerance is reached. These steps are outlined below for a pair of convex (sub-)polygons/polyhedra, while full details of the algorithm can be found in [56].

1) Initialisation: Let $\mathbf{c}_A$ and $\mathbf{c}_B$ be the centroids of the two polygons/polyhedra, A and B. The initial CP is located at their midpoint (i.e. initial reference point): $\mathbf{c}_0 = (\mathbf{c}_A + \mathbf{c}_B)/2$ with their link as the normal: $\mathbf{n}_0 = \mathbf{c}_B - \mathbf{c}_A$. Compute the distances of A and B: d_A and d_B, and the initial overlap $d_0 = d_A + d_B$.

2) Update at iterative step i:

   - Position (reference point) update: $\mathbf{c}_i = \mathbf{c}_{i-1} + d_0\,\mathbf{n}_{i-1}/2$
   - Orientation update: i) Identify two arbitrary orthogonal axes on the current CP,

and set a (small) angle perturbation size θ to alter the normal $\mathbf{n}$ around these two axes, each with both positive and negative senses in a total of four separate perturbations. Recompute the total overlap in each perturbation; ii) If the overlap decreases most in one perturbation, accept the perturbed normal as $\mathbf{n}_i$ and also the correspond overlap as d_i, set $i = i + 1$ and go back to 2) for next iteration; iii) Otherwise when the overlap does not decrease for all four perturbations, half θ, and repeat the current step; iv) The whole iteration stops when the perturbation size θ at the current step is reduced to be smaller than the allowed minimum value θ_{min}. Then the latest normal, reference point and overlap found are taken to be the final contact geometric features. The iteration can also terminate if the overlap becomes zero or negative, meaning a non-contact case.

Clearly the above numerical procedure to determine the CP is of an iterative nature. When a contact between two polygons/polyhedra has been established, the CP from the previous step can be taken as the initial CP for the current time step to reduce the iterations. Some additional measures are also introduced [56] to make the algorithm more robust.

This CP method is a significant advance over other algorithms [59, 60] at the time as it reduces the polygon/polyhedron contact detection problem to a much simpler plane-to-polygon/polyhedron contact detection problem, and those non-trivial contact situations such as vertex-to-vertex, vertex-to-edge or edge-to-edge contacts in 3-D cases could now be tackled in an easy and systemic way.

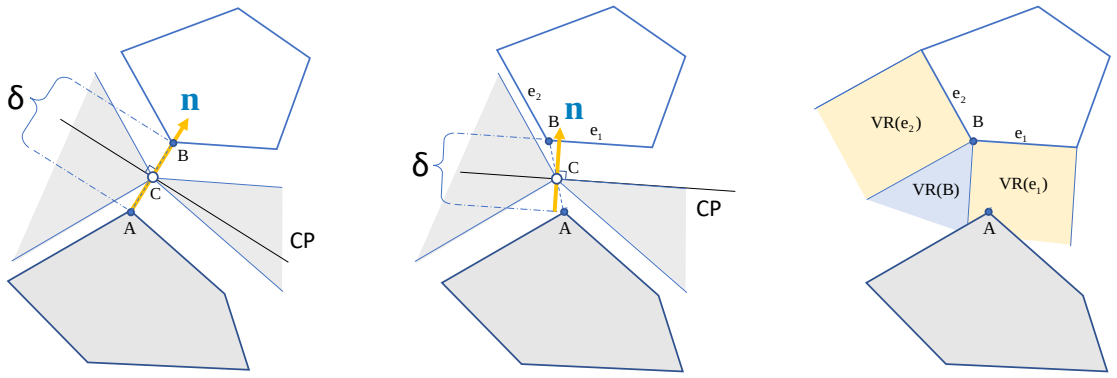
However, the above algorithm to locate the CP is also of an *heuristic* nature. There is no theoretical guarantee that a correct CP can always be found using the procedure outlined above, as pointed in [58]. There is also no guarantee that the iterations can stop within in a certain bound. In addition, the result may be affected by the choice of the initial CP. Furthermore, the total number of iterations will depend on the initial CP, the maximum size of angle perturbation and the allowed minimum perturbation value. In general, a large number of iterations is normally expected, and can be especially high when the CP is determined for the first time for a pair of polygons/polyhedra and, at the same time, very different from the actual CP.

4.2.2 The Fast Common-Plane Method

An improved CP method, called the *fast CP method*, is proposed in [120], which is based on the proof that when two polygons are *not in contact*, the correct CP can be obtained in a limited number of steps. In particular, when the two closest vertices among all the vertex pairs, one from each polygon, are obtained (in no more than $n+m$ iterations, where n and m are the numbers of vertices of the polygons respectively), the correct CP can be determined in no more than five steps. A similar conclusion is also drawn for polyhedra.

This fast CP method is a major advance over Cundall's original version [56] as it significantly reduces the number of iterations required to obtain the CP, and more importantly it is guaranteed that the correct CP can always be found. Figure 13(a) and (b) illustrate the key observation made in [120] that when A and B are the two closest points, the CP must lie in the shadowed area formed by the four edges incident to A and B. Then the CP can be found by checking no more than five cases. The procedure to deal with two convex polyhedra can also be found in [58].

Nevertheless, the fast CP method can be easily understood geometrically almost without invoking the rather tedious proof presented in [120]. Locating the CP is essentially equivalent



(a) (b) (c)

Figure 13: The fast common-plane method for two convex polygons: (a) the PB of AB is the CP; (b) the CP parallels side e_1; (c) Voronoi regions of vertex B and sides e_1, and e_2

to finding two points on the two polygons each that their distance is the shortest distance between the two polygons. As A and B are the two vertices that have the shortest distance among all vertices from both polygons, there are five possible cases in total: 1) A and B are the closest points between the two polygons; 2) two cases - A is the closest to one of the two edges incident to B; or 3) two cases - B is the closest to one of the two edges incident to A. In case 1), the CP is the perpendicular bisector (PB) of the segment from A to B; while in the next two cases, the CP is the PB of A and its orthogonal projection on one of the edges of B; while in the last two cases, the CP is the PB of B and its orthogonal projection on one of the edges of A. In Figure 13(a), the CP is the PB of segment AB, while in Figure 13(b), the CP is the PB of vertex A and its orthogonal projection on side e_1.

Moreover, the above observations from a geometric perspective had already been fully exploited in the 1990s for collision detection between (convex) polygons/polyhedra in other fields, including robotic planning, computer games/graphics and computational geometry. Many highly effective algorithms were proposed then, including the Lin-Canny algorithm [76, 77] and the V-Clip [78], to name a few. Central to the success of these algorithms is using the concepts of *Voronoi regions and planes*. By definition [78], the Voronoi region of a feature X (vertex, edge or face) of a polygon/polyhedron, denoted $VR(X)$, is the region outside the polygon/polyhedron that is as close to $\mathbf{x}$ as to any other feature on the polygon/polyhedron. The Voronoi plane $VP(X;Y)$ between two neighbouring features X and Y is the plane containing $VR(X) \cap VR(Y)$.

The Voronoi regions for the top polygon shown in Figure 13(b) are also illustrated in Figure 13(c). As vertex A is within the Voronoi region of side e_1, $VR(e_1)$, the closest feature pair must be vertex A and side e_1, and the actual CP is the same as obtained in Figure 13(b). Thus it is easy to understand that the fast CP method is simply one possible implementation to determine the closest features of the two polygons.

Although the fast CP method is a much improved method for locating the CP, it works only for the two polygons/polyhedra that are *not in contact*, which is a case that is not particular interested in DEM. In order to be able to handle cases of contacting polygons/polyhedra, the above method is extended in [120] by first separating the two shapes apart in a proper distance, so that now the resulting non-contact case can be tackled by the original procedure. It is proposed that the two shapes should be moved away based on their velocities in the previous time instance. After the CP is located for this non-contact situation, the pair are moved back to their original positions to compute the contact features required.

This extended version for treating contacting cases seems reasonable but it may fail, since the obtained CP and corresponding contact features strongly depend on the relative direction along with the two shapes are moved away and also how far the movement is. Consequently the obtained contact features may no longer be correct. Another related scheme is proposed in [121], but the problem concerned remains unsolved. No further elaboration on the issue will be provided here though.

Another non-iterative algorithm for finding the common plane is also proposed in [58] based on two analytical developments. It is claimed [58] that the proposed method inherits all the key features of the common plane approach, but overcomes its iterative weakness. Also the method can resolve contact detection problems for both overlapping and non-overlapping conditions accurately based on region partition similar to the Voronoi regions. The proposed computational procedure seems complicated though.

4.2.3 The Method Based on Potential Particles

A special contact detection method for convex polygons/polyhedra is proposed in [118], where three numerical techniques are presented to address three issues of determining the contact features of two polygons/polyhedra. The method is outlined below. The reader is referred to [118] for full details.

As already introduced in Section 2.3, a polygons/polyhedra can be represented by a set of linear inequality equations with each representing a half-space:

$$\mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) \leq 0 \quad (i = 1, \cdots, N) \tag{36}$$

where N is the number of sides/faces of the polygon/polyhedron, $\mathbf{n}_i$ is the unit outer normal of the i-th side/face and $\mathbf{p}_i$ a point on it.

Determining the contact state between two polygons/polyhedra A and B, i.e. if they are in contact or not, can be formulated as a linearly constrained optimisation problem:

$$\begin{cases} \min\ s \\ \text{s.t.} \quad \mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) < s, \quad i = 1, \cdots, N_A + N_B \end{cases} \tag{37}$$

where s is a scalar, and N_A and N_B are the numbers of sides/faces of the two shapes respectively. This optimisation is also a standard linear programming (LP) problem and can be solved by the classic simplex method, for instance. A contact exists between the shapes if the minimal objective value obtained $s < -\epsilon$, where ϵ is a prescribed positive tolerance for the overlap between the two particles.

When the contact is established, the next step is to find the contact point $\mathbf{x}_c$. The point is chosen to be the so-called 'analytical centre' which is the solution to another optimisation problem:

$$\mathbf{x}_c = \arg\max_{\forall \mathbf{x}} \sum_{i=1}^{N_A+N_B} \log(-\mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i)) \tag{38}$$

The sum of the log-function above is a concave function which could be interpreted as a potential associated with the total repelling force field generated by the constraints. The analytic centre is the point which maximises the potential or minimises the repelling forces. Although this optimisation can be converted into another LP problem, finding the solution is not straightforward, and many numerical issues need to be addressed. The details can be found in [118].

After $\mathbf{x}_c$ is obtained, the contact normal is assumed to be the gradient of a potential function f at the point:

$$\mathbf{n} = \nabla f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_c} \tag{39}$$

Following the concept of a 'potential particle' introduced in [116], the function f is chosen to be the material property weighted sum of the potentials of the two polygons/polyhedra but modified by introducing a constant term r. The modified potential for one polygon/polyhedron has the form

$$f(\mathbf{x}) = \sum_{i=1}^{N} < \mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) + r > \tag{40}$$

Some details of this potential will be explained in Section 4.2.3. When the contact normal is evaluated, *"the overlap distance between the two particles can be obtained by using bracketing algorithms along the direction of the contact normal"* [118].

Although a few new ideas and concepts are introduced in this very special method, there is a need to solve two optimisation problems (37) and (38), and also to appropriately choose the parameter r to construct a valid potential function for determining the contact normal. In addition, some numerical issues as mentioned in [118] also need to be properly addressed. These concerns make this method less attractive.

4.2.4 Non-overlap Based Contact Modelling Methods

The original and fast CP methods and also the potential particle based method aim to define a set of overlap oriented contact features, and thus are all classified as overlap based. For convex polygons/polyhedra, a number of contact detection methods were also proposed in around the 2000s, which are based on non-overlap contact features.

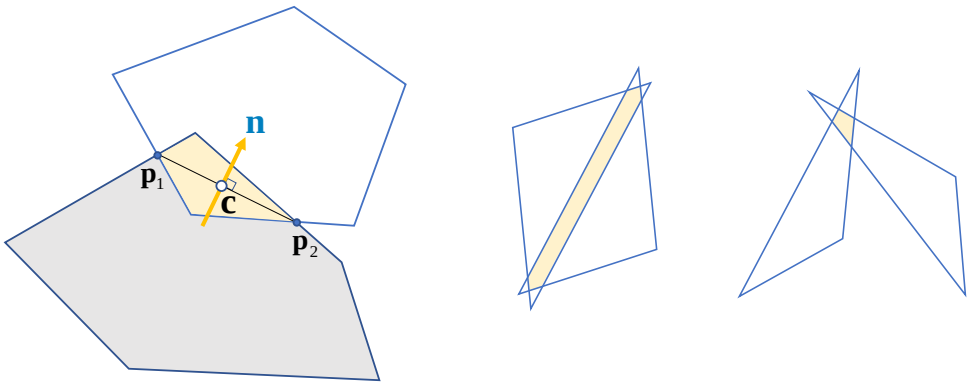
1) Contact-area based methods for polygons.

The first contact area based contact modelling scheme is introduced in [66], where the two intersection points, as shown in Figure 14(a), are used to define the tangential direction, and hence the normal direction as the perpendicular direction. The contact force is assumed to be proportional to the contact area formed by the two polygons, and acts at the contact point chosen either as the midpoint of the two intersection points, or the centroid of the contact area. The numerical examples conducted show not much difference between these two choices. Although this scheme appears to work well for general polygon/polygon contact scenarios as shown in Figure 14(a), it fails for the two undesirable scenarios illustrated in Figure 14(b) and (c) where the number of intersection points in each case is four so the normal or tangential direction cannot be readily defined.

Mainly concerned with a possible contact normal jump for a corner-conner contact of two polygons between two consecutive time instances, an energy potential/function based contact modelling scheme is proposed in [98]. It assumes that the current contact status is fully determined by an energy potential w which is a function of the overlap area A_c, and that the normal contact force $\mathbf{F}_n$ is the negative gradient of w, i.e.

$$\mathbf{F}_n = -\nabla w(A_c) \tag{41}$$

Following a tedious derivation [98], the contact normal $\mathbf{n} = \mathbf{F}_n/\|\mathbf{F}_n\|$ is found to be perpendicular to the segment linking the two intersection points, and the contact point should be the midpoint. This essentially ratifies the (most) choice of contact features in the previous method [66]. Also, the magnitude of the normal contact force is proportional to the contact width (the distance between the two intersection points) and/or contact area depending on



(a) (b) (c)

Figure 14: The contact area based method for two convex polygons: (a) General case; (b) first undesirable case; (c) second undesirable case.

the choice of the energy function. It is the first time that the contact width appears in determining the contact force.

Because the contact normal obtained is now orthogonal to the line passing the two intersection points, it will evolve smoothly without jumping when one polygon moves from one side to the other. This new contact area based model can also naturally handle the two undesirable scenarios in Figure 14(b) and (c) without introducing any arbitrary measures. Another significance of this contact model is that it can be extended to any (convex) 2D shapes as any shape can be represented by a polygon with an infinite number of vertices/sides. Also because the required operations associated with polygons can be done very efficiently [99], this contact model for polygons is used to approximately represent super-quadrics in [101] to achieve a higher computational efficiency than the methods introduced in the previous section for CFR shapes.

This initial idea of introducing a contact energy function to characterise a contact has been fully exploited afterwards. It is first extended to convex polyhedra in [100] and then developed into a general contact model for arbitrarily shaped particles in [102]. Finally it leads to the establishment of a general energy-conserving contact theory in [67]. This theory and resulting contact models will be fully covered in Section 6.2.

2) Contact-volume based methods for polyhedra.

The extension of the above contact area based model [98] from 2D to 3D is briefly discussed here. The contact-volume based method is first proposed in [100] for convex polyhedra and further improved in [102]. Similar to the polygon case, it is now assumed that the current contact status between two polyhedra is fully determined by an energy potential w which is a function of the contact volume V_c, and that the normal contact force $\mathbf{F}_n$ is the negative gradient of w, i.e.

$$\mathbf{F}_n = -\nabla w(V_c) \tag{42}$$

Then the contact normal direction and point can be obtained following the procedures outlined in [102]. This contact volume based model is also adopted in [111, 112]. However, the contact point in [111] is taken as the centroid of the overlap volume. More details and issues related to this contact model are fully described in Section 6.2

4.3 Contact Modelling for Dilated and Minkowski-Sum Shapes

As mentioned in Section 2.4, dilated shapes are first introduced in [106] into DEM. The most commonly used dilated shape is the dilated polyhedra with spheres, or sphero-polyhedra. This can also be viewed as a way to overcome the difficulties associated with the singularity of polyhedra when modelling their contact. The sharpness of vertices and edges can be adjusted by varying the dilation radius. As each vertex is dilated into a sphere segment and an edge becomes a cylinder segment, or into a capsule when considering the spherical segments of its two end vertices. Each flat face/surface of the polyhedron remains to be a flat face/surface but having been shifted and tangentially connecting the dilated cylinders/capsules of its associated edges.

Contact modelling between two sphero-polyhedra can now be reduced to a few simple contact cases: sphere-sphere, sphere-capsule, capsule-capsule, sphere-plane, capsule-plane, and plane-plane. These can all be dealt with relatively easily and some closed-form solutions exist for some of these cases. Note that cylinders have been replaced by capsules in the above contact detection to reduce the complexity of the contact between two cylinders. The contact between cylinders is surprisingly non-trivial because of the presence of sharp edges/end discs. Contact modelling methods for cylinders will not be reviewed in detail, but some relevant work can be found in [143, 144, 145, 104].

When the dilating shape is not a sphere but another convex shape, the contact between two such dilated shapes have to be grouped into a number of simpler cases depending on both the basic and dilating shapes concerned. If dilated shapes can be constructed using Minkowski-sum, as outlined in Section 2.4, a generic contact modelling may be used to tackle the contact between such shapes, provided that a support function can be defined for both the basic and dilating shapes, or for each constitutive shape in a hierarchical Minkowski-sum shape. This general contact modelling approach will be reviewed in Section 4.5.2.

4.4 Contact Modelling of Potential Particles

The *'potential particle'* is proposed in [116] which is conceptually equivalent to a particle with a defined geometric potential. Thus potential particles will be no different from basic CFR shapes, such as ellipses, ellipsoids and super-quadrics. The contact detection algorithms outlined in 4.1.2 remain to be valid for such potential particles. The main contribution of this potential particle concept is that it provides a means to define geometric or potential functions for convex polygons/polyhedra and also an ability to define rounded polygons/polyhedra that are different from sphero-polyhedra covered above.

Firstly, for a half-space defined by the plane $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$ where $\mathbf{n}$ is the unit outer normal and $\mathbf{p}$ is a point on the plane, its potential is defined as [116]

$$f(\mathbf{x}) =< \mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) > \tag{43}$$

where <> are Macaulay brackets with the property: $< x >= x$ if $x > 0$, and $< x >= 0$ otherwise. Thus for any point $\mathbf{x}$ on and within the half space, $f(\mathbf{x}) = 0$, while for an outside point $\mathbf{x}$, $f(\mathbf{x})(> 0)$ is its distance to the plane.

Then the (geometric) potential of a polygon/polyhedron can be expressed as

$$f(\mathbf{x}) = \sum_{i=1}^{N} < \mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) > \tag{44}$$

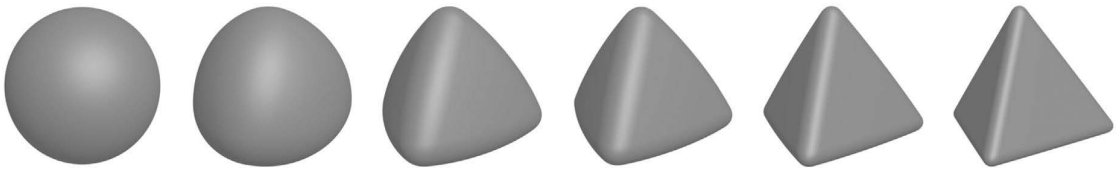


Figure 15: Evolution of rounded sphero-tetrahedral shapes when the control parameter k changes from 1.0 to 0.001 [108]

where N is the number of sides/faces of the polygon/polyhedron, $\mathbf{n}_i$ and $\mathbf{p}_i$ are the unit outer normal of and a point on the i-th side/face, respectively.

Now a rounded potential polyhedron with dilation radius r can be defined by the potential

$$f(\mathbf{x}) = \sum_{i=1}^{N} < \mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) >^2 - r^2 \tag{45}$$

Note that this rounded potential polyhedron is almost identical to a dilated polyhedron using the same basis polyhedron and the dilating sphere, except at the rounded corners, which now become ellipsoids instead of spheres in the dilated polyhedron. Although a small change, this significantly increases the difficulty in resolving contact between rounded potential polyhedral particles because the spheres in dilated polyhedra become ellipsoids now. This is also applied to the following rounded-sphero-polyhedral shapes.

The flat surfaces of the above rounded potential polyhedron can also be rounded by adding the weighted quadratic potential of a 'shadow' sphere of radius R in the form

$$f(\mathbf{x}) = (1-k)\left(\sum_{i=1}^{N} \frac{< \mathbf{n}_i \cdot (\mathbf{x} - \mathbf{p}_i) >^2}{r^2} - 1\right) + k\left(\frac{\mathbf{x} \cdot \mathbf{x}}{R^2} - 1\right) \tag{46}$$

where parameter $0 \le k \le 1$ controls the fraction of the 'sphericity' to be added. The resulting shape is called a *rounded-sphero-polyhedron*. Figure 15 illustrates how a rounded potential tetrahedron changes (from the right to left) almost to a sphere when the parameter k increases from 0.001, 0.01, 0.05, 0.1, 0.5 to 1.

As demonstrated in [107, 108], the contact between two rounded-sphero-polyhedra A and B can be formulated as the constrained optimisation problem (31) and solved by the same midway or equal potential method as introduced in 4.1.2. However, a different approach is taken in [119]. Because of the discontinuity in the second derivative due to the Macaulay brackets, the problem (31) is reconstructed into a second-order cone programming for a new minimisation problem but with the presence of both equality and in-equality constraints [119]. After the contact point $\mathbf{x}_c$ is found as the solution to the problem (31), the contact normal $\mathbf{n}$ is taken as the gradient of the potential function f_A of (46) at the point:

$$\mathbf{n} = \nabla f_A(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_c} \tag{47}$$

Next, find the two intersections of the normal contact line with the two shapes $f_A(\mathbf{x}_A) = 0$ and $f_B(\mathbf{x}_B) = 0$. Then the overlap is the distance between these two points: $\delta = \|\mathbf{x}_A - \mathbf{x}_B\|$.

The full details of the algorithm are given in [119]. A slightly different contact detection approach is proposed in [117].

4.5 Contact Modelling of Mixed Pairs of Shapes

The contact modelling methods having been reviewed so far are mainly for shapes of the same type. When two different shapes are in contact, a special contact modelling scheme often needs to be developed to account for the geometric uniqueness of the shapes concerned and/or for the efficiency of the resulting algorithm.

The first two contact modelling methods covered in Section 4.2, i.e. geometric potential and common-normal based, are sufficiently generic to be applicable to any two (P)CFR shapes and thus greatly simplify and unify the contact detection for (P)CFR shapes, if implemented using a templating computer language, such as C++. Nevertheless, their individual shapes are often explicitly considered in the DEM code implementation of the contact method in order to achieve a high numerical efficiency, which inevitably results in a large number of possible combinations of contact cases to be included, such as sphere-sphere, sphere-ellipsoids, sphere-superquadric, etc.

When further considering the contact between a pair of mixed shapes, such as a sphere-polyhedron, a special contact model explicitly utilising the geometric properties of the shapes has to be developed. As the total number of combined cases increases exponentially with the number of different shapes considered, it makes the development of a general DEM code with a capability of including many different shapes very complicated and tedious.

The situation becomes almost untenable when further considering all possible wall boundaries allowed for DEM simulations in real applications, as wall geometric forms can be much diverse and almost infinitely many so that all the contact combinations between different particles and these wall forms need to be dealt with separately.

Clearly, more manageable solutions must be sought. Two possible solutions are briefly described below.

4.5.1 Unified Polygonal/Polyhedral Contact Modelling Approach

The first possible solution, if computational efficiency is not the primary concern, is to discretise those non-circular/spherical and non-polygonal/polyhedral shapes into polygons/polyhedra. Then only two types of contact need to be considered: 1) disc/sphere-polygon/polyhedron, and 2) two polygons/polyhedra. The second type clearly provides a unified contact modelling strategy for non-spherical shapes of the same or different types.

The same strategy is also applicable to wall boundaries which are often discretised into surface triangular meshes. The contact modelling of spheres and surface triangulated walls is first conducted in [53, 54], where each surface triangle is termed a 'wall element', and various sphere-triangle contact cases are discussed. Note, however, that because a wall can often be non-convex, the contact between a sphere and multiple triangles is a non-convex contact case and thus is not trivial to handle. A contact area ratio weighted approach is proposed in [55] where all possible overlapping cases between a sphere and a triangle are identified and the corresponding contact areas are formulated. This approach is actually the special case of the contact-volume based energy-conserving model proposed in [102, 68] and is reviewed later in Section 6.2.

If the polygons/polyhedra involved in both contact types are convex, the next solution to be discussed below can be employed, while if at least one shape is concave, a few approaches to be reviewed in Section 5 can be applied.

4.5.2 Minkowski-difference Based Contact Modelling Approach

The second possible solution is termed the *Minkowski-difference based contact modelling approach* for convex shapes. The following outline the key components involved in this approach.

i) Minkowski difference. Consider two shapes A and B. First to define their *Minkowski difference*, denoted as $A \ominus B$, to be the Minkowski sum of A and $(-B)$:

$$A \ominus B = \{\mathbf{a} - \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\} \tag{48}$$

The boundary of $A \ominus B$ is denoted as $\partial(A \ominus B)$. We have the following properties: the Minkowski sum/difference of two convex shapes is also convex, and the Minkowski sum/difference of two convex polygons/polyhedra is a convex polygon/polyhedron. In addition, the boundary of $A \oplus B$ or $A \ominus B$ is the convex hull of the boundary/vertex sum of A and B or A and $-B$. This property provides an effective approach to explicitly construct the Minkowski sum or difference of two convex polygons/polyhedra if needed. Furthermore, every boundary geometric feature (vertex, edge and face for 3D) of $A \ominus B$ is formed by two boundary features, one from each shape.

An important statement, which is valid for both convex and concave shapes, is that the contact state of two shapes A and B, in contact or not, is equivalent to determining if the origin $\mathbf{o}$ belongs to their Minkowski difference $A \ominus B$ or not, thereby converting the contact detection of A and B into an equivalent but often simpler point inclusion test:

$$A \cap B \neq \varnothing \iff \mathbf{o} \in A \ominus B \tag{49}$$

Thus, the Minkowski difference provides a conceptually simple and universal approach to determining the contact state of any two shapes. Based on this equivalency, an elegant and often effective algorithm, called the GJK algorithm [80], can be employed to detect if two shapes are in contact or not. One key component of this algorithm is to utilise the so called *support point or function* of a (convex) shape.

ii) Support point/function. For a given vector $\mathbf{v} \neq 0$ as the search direction, the support point (or function) of shape A along $\mathbf{v}$, denoted as $\mathbf{p}_s(A, \mathbf{v})$, is the furthest point of A along the direction $\mathbf{v}$:

$$\mathbf{p}_s(A, \mathbf{v}) = \mathbf{p} \in A : \ \mathbf{p} \cdot \mathbf{v} \geqslant \mathbf{q} \cdot \mathbf{v}, \ \forall \mathbf{q} \in A \tag{50}$$

The support point for the Minkowski difference of A and B can be expressed as:

$$\mathbf{p}_s(A \ominus B, \mathbf{v}) = \mathbf{p}_s(A, \mathbf{v}) + \mathbf{p}_s(-B, \mathbf{v}) = \mathbf{p}_s(A, \mathbf{v}) - \mathbf{p}_s(B, -\mathbf{v}) \tag{51}$$

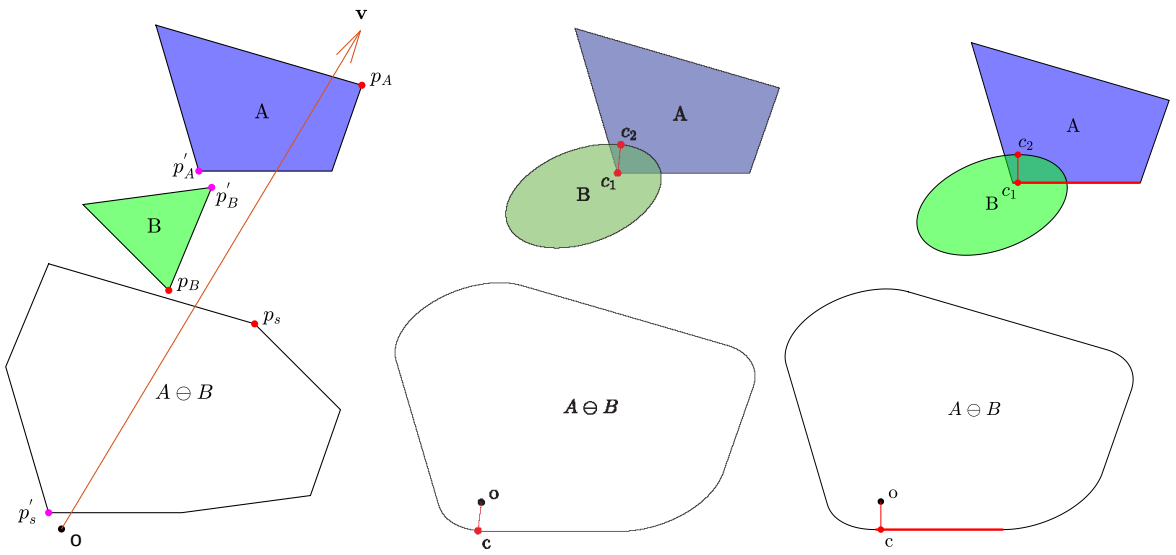
This expression is significant in the sense that there is no need to explicitly construct $A \ominus B$ in order to compute its support point. In other words, as long as the support points of A and B each can be defined, the support point of $A \ominus B$ can be readily obtained from the support points of A and B. This entirely avoids the often very costly explicit generation of $A \ominus B$.

For the Minkowski sum of A and B, it also has

$$\mathbf{p}_s(A \oplus B, \mathbf{v}) = \mathbf{p}_s(A, \mathbf{v}) + \mathbf{p}_s(B, \mathbf{v}) \tag{52}$$

If shape A is define by a (hierarchical) Minkowski sum of m constituent shapes $A = A_1 \oplus A_2 \oplus \cdots \oplus A_m$, it has

$$\mathbf{p}_s(A, \mathbf{v}) = \sum_{i=1}^{m} \mathbf{p}_s(A_i, \mathbf{v}) \tag{53}$$



(a) Support points (b) Vertex-Edge (c) Edge-Edge

Figure 16: Support points and definition of contact features and types: (a) The support points of A, $-B$ and $A \ominus B$ along search directions $\mathbf{v}$ $(p_A, p_B, p_s = p_A - p_B)$ and $-\mathbf{v}$ $(p'_A, p'_B, p'_s = p'_B - p'_A)$; (b) The contact features and types between A and B in two cases: contact point $\mathbf{c}$ on $A \ominus B$, contact points $\mathbf{c}_1$ and $\mathbf{c}_2$ on A and B respectively; $\mathbf{oc}$ defines the contact normal $\mathbf{n}$ and is perpendicular to $\partial(A \ominus B)$; while $\mathbf{c}_1\mathbf{c}_2$ is the normal of B at point $\mathbf{c}_2$ in (a) and (b), and also normal to the bottom edge of A in (b)

i.e the support point of A can be obtained additively from the support functions of the individual constituent shapes without explicitly constructing A, hence greatly reducing the complicity of representing complex shapes using Minkowski sum. Further from both properties (51) and 52), the support point of the Minkowski difference of two shapes of a Minkowski-sum type, can also be readily attained if the individual support points concerned are available.

Figure 16(a) illustrates the support points of a quadrilateral, a triangle and their Minkowski difference along two search directions $\mathbf{v}$ and $-\mathbf{v}$.

One very important property of $A \ominus B$ is that A and B are not in contact if there exists a search direction $\mathbf{v}$ such that

$$\mathbf{v} \cdot \mathbf{p}_s(A \ominus B, \mathbf{v}) < 0 \tag{54}$$

or

$$\mathbf{v} \cdot \mathbf{o} > \mathbf{v} \cdot \mathbf{p}_s(A \ominus B, \mathbf{v}) \tag{55}$$

i.e. if the origin $\mathbf{o}$ is even further in the search direction $\mathbf{v}$ than the support (i.e the furthest) point of the Minkowski difference $A \ominus B$ along $\mathbf{v}$, the origin must not be in $A \ominus B$ and thus no contact exists between A and B. In Figure 16(a), the origin $\mathbf{o}$ is further in the search direction $-\mathbf{v}$ than the support point $\mathbf{p}'_s$ of the Minkowski difference $A \ominus B$ along the direction $-\mathbf{v}$, so A and B must not be in contact.

The property (54) severs as an early termination criterion when a non-contact case is detected in the following GJK algorithm [80].

iii) GJK algorithm. The GJK algorithm is an iterative procedure that attempts to dynamically find a simplex (triangle in 2D and tetrahedron in 3D) from the boundary or vertices of $\partial(A \ominus B)$ that contains the origin, or report a non-contact case when the criterion (54) is met.

The algorithm starts from an (arbitrary) initial search direction with an empty simplex $\mathbf{S} = \varnothing$. Then at each subsequent iteration, $\mathbf{S}$ is updated either by adding a vertex to it or reducing it by discarding vertices to the lowest dimension possible that still contains the closest feature to the origin, and followed by computing a new search direction that is towards the origin normal to the new simplex. The iteration continues until a simplex is formed or a non-contact cases is encountered.

The relevant algorithmic detail can be found elsewhere [80, 82, 73]. An effective implementation requires a systematic approach that can reduce the cases to be examined to a minimum for the inclusion check of the origin, and voids any unnecessary numerical operations. There are several (slightly) different implementations available in the literature or online sources. The GJK algorithm can, in principle, be applied to any two convex shapes as long as the support function of each shape can be (numerically) defined to compute the furthest boundary point along a given direction.

When applied to convex polygons/polyhedra, the iteration procedure can terminate in a limited number of iterations, proportional to the sum of the numbers of vertices of the two shapes concerned. Another favourable feature is that the support point can be obtained from the vertices of the polygons/polyhedra, i.e. no other topological properties, such as edges and faces, are needed. Thus the support function for a polygon/polyhedron can be trivially implemented. A higher efficiency may be achieved, however, if the topological properties are also utilised to find the support point of polygon/polyhedron [82], particularly when the number of vertices is large.

Overall, the GJK algorithm offers a unified approach with greater versatility for detecting if two convex shapes are in contact or not. Also, the outputted simplex when a contact is detected can be utilised to construct a good initial guess for computing the overlap using other methods, for instance the ones mentioned in Section 4.1 for PCFR shapes.

iv) Overlap and the other contact features. From the Minkowski difference of two (convex) shapes A and B in contact, the contact overlap δ can be readily defined as the *shortest distance* from the origin $\mathbf{o}$ to the boundary of $A \ominus B$:

$$\delta = \min_{\mathbf{x} \in \partial(A \ominus B)} \|\mathbf{x} - \mathbf{o}\| = \|\mathbf{c}\| \tag{56}$$

where $\mathbf{c}$ is the orthogonal projection of $\mathbf{o}$ onto $\partial(A \ominus B)$, from which the normal direction can be attained $\mathbf{n} = \mathbf{c}/\delta$. The two corresponding contact points associated with $\mathbf{c}$ on the boundaries of A and B are respectively $\mathbf{c}_1$ and $\mathbf{c}_2$, from which the contact point $\mathbf{c}$ can be chosen as the midpoint

$$\mathbf{c} = (\mathbf{c}_1 + \mathbf{c}_2)/2 \tag{57}$$

There is a theoretical justification [97] for such a choice if A and B have the same elastic material properties. It also has the relation: $\mathbf{c} = \mathbf{c}_2 - \mathbf{c}_1 = \delta\mathbf{n}$.

Figures 16(b) and (c) illustrate the contact features defined for the contact of a triangle and an ellipse at two positions, where the point $\mathbf{c}$ is the orthogonal projection of the origin $\mathbf{o}$ onto the boundary $\partial(A \ominus B)$ in the two cases, but the corresponding contact types are different. The significance of defining the overlap and other contact features this way will be revealed in Section 6.3.

It is highlighted that the above definitions of the contact features are valid not only for convex shapes but also for concave shapes and hence are generic. Consequently the contact detection or the contact feature evaluation of two shapes can be reduced to solving the minimal problem (56). Available solution methods for convex shapes are briefly reviewed below, while the review for concave cases will be done in Section 5.

v) Evaluations of Overlap and the other contact features. For convex polygons/polyhedra, partnered with GJK, the *expanding polytope algorithm* (EPA) [81] has proved to be an effective solution procedure to evaluate the contact features defined above, where the support point also plays the central role. The main idea of EPA is to iteratively construct an inscribed polytope $\mathbf{P}$ to the Minkowski difference of two shapes A and B, so that the shortest distance of the origin to one of the flat faces of the polytope is taken as the (approximate) contact overlap.

EPA starts with the simplex outputted from GJK as the initial polytope $\mathbf{P}$. The face closest to the origin in $\mathbf{P}$ is found and its outward normal is used as the next search direction to compute a new support point $\mathbf{a}$ on $\partial(A \ominus B)$. Use this new point to update or expand $\mathbf{P}$ by removing some old faces based on their visibility from $\mathbf{a}$ and adding some new faces. Then check if the new shortest distance computed after the new faces added has converged or not. This procedure repeats until the face closest to the origin is also a face of $\partial(A \ominus B)$ for (convex) polygon/polyhedron cases, or is within a prescribed tolerance for other convex shapes.

A nice feature of EPA is that the iterative procedure is guaranteed to terminate for convex polygons/polyhedra within a finite number of iterations proportional to the total number of vertices of the two shapes. This limited maximum iteration property is, however, no longer valid for curved/smooth shapes, and the overall computational efficiency may not be high as the convergence can be slow. In this case, EPA can be coupled with other solution procedures suitable for the minimisation problem (56). This coupled strategy is outlined in [74].

Although the GJK algorithm has been extensively used for collision detection in computer games and many other areas since the 1990s, its use for contact detection of convex shapes in DEM is first introduced in [72], where very similar definitions of overlap, normal direction and contact point are also used. However, the GJK algorithm used in [72] is different from what is described above (and used in [73]) in a few important aspects.

Firstly, the criterion (54) is not checked in [72] to stop the iteration when a non-contact case is encountered. The second and also major difference is that the GJK algorithm is also employed in [72] to compute the overlap when a contact state is established. Because the GJK algorithm cannot effectively evaluate the overlap or distance when the origin is inside the Minkowski difference, the two shapes first have to be sufficiently shrunk with the same scale to separate them, then GJK is used to compute the gap or distance of the origin to the Minkowski difference. Finally the shapes are scaled up back to their original sizes to reconstruct the overlap distance and the normal vector at the contact point based on the result of the GJK. The problem with this approach, as explained in [72], is that the reconstructed overlap, normal direction and contact point are only approximations of the true values and are also dependent on the shrunk scale. On the contrary, the true contact features can be accurately obtained by EPA.

Instead of using EPA for non-polygonal/polyhedral particles, other solution procedures have been proposed. In fact, the numerical procedure proposed in [83] for super-quadrics and mentioned in Section 4.1.3 is essentially solving a problem equivalent to the minimal problem (56) but using the combined Newton-Raphson and Levenberg-Marquardt approach instead. The key point of the solution procedure is for each approximate contact normal $\mathbf{n}$, the two points on each super-quadric are explicitly obtained such that they have the same outer normal as $\mathbf{n}$. This is precisely equivalent to finding the two support points of the super-quadrics involved. The Newton-Raphson leads to good convergence while the Levenberg-Marquardt makes the method more robust. Thus this method implicitly utilises the support point.

In the solution procedure presented in [110] for poly-super-ellipsoids and also mentioned in Section 4.1.3, the GJK is used to check the contact state of two poly-super-quadrics first, and then the same Newton-Raphson and Levenberg-Marquardt approach as in [83] is adopted to find the solution of an equivalent problem to (56), but the support point is explicitly utilised in the solution procedure.

vi) Short summary. The overlap and corresponding contact features defined in this subsection are significant in several aspects: 1) they are universal for both convex and concave shapes; 2) their evaluation is converted into solving a minimal problem (56); and 3) different solution procedures can be adopted and can also be tailored to achieve the best performance for the shapes concerned. More fundamental properties associated with the definition of this set of contact features will be discussed in Section 6.3.

GJK provides an elegant, effective and unified approach for checking if any two convex shapes are in contact or not. The actual contact state can be found often in a small number of iterations. The algorithm can be implemented in a generic manner as long as the support point/function for each shape concerned can be defined, and therefore it can significantly simplify the code implementation to include a wide range of mixed shapes in DEM. The simplex **S** outputted from GJK when a contact case is found, or the polytope **P** obtained using EPA with a few iterations, can serve as a good initial value for the subsequent iterative method to evaluate the contact features for non-polygonal/polyhedral shapes. This would generally improve both efficiency and robustness of the whole contact modelling procedure.

The combination of GJK and EPA appears to be a preferred contact detection approach for convex polygons/polyhedra. The details of the numerical procedure and associated issues can be found in [115, 70]. The combined strategy has also be employed in [75] for polyhedra where comparing with the fast common-plane method has been made.

5 Contact Modelling for Non-Convex Shapes

As mentioned in Section 3, the existence of (associated) multiple contact cases imposes considerable challenges on developing appropriate contact modelling approaches for non-convex shapes in DEM. Several contact detection methods that have been developed and used are briefly covered in the following subsections.

5.1 Clumped Sphere Contact Modelling

Using a cluster of (overlapping) spheres to represent any shape has been the most successful approach to modelling the contact of non-convex shapes in DEM since the 1990s. The shape representation aspect has been covered in Section 2.2.1. The contact modelling of two such shapes can be broken down to individual sphere-sphere contact cases in a straightforward manner. Any suitable contact interaction constitutive law can be applied for each sphere-sphere contact pair. The contact forces obtained from the individual contact pairs are assembled and applied to the mass centre of each clumped shape involved. However, for frictional forces modelled with the Coulomb friction law which is history-dependent, the tangential displacement at each contact sphere-sphere pair should be stored locally for computing the correct friction force between the pair if their contact persists at the next time instance.

Thus the contact modelling of clumped spheres can naturally handle multiple contact scenarios, and treats the friction between two shapes as a distributed force acting on multiple places.

Nevertheless, in addition to the geometric related issues as already discussed in Section 2.2.1, another drawback of clumped spheres is that as the contact history at each contacting pair has to be stored and retrieved, the extra memory and computational requirements can be high, in particular when the number of constituent spheres in the clumped shape is large. However, the main issue with clumped-sphere shapes is that these shapes may not be able to effectively represent real non-convex shapes both geometrically and physically, as already mentioned.

5.2 Convex Primitive Decomposition Based Contact Modelling

When a complex and non-convex shape can be represented by the combination or cluster of a number of basic convex primary shapes, such as composite shapes, a natural approach to handling the contact between these shapes is to decompose each shape into individual primary components, and then the contact can be modelled between pairs of basic convex shapes, assuming their contacts can be treated properly. For general cases, two schemes are mentioned below.

1) 2D Multiple Arcs. For 2D arbitrarily shaped particles obtained, for instance, from CT scanning, an arcs-based contact modelling scheme is proposed in [129], where the circle-growing technique and the least squares method are combined to establish the multi-connected inward and outward arcs that represent the whole particle profile. The contact detection algorithm is developed to identify overlapping arcs from contacting shapes and the area-based contact model is used to account for the contact force between each overlapping region. Because the basic component of the shape is an arc, all the contact features can be evaluated fairly effectively, and multiple contact scenarios can be handled without any difficulties. This arc-based contact model can be viewed as an extension of the Polyarc scheme proposed in [128] from convex to general non-convex cases.

2) Polygons/Polyhedra. More universally applied contact modelling approaches are polygons/polyhedra based as any shapes, including wall boundaries, can be discretised and (approximately) represented by polygonal/polyhedral meshes. In theory, a non-convex polygon/polyhedron can be decomposed into sub-convex polygons/polyhedra. Then the contact between these sub-polygons/polyhedra can be modelled independently using the methods mentioned in Section 4.2, for instance.

However, it is practically not easy to decompose an arbitrary polyhedron into a *minimal* number of convex sub-blocks in order to reduce possible contact pairs. Nevertheless, any concave polygon/polyhedron can be split into individual triangles/tetrahedra which are convex. As a consequence, the contact of any two shapes can be reduced into *contacts of triangles/tetrahedra* from both shapes, which can again be done using any approach mentioned in Section 4.2.

The main issue is that the number of triangles/tetrahedra required to realistically represent a real shaped particle can be very high. Hence the number of contact pairs involved for two such particles can be substantial and the corresponding computational costs can also be extremely expensive. In this case, the GPU can be used to parallelize these contact pairs as their computations are perfectly concurrent.

The contact between spherical harmonic shapes are mainly modelled in this way, as done in [90, 96, 91]. This type of contact involving spherical harmonic particles can also be handled by the level-set based method [37] below, and by the contact volume based energy-conserving model demonstrated in [71] as will be covered in Section 6.

5.3 Level-set Based Contact Modelling

The use of level-sets to describe any shaped particles is covered in Section 2.5. The two key ingredients of this shape representation for contact modelling include: 1) use a (regular) grid to approximately represent the chosen field of the shape, $f(\mathbf{x})$, where the values of the field at grid nodes are attained; and 2) the boundary of the shape is discretised, independently from the gird, by a surface mesh/patch or simply a set of discrete points (or point cloud). Figure 17 shows such a representation for a spherical harmonic representation where the signed distance is chosen to be the level-set function.

The field f can be an explicitly or implicitly defined function, and its value f and gradient ∇f at a point $\mathbf{x}$ can be readily obtained if f is explicitly expressed, or interpolated from the grid nodes if f is implicitly defined. The relative position of a point $\mathbf{x}$ can be checked against the shape using the following convention: the point is inside/on/outside the shape if $f(\mathbf{x}) < 0, = 0$, or > 0.

In [133, 134] which first introduce the concept of level set into DEM, the function f is taken as the implicit function representing the shape (such as super-quadrics) and is used to readily check if a boundary point of a shape is in contact with another shape. Although no detail is explicitly provided about how the normal contact force is related to this function, it is suspected that either node-surface or surface-surface contact treatments used in finite element analysis of contact mechanics [49] are employed.

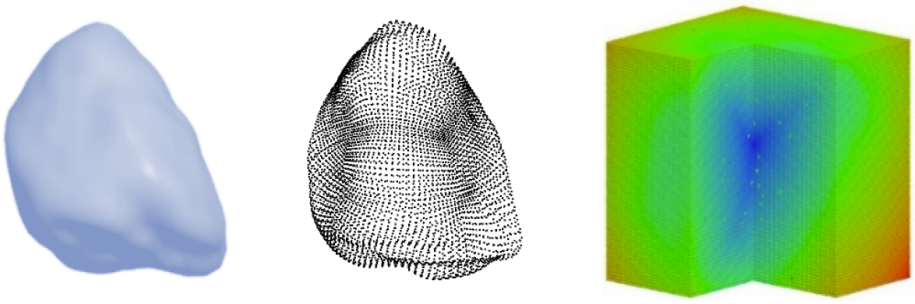
The remainder of this subsection will outline a recent contact modelling method for level-set shapes, termed LS-DEM and proposed in [139] from which more details can be found. In this method, the level-set function $\phi(\mathbf{p})$ at point $\mathbf{p}$ is chosen to be the signed (minimal or shortest) distance of the point to the boundary of the shape. Thus, except for a very few special shapes, the function has to be implicitly defined using a background grid where its values at the grid nodes need to be numerically obtained. Then the value of ϕ at any point can be interpolated by the nodal values of the cell surrounding the point. The gradient of the function, $\nabla\phi$, at the point can also be evaluated in a similar interpolation fashion. In addition, the boundary of each shape is separately sampled by discrete points.

When modelling the contact between two level-set shapes i and j, for each boundary point $\mathbf{p}_k$ of shape i, obtain its level-set value or distance at shape j, $d_k = \phi_j(\mathbf{p}_k)$. If $d_k < 0$, the point penetrates into shape j, and the normal contact force $\mathbf{F}_n^k$ acting at the point can be computed as

$$\mathbf{F}_{n,k} = -k_n d_k \mathbf{n}_k \tag{58}$$

where $\mathbf{n}_k = \nabla\phi(\mathbf{p}_k)/\|\nabla\phi(\mathbf{p}_k)\|$ is the normal direction. The corresponding frictional force at the point can also be computed, if needed. Both normal and tangential forces are assembled at the mass centre of shape i. The corresponding reaction forces on shape j can be processed similarly.

The main computational cost occurs at evaluating the level-set function and its gradient at each boundary point and are proportional to the number of discrete boundary points used. Because the background grid has already been prepared *in priori*, the cost involved in retrieving the function and its gradient for one point is only associated with some local interpolation operations in a cell which is in-significant and also independent of the grid resolution. Thus the LS-DEM is very efficient at the contact resolution stage if the number of discrete boundary points used is not excessive. The method can also benefit greatly from GPU parallelisation because the contact detection for all boundary points of a pair of contacting particles can be done perfectly in parallel.



(a) (b) (c)

Figure 17: A signed distance function represented spherical harmonic shape: (a) shape; (b) boundary points; (c) Level-set representation on a grid [37]

Obviously, the resolution or the total number of nodes of the background grid and the number of boundary points for each level-set shape both play an important part in LS-DEM in terms of memory requirement, accuracy and efficiency. Millions of grid nodes are often required for a shape. Constructing the background grid of such a size for a shape can be computationally intensive, although GPU parallelisation can speed-up the procedure. Thus, the number of different shapes allowed in a DEM simulation may not be very large if the computer resource available is limited. The influence of the grid resolution and the number of boundary points on the overall solution accuracy has been evaluated in [140].

In [139], the background grid of each shape is directly obtained from the scanned image of a shape [138]. Construction of background grids using the signed distance function for a wide range of complex and concave shapes, including concave super-quadrics and spherical harmonic shapes, with surface triangular meshes has been discussed in detail in [37], where the contact force at each contact point is also scaled by the total number of contacting points between the two shapes at this time instance. This simple treatment largely reduces the impact of using different boundary points on the particle behaviour when the other parameters are unchanged. More thorough examinations of the contact behaviour of the LS-DEM have also been conducted in [37]. A very recent work [142] also presents a similar scheme but the signed distance is extended to general distances depending on shapes.

Finally, compared to all the contact modelling methods mentioned in the previous section, because each particle is discretised by boundary points, level-set based contact modelling methods essentially follow the 'node-surface' contact approach originally adopted in finite element modelling of contacts [49]. As each discrete boundary point can be an independent contact point, level-set based methods naturally resolve the multiple contact issue that largely hinders proper modelling of contacts between concave particles, but with the expenses of compromised accuracy and increased memory costs for the store and retrieval of history related data at discrete boundary points.

6 Energy-Conserving Contact Theory and Contact Models for Arbitrarily Shaped Particles

The previous two sections have reviewed a diverse range of most commonly used contact modelling methods or models for non-spherical particles, but these models share a common feature that they are established mainly by proposing a set of rules, based on which the main contact features in each method are defined and determined individually or collectively. Thus

the conventional contact methods are *rule-based*. The main difference among these methods is that the nature and number of introduced rules vary. In some cases, one rule may be introduced for determining one particular feature, overlap, normal direction or contact point, while in other cases, one rule can lead to the determination of more than one feature.

The main issue with the rule-based contact modelling strategy is that those rules introduced may not be based on a sound physical principle but rather in a heuristic manner. In a worse case when multiple rules are presented, their logical and physical compatibility may not be guaranteed and even some basic physical principles may be violated.

The fundamental reason behind these situations is that there has been a lack of sound contact theories for discrete element modelling of non-spherical particles. The classic DEM proposed by Cundall and Stack in [5] provides a computational modelling framework initially for discs and spheres. The Hertz contact theory for linear elastic spheres can be viewed as the underlying contact theory of the classic DEM. The extension of DEM to non-spherical cases initially relied on the applicability of the Hertz contact theory to these cases where shapes are (locally) smooth, such as ellipses/ellipsoids and super-quadrics.

This applicability immediately faces difficulties for shapes with sharp edges/corners, such as cylinders, and more seriously for polygons/polyhedra, as the contacts involving edges/corners cannot be approximated by the Hertz contact theory at all. The earlier contact modelling strategy for polygons/polyhedra bears some similarity to the contact modelling approaches used in FEM [49], where the penetration or overlap based penalty method has been dominant.

Under this view, the DEM can also be regarded as a penalty based method, where the interaction law between contacting particles acts as the penalty function. This equivalency of both DEM and FEM, however, is incorrect in a fundamental way: in FEM, the penalty function is only used to (approximately) impose the impenetration condition where the deformation of the contact region is accounted for by the finite element formulation, while in DEM, the interaction law in principle mimics the actual contact forces between the particles without explicitly considering their deformation.

Thus the difficulties faced in DEM are twofold: 1) the Hertz contact theory is no longer valid for non-smooth shapes, and 2) contact interaction laws for non-spherical particles in DEM have to represent, to a first degree, the actual contact behaviour of particles but treating them as rigid. These difficulties can only be overcome by establishing new contact theories for DEM. Such a contact theory should provide a coherent framework where different contact modelling methods can be derived in a systematic manner without violating some basic physical principles.

It is by following the above considerations that an energy-conserving contact theory for arbitrarily shaped particles is proposed in [67], from which a general energy-conserving contact model is derived. By choosing two different contact energy functions/potentials two different contact models are realised and are guaranteed to be energy-conserving for elastic impact. The following subsections will briefly described these developments and related issues. The connection between these energy-conserving contact models with some of conventional contact methods reviewed in Sections 4 and 5 will also be provided.

6.1 The Energy-Conserving Contact (ECC) Theory and the General Contact Model

The initial idea of this ECC theory can be traced back to the contact area based model proposed for polygons in [98] and its later extension to polyhedra and general shapes in

[100, 102]. A concise general form is presented in [103]. The theory is fully established in a more fundamental and rigorous manner in [67], from which all the details can be found. The following only provide the outline of the theory.

Th ECC theory consists of two key components - one principle and one assumption:

- *Principle*: The total mechanical energy of two particles is always conserved before and after any elastic impact.
- *Assumption*: The current contact state can be fully described by an contact energy potential w as the function of the current positions, orientations and possibly some history related variables of the two particles:

$$w = w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{\Psi}) \tag{59}$$

  where $\mathbf{x}_o$ and $\mathbf{x}'_o$ are the centroids of the two particles, $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ their orientations, and $\boldsymbol{\Psi} = \{\psi_1, \cdots, \psi_m\}$ are m internal variables.

The contact energy w is also equal to the total work done when the particles are bought into the current positions from an initially non-contact state [67]. The principle of energy conservation requires that for the elastic impact of two particles shown in Figure 18, the total mechanical energy before the impact (a) must be the same after the impact (d).

The principle of mechanical energy conservation leads to two important conclusions about the contact potential/function w:

- *Time and history independency*: The energy potential/function must only be a function of the current positions and orientations:

$$w = w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}') \tag{60}$$

  i.e. w is independent of time t, velocities/accelerations and any history related variables;

- *Conservative field*: The energy potential/function is conservative in the configuration space of the two particles so that the contact forces $\mathbf{F}_n, \mathbf{F}'_n$ and moments $\mathbf{M}_n, \mathbf{M}'_n$ between these two particles can be determined by

$$\mathbf{F}_n = -\frac{\partial w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \mathbf{x}_o}; \quad \mathbf{M}_n = -\frac{\partial w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \boldsymbol{\theta}} \tag{61}$$

$$\mathbf{F}'_n = -\frac{\partial w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \mathbf{x}'_o}; \quad \mathbf{M}'_n = -\frac{\partial w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \boldsymbol{\theta}'}; \tag{62}$$

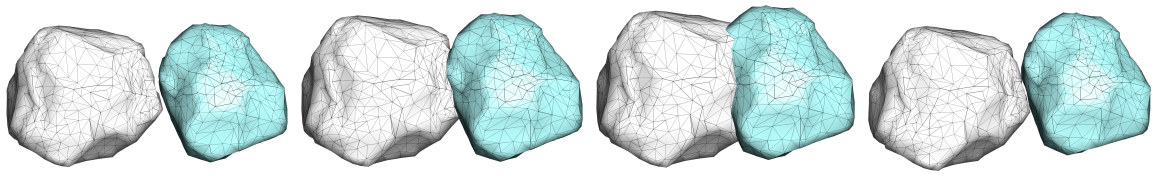
  where two pairs $(\mathbf{F}_n, \mathbf{M}_n)$ and $(\mathbf{F}_n, \mathbf{M}_n)$ act at the centroids of the two particles respectively.

The coordinate invariance of the energy function w gives rise to the expected property [67]:

$$\mathbf{F}_n = -\mathbf{F}'_n \tag{63}$$

From $\mathbf{F}_n$ (or $\mathbf{F}'_n$), the normal contact direction $\mathbf{n}$ and the magnitude F_n can be identified as

$$\mathbf{n} = \mathbf{F}_n / F_n; \quad F_n = \|\mathbf{F}_n\| \tag{64}$$



(a) (b) (c) (d)

Figure 18: Elastic impact of two particles: (a) before impact; (b) at the increasing penetration stage; (c) at the maximum penetration; and (d) after impact

Further derivations yield the formal definition of the contact point/line as stated: the contact point $\mathbf{x}_c$ is a common position that when the contact forces $\mathbf{F}_n$ and $\mathbf{F}'_n$ act at this point while the contact moments vanish, and can be determined by

$$\mathbf{x}_c = \mathbf{x}_o + \frac{\mathbf{n} \times \mathbf{M}_n}{\|\mathbf{F}_n\|} + \lambda \mathbf{n} \tag{65}$$

where λ is a parameter that can take any value, meaning that for normal contact, the contact point is not unique but any point on the contact line defined by this equation can be taken as the contact point, in principle.

Thus, when the contact energy potential function w is specified, the contact forces can be fully determined in terms of the normal direction $\mathbf{n}$, magnitude F_n and contact point/line $\mathbf{x}_c$. It is guaranteed that the mechanical energy will be conserved under any contact condition.

Thus the ECC theory naturally leads to the following *general energy-conserving normal contact model* [102, 67]:

For two (rigid) bodies in contact, as shown in Figure 19, assume that their positions and orientations can be described by $(\mathbf{x}_o, \boldsymbol{\theta})$ and $(\mathbf{x}'_o, \boldsymbol{\theta}')$ respectively. The following normal contact model will be energy-conserving if there exists a *scalar* potential function $w(\mathbf{x}_o, \mathbf{x}'_o, \boldsymbol{\theta}, \boldsymbol{\theta}')$ such that

1) The normal contact force, $\mathbf{F}_n$, acting on Body 1 from Body 2 is determined by

$$\mathbf{F}_n = -\frac{\partial w(\mathbf{x}_o, \boldsymbol{\theta}, \mathbf{x}', \boldsymbol{\theta}')}{\partial \mathbf{x}_o} \tag{66}$$

The contact force applied to Body 2 from Body 1 is:

$$\mathbf{F}'_n = -\mathbf{F}_n \tag{67}$$

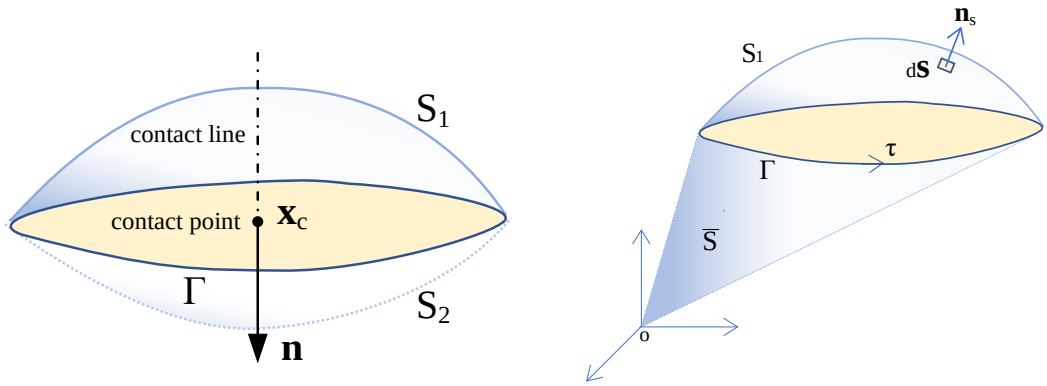
2) The contact normal $\mathbf{n}$ can be deduced as the unit vector of $\mathbf{F}_n$

$$\mathbf{n} = \mathbf{F}_n / F_n \tag{68}$$

where $F_n = \|\mathbf{F}_n\|$ is the magnitude of $\mathbf{F}_n$.

3) The contact point $\mathbf{x}_c$, at which only the contact force $\mathbf{F}_n$ needs to be applied, is given by (65). The normal contact line is formed by $\mathbf{x}_c$ as the linear function of λ along the normal direction $\mathbf{n}$.

4) The tangential plane $\boldsymbol{\pi}$ is the plane perpendicular to the contact normal $\mathbf{n}$ and passes through the chosen contact point.



(a) (b)

Figure 19: Two arbitrarily shaped bodies in contact: (a) the contact region formed by two contact surfaces S_1 and S_2; (b) the contact surface S_1 with the boundary Γ and one 'conical' replacement surface $\bar{S}$ [68]

Clearly, different energy-conserving contact models can be obtained by choosing different energy functions w. The key to developing such contact models is therefore to choose appropriate energy functions. When such a function is specified, the corresponding contact model automatically follows, and the energy-conserving property of the model ensures that a more robust and stable discrete element modelling can be achieved under any complex circumstance.

To facilitate the construction of an energy function, the coordinate invariance [67] suggests that w should be a function of some geometric features of the contact region Ω_c of the two particles that are coordinate invariant:

$$w = w(g_1, g_2, \cdots, g_{n_g}) \tag{69}$$

Geometric features that are coordinate invariant include overlap, contact area/volume.

This energy-conserving contact framework can handle both distinct and associated multiple contact cases without requiring any additional treatment. The proposed general contact model provides both flexibility and stability for any complex particle contact scenario that may be encountered. As commended in [102] both Hertzian and linear spring models for discs/spheres are energy-conserving. It can also explain why the widely used clumped sphere contact model is also energy-serving for any non-spherical contact cases, thereby providing a rational reason for its success.

6.2 The Contact-Volume Based Energy-Conserving Contact Model

The first energy-conserving normal contact model is realised by assuming that the energy potential function is a monotonically increasing function of the contact volume V_c:

$$w = w(V_c) \tag{70}$$

Then the normal force $\mathbf{F}_n$ can be obtained as

$$\mathbf{F}_n = -\frac{w(V_c)}{\partial \mathbf{x}_o} = -w'(V_c)\,\mathbf{S}_n \tag{71}$$

where $\mathbf{S}_n$ is the vector area of the contact surface S_1

$$\mathbf{S}_n = \int_{S_1} d\mathbf{S} \tag{72}$$

with the projected or scalar contact area $S_n = |\mathbf{S}_n|$. The magnitude of the force is $F_n = w'(V_c)S_n$, and the normal direction is obtained by $\mathbf{n} = \mathbf{F}_n/F_n$. The contact point/line $\mathbf{x}_c$ is determined by

$$\mathbf{x}_c = \frac{\mathbf{n} \times \mathbf{G}_n}{S_n} + \lambda \mathbf{n}, \quad \text{with } \mathbf{G}_n = \int_{S_1} \mathbf{x} \times d\mathbf{S} \tag{73}$$

Thus all the contact features are fully determined, and the resulting contact model is termed the *contact volumed based energy-conserving model* [68].

The key to the development of the above model is the derivation of the following relation

$$\frac{\partial V_c}{\partial \mathbf{x}} = \mathbf{S}_n = \int_{S_1} d\mathbf{S} \tag{74}$$

which is obtained by applying the classic Leibniz integral rule with variable limits [151], or the Reynolds transport theorem [152], to an integral with a moving boundary.

Both relations (72) and (73) indicate that the contact surface S_1 (and equally S_2) solely determines the contact normal direction and the contact point or line. Further derivations lead to alternative but simpler forms of $\mathbf{S}_n$ and $\mathbf{G}_n$

$$\mathbf{S}_n = \int_{S_1} d\mathbf{S} = \frac{1}{2} \oint_{\Gamma} \mathbf{x} \times d\mathbf{\Gamma}; \quad \mathbf{G}_n = \int_{S_1} \mathbf{x} \times d\mathbf{S} = -\frac{1}{3} \oint_{\Gamma} \mathbf{x} \cdot \mathbf{x}\, d\mathbf{\Gamma} \tag{75}$$

where Γ is the interaction line of the two shapes. These new forms in (75) further indicate that both $\mathbf{S}_n$ and $\mathbf{G}_n$, and also the normal direction $\mathbf{n}$ and the contact point or line $\mathbf{x}_c$, are solely determined by the intersection of the two contacting shapes, and that contact surfaces S_1 and S_2 are no need to be explicitly obtained. Although $\mathbf{S}_n$ and $\mathbf{G}_n$ can be evaluated from the intersection line Γ, the magnitude of the force may depend on the contact volume V_c. Thus in general, it may have to explicitly obtain the whole contact region Ω_c in order to compute V_c. This is clearly a challenging task to fulfil for most geometric shapes.

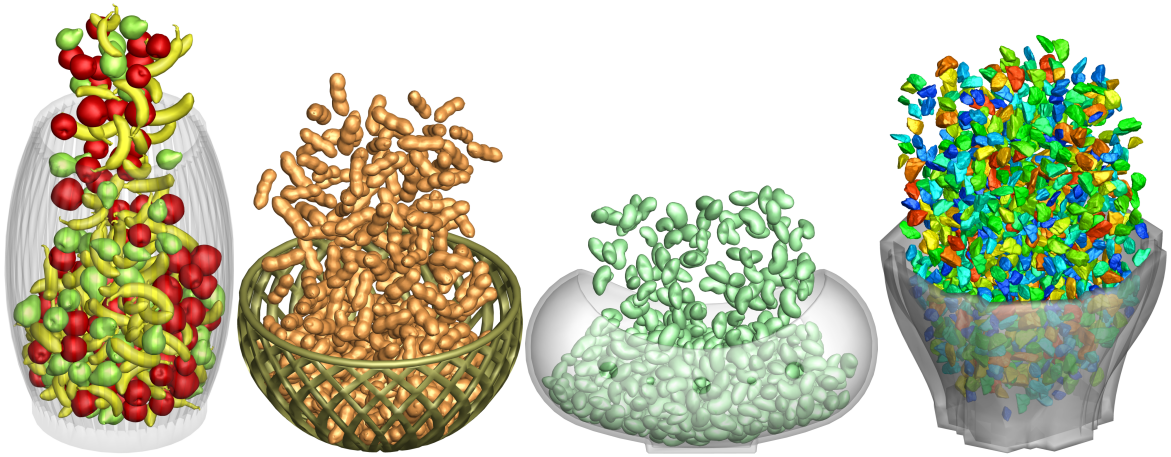
However, when w is assumed to a linear function of V_c, F_n becomes

$$F_n = k_n S_n \tag{76}$$

Now, F_n is only proportional to S_n which can be obtained solely from Γ and crucially, V_c disappears. Hence, neither the contact region Ω_c nor the contact surface S_1 or S_2 are involved in $\mathbf{F}_n$, and only the intersection Γ needs to be explicitly constructed. This feature leads to a much simpler and computationally effective normal contact model. This particular contact model, referred to as the *contact volume-based linear model* or the *linear contact model* for short, is recommended as the default model within the contact volume based modelling framework.

Because V_c, $\mathbf{S}_n$ and $\mathbf{G}_n$ are all integrals of geometric domains (regions, surfaces and lines), they are therefore naturally additive, i.e. when the original domains of the two particles are divided into sub-domains, the required values can be obtained by summating all the sub-domain values. This additive property makes the current contact model flexible and highly parallelisable.

When applying this contact volume based model to non-spherical shapes, the real computational challenge is effectively evaluating $\mathbf{S}_n$, $\mathbf{G}_n$ and V_c (when necessary) for various shapes.



(a) (b) (c) (d)

Figure 20: Random deposition of various concave particles in containers using the contact-volume based ECC model

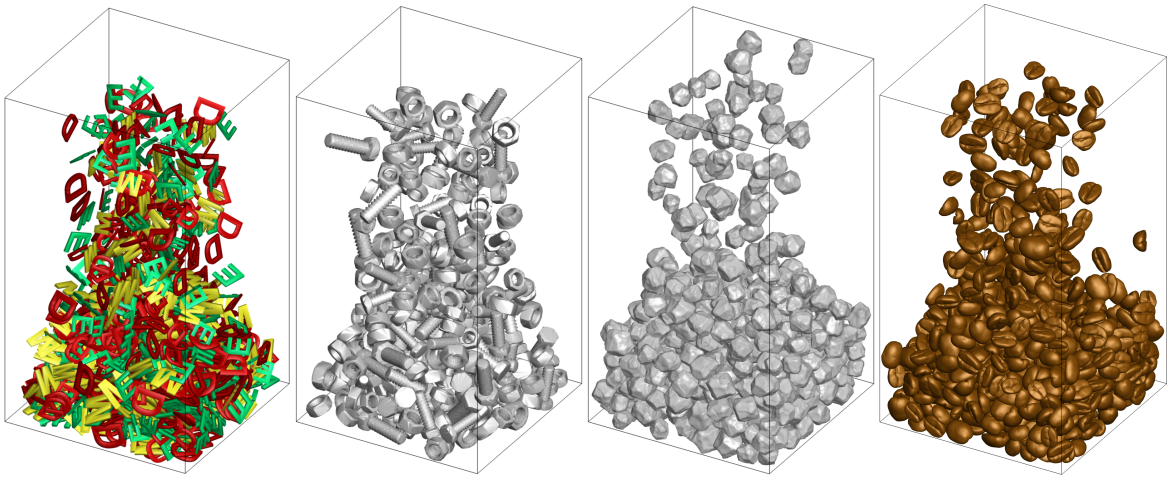
A universal solution is to discretise these shapes using either volumetric or surface triangular meshes. Then contact modelling of these shapes become the modelling of contacts between volumetric or surface triangular meshes from which V_c, $\mathbf{S}_n$ and $\mathbf{G}_n$ can be computed. However, the computational procedures for volumetric and surface triangular meshes are different, resulting in different performance and applicability.

For volumetric meshed shapes, the contact can be decomposed into the contacts between (convex) polyhedra that form the volumetric meshes. Three different approaches under the current energy-conserving contact framework are discussed in [112]. The second 'ray tracing' approach is used in [68].

The third 'dual approach' is more appealing. It is based on a technique proposed in [113] that transfers a convex polyhedron P into its dual polyhedron P^d in a dual space. Then the overlapping region of two polyhedra is also a polyhedron that is the dual to the convex hull of the two dual polyhedra. When the convex hull is constructed, the overlapping polyhedron is obtained by transforming the hull back to the primary space, from which $\mathbf{S}_n$, $\mathbf{G}_n$ and V_c can be obtained.

The key step in this dual approach is to find an initial point that is located inside the overlapping region. However, it is not trivial to determine if two given polyhedra are in contact or not. Thus a fairly sophisticated iterative procedure [114] is often needed to find an initial point, if exists. An ideal solution, which is recommended here, is combining GJK with the dual approach, where GJK is used to obtain the contact state of two polyhedra. If the contact exists, the centroid of the simplex from GJK can be taken as the interior point for construct the dual polyhedra.

It is argued in [68] that the surface triangular mesh representation for non-spherical particles is superior to the volumetric representation, and thus is recommended. The corresponding numerical procedures are discussed in detail in [68, 70]. The associated core operation involves the determination of the intersections between two sets of triangles from the two shapes concerned. Based on the additive property of the proposed contact model, the modelling of two 3D shapes in contact can be conducted fully in parallel without incurring much communication overhead. An automatic triangular mesh simplification is also proposed in [71] to reduce the mesh size to any desired level in an optimal fashion.



(a) (b) (c) (d)

Figure 21: Random deposition of various concave particles in boxes using the contact-volume based ECC model [68, 70]

A large number of numerical examples are conducted in [68, 70, 71] to verify the energy-conserving property of the proposed contact model for a wide range of shapes and contact scenarios. Additional examples are also provided to illustrate the robustness and applicability of the model for realistic problems. Some of these examples are shown in Figures 20 and 21.

As the form of the energy potential function plays a crucial role, ways to properly choose the material property related parameters in the energy potential function are also discussed in [69].

6.3 The Contact-Overlap Based Energy-Conserving Contact Model

As the overlap has been the dominant contact feature in existing contact models, particularly for convex particles, it would be natural to consider developing a generic contact overlap based energy-conserving contact model for all particle shapes, alternative to the above contact volume based model [68]. However, there are technical challenges that need to be overcome before this goal can be achieved within the framework of the energy-conserving contact theory [67].

As reviewed in the last two sections, many different definitions of overlap have been proposed in different contact models for convex particles. The first challenge is to know if any of these contact models are already energy-conserving. The second challenge is how to define an overlap for concave shapes. Should it be shape-dependent as this appears to be the case for convex shapes? If so, this would be bad news because in contrast to the above contact volume based model, there does not exist a single contact model that can apply to all shape cases.

Fortunately, all these technical challenges have been resolved in [74], where a *unified* overlap, termed the *Minkowski overlap*, is proposed which is applicable to both convex and concave shapes. It further theoretically proves a key property of the Minkowski overlap which leads to the establishment of an overlap based energy-conserving contact model for all shapes.

6.3.1 Minkowski Overlap and its Properties

The Minkowski overlap is essentially the extension of the overlap defined for two shapes A and B based on their Minkowski difference $A \ominus B$ in Section 4.5.2. It can be formally defined as the minimal (vector) distance that one shape can be moved in translation only such as the two shapes are just in touch.

If denoting the translated shape A by a (vector) distance $\mathbf{v}$ as

$$A^{\mathbf{v}} \equiv A \oplus \{\mathbf{v}\} = \{\mathbf{a} + \mathbf{v} : \mathbf{a} \in A\} \tag{77}$$

the Minkowski (vector) distance between A and B, denoted as $\mathbf{d}_M$, can be expressed as

$$\mathbf{d}_M = \underset{\mathbf{v}\in\mathbb{R}^n}{\operatorname{argmin}}\{\|\mathbf{v}\| : \mathbf{o} \in \partial(A \ominus B^{\mathbf{v}})\} \tag{78}$$

Equivalently, $\mathbf{d}_M$ is also the minimum vector distance from the origin $\mathbf{o}$ to the boundary of $A \ominus B$:

$$\mathbf{d}_M = \min_{\mathbf{v}\in\partial(A\ominus B)} \|\mathbf{v}\| \tag{79}$$

The vector distance $\mathbf{d}_M$ can be written as

$$\mathbf{d}_M = \delta_M \mathbf{n}_M \tag{80}$$

where $\delta_M = \|\mathbf{d}_M\|$ is the Minkowski overlap which is exactly the overlap defined in (56), while $\mathbf{n}_M$ is the contact normal direction.

Refer to Figure 16 again. Let $\mathbf{c}$ be the point on $\partial(A \ominus B)$ that has the shortest distance with the origin $\mathbf{o}$, and the two corresponding contact points associated with $\mathbf{c}$ on the boundaries of A and B are respectively $\mathbf{c}_1$ and $\mathbf{c}_2$. Then the contact point $\mathbf{c}_M$ can be chosen as their midpoint: $\mathbf{c}_M = (\mathbf{c}_1 + \mathbf{c}_2)/2$. It is easy to show

$$\mathbf{d}_M = \delta_M \mathbf{n} = \mathbf{c} - \mathbf{o} = \mathbf{c}_1 - \mathbf{c}_2 \tag{81}$$

with $\mathbf{n}_M = \mathbf{c}/\|\mathbf{c}\| = (\mathbf{c}_1 - \mathbf{c}_2)/\|\mathbf{c}_1 - \mathbf{c}_2\|$, $\delta_M = \|\mathbf{c}\| = \|\mathbf{c}_1 - \mathbf{c}_2\|$. Obviously, $\mathbf{n}_M$ is also the normal of $\partial(A \ominus B)$ at $\mathbf{c}$. Furthermore, both $\mathbf{n}_M$ and the line $\mathbf{c}_1\mathbf{c}_2$ are the common normal to both A and B if $\mathbf{c}_1$ and $\mathbf{c}_2$ are not vertices.

When considering B in a translation motion with $\mathbf{x}$ being a reference point on B, because δ_M is the shortest distance between $\mathbf{o}$ and $\partial(A \ominus B)$, it can theoretically prove [74]

$$\frac{\partial \mathbf{c}}{\partial \mathbf{x}} = -\mathbf{I} \quad \text{(identity matrix)} \tag{82}$$

which leads to

$$\frac{\partial \delta_M}{\partial \mathbf{x}} = -\mathbf{n}_M \tag{83}$$

This is the most important property of the proposed Minkowski overlap which ensures that the resulting overlap based contact model is guaranteed to be energy-conserving. It also has a clear geometric explanation that the most effective way to move B to reduce the overlap δ_M must be along the negative direction of the contact normal $\mathbf{n}_M$.

Note that the definition of the Minkowski overlap by (78) indicates that the contact overlap may not longer be determined by the boundaries of the two shapes in the vicinity of the contact region, but possibly by the entire shapes. It is also emphasised that the Minkowski overlap (79) and its property (83) are valid for both convex and concave shapes. As a final

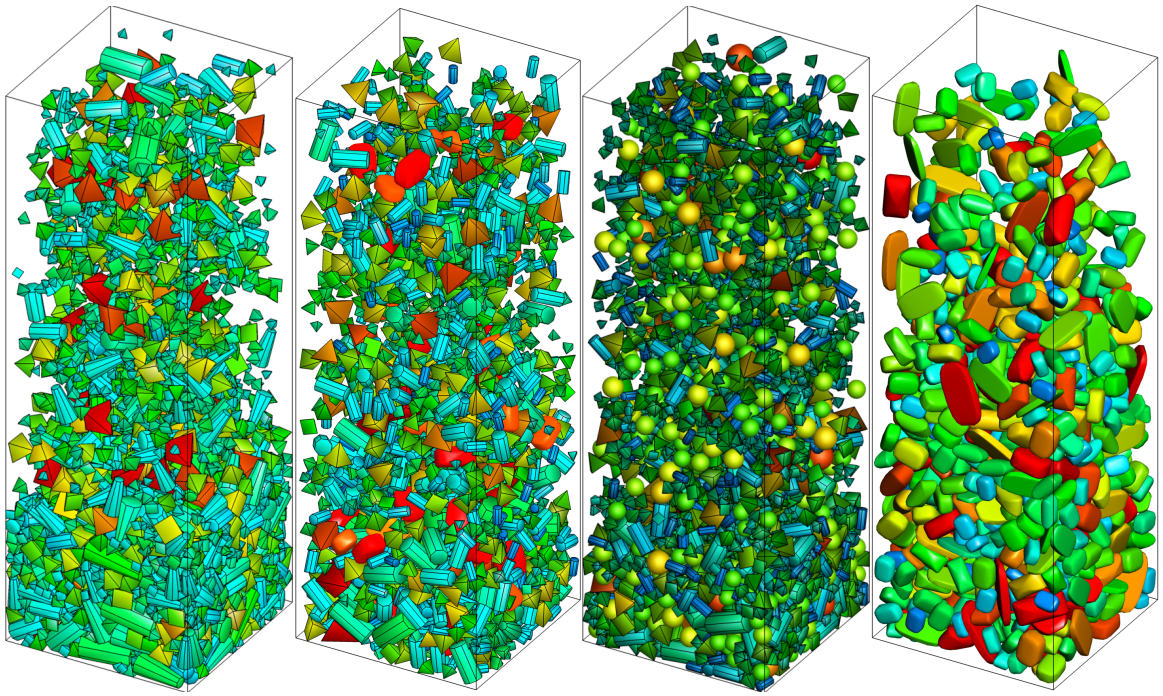


Figure 22: Random deposition of various convex particles into boxes using the Minowski-overlap based ECC model [73, 74]

note, the Minkowski distance δ_M is also equivalent to the (positive) signed distance of the origin to the boundary of $A \ominus B$, $\partial(A \ominus B)$, and thus should satisfy the so called Eikonal equation

$$\left\| \frac{\partial \delta_M}{\partial \mathbf{x}} \right\| = 1 \tag{84}$$

where the origin is conceptionally treated as a moving point relative to the now stationary $\partial(A \ominus B)$. Clearly the derived property (83) indeed satisfies this Eikonal equation.

6.3.2 Minkowski Overlap Based Energy-Conserving Contact Model

Let the contact energy potential w be a function of the Minkowski overlap δ_M defined above:

$$w = w(\delta_M) \tag{85}$$

Following the general energy-conserving model, and utilising (83), the normal force $\mathbf{F}_n$ should be

$$\mathbf{F}_n = -\frac{\partial w(\delta_M)}{\partial \mathbf{x}} = -w'(\delta_M)\frac{\partial \delta_M}{\partial \mathbf{x}} = w'(\delta_M)\mathbf{n}_M \tag{86}$$

which means that $\mathbf{n}_M$ is indeed the direction of $\mathbf{F}_n$, and the magnitude $F_n = w'(\delta_M)$. The point $\mathbf{c}_M$ is the contact point. Thus, this Minkowski overlap based contact model is guaranteed to be energy-conserving for elastic contact cases for all shapes.

Choosing different forms of the contact potential function can lead to the contact force with different magnitudes. In general, $w(\delta_M)$ can be chosen as a power-law function of the overlap

$$w(\delta_M) = \frac{1}{n+1} k_n \delta_M^{n+1} \tag{87}$$

where k_n is the (nominal) stiffness and $n \geqslant 1$. The corresponding force magnitude has the form

$$F_n = k_n \delta_M^n \tag{88}$$

from which both classic linear $(n = 1)$ and Hertz-type $(n = 3/2)$ models can be recovered.

Computationally, evaluating the Minkowski contact features becomes the key to the success of this contact model. For convex shapes, the procedures mentioned in Section 4.5.2 are applicable. As highlighted there, for polygons/polyhedra the combined GJK and EPA offers the most elegant and robust numerical solution. Further related issues can be found in [70]. Some numerical examples are conducted in [73, 74] to demonstrate the robustness and applicability of the proposed computational procedure and the Minkowski overlap based contact model. Some of these examples are illustrate in Figure 22.

For non-convex shapes, using surface triangulated meshes for the contact modelling appears to be the most generic approach, following similar reasons mentioned in the previous subsection. However, the combined GJK and EPA is no longer applicable. The real challenge is that as the Minkowski difference of two such shapes is not convex, much more complicated computational procedures have to be employed to obtain the Minkowski contact features. The related work has not be published yet.

6.3.3 On Existing Overlap Based Contact Models

A few interesting observations can be made on existing overlap based contact models covered in Sections 4 and 5. A pertinent question is: are any of these contact models energy-conserving? The criterion that can be applied to find the answer to the question is the property (83), which itself is the consequence of the definition of the Minkowski overlap. Thus, it is not difficult to come to the conclusion that *all common-normal contact models are energy-conserving.* Furthermore, the level-set method based on the signed distance function, such as LS-DEM mentioned in Section 5.3, is also energy conserving.

On the other hand, the original common-plane and fast common-plane methods mentioned in Section 4.2 cannot guarantee that the obtained overlap for polygons/polyhedra satisfies the property (83), and thus may not always conserve energy. As a consequence, the occurrence of numerical instability in simulations cannot be ruled out.

7 Concluding Remarks

Some major developments in the contact modelling of non-spherical particles in DEM over the last three decades have been reviewed. Instead of conducting an exhaustive survey of all the developments on the relevant topics, a more selective approach has been followed, in the sense that only pieces of work that have made contributions to the methodological and computational aspects have been included. In addition, a structured review has been undertaken within a purposefully designed framework that reflects, to a certain extend, the underlying logical connections of those methodologies, with the wish that this review could provide a better overall picture of the field. Also, different from many review articles, some key methodologies have been covered with more technical detail presented in an attempt to help understand possible in-depth connections between different but related methods.

On a technical level, because the focus has been placed on the geometric aspects of each contact modelling method, i.e. how the three contact geometric features are defined and

numerically evaluated, the physical aspect of contact models, i.e. contact constitutive laws for non-spherical particles has been entirely omitted. The main reason behind this omission is that different constitutive laws for non-spherical particles cannot be rigorously accessed currently. Although a great amount of work has been done to numerically and experimentally validate various contact constitutive laws for a wide range of particle shapes, there still severely lacks some widely agreed simple cases that can serve as reliable benchmarks to quantitatively access the accuracy of contact models for non-spherical particles when both geometric and physical aspects of the contact model are considered.

Acknowledgement

The author would like express his sincere thanks to Professor Shunying Ji and Dr Shiqiang Wang for kindly sharing with him their large collections of DEM literature on the topics covered, without which this review could not be accomplished in the current form.

This work is partially supported by National Natural Science Foundation of China under Grant No. 12072217. This support is greatly acknowledged.

References

[1] P. A. Cundall. A computer model for simulating progressive, large-scale movements in blocky rock systems, Proc. Int. Symp. on Rock Fracture, pp. 11-18, 1971.

[2] P. A. Cundall. A computer model for rock-mass behaviour using interactive graphics for the input and output of geometric data. Report AD/A-001, 602, U.S. National Technical Information Services, 1974.

[3] O. D. L. Strack, P. A. Cundall. The distinct element method as a tool for research in granular media. Report to National Science Foundation from Department of Civil and Mineral Engineering, University of Minnesota, USA, 1978.

[4] P. A. Cundall. BALL-A program to model granular media using the distinct element method. Technical Note, Advanced Technology Group, Dames & Moore, London, 1978.

[5] P. A. Cundall, and O. D. L. Strack. A discrete numerical model for granular assemblies. *Géotechnique*, 29(1):47-65, 1979.

[6] A. Munjiza, D.R.J. Owen, N. Bicanic. A combined finite-discrete element method in transient dynamics of fracturing solids. *Engineering Computations*, 12(2):145-74, 1995.

[7] D. R. Owen, Y. T. Feng, E. de Souza Neto, F. Wang, M. G. Cottrel, F. A. Pires, J. Yu. The modelling of Multi-fracturing Solids and Particulate Media. *International Journal for Numerical Methods in Engineering*. 60(1): 317-340, 2004.

[8] A. Munjiza. *The Combined Finite-Discrete Element Method.* England, Wiley & Sons, 2004.

[9] PFC - Particle Flow Code, Version 6.0. Itasca Consulting Group, Inc. Minneapolis:Itasca, 2018.

[10] EDEM 2019 User Guide. DEM Solutions Ltd. https://www.edemsimulation.com/

[11] StarCCM+: https://www.plm.automation.siemens.com/global/en/products/simcenter/STAR-CCM.html

[12] LIGGGHTS: PUBLIC Documentation, Version 3. DCS Computing GmbH. 2016, https://www.cfdem.com/media/DEM/docu/Manual.html

[13] Yade Documentation, 2nd ed. (2015). The Yade Project. http://yade-dem.org/doc/

[14] MatDEM: Fast GPU Matrix Computation of Discrete Element Method. http://matdem.com/index.asp?lg=en

[15] L. Jing, O. Stephansson. *Fundamentals of discrete element methods for rock engineering: theory and applications.* Elsevier, 2007.

[16] C. O'Sullivan. *Particulate discrete element modelling: a geomechanics perspective.* CRC Press, 2011.

[17] H.-G. Matuttis, J. Chen. *Understanding the Discrete Element Method Simulation of Non-Spherical Particles for Granular and Multi-Body Systems.* John Wiley & Sons, 2014.

[18] S. Ji, L. Lu. *Computational Granular Mechanics and Its Engineering Applications.* Springer, 2020. ISBN 978-981-15-3304-4.

[19] C. Hogue. Shape representation and contact detection for discrete element simulations of arbitrary geometries. *Engineering Computations.* 15(2-3):374-390, 1998

[20] N. Bićanić. Discrete Element Methods. *Encyclopedia of Computational Mechanics* (Second Edition), Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes. John Wiley & Sons, Ltd., 2007

[21] G. Lu, J. R. Third, C. R. Müller. Discrete element models for non-spherical particle systems: From theoretical developments to applications. *Chemical Engineering Science*, 127:425-465, 2015.

[22] Jalal Kafashan, Joanna Wiacek, Noorhazlinda Abd Rahman, Jieqing Gan. Two-dimensional particle shapes modelling for DEM simulations in engineering: a review. *Granular Matter.* 21:80, 2019.

[23] R. Dobry, T-T. Ng. Discrete modelling of stress-strain behaviour of granular media at small and large strains. *Engineering Computations.* 9:129–143, 1992.

[24] H. P. Zhu, Z. Y. Zhou, R. Y. Yang, A. B. Yu. Discrete particle simulation of particulate systems: theoretical developments. *Chemical Engineering Science.* 62:3378-3396, 2007.

[25] L. Rothenburg, R. J. Bathurst. Numerical simulation of idealized granular assemblies with planar elliptical particles. *Computers and Geotechnics*, 11(31):5-329, 1991.

[26] L. Rothenburg, R. J. Bathurst. Micromechanical features of granular assemblies with planar elliptical particles. *Géotechnique*, 42(1):79-95, 1992.

[27] J. M. Ting. A robust algorithm for ellipse-based discrete element modelling of granular materials. *Computers and Geotechnics.* 13(3):175-186, 1992.

[28] J. M. Ting, M. Khwaja, L. R. Meachum, J. D. Rowell. An ellipse-based discrete element model for granular materials. *International Journal for Numerical and Analytical Methods in Geomechanics.* 17:603-623, 1993.

[29] T.-T. Ng. Numerical simulations of granular soil using elliptical particles. *Computers and Geotechnics*. 16(2):153-169, 1994.

[30] X. Lin, T.-T. Ng. Contact detection algorithms for three-dimensional ellipsoids in discrete element modelling. *International Journal for Numerical and Analytical Methods in Geomechanics.* 19(9):653-659, 1995.

[31] J. R. Williams, A. Pentland. Superquadrics and model dynamics for discrete elements in interactive design. *Engineering Computations*. 9:115-128, 1992.

[32] P. W. Cleary, M. L. Sawley. DEM modelling of industrial granular flows: 3D case studies and the effect of particle shape on hopper discharge. *Appl. Math. Model.* 26(2):89-111, 2002.

[33] P. W. Cleary. DEM prediction of industrial and geophysical particle flows. *Particuology.* 8:106-118, 2010.

[34] G. Lu, J. R. Third, C. R. Müller. Critical assessment of two approaches for evaluating contacts between super-quadric shaped particles in DEM simulations. *Chemical Engineering Science*, 78(34):226-235, 2012.

[35] A. Podlozhnyuk, S. Pirker, C. Kloss. Efficient implementation of superquadric particles in discrete element method within an open-source framework. *Comp. Part. Mech.*, 4:101-118, 2017.

[36] S. Wang, Y. Fan, S. Y. Ji. Interaction between super-quadric particles and triangular elements and its application to hopper discharge. *Powder Technology.* 339:534-549, 2018.

[37] S. Wang, S. Ji. A unified level set method for simulating mixed granular flows involving multiple non-spherical DEM models in complex structures. *Computer Methods in Applied Mechanics and Engineering.* 393:114802 2022.

[38] A. H. Barr. Superquadrics and Angle-Preserving Transformations. *IEEE Computer Graphics and Applications.* 1(1):11-23, 1981.

[39] E. T. Bowman, K. Saga, T. W. Drummond. Particle shape characterization using Fourier analysis. *Géotechnique.* 51(6):545-554, 2001.

[40] W. Zhou, Y. Huang, T.-T. Ngb, G. Ma. A geometric potential-based contact detection algorithm for egg-shaped particles in discrete element modeling *Powder Technology.* 327:152-162, 2018.

[41] P. E. Danielsson. Euclidean distance mapping. *Computer Graphics and Image Processing.* 14(3):227-248, 1980.

[42] Y. Ge and J. M. Fitzpatrick. On the Generation of Skeletons from Discrete Euclidean Distance Maps. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 18(11):1055-1066, 1996.

[43] L. Wang, J.-K. Park, Y. Fu. (2007). Representation of real particles for DEM simulation using X-ray tomography. *Construct. Build. Mater.* 21(2):338-346, 2007.

[44] X. Garcia, J.-P. Latham, J. Xiang, J. P. Harrison. A clustered overlapping sphere algorithm to represent real particles in discrete element modelling. *Géotechnique*, 59(9):779-784, 2009.

[45] Cheng-Qing Li, Wen-Jie Xu, Qing-ShanMeng Multi-sphere approximation of real particles for DEM simulation based on a modified greedy heuristic algorithm. *Powder Technology.* 286:478-487, 2015.

[46] H. Kruggel-Emden, S. Rickelt, S. Wirtz, V. Scherer. A study on the validity of the multi-sphere discrete element method. *Powder Technology.* 188:153-165, 2008.

[47] D. Markauskas, R. Kacianauskas, A. Džiugys, R. Navakas. Investigation of adequacy of multi-sphere approximation of elliptical particles for DEM simulations. *Granular Matter.* 12:107-123, 2010.

[48] K. J. Johnson. *Contact Mechanics.* Cambridge University Press, 1985.

[49] P. Wriggers. *Computational Contact Mechanics.* (2nd ed.), Springer-Verlag, Berlin, Germany, 2006.

[50] P. M. Hubbard Approximating polyhedra with spheres for time-critical collision detection, *ACM Trans. Graph.* 15:179-210, 1996.

[51] O. R. Walton, R. L. Braun. Simulation of rotary-drum and repose tests for frictional spheres and rigid sphere clusters. Joint DOE/NSF workshop on flow of particles and fluids, Ithaca, NY, USA, 1993.

[52] J. F. Favier, M. H. Abbaspour, M. Kremmer, A. Raji. Shape representation of axisymmetrical, non-spherical particles in discrete element simulation using multi-element model particles. *Engineering Computations.* 16(4):467-480, 1999.

[53] M. Kremmer, J. F. Favier. A method for representing boundaries in discrete element modelling—part I: Geometry and contact detection. *International Journal for Numerical Methods in Engineering.* 51:1407-1421, 2001.

[54] M. Kremmer, J. F. Favier. A method for representing boundaries in discrete element modelling—part II: Kinematics. *International Journal for Numerical Methods in Engineering.* 51:1423-1436, 2001.

[55] Q. Zhou, W. J. Xu, G. Y. Liu. A contact detection algorithm for triangle boundary in GPU-based DEM and its application in a large-scale landslide. *Computer and Geotechnics.* 138:104371, 2021.

[56] P. A. Cundall. Formulation of a three-dimensional distinct element model - Part I. A scheme to detect and represent contacts in a system composed of many polyhedral blocks. *International Journal of Rock Mechanics*, 25(3):107-116, 1988.

[57] R. Hart, P. A. Cundall, J. Lemos. Formulation of a three-dimensional distinct element model - Part II. Mechanical calculations for motion and interaction of a system composed of many polyhedral blocks, *International Journal of Rock Mechanics*, 25(3):117-125, 1988.

[58] S.-W. Chang, C.-S. Chen. A non-iterative derivation of the common plane for contact detection of polyhedral blocks. *International Journal for Numerical Methods in Engineering.* 74:734-53, 2008.

[59] R. E. Barbosa. *Discrete element models for granular materials and rock masses*, PhD Thesis, University of Illinois at Urbana-Champaign, IL., 1990.

[60] J. Ghaboussi. and R. E. Barbosa. Three-dimensional discrete element method for granular materials. *International Journal for Numerical and Analytical Methods in Geomechanics*, 14:451-472, 1990.

[61] J. P. Latham, Y. Lu, A. Munjiza. A random method for simulating loose packs of angular particles using tetrahedra. *Ge´otechnique.* 51(10):871-879, 2001.

[62] J. P. Latham, A. Munjiza. The modeling of particle systems with real shapes. *Phil. Trans. R. Soc. Lond. A*, 362:1953-1972, 2004.

[63] P. W. Cleary, N. Stokes, J. Hurley. Efficient Collision Detection for Three Dimensional Super-ellipsoidal Particles. Proceedings of 8th international conference on field programmable logic and applications, pp 1-7, 1997.

[64] J. F. Peters, M. A. Hopkins, R. Kala, R. E. Wahl. A poly-ellipsoid particle for non-spherical discrete element method. *, Engineering Computations.* 26(6):645-657, 2009.

[65] B. Zhang, R. Regueiro, A. Druckrey, K. Alshibli. Construction of poly-ellipsoidal grain shapes from SMT imaging on sand, and the development of a new DEM contact detection algorithm. *Engineering Computations.* 35:733-771, 2018.

[66] H. G. Matuttis, S. Luding, H. J. Herrmann. Discrete element simulations of dense packings and heaps made of spherical and non-spherical particles. *Powder Technology.* 109(1-3):278-292, 2000.

[67] Y. T. Feng. Energy-conserving contact interaction models for arbitrarily shaped discrete elements: Basic Framework and General Contact Model. *Computer Methods in Applied Mechanics and Engineering.* 373:113454, 2021.

[68] Y. T. Feng. An energy-conserving contact theory for discrete element modelling of arbitrarily shaped discrete elements: Contact volume based model and computational issues. *Computer Methods in Applied Mechanics and Engineering.* 373:113493, 2021.

[69] Ting Qiao, Ji Li, Shunying Ji. A modified discrete element method for concave granular materials based on energy-conserving contact model *Theoretical and Applied Mechanics Letters.* 12(2):100325, 2022.

[70] Y. T. Feng. A Generic Energy-Conserving Discrete Element Modelling Strategy for Concave Particles Represented by Surface Triangular Meshes. *International Journal for Numerical Methods in Engineering.* 122(10):2581-2597, 2021.

[71] Y. T. Feng. An Effective Energy-Conserving Contact Modelling Strategy for Spherical Harmonic Particles Represented by Surface Triangular Meshes with Automatic Simplification. *Computer Methods in Applied Mechanics and Engineering.* 379:113750, 2021.

[72] A. Wachs, L. Girolami, G. Vinay, G. Ferrer. Grains3D, a flexible DEM approach for particles of arbitrary convex shape — Part I: Numerical model and validations *Powder Technology.* 224:374-389, 2012.

[73] Y. T. Feng, Y. Tan. On Minkowski difference-based contact detection in discrete/ discontinuous modelling of convex polygons/polyhedra: Algorithms and implementation. *Engineering Computations.* 37:54-72, 2020.

[74] Y. T. Feng, Y. Tan. The Minkowski overlap and the energy-conserving contact model for discrete element modelling of convex nonspherical particles. *International Journal for Numerical Methods in Engineering.* 122:6476-6496, 2021.

[75] Y. Descantes, F. Tricoire, P. Richard. Classical contact detection algorithms for 3D DEM simulations: Drawbacks and solutions. *Computers and Geotechnics.* 114:103134, 2019.

[76] M. C. Lin, J.F. Canny. A fast algorithm for incremental distance calculation. Proceedings of the 1991 IEEE international conference on robotics and automation. 1991. p. 1008-1014.

[77] J. Cohen, M. Lin, D. Manocha, K. Ponamgi. I-COLLIDE: An Interactive and Exact Collision Detection System for Large Scale Environments. *ACM Int. 3D Graphics Conference.* pp 189-196, 1995.

[78] B. Mirtich V-clip: Fast and Robust Polyhedral Collision Detection. Mitsubishi Electric Research Laboratory TR-97-05, July 1997.

[79] Barki H, Denis F, Dupont F. Contributing vertices-based Minkowski sum computation of convex polyhedra. *Computer-Aided Design.* 41(7):525-538, 2009.

[80] E. G. Gilbert, D. W. Johnson and S. S. Keerthi. A Fast Procedure for Computing the Distance between Complex Objects in Three-Dimensional Space. *IEEE Trans. Robotics and Automation.* 4(2):193-203, 1988.

[81] G. van de Bergen. Proximity queries and penetration depth computation on 3d game objects. Game Developers Conference, 2001.

[82] C. Cameron. Enhancing GJK: Computing Minimum and Penetration Distances between Convex Polyhedra. *Int Conf Robotics and Automation*, pp 3112-3117, April 1997.

[83] C. Wellmann, C. Lillie, and P. Wriggers. A contact detection algorithm for superellipsoids based on the common-normal concept. *Engineering Computations.* 25(5):432-442, 2008.

[84] J. Andrade, K. Lim, C. Avila, I. Vlahini´c. Granular element method for computational particle mechanics. *Computer Methods in Applied Mechanics and Engineering.* 241:262-274, 2012.

[85] K. Lim, J. Andrade. Granular element method for three-dimensional discrete element calculations. *International Journal for Numerical and Analytical Methods in Geomechanics.* 38(2):167-188, 2014

[86] T. J. Hughes, J. A. Cottrell, Y Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering.* 194(39-41):4135-4195, 2005.

[87] W. Gao, J. Wang, S. Yin, Y. T. Feng. A coupled 3D isogeometric and discrete element approach for modeling interactions between structures and granular matters. *Computer Methods in Applied Mechanics and Engineering.* 354:441-463, 2019.

[88] W. Gao, Y. T. Feng. A coupled 3D discrete elements/isogeometric method for particle/structure interaction problems. *Computational Particle Mechanics.* 7:869-880, 2020.

[89] D. H. Ballard, C. M. Brown. *Computer Vision.* Englewood Cliffs, NJ:Prentice-Hall; 1982.

[90] E. J. Garboczi, J. W. Bullard. Contact function, uniform-thickness shell volume, and convexity measure for 3D star-shaped random particles. *Powder Technology*, 237:191-201, 2013.

[91] Z. Zhu, H. Chen, W. Xu, L. Liu. Packing simulation of three-dimensional multi-sized star-shaped particles. *Modelling and Simulation in Materials Science and Engineering*, 22(3), article id. 035008, 2014.

[92] C. Brechbuhler, G. Gerig, O. Kubler. Parametrization of closed surfaces for 3D shape description. Computer Vision and Image Understanding, 61(2):154-70, 1995.

[93] J. K. Anochie-Boateng, J. J. Komba, G. M. Mvelase. Three-dimensional laser scanning technique to quantify aggregate and ballast shape properties. *Construction and Building Materials.* 43:389-398, 2013.

[94] X. Jia, R. A. Williams. A packing algorithm for particles of arbitrary shapes *Powder Technology.* 120:175-186, 2001.

[95] X. Jia, M. Gan, R. A. Williams, D. Rhodes. Validation of a digital packing algorithm in predicting powder packing densities. *Powder Technology.* 174:10-13, 2007.

[96] E. J. Garboczi, J. W. Bullard. 3D analytical mathematical models of random star-shape particles via a combination of X-ray computed microtomography and spherical harmonic analysis. *Advanced Powder Technology*, 28(2):325-339, 2017.

[97] Y. T. Feng, W. Gao. On the strain energy distribution of two elastic solids under smooth contact *Powder Technology.* 389:376-382, 2021.

[98] Y. T. Feng, D. R. J. Owen. A 2D polygon/polygon contact model: algorithmic aspects. *Engineering Computations*, 21:265-277, 2004.

[99] F. P. Preparata, M. I. Shamos. *Computational Geometry. An Introduction.* Springer: Berlin, Germany, 1985.

[100] Y. T. Feng, K. Han and D. R. J. Owen. An energy based polyhedron-to-polyhedron contact model. *Proceeding of 3rd M.I.T. Conference of Computational Fluid and Solid Mechanics*, pp210-214, MIT, USA, 14-17 June, 2005

[101] K. Han, Y. T. Feng, D. R. J. Owen. Contact resolution for non-circular discrete objects. *International Journal for Numerical Methods in Engineering.* 66(3):485-501, 2006.

[102] Y. T. Feng, K. Han, D. R. J. Owen. Energy-conserving contact interaction models for arbitrarily shaped discrete elements. *Computer Methods in Applied Mechanics and Engineering.* 205-208:169-177, 2012.

[103] Y. T. Feng. A general contact theory for non-spherical particles. In: X. Li, Y. Feng, G. Mustoe (eds), *Proceedings of 7th International Conference on Discrete Element Methods (Dem 2016).* Springer, pp 29-35, 2017.

[104] Y. T. Feng, K. Han, D. R. J. Owen. A generic contact detection framework for cylindrical particles in discrete element modelling. *Computer Methods in Applied Mechanics and Engineering.* 315:632-651, 2017.

[105] J. W. Perram, M. S. Wertheim. Statistical mechanics of hard ellipsoids. I. Overlap algorithm and the contact function. *Journal of Computational Physics.* 58:409-416, 1985.

[106] M. A. Hopkins. Discrete element modeling with dilated particles. *Engineering Computations.* 21(2/3/4):422-430, 2004

[107] L. Liu, S. Y. Ji. A fast detection algorithm based on the envelope function of dilated polyhedron (in Chinese). *Sci. Sin-Phys. Mech. Astron.*, 49:064601, 2019.

[108] L. Liu, S. Ji. A new contact detection method for arbitrary dilated polyhedra with potential function in discrete element method. *International Journal for Numerical Methods in Engineering.* 121:5742-5765, 2020.

[109] Z. Zhu, W. Xua, H. Chen. The fraction of overlapping interphase around 2D and 3D polydisperse non-spherical particles: Theoretical and numerical models. *Computer Methods in Applied Mechanics and Engineering.* 345:728-747, 2019.

[110] S. Zhao, J. Zhao. A poly-superellipsoid-based approach on particle morphology for DEM modeling of granular media. *International Journal for Numerical and Analytical Methods in Geomechanics.* 43:2147-2169, 2019.

[111] B. Nassauer, T. Liedke and M. Kuna. Polyhedral particles for the discrete element method: Geometry representation, contact detection and particle generation. *Granular Matters*, 15:85-94, 2013.

[112] N. Govender, N. Daniel, C.-Y. Wu, J. Khinast, P. Pizette, X. Wenjie. Hopper flow of irregularly shaped particles (non-convex polyhedra): GPU-based DEM simulation and experimental validation. *Chemical Engineering Science.* 188:34-51, 2018.

[113] D. E. Muller, F. P. Preparata. Finding the intersection of two convex polyhedra. *Theor. Comput. Sci.* 7(2): 217-236, 1978.

[114] S. Zhao, X. Zhou, W. Liu. Discrete element simulations of direct shear tests with particle angularity effect. *Granular Matter.* 17:793-806, 2015.

[115] Y. T. Feng, Y. Tan. On Minkowski difference-based contact detection in discrete-discontinuous modelling of convex polygons/polyhedral: Algorithms and implementation. *Engineering Computation*, 37:54-72, 2020.

[116] G. T. Houlsby. Potential particles: a method for modelling non-circular particles in DEM. *Computers and Geotechnics*, 36(6):953-959, 2009.

[117] J. Harkness. Potential particles for the modelling of interlocking media in three dimensions. *International Journal for Numerical Methods in Engineering.* 80:1573-1594, 2009.

[118] C. W. Boon, G. T. Houlsby, S. Utili. A new algorithm for contact detection between convex polygonal and polyhedral particles in the discrete element method. *Computers and Geotechnics*, 44(1):73-82, 2012.

[119] C. W. Boon, G. T. Houlsby, S. Utili. A new contact detection algorithm for three-dimensional non-spherical particles. *Powder Technology.* 248:94-102, 2013.

[120] E. G. Nezami, Y. M. A. Hashash, D. Zhao D. A fast contact detection algorithm for 3-D discrete element method. *Computers and Geotechnics*, 31(7):575-587, 2004.

[121] E. G. Nezami, Y. M. A. Hashash, D. Zhao. Shortest link method for contact detection in discrete element method. *International Journal for Numerical and Analytical Methods in Geomechanics*, 30(8):783-801, 2006.

[122] S. A. Galindo-Torres, D. M. Pedroso Molecular dynamics simulations of complex-shaped particles using voronoi-based spheropolyhedra. *Physics Review E*, 81:061303, 2010.

[123] D. Su, W. M. Yan. 3D characterization of general-shape sand particles using microfocus X-ray computed tomography and spherical harmonic functions, and particle regeneration using multivariate random vector. *Powder Technology*, 323:8-23, 2018.

[124] D. Su, X. Wang. *Characterization of general shaped particles and discrete element simulations* (in Chinese). Tsinghua University Press, Beijing, China, 2022. ISBN 978-7-302-60093-0

[125] G. Mollon, J. Zhao. 3D generation of realistic granular samples based on random fields theory and Fourier descriptors. *Computer Methods in Applied Mechanics and Engineering*, 279:46-65, 2014.

[126] C.-Y. Wang, V.-C. Liang. A packing generation scheme for the granular assemblies with planar elliptical particles. *International Journal for Numerical Methods in Engineering.* 21:347-358, 1997.

[127] A. V. Potapov, C. S. Campbell. A fast model for the simulation of non-round particles. *Granular Matter.* 1:9-14, 1998.

[128] P. Fu, O. Walton, J. Harvey. Polyarc discrete element for efficiently simulating arbitrarily shaped 2D particles *International Journal for Numerical Methods in Engineering.* 89(5):599-617, 2012.

[129] Xiang Wang, Zhen-Yu Yin, Dong Su, Hao Xiong, Y. T. Feng. A novel Arcs-based discrete element modeling of arbitrary convex and concave 2D particles, *Computer Methods in Applied Mechanics and Engineering*, 386:114071, 2021.

[130] Xiang Wang, Zhen-Yu Yin, Hao Xiong, Dong Su, Y. T. Feng. A spherical-harmonic-based approach to discrete element modeling of 3D irregular particles *International Journal for Numerical Methods in Engineering.* 122(20):5626-5655, 2021.

[131] C.-Y. Wang, V.-C. Liang. A packing generation scheme for the granular assemblies for 3D ellipsoidal particles. *International Journal for Numerical and Analytical Methods in Geomechanics.*, 23:815-828, 1999.

[132] G. Varadhan, D. Manocha. Accurate Minkowski sum approximation of polyhedral models. *Graphical Models*, 68(4):343-355, 2006.

[133] J. R. Williams, R. O'Connor. A linear complexity intersection algorithm for discrete element simulation of arbitrary geometries. *Engineering Computations*, 12:185-201, 1995.

[134] J. R. Williams, R. O'Connor. Discrete Element Simulation and the Contact Problem. *Archives of Computational Methods in Engineering.* 6(4):279-304, 1999.

[135] F. Thomasset, A. Dervieux. A finite element method for the simulation of a Rayleigh-Taylor instability, *Lectures Notes in Mathematics.* 771:145-158, 1979.

[136] D. S. Stafford, T. L. Jackson. Using level sets for creating virtual random packs of non-spherical convex shapes. *Journal of Computational Physics.* 229:3295-3315, 2010.

[137] B. Sonon, B. François, T. J. Massart. A unified level set based methodology for fast generation of complex microstructural multi-phase RVEs *Computer Methods in Applied Mechanics and Engineering.* 223-224:103-122, 2012.

[138] I. Vlahini´c, E. Andò, G. Viggiani J. Andrade. Towards a more accurate characterization of granular media: extracting quantitative descriptors from tomographic images. *Granular Matter.* 16:9-21, 2014.

[139] R. Kawamoto, E. Andò, G. Viggiani, J. Andrade. Level set discrete element method for three-dimensional computations with triaxial case study. *Journal of the Mechanics and Physics of Solids.* 91:1-13, 2016.

[140] J. Duriez, S. Bonelli. Precision and computational costs of level set-discrete element method (LS-DEM) with respect to DEM. *Computers and Geotechnics.* 134:104033, 2021.

[141] Z. Lai, Q. Chen, L. Huang. Fourier series-based discrete element method for computational mechanics of irregular-shaped particles. *Computer Methods in Applied Mechanics and Engineering.* 362:112873, 2020.

[142] Z. Lai, S. Zhao, J. Zhao, L. Huang. Signed distance field framework for unified DEM modeling of granular media with arbitrary particle shapes *Computational Mechanics.* 2022. https://doi.org/10.1007/s00466-022-02220-8

[143] M. Kodam, R. Bharadwaj, J. Curtis, B. Hancock, C. Wassgren. Cylindrical object contact detection for use in discrete element method simulations: part I - Contact detection algorithms. *Chemical Engineering Science* 65(22): 5852-5862, 2010.

[144] M. Kodam, R. Bharadwaj, J. Curtis, B. Hancock, C. Wassgren. Cylindrical object contact detection for use in discrete element method simulations: part II - Experimental validation. *Chemical Engineering Science* 65(22): 5863-5871, 2010.

[145] Y. Guo, C. Wassgren, W. Ketterhagen, B. Hancock, J. Curtis. Some computational considerations associated with discrete element modeling of cylindrical particles. *Powder Technology* 228: 193-198, 2012.

[146] M .A. Hopkins, J. Tuhkuri. Compression of floating ice fields. *Journal of Geophysical Research Atmospheres.* 1041:15815-15826, 1999.

[147] B. Ny, A. V. Kulchitsky, J. B. Johnson. Intersecting dilated convex polyhedra method for modeling complex particles in discrete element method. *International Journal for Numerical and Analytical Methods in Geomechanics.* 38(9):978-990, 2014.

[148] S. Ji, S. Sun, Y. Yan. Discrete element modeling of rock materials with dilated polyhedral elements. Procedia Engineering. 102:1793-1802, 2015.

[149] S. Ji, S. Sun, Y. Yan. Discrete element modeling of dynamic behaviors of railway ballast under cyclic loading with dilated polyhedra. *International Journal for Numerical and Analytical Methods in Geomechanics.* 41:180-197, 2017.

[150] D. Meagher. Octree encoding: A new technique for the representation, manipulation and display of arbitrary 3-D objects by computer. Rensselaer Polytechnic Institute, Technical Report IPL-TR-80-111, October 1980.

[151] W. Kaplan. *Integrals Depending on a Parameter - Leibnitz's Rule. Advanced Calculus* 2nd ed, Addison-Wesley. pp. 285-288, 1973.

[152] L. Leal. *Advanced transport phenomena: fluid mechanics and convective transport processes.* Cambridge University Press. ISBN 978-0-521-84910-4, 2007.