



Available online at www.sciencedirect.com

ScienceDirect

Comput. Methods Appl. Mech. Engrg. 373 (2021) 113454

**Computer methods
in applied
mechanics and
engineering**

www.elsevier.com/locate/cma

An energy-conserving contact theory for discrete element modelling of arbitrarily shaped particles: Basic framework and general contact model

Y.T. Feng

Zienkiewicz Centre for Computational Engineering, Swansea University, UK

Received 6 July 2020; received in revised form 10 August 2020; accepted 21 September 2020

Available online 5 October 2020

Abstract

The first paper of this series establishes a unified theoretical framework that lays a solid foundation for developing energy-conserving normal contact models for arbitrarily shaped bodies in the discrete element method. It is derived based solely on the requirement that the potential energy must be conserved for an elastic impact of two shapes under any condition. The resulting general energy-conserving contact model states that the normal force as a vector must be the gradient of a contact potential field. When such a contact potential or energy function is specified, a complete normal contact model for a pair of arbitrarily shaped particles, including the contact normal direction, contact point/line and force magnitude, will be automatically followed without introducing any additional assumptions. In this framework, the contact geometry and contact force are indispensably related and are evaluated in a consistent manner. Due to the paramount role that the energy function plays in the current theory, its fundamental properties are discussed, which serve as general guidance for choosing a valid energy function. In addition, both single and multiple contacts and their evolution can be handled in a seamless way. Some symmetric properties of particle shapes can also be utilised to simplify the contact models. Within the proposed theoretical framework, different choices or combinations of geometric features as variables for the contact energy function can give rise to unique types of energy-conserving contact models with distinct characteristics and features. Two such functions using only one primary feature, which lead to two specialised energy-conserving contact models, will be presented in the subsequent papers of this series.

© 2020 Elsevier B.V. All rights reserved.

Keywords: Discrete element; Unified contact theory; Convex and concave shapes; Energy-conserving contact model; Multiple contacts

1. Introduction

The classic discrete element methods (DEM) [1] have been firmly established as one of the most effective computational modelling techniques for modelling systems exhibiting discrete or particulate behaviour in many diverse scientific and engineering fields. These applications range from granular matters in physics to rock and soil mechanics, from processing and chemical engineering, civil, geo or mining engineering, to agricultural or environmental engineering, from the manufacturing sector to the defence industry.

DEM has been coupled with other continuum based methods (such as the finite element method (FEM) [2] and the material point method (MPM) [3,4]), and also with other physical phases (such as fluids, thermal and

E-mail address: y.feng@swansea.ac.uk.

magnetic fields) in conjunction with the corresponding numerical techniques, including various CFD methods [5]. This coupling approach greatly extends the applicability of DEM to much wider applications. A large body of literature exists that has covered many aspects of DEM from theoretical developments, computational techniques and implementation to applications. See monographs on DEM and its applications in various engineering fields [2,6–9] and the references therein.

The growing popularity of DEM, particularly in industrial applications, may be attributed to the following reasons: (1) the discrete and rigid nature of particles; (2) a simple computational framework with discs/spheres as the principle geometric primitives; (3) a set of explicitly defined contact laws replacing complicated contact interactions between solid particles; and (4) the availability of commercial software (such as PFC [10], EDEM [11], StarCCM+ [12]) and open-source codes (such as LIGGGTHS [13], Yade [14] and MatDEM [15]), coupled with the use of more accessible parallel computing platforms (GPU or GPGPU for instance).

Nevertheless, there are still many challenging issues to be resolved in discrete element modelling. There is an increasing need to develop more fundamental contact models for general non-disc/spherical (convex and concave) particles. Particle shape has long been recognised as one of the most important factors that affect the physical behaviour of granular systems in real applications. Since the 1980s, various non-disc/spherical or irregular particle shapes have been proposed in DEM including ellipses [16,17], ellipsoids [18], super-quadratics [19], cylinders [20], dilated or Minkowski-sum generated shapes [21–23], poly-ellipsoids [24,25] and poly-super-ellipsoids [26], polygons [27,28] and polyhedra [29–32]. More recent inclusions are level-set defined shapes based on CT scanned images [33], and analytical function represented smooth shapes, such as Fourier series-based for 2D shapes [34], NURBS (or B-spline) based [35] 3D shapes, spherical harmonic represented 3D objects [36,37], and iso-geometric representation of 3D surfaces [38,39]. No doubt, these shapes significantly enhance the shape representation capability of DEM for industrial applications.

A complete contact model for a pair of shapes consists of two aspects: geometric and physical. The geometric aspect defines all the geometric features associated with a contact, including the contact region/surfaces, contact volume/area/overlap, contact normal direction, and contact point; while the physical aspects defines the resulting normal contact force (or its magnitude) in terms of some of these contact geometric features. By combining these two aspects, a contact model should unambiguously define the direction, contact point and magnitude of the normal contact force for any possible configuration of the two shapes concerned.

Typically in DEM, the geometric aspect is treated as a contact detection or local resolution issue, while the physical aspect is specified by various contact interaction or constitutive laws. In almost all existing contact models developed for non-disc/spherical contact models, the contact geometric features are the focus and are evaluated independent of the contact laws used. In the other words, the geometric and physical aspect of a contact model for a particular pair of shapes is essentially decoupled. This decoupled contact modelling framework stems from discs and spheres where the contact geometric features are well defined and the interaction law only specifies the magnitude of the normal contact force as a linear or nonlinear function of the overlap.

However, for irregular particles, their contact geometric features are no longer uniquely defined, and different definitions may exist for a particular geometric feature. For instance, for a vertex-vertex contact situation between two polygons/polyhedra [27,29,31], defining the contact normal direction may be a challenging issue because of the boundary normal discontinuity of polygons/polyhedra at their vertices and edges. The same argument can equally apply to the overlap and the contact point. This non-uniqueness for one particular geometric feature alone will not cause any concern, but potential problems arise when the normal direction, overlap and contact point are collectively defined not in a consistent but an “*ad-hoc*” fashion. For instance, in the cylindrical particle model proposed in [20], the overlap of the two cylinders is defined independent of the contact normal. Later tests reveal that this rather inconsistent definition of the overlap with the contact normal may result in an energy increase for an elastic impact and therefore becomes a source of potential numerical instability. Similar inconsistencies between different contact geometric features can be found in many existing irregular particle contact models. Moreover, the decoupling between the geometric and physical aspects of a contact model for non-spherical particles can also give rise to the potential problem of an energy increase. This claim is not obvious but much subtle, which will be highlighted in Section 4.

In summary, contacts between non-spherical particles are much more complex than currently assumed. The issues mentioned above can only be adequately addressed from a more fundamental consideration of how to properly define the contact geometric features and force magnitude in a more systematic way. The aim of the current work

of this series is to establish such a theoretical framework purely based on one condition: *the potential energy of the two contact particles should be conserved for an elastic impact*. As will be shown later, by introducing a contact potential energy function, a complete normal contact mode for any shapes can be derived automatically without any additional assumptions, and the model is guaranteed to be energy-conserving. This provides a unified framework to establish various normal contact models based on different choices of the contact energy function. Due to the energy-conserving property, the resulting normal contact models ensure that no superficial energy will be generated, thereby leading to much more robust and stable DEM simulations for complex contact scenarios.

It is noted that a similar idea of using the concept of force potential for contact is also proposed in [2,40,41] for triangle/tetrahedron contacts in the context of the combined finite element/discrete element methodology [2]. However, the force potential used is (discretised) element based rather than the contact region based in the current work.

In the next section, the contact geometry, contact forces of two particles and interaction laws are briefly reviewed to allow for the development of a **general energy-conserving contact theory** in a fundamental way in Section 3. The theory is based on a proper definition of the incremental contact work done by a pair of contact force and moment over an infinitesimal path and examining the conservation property of the total contact work done when moving two contacting particles in their contact configuration space. **The energy-conserving requirement leads to the statement that the normal contact force must be the gradient of a scalar potential field, or the contact energy function.** By specifying such an energy function, a general energy-conserving normal contact model is fully defined, as described in Section 4. The theory clearly demonstrates that all the contact geometry features and contact force of a contact are indispensably related and should be evaluated in a consistent manner. Due to the paramount role that the energy function plays in the current theory, its fundamental properties are discussed, which will serve as general guidance for choosing a valid energy function. Both single or multiple contacts and their evolution can be handled in a seamless way. Some symmetric properties of particle shapes can also be utilised to simplify the contact models. These two issues are discussed in detail in Section 5.

Within the proposed theoretical framework, different choices or combinations of geometric features as variables for the contact energy function can give rise to unique types of energy-conserving contact models with distinct characteristics and features. Two such functions which use only one primary feature lead to two specialised energy-conserving contact models, and will be presented in the subsequent papers of this series [42,43]. The corresponding computational procedures will also be described in detail. Numerical examples involving a wide variety of 3D concave shapes will be included to verify the energy-conserving property of the contact models, and also to illustrate their robustness and applicability in more complex and realistic problems. As will be demonstrated in a later paper [43], the existing contact models (linear or Hertz) for discs/spheres can be recast as a special case within this framework and are energy-conserving as expected.

2. Preliminary: Contact geometry, forces, and interaction laws

2.1. Contact geometry

Consider, at a time instance t , two arbitrarily shaped rigid bodies Ω_1 and Ω_2 with an overlap in 3D space, as shown in Fig. 1. In a fixed global coordinate system, choose two arbitrary points, with coordinates \mathbf{x} and \mathbf{x}' at the two bodies respectively, as the two reference points. These points are often chosen to be the **mass or geometric centroids** of the bodies. The angular orientations of the bodies are represented by **two vectors θ and θ'** respectively.

The intersection of the two bodies, shown in Fig. 1, forms an **overlap or contact region** Ω_c

$$\Omega_c \equiv \Omega_1 \cap \Omega_2 \neq \emptyset$$

Let $S_i \equiv \partial\Omega_i$ ($i = 1, 2$) be the **boundary surface of the body** i , and $S_c \equiv \partial\Omega_c$ the boundary of the contact region. S_c is generally well defined, but can be a multi-connected domain. Divide S_c into two parts:

$$S_c = S_1^c + S_2^c \quad (1)$$

where S_i^c ($i = 1, 2$) is part of the contact boundary that belongs to the body i :

$$S_i^c = S_c \cap S_i$$

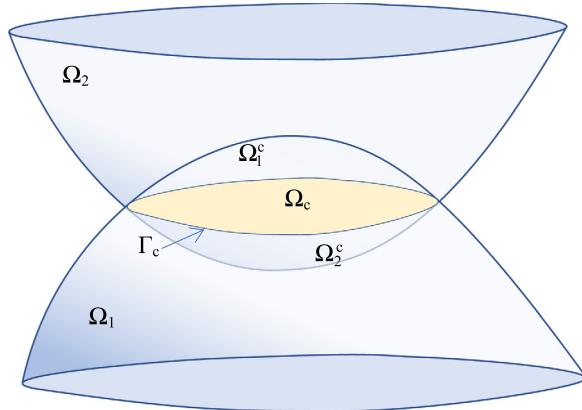


Fig. 1. Two generally shaped bodies Ω_1 and Ω_2 in contact and their contact region.

Both surfaces S_1^c and S_2^c are open and share a common and properly oriented boundary, denoted as Γ^c . In general Γ^c is a closed curve but may consist of multiple disjointed closed curves.

The contact region Ω_c normally determines the geometric aspect of the contact between the two bodies concerned. Define g_k , ($k = 1, \dots, n_g$) to be n_g characteristics or geometric features of the contact region Ω_c :

$$g_k = |\Omega_c|_k \quad (k = 1, \dots, n_g)$$

where $|\Omega|_k$ denotes a particular measure of a domain Ω . These measures, i.e. geometric features, must be coordinate invariant, as will be addressed later. The commonly used contact geometric features include: overlap or penetration depth, (contact) area or volume. However, some geometric features may be associated not only with the contact region Ω_c , but also with its vicinity of Ω_1 and Ω_2 . This particular issue will be further exploited and explained in another paper of this series [43].

In what follows, where only the contact region and the corresponding contact surfaces are considered, the superscript c will be dropped from S^c , S_1^c , S_2^c and Γ^c for simplicity.

It is also noted that both \mathbf{x} and \mathbf{x}' introduced earlier will specifically represent the coordinates of the two default reference points in the next section for simplicity. However, \mathbf{x} will be used elsewhere to represent the coordinates of an arbitrary point, and at the same time $\mathbf{r} = \mathbf{x} - \mathbf{o} = \mathbf{x}$ is adopted to denote its position vector from the origin. Both \mathbf{x} and \mathbf{r} are mathematically equivalent but may have slightly different geometric or physical meanings. Therefore \mathbf{x} and \mathbf{r} will be selectively chosen depending on the circumstances.

For further simplification (but without altering any conclusion), the second body Ω_2 is always assumed stationary, unless stated otherwise.

2.2. Contact forces

In the context of continuum mechanics, when two solid bodies Ω_1 and Ω_2 are brought into contact, they are deformed at the vicinity of the contact zone, and a pair of equal but opposite distributed contact tractions exist over the contact area on each body. In contact mechanics, these tractions are determined, together with the deformation, by solving the contact problem concerned, typically by penalty or Lagrangian methods [44].

In classic DEM, all solid bodies are rigid and cannot deform. Thus, the real contact zone between two bodies has to be approximated by allowing the two bodies to overlap and form a contact region. There is no underlying mechanism that can obtain the contact traction. Instead, only the resultant force \mathbf{F} of the traction is assumed to be applied at a point in the overlapping zone on one contacting body. An equal but opposite force $\mathbf{F}' = -\mathbf{F}$ is applied at the same point but on the other body. The resultant contact moment about the point is ignored in general, even if it exists.

The force \mathbf{F} is further resolved into a normal component, \mathbf{F}_n , and a tangential component, \mathbf{F}_τ . The *contact direction, point and magnitude* are three essential elements of the contact force \mathbf{F} , and are collectively termed the *contact force features* in this work.

For discs or spheres, the normal direction of a contact is well defined. However, this may not be the case for most non-spherical shapes where the normal contact direction is often defined on an *ad-hoc* basis. Furthermore, the tangential contact force is generally considered as a frictional force, although this is also the resultant of many contact forces due to asperities at a microscopic level.

2.3. Contact interaction laws

As all elements in DEM are rigid and no deformation is allowed, the contact forces cannot be evaluated as part of a solution procedure. Thus they have to be pre-defined by *contact interaction laws*. A complete contact interaction law should explicitly define the relationship between the contact force and the contact geometric features together with some history variables of the contact, if necessary. The force can be normal or tangential, and all the contact features of the force, i.e. direction, point and magnitude, must also be defined unambiguously by the interaction law.

As interaction laws are deemed to represent the true contact mechanics behaviour of the two bodies to a certain degree, they should be physically based, i.e. either theoretically developed based on underlying physical principles or obtained experimentally. An inappropriate contact law can lead to the erroneous physical behaviour of the particle system being modelled by DEM.

Historically, the most widely used interaction laws are proposed for discs and spheres. For these shapes, the overlap as a contact geometric feature can be readily calculated — the contact direction is fixed and the contact point or centre can be specified within a range. Thus, an interaction law mainly defines the relationship between the magnitude of the force and the overlap. For mechanical contact without considering adhesion/cohesion, the Hertzian contact model [45] is theoretically derived for two linearly elastic spheres under a normal contact, while the Hertz–Mindlin model [45,46] is for elastic spheres in an oblique contact. All other normal contact laws, including the linear spring model, are of an approximation nature. The classic Coulomb friction model is the most simple and widely used tangential interaction law.

The Hertzian contact theory is often extended to non-spherical but elastic shapes, such as ellipsoids [18] or super-quadratics [9], where their surfaces can be described by analytical functions. However, the resulting model cannot be expressed in a simple explicit form, and the force values need to be evaluated numerically, unless further assumptions are made.

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. The same applies to realistic shapes obtained digitally and represented by meshes or other forms [35]. In this sense, existing contact models for non-spherical particles can all be classified as numerically-based.

As already mentioned in the introduction, almost all existing contact models independently define the contact geometry from the magnitude of the normal contact force. This decoupled approach can cause non-physical behaviours in DEM simulations. Hence there has a growing need to develop a general contact theory for arbitrarily shaped particles that can provide a unified approach to treat all the contact geometric features and the normal contact force in a consistent manner, such that robust normal contact models for these particles can be derived in a systemic way.

Note that the resulting contact model may not be physically accurate as this is not realistic, but it should possess certain properties that can reflect fundamental features of a granular system. The conservation of elastic potential energy is the most fundamental property for an elastic impact. Purely based on the requirement of energy conservation, the next section will develop a comprehensive normal contact theory that guarantees the conservation of elastic energy for arbitrarily shaped particle collisions. Based on this theory, a general energy-conserving contact model will naturally follow to define all the contact features uniquely.

Under this new framework, the interaction law will describe the normal contact force as a vector in terms of some contact geometric features. The contact geometric features depend on the positions, orientations and shapes of the two contacting bodies. Thus, the most general form of an interaction law for the normal contact force can be expressed as a function of the positions and velocities of the contact pair, but may also depend on some internal variables, or may be explicitly related to time t :

$$\mathbf{F}_n = \mathbf{F}_n(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}, \boldsymbol{\theta}', \dot{\mathbf{x}}, \dot{\mathbf{x}}', \dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\theta}}', t, \Psi) \quad (2)$$

where $\Psi = \{\psi_1, \dots, \psi_{n_\psi}\}$ denotes a set of internal variables ψ_i . From \mathbf{F}_n , the magnitude F_n and the unit normal direction \mathbf{n} can be obtained as

$$F_n = \|\mathbf{F}_n\|; \quad \mathbf{n} = \mathbf{F}_n / F_n \quad (3)$$

Note that the normal force should be acting at the contact point, but this point as a contact geometric feature is not defined by (2), and therefore needs to be determined based on an additional condition.

In the next section, before the contact point is formally defined, the normal contact force \mathbf{F}_n is technically transferred from the (unknown) contact point to the reference point \mathbf{x} but must be accompanied by a moment \mathbf{M}_n . This is to follow the transmissibility principle in statics to ensure that the force–couple pair $(\mathbf{F}_n, \mathbf{M}_n)$ acting at the reference point is equivalent to a single force \mathbf{F}_n acting at the contact point.

3. An energy-conserving contact theory

The main idea of this theory is proposed in [27] for polygons and is then extended to polyhedra and general shapes in [30,31]. A concise general form is presented in [47]. In this section, the theory will be re-established in a more fundamental and rigorous manner, and many of its new aspects will also be exploited.

3.1. Configuration space, motion, contact forces and contact energy

Both the position and orientation of a rigid body are fully determined by \mathbf{x} and $\boldsymbol{\theta}$ at the chosen reference point, and can be represented by a point in the $(\mathbf{x}, \boldsymbol{\theta})$ space which is termed the *configuration* space and denoted as \mathcal{C} . The domain occupied by the body can be expressed as a function $\Omega(\mathbf{x}, \boldsymbol{\theta})$. The motion of a rigid body in the physical space corresponds to a path in its configuration space.

In association with the reference point $(\mathbf{x}, \boldsymbol{\theta})$ of Body 1 in its configuration space, we can define the corresponding *contact force* (vector) \mathbf{F} and *contact moment* (vector) \mathbf{M} acting at the reference point due to the contact with Body 2 (which is assumed fixed). The force–moment pair (\mathbf{F}, \mathbf{M}) defines the so called *conjugate* space \mathcal{F} such that at each point in \mathcal{C} , we can define an incremental *contact work done* along any infinitesimal path, defined by $(d\mathbf{x}, d\boldsymbol{\theta})$, as

$$dW = -(\mathbf{F}, \mathbf{M}) \cdot (d\mathbf{x}, d\boldsymbol{\theta}) = -(\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) \quad (4)$$

Further define a total *contact work done*, W , along a path $\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}}$ in the configuration space from a starting point \mathbf{a} to an end point \mathbf{b} as

$$W(\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}}) = \int_{\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}}} dW = \int_{\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}}} -(\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) \quad (5)$$

which is equal to the total external work required to move Body 1 against Body 2 along the path from \mathbf{a} to \mathbf{b} . This work done can also be considered as the total contact energy stored in the two-body system after the motion.

As there are an infinite number of paths that can be taken from \mathbf{a} to \mathbf{b} , $W(\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}})$ may be path-dependent or multi-valued. This path dependency depends solely on how \mathbf{F} and \mathbf{M} are defined and if they are conservative fields.

3.2. Contact energy generation, conservation and dissipation

Consider that Body 1 moves in the configuration space along a closed path $\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{a}}$ that starts and ends at the same point \mathbf{a} . The total external work done along the path is

$$W(\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{a}}) = - \oint_{\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{a}}} (\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta})$$

There are three possible cases: $W > 0$, $= 0$ or < 0 which are discussed below.

- Case 1: $W > 0$. Some contact energy has been dissipated. Extra net energy needs to be inputted to the system in order for Body 1 to return back to its original state. When the same path is followed repeatedly, more energy needs to be inputted, and therefore the contact system must dissipate energy and exhibits a dissipative nature. From a computational point of view, the system will eventually converge to a stationary state if no extra energy is put into the system.

- Case 2: $W = 0$. The contact energy is conserved. Both the net external and internal energy are zero, and no extra energy is inputted or gained from the contact system. Thus the system returns back to its original state after one complete motion cycle.
- Case 3: $W < 0$. Extra energy has been generated. The net energy input required to return the contact system back to its original state is negative, thus the contact force and moment must have produced extra energy in one complete cycle. If the cycle repeats, extra energy will be continuously generated and will accumulate in the system, which indicates that the system exhibits an unrealistic and non-physical behaviour which will eventually cause numerical instability.

Thus, to avoid the non-physical behaviour of a contact system, the contact energy W associated with \mathbf{F} and \mathbf{M} for a closed path must be dissipative or at least conserved, i.e. $W \geq 0$. This applies to any closed path through any point in the configuration space. Consequently, a pair of given contact force and moment fields (\mathbf{F}, \mathbf{M}) are computationally acceptable only if the total work done along any closed path \mathcal{L} is non-negative:

$$W(\mathcal{L}) = - \oint_{\mathcal{L}} (\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) \geq 0; \quad \forall \mathcal{L} \text{ (closed path)} \in \mathcal{C} \quad (6)$$

However, if \mathbf{F} and \mathbf{M} are functions of \mathbf{x} and $\boldsymbol{\theta}$ only, they cannot be dissipative, as will be proved below. Suppose that \mathbf{F} and \mathbf{M} are dissipative for a closed path \mathcal{L} for a certain direction of travel, and that the total work done is

$$W(\mathcal{L}) = - \oint_{\mathcal{L}} (\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) > 0$$

Consider a new closed path $\mathcal{L}' = -\mathcal{L}$, which is the reverse of the original path. \mathbf{F} and \mathbf{M}_c on \mathcal{L}' remain the same as on \mathcal{L} , but the path travel direction ($d\mathbf{x}, d\boldsymbol{\theta}$) is reversed. Thus the work done on the new path must be positive:

$$W(\mathcal{L}') = - \oint_{\mathcal{L}'} (\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) = \oint_{\mathcal{L}} (\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) = -W(\mathcal{L}) < 0$$

so \mathbf{F} and \mathbf{M} are not dissipative which contradicts the original assumption.

Hence, it can be concluded that *the contact system cannot dissipate energy if \mathbf{F} and \mathbf{M} are purely determined by \mathbf{x} and $\boldsymbol{\theta}$, or in general by the contact geometric features of the two bodies*, unless the contact force and moment are also affected by other factors, such as velocity, path or time-dependent internal or state variables.

Furthermore, when developing elastic contact interaction models based on contact geometric features only, the models have to conserve energy. Otherwise, numerical instability during DEM simulations may occur due to the accumulation of superficially generated contact energy. Thus, energy conservation should be the most basic requirement for any elastic contact interaction law in DEM.

3.3. Energy conservative contact force and moment

It is of both theoretical and computational importance in DEM to establish the combined force–moment vector field (\mathbf{F}, \mathbf{M}) that can lead to energy conservation, i.e. satisfies the condition

$$W(\mathcal{L}) = \oint_{\mathcal{L}} (\mathbf{F} \cdot d\mathbf{x} + \mathbf{M} \cdot d\boldsymbol{\theta}) = 0; \quad \forall \mathcal{L} \text{ (closed path)} \in \mathcal{C} \quad (7)$$

This is also equivalent to stating that the work done along any path between two points \mathbf{a} and \mathbf{b} in the configuration space is a path-independent constant:

$$W(\mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}}) = \text{const}; \quad \forall \mathcal{L}_{\mathbf{a} \rightarrow \mathbf{b}} : \mathbf{a}, \mathbf{b} \in \mathcal{C} \quad (8)$$

According to the gradient theorem in calculus [48], the above two equivalent conditions (7) and (8) can be satisfied if (\mathbf{F}, \mathbf{M}) is a potential vector field, i.e. there exists a *scalar* potential function, termed the *contact energy function*, $w(\mathbf{x}, \boldsymbol{\theta})$, such that the (\mathbf{F}, \mathbf{M}) is the negative gradient of the function:

$$(\mathbf{F}, \mathbf{M}) = -\nabla w(\mathbf{x}, \boldsymbol{\theta}) \quad (9)$$

or in component form

$$\mathbf{F} = -\nabla_x w(\mathbf{x}, \boldsymbol{\theta}) \equiv -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta})}{\partial \mathbf{x}}; \quad \mathbf{M} = -\nabla_{\boldsymbol{\theta}} w(\mathbf{x}, \boldsymbol{\theta}) \equiv -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (10)$$

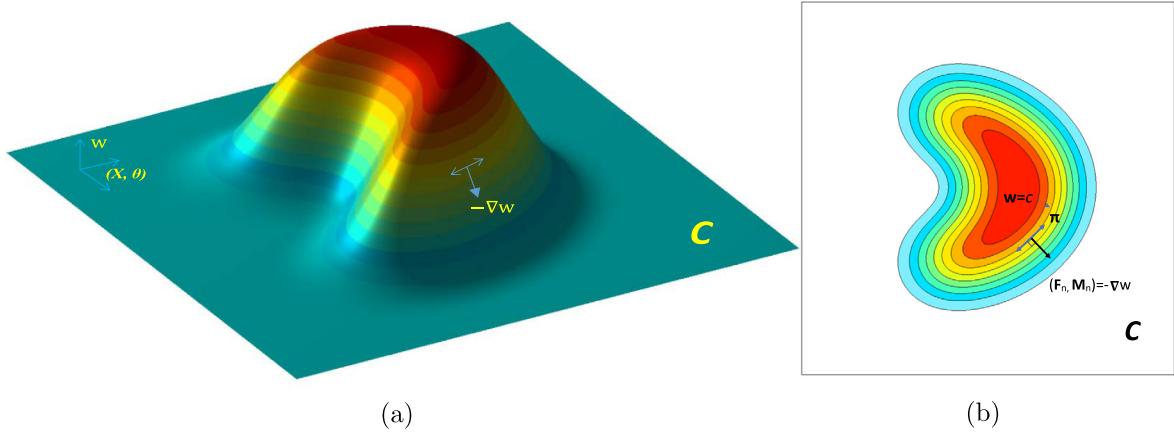


Fig. 2. An illustrative energy function w : (a) the *energy mountain* in the configuration space \mathcal{C} ; and (b) the contour plot.

The above conclusion has a sound physical explanation: if the current contact state is fully described by a contact energy function w , the true contact force and moment at the state will reduce the contact energy most effectively or at the greatest rate, i.e. along the negative gradient direction of w .

The contact energy function w for the energy conservative vector field (\mathbf{F}, \mathbf{M}) is not unique, but up to an arbitrary constant. This can be resolved by setting zero as the reference value for any point in the non-contact region.

Thus, $w(\mathbf{x}, \boldsymbol{\theta})$ defines a global function in the configuration space \mathcal{C} , and it can be represented by an 'energy mountain' or iso-energy contours, as illustrated in Fig. 2. The negative gradient direction of the energy at each point on the energy map is normal to the iso-energy line, and therefore determines the direction of the force and moment at the position. The tangential plane π at a point on an iso-energy line defines all possible directions of motion at the vicinity of the point along which no contact force/moment will be experienced, as the change in energy is zero.

Hence, the force \mathbf{F} defined above must be the *normal contact force*. So from now on, \mathbf{F} is denoted as \mathbf{F}_n , and \mathbf{M} as \mathbf{M}_n , with

$$\mathbf{F}_n = -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta})}{\partial \mathbf{x}} \quad (11)$$

$$\mathbf{M}_n = -\nabla_{\boldsymbol{\theta}} w(\mathbf{x}, \boldsymbol{\theta}) \equiv -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (12)$$

Consequently, the explicit expression (11) for \mathbf{F}_n leads to the formal definition of the *magnitude* $F_n = \|\mathbf{F}_n\|$ and the *normal contact direction* \mathbf{n} as the unit vector of \mathbf{F}_n

$$\mathbf{n} = \mathbf{F}_n / F_n \quad (13)$$

3.4. Normal contact force and moment for a new reference point

Both the normal contact force \mathbf{F}_n and moment \mathbf{M}_n established in the preceding section act on the reference point \mathbf{x} of Body 1. When choosing a new reference point $\bar{\mathbf{x}}$ instead of \mathbf{x} for the body, while adopting the same $\boldsymbol{\theta}$ for orientation, the normal contact force and moment will be applied at this new point. Let the new contact force and moment be denoted $\bar{\mathbf{F}}_n$ and $\bar{\mathbf{M}}_n$. The relationship between $(\mathbf{F}_n, \mathbf{M}_n)$ and $(\bar{\mathbf{F}}_n, \bar{\mathbf{M}}_n)$ will be established below. This relationship will be used in the next subsection to formally define the contact point.

Define the relative position vector of $\bar{\mathbf{x}}$ to \mathbf{x} to be $\bar{\mathbf{r}} = \bar{\mathbf{x}} - \mathbf{x}$, so

$$\bar{\mathbf{x}} = \mathbf{x} + \bar{\mathbf{r}}$$

First consider the differential of $\bar{\mathbf{x}}$ (which has a fixed distance to \mathbf{x}) due to the infinitesimal changes of both \mathbf{x} and $\boldsymbol{\theta}$ at the original reference point:

$$d\bar{\mathbf{x}} = d\mathbf{x} + d\boldsymbol{\theta} \times \bar{\mathbf{r}} \quad (14)$$

from which we can identify the relations

$$\frac{\partial \bar{\mathbf{x}}}{\partial \mathbf{x}} = \mathbf{I}_3$$

and

$$\frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\theta}} = \mathbf{I}_3 \times \bar{\mathbf{r}} \quad (15)$$

where \mathbf{I}_3 is the identity matrix of order 3. Note that $\mathbf{I}_3 \times \bar{\mathbf{r}}$ is a skew-symmetric matrix associated with $\bar{\mathbf{r}} = (r_x, r_y, r_z)$ as

$$\mathbf{I}_3 \times \bar{\mathbf{r}} = -\bar{\mathbf{r}} \times \mathbf{I}_3 = \begin{pmatrix} 0 & r_z & -r_y \\ -r_z & 0 & r_x \\ r_y & -r_x & 0 \end{pmatrix} \quad (16)$$

Thus the contact energy function w can be re-written in terms of $\bar{\mathbf{x}}$ and $\boldsymbol{\theta}$ as

$$w(\mathbf{x}, \boldsymbol{\theta}) = w(\bar{\mathbf{x}} + \mathbf{r}, \boldsymbol{\theta}) = \bar{w}(\bar{\mathbf{x}}, \boldsymbol{\theta}) \quad (17)$$

From the energy conserving definition (11), the normal force associated with this new reference point is

$$\bar{\mathbf{F}}_n = -\frac{\partial \bar{w}}{\partial \bar{\mathbf{x}}} = -\frac{\partial w}{\partial \mathbf{x}} = \mathbf{F}_n \quad (18)$$

This indicates that the normal force \mathbf{F}_n remains the same and thus is independent of the reference point chosen, as expected. As $\boldsymbol{\theta}$ is also considered as a rotation about the new point, the associated moment is by definition (12)

$$\bar{\mathbf{M}}_n = -\frac{\partial \bar{w}}{\partial \boldsymbol{\theta}} \quad (19)$$

while the moment at the original reference point is

$$\mathbf{M}_n = -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial \bar{w}(\bar{\mathbf{x}}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial \bar{w}}{\partial \bar{\mathbf{x}}} \cdot \frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\theta}} - \frac{\partial \bar{w}}{\partial \boldsymbol{\theta}} = \mathbf{F}_n \cdot \frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\theta}} + \bar{\mathbf{M}}_n \quad (20)$$

Thus, utilising (15), we have

$$\mathbf{M}_n = \frac{\partial w}{\partial \boldsymbol{\theta}} + \bar{\mathbf{M}}_n = \mathbf{F}_n \cdot (\mathbf{I}_3 \times \bar{\mathbf{r}}) + \bar{\mathbf{M}}_n = \bar{\mathbf{r}} \times \mathbf{F}_n + \bar{\mathbf{M}}_n \quad (21)$$

or

$$\bar{\mathbf{M}}_n = \mathbf{M}_n - \mathbf{r} \times \mathbf{F}_n \quad (22)$$

Consequently, both the normal contact force (18) and the moment (22) at the new reference point exactly follow the principle of transmissibility for a force–couple in statics of mechanics. Thus the reference point can be chosen at any position in principle.

3.5. Contact point and normal contact line

The contact moment \mathbf{M}_n is introduced to account for the resistance that one body experiences against the other in a rotational motion about the reference point \mathbf{x} . However, this moment is '*nominal*' in that no contact moment is explicitly considered in the contact region in existing contact models. Typically, the normal contact force is applied at a point in the region without considering any contact moment. This point is usually called the *contact point*.

As discussed in the introduction, in addition to the contact normal direction, the contact point is also defined in an *ad-hoc* manner in conventional discrete element interaction models. For a non-disc/sphere contact model, if the contact point is selected arbitrarily, there is no guarantee that energy conservation will hold for a pure elastic impact.

However, energy conservation is ensured when both \mathbf{F}_n and \mathbf{M}_n are determined by (11) and (12) for the reference point \mathbf{x} within the current normal contact model framework. This reference point could also be considered as the contact point.

As proved in the previous section, the choice of the reference point is arbitrary, and any point can serve such a role. However, a general reference point may not be located in the contact zone and an accompanying moment

will need to be applied. Thus such a reference point may not be an appropriate candidate for the contact point, and particularly when a tangential force also needs to be applied at the point. These considerations lead to the formal definition of the contact point.

The establishment of (22) for the new contact moment at a different reference point suggests that there must exist a point where the corresponding contact moment will disappear. Thus this particular point, if exists, is defined as the *contact point*. Denote \mathbf{x}_c to be its coordinates and $\mathbf{r}_c = \mathbf{x}_c - \mathbf{x}$ its position vector.

Eq. (22) then gives rise to the following condition that can be used to determine the contact point \mathbf{x}_c

$$\mathbf{M}_c = \mathbf{M}_n - \mathbf{r}_c \times \mathbf{F}_n = 0 \quad (23)$$

or

$$(\mathbf{x}_c - \mathbf{x}) \times \mathbf{F}_n = \mathbf{M}_n \quad (24)$$

from which \mathbf{x}_c can be found to be

$$\mathbf{x}_c = \mathbf{x} + \frac{\mathbf{n} \times \mathbf{M}_c}{F_n} + \lambda \mathbf{n} \quad (25)$$

where \mathbf{x} is the coordinates of the reference point and λ is a free parameter to be determined.

The presence of λ defines a fixed line along the contact normal direction, which implies that the contact point \mathbf{x}_c is not unique and can be taken as any point along the line. This line, which is spatially fixed, is therefore termed *the normal contact line*. From the point of view of the normal contact, any point on the line will result in the same normal motion of the bodies. However, the position of this point will have different effects on the tangential force when it has to be considered. As the current work focuses on the normal force, this issue of determining a unique contact point will not be addressed here. *However, by introducing a special condition, a fixed contact point can be found, as will be discussed in another paper of this series [42].*

4. A general energy conserving normal contact model

As discussed in the previous section, the contact energy function w is a function of the position and orientation of the first body. This is under the assumption that the second body is stationary and therefore its position and orientation $(\mathbf{x}', \boldsymbol{\theta}')$ are not presented as additional variables for the energy function. In general, both bodies are movable so w should be expressed as a function of both $(\mathbf{x}, \boldsymbol{\theta})$ and $(\mathbf{x}', \boldsymbol{\theta}')$. However, all the conclusions drawn above remain valid.

All the preceding discussions lead to a general energy conserving normal contact model for a given energy function w . The details of this general model are summarised below.

Model 1 (General Energy-Conserving Normal Contact Model). *For two (rigid) bodies in contact, assume that their positions and orientations can be described by $(\mathbf{x}, \boldsymbol{\theta})$ and $(\mathbf{x}', \boldsymbol{\theta}')$ respectively. The following normal contact model will be energy-conserving if there exists a scalar potential function $w(\mathbf{x}, \boldsymbol{\theta}, \mathbf{x}', \boldsymbol{\theta}')$ such that*

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

$$\mathbf{F}_n = -\nabla_{\mathbf{x}} w(\mathbf{x}, \boldsymbol{\theta}, \mathbf{x}', \boldsymbol{\theta}') = -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta}, \mathbf{x}', \boldsymbol{\theta}')}{\partial \mathbf{x}} \quad (26)$$

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

$$\mathbf{F}'_n = -\mathbf{F}_n \quad (27)$$

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

$$\mathbf{n} = \mathbf{F}_n / F_n \quad (28)$$

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

(3) *The contact point c , at which only the contact force \mathbf{F}_n needs to be applied, is given by*

$$\mathbf{x}_c = \mathbf{x} + \frac{\mathbf{n} \times \mathbf{M}_n}{F_n} + \lambda \mathbf{n} \quad (29)$$

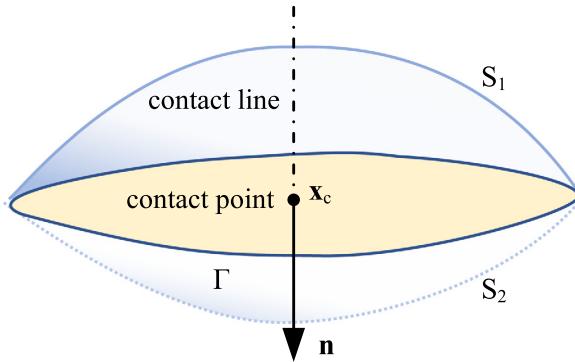


Fig. 3. Illustration of the contact normal, point and line of the body Ω_1 .

where λ is a free parameter; and \mathbf{M}_n is a nominal contact moment that is not applied to the contact point and is defined as

$$\mathbf{M}_n = -\nabla_{\theta} w(\mathbf{x}, \boldsymbol{\theta}, \mathbf{x}', \boldsymbol{\theta}') = -\frac{\partial w(\mathbf{x}, \boldsymbol{\theta}, \mathbf{x}', \boldsymbol{\theta}')}{\partial \boldsymbol{\theta}} \quad (30)$$

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 π is the plane perpendicular to the contact normal \mathbf{n} and passes through the chosen contact point.

[Fig. 3](#) illustrates the contact normal, point and line for Body 1 defined by the above model. Both (26) and (30) clearly indicate that the contact force and moment are indispensably related to the contact geometry in a much complex manner and can constantly evolve as the contact geometry changes.

The remainder of this section will discuss some theoretical issues inherent in this general energy-conserving normal contact model.

4.1. Fundamental properties of the contact energy function

The contact energy function w plays a pivotal role in the development of the above general energy conserving contact theory and the general contact model. This subsection will further establish some basic properties that the energy function should satisfy. The consequences of such properties will provide general guidelines for choosing a valid energy function.

4.1.1. Coordinate invariance

The most fundamental property of the energy function is that it should be independent of the chosen global coordinate system. Thus, when considering the contact geometry of two bodies, a local, rather a global system, convenient to this pair of bodies, should be used. Equivalently, the energy function should remain unchanged when the coordinate system is subjected to any translation and/or rotation. This coordinate system invariance of the energy function can be further exploited to establish other important relationships that the current contact theory implies.

Note that in the previous section, the reference position \mathbf{x}' and orientation $\boldsymbol{\theta}'$ of Body 2 are assumed constant, so \mathbf{x}' and $\boldsymbol{\theta}'$ are omitted from w and the subsequent derivations. In the following discussion, \mathbf{x}' and $\boldsymbol{\theta}'$ will be treated as variables under a coordinate transformation. Thus, we explicitly assume that w is also a function of \mathbf{x}' and $\boldsymbol{\theta}'$ when necessary.

1. Translation invariance: The function w remains the same when the coordinate system is subjected to any pure translation.

Then, ignoring the dependency of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, w must have the property:

$$w(\mathbf{x}, \mathbf{x}') = \bar{w}(\mathbf{x} - \mathbf{x}') \quad (31)$$

This can be proved as follows. Consider an arbitrary translation \mathbf{u} of the coordinate system. The translation invariance requires

$$w(\mathbf{x}, \mathbf{x}') = w(\mathbf{x} + \mathbf{u}, \mathbf{x}' + \mathbf{u})$$

Set $\mathbf{u} = -\mathbf{x}'$, it follows

$$w(\mathbf{x}, \mathbf{x}') = w(\mathbf{x} - \mathbf{x}', 0) = \bar{w}(\mathbf{x} - \mathbf{x}')$$

which is the stated property (31). Thus, the energy function can be condensed to $w(\mathbf{x} - \mathbf{x}', \boldsymbol{\theta}, \boldsymbol{\theta}')$ with three sets of variables instead of four.

Following (11),

$$\mathbf{F}_n = -\frac{\partial w}{\partial \mathbf{x}} = \frac{\partial w}{\partial \mathbf{x}'} = -\mathbf{F}'_n; \quad \text{or } \mathbf{F}_n + \mathbf{F}'_n = 0 \quad (32)$$

i.e. \mathbf{F}_n and \mathbf{F}'_n must be a action and reaction pair, as expected.

2. Rotation invariance: When the default global coordinate system is subjected to a rotation of any angle, w remains unchanged.

The full differential of the function w for a fixed coordinate system is

$$dw = \frac{\partial w}{\partial \mathbf{x}} \cdot d\mathbf{x} + \frac{\partial w}{\partial \mathbf{x}'} \cdot d\mathbf{x}' + \frac{\partial w}{\partial \boldsymbol{\theta}} \cdot d\boldsymbol{\theta} + \frac{\partial w}{\partial \boldsymbol{\theta}'} \cdot d\boldsymbol{\theta}' \quad (33)$$

Consider the case where the origin of the coordinate system coincides with the reference point \mathbf{x} of Body 1, and the system is subjected to an infinitesimal rotation of $d\boldsymbol{\theta}^*$ about the origin. Such a rotation is equivalent to rotating both bodies about the origin with the same angle but in the opposite direction. Thus, we have the following relations:

$$d\mathbf{x} = 0; \quad d\boldsymbol{\theta} = d\boldsymbol{\theta}^*; \quad d\mathbf{x}' = -d\boldsymbol{\theta}^* \times \mathbf{r}'; \quad d\boldsymbol{\theta}' = -d\boldsymbol{\theta}^* \quad (34)$$

where $\mathbf{r}' = \mathbf{x}' - \mathbf{x}$. As w remains unchanged under any rotation, $dw = 0$. Thus (33) reduces to

$$\frac{\partial w}{\partial \mathbf{x}'} \cdot d\mathbf{x}' + \frac{\partial w}{\partial \boldsymbol{\theta}} \cdot d\boldsymbol{\theta} + \frac{\partial w}{\partial \boldsymbol{\theta}'} \cdot d\boldsymbol{\theta}' = 0; \quad (35)$$

By utilising (11), (12) and (15), we have

$$\mathbf{F}'_n \cdot (d\boldsymbol{\theta}^* \times \mathbf{r}) + \mathbf{M}_c \cdot d\boldsymbol{\theta}^* + \mathbf{M}'_n \cdot d\boldsymbol{\theta}^* = 0 \quad (36)$$

which, due to the arbitrary nature of $d\boldsymbol{\theta}^*$, yields

$$-\mathbf{r}' \times \mathbf{F}_n + \mathbf{M}_n + \mathbf{M}'_n = 0 \quad (37)$$

The first term $-\mathbf{r}' \times \mathbf{F}_n$ is the moment produced by the couple \mathbf{F}_n and $\mathbf{F}'_n (= -\mathbf{F}_n)$. Thus, the above equation simply states the moment equilibrium of the two bodies as a whole system.

In summary, both translational and rotational invariances lead to (32) and (37) which are simply the standard equilibrium conditions that both contact forces and moments should satisfy. However, these conditions are previously imposed on the contact model, while the conditions are naturally derived here from the requirement of the coordinate system invariance for the energy function w , i.e. as long as w is coordinate system independent, the equilibrium of the contact forces and moments is guaranteed.

4.2. Forms

The above coordinate invariance suggests that the energy function w should be a function of some geometric features of the contact region Ω_c that are coordinate invariant. Let g_1, g_2, \dots, g_{n_g} be n_g such geometric features. w can be expressed in a general form:

$$w = w(g_1, g_2, \dots, g_{n_g}) \quad (38)$$

Geometric features that are coordinate invariant include overlap, contact area and volume, and the invariants or principal moments of inertia of the contact region. The simplest option for w is to assume that it only depends on one primary geometric feature g :

$$w = w(g) \quad (39)$$

Different choices or combinations of geometric features as variables for the contact energy function can give rise to unique types of energy-conserving contact models with distinct characteristics and features. Two such functions with one primary feature, leading to two specialised energy-conserving contact models within the current theoretical framework, will be presented in the subsequent papers of this series [42,43].

4.3. Smoothness and monotonicity

As the normal force and moment are the gradient of the energy function, w should be at least C^1 continuous on the space \mathcal{C} , but the gradient is allowed to be non-smooth at lower dimensional domains. For instance, at the boundary of the contact domain in \mathcal{C} , the gradient of w may be discontinuous.

If w only depends on one geometric feature g , it should be monotonically increasing

$$w(g_a) > w(g_b), \quad \forall g_a > g_b$$

If w depends on multiple geometric features $g_i (i = 1, \dots, m)$, it should be a generalised monotonically increasing function

$$w(g_{1a}, \dots, g_{ma}) > (g_{1b}, \dots, g_{mb}), \quad \forall g_{ia} > g_{ib} (i = 1, \dots, m)$$

4.4. Geometric property consistency

For shapes having certain geometric properties, such as symmetry, the energy function should also preserve these properties. For instance, consider the contact between a cylinder and any other shape: any rotation about the axis of the cylinder will not alter the contact region so that all the contact geometric features should remain the same. It is therefore expected that the energy function also exhibits this property, i.e. the energy function should remain unchanged under such motion. Hence, invariants that are induced by these geometric properties must be presented in the energy function either implicitly or explicitly.

The consequence of this consistency of w with the underlying geometric natures of the shapes concerned will be explored in Section 5.2, where for a few special geometric shapes, the contact normal and/or the contact point/line may be determined without any prior knowledge of the actual form of the energy function.

4.5. Energy conservation and dissipation

As discussed in Section 4.2, the form of the energy function w can be simple, or very complicated. The important feature of w is that it is purely a function of the position \mathbf{x} and orientation $\boldsymbol{\theta}$ of the reference point/body. Thus the normal force and moment are also functions of these two vector variables, and so cannot depend on any other variables. Otherwise, the conservative nature of the force–moment field will not be held and energy conservation will be violated. This also implies that no history-dependent and internal state variables can be included in the expression of w . \mathbf{F}_n and \mathbf{M}_n must be determined based only on the current configurations of the two contacting bodies. Other qualities such as velocity and acceleration cannot be present.

When the energy function is described in a way that satisfies all the conditions outlined in the previous subsections, the whole contact model will follow naturally without any further assumption. As noted earlier, different energy functions will lead to different normal contact models in terms of their contact geometric and force features. The corresponding computational procedures involved will be different and thus may have various numerical efficiencies, so they exhibit different advantages and disadvantages. Nevertheless, these specialised contact models can be viewed as individual members of the same general contact model that is derived from the current energy-conserving contact theory.

When energy dissipation mechanisms, such as plastic deformation or other forms of energy loss, need to be considered in the contact, relevant energy dissipation terms or models can be added to complement the current normal contact model. Dissipative behaviour can also be considered in the current framework by modifying the energy function to include additional terms that depend on internal or history-dependent state variables, or time-dependent quantities, such as velocity and acceleration.

For instance, analogous to the conventional viscous damping model, a viscous damping induced energy due to the relative contact velocity can be added to the energy function. Alternatively, some parameters in the function

can be history-dependent, similar to the Walton–Braun hysteresis model [49] (where different stiffness coefficients are used for loading and unloading, but these two stages are distinguished by the maximum penetration which is a history variable). Furthermore, friction can be taken into account by adding friction and tangential motion related terms into the energy function. As the current work focuses on the elastic energy conservation of the normal contact, energy dissipative contact issues will not be discussed further.

5. Multiple contacts and geometric considerations for simple shapes

5.1. Multiple contacts

A distinct contact region between two bodies is regarded as one contact. The contact between any two convex shapes always results in a single contact. Multiple contacts can only occur when at least one contacting body is concave.

Multiple contacts happen in two different cases. In the first case, an initially single contact region develops into multiple contacts as the contact progresses, but can merge into or restore back to a single region at a later or bounce back stage. These multiple contacts are classified as the *associated* multiple contacts. For this type of multiple contacts at a particular time instance, the contact region between the two bodies has disjointed contact domains. The corresponding energy function will consider the overall contribution from these multiple domains. Thus, the current contact framework can handle the associated multiple contact case without requiring any additional treatment.

The second case concerns with the presence of multiple distinct contact regions that may evolve or disappear during the course of contact but cannot merge into fewer contact regions. For this multiple-contact case, the proposed contact model can be applied to each contact region independently. For the i th contact region, define a contact energy function w_i , and let \mathbf{x}_c^i be the contact point of this contact where the contact moment disappears ($\mathbf{M}_n^i = 0$). The normal contact force \mathbf{F}_n^i is expressed as

$$\mathbf{F}_n^i = -\nabla_{\mathbf{x}} w_i \quad (40)$$

Now define the total energy function w for the whole contact as

$$w = \sum_i w_i \quad (41)$$

Let \mathbf{x} and $\boldsymbol{\theta}$ be the coordinates and orientation angles of the reference point of Body 1. Thus, the total normal contact force, by definition (26), is

$$\mathbf{F}_n = -\nabla_{\mathbf{x}} w = \sum_i (-\nabla_{\mathbf{x}} w_i) = \sum_i \mathbf{F}_n^i \quad (42)$$

which is the resultant force of all the individual normal forces. The overall normal direction can be obtained as $\mathbf{n} = \mathbf{F}_n / |\mathbf{F}_n|$.

Utilising the formula (22) for a new reference point, the total contact moment about the reference point \mathbf{x} of the body is

$$\mathbf{M}_n = -\nabla_{\boldsymbol{\theta}} w = \sum_i [\mathbf{M}_n^i - \mathbf{F}_n^i \times (\mathbf{x} - \mathbf{x}_c^i)] = \sum_i \mathbf{F}_n^i \times (\mathbf{x}_c^i - \mathbf{x}) \quad (43)$$

The combined or overall contact point \mathbf{x}_c can be determined by the formula (29).

The above process shows that treating multiple contacts independently or collectively will yield the identical result. Another feature is that individual energy functions for different contacts can be chosen independently. As will be explored in other papers in this series [42,43], different energy functions or rather which contact geometric features will be adopted as the variables for the energy function, can lead to various types of energy-conserving contact models with distinct computational procedures. It may even be computationally beneficial to select different contact models. For instance, for clumped particles that consist of different geometric entities, distinct models can be employed based on the shapes of the contact pair. In any case, the current contact theory guarantees that the total elastic energy will be conserved if each contact conserves energy. Thus, the proposed general contact model provides both flexibility and stability for any complex particle contact scenario that may be encountered.

Note that the above discussion explains why the widely used clumped sphere contact model will conserve energy if energy-conserving contact models (such as Hertzian and linear spring models) are used for all of the underlying pair-wise sphere contacts.

5.2. Shapes with symmetric properties

Discs and spheres are the simplest and most commonly used geometric entities in DEM simulations. A flat face or its triangular discretisation is often employed to represent a wall or part of it. Slightly more complex non-circular/spherical shapes include ellipse, ellipsoid, cylinder, and some special super-quadratics. These shapes all possess some symmetric properties.

For contact between these simple shapes, the contact overlap, normal direction or contact point may be easily determined, and their evaluations could be separated from the normal force, while its magnitude is specified by a contact interaction law.

For instance, the normal contact direction between a planar face (plane) and any shape must be the normal of the face; the normal contact force between a sphere and any shape must pass through the centre of the sphere. For the cylinder–cylinder contact developed in [20], the normal contact direction for most contact scenarios must pass through both axes of the cylinders.

The current contact theory provides a unified framework that can tackle contacts involving any arbitrarily shaped bodies and any contact type, i.e. convex–convex, convex–concave or concave–concave. For contacts of shapes with certain symmetric properties, the theory can recover certain geometric aspects of the contact features without specifying the energy function. The resulting contact models, complemented by a force interaction law, form a special theory for the shapes concerned. The detail is discussed below.

Suppose that the energy function depends on one contact geometric feature g , for simplicity, but its actual form is not given. Also assume that the shape of Body 1 in contact has certain symmetric properties, while Body 2 is stationary and can have any shape.

By definition, the normal force and moment are

$$\mathbf{F}_n = -\frac{\partial w(g)}{\partial \mathbf{x}} = -\frac{dw}{dg} \cdot \frac{\partial g}{\partial \mathbf{x}}; \quad \mathbf{M}_n = -\frac{\partial w(g)}{\partial \boldsymbol{\theta}} = -\frac{dw}{dg} \cdot \frac{\partial g}{\partial \boldsymbol{\theta}} \quad (44)$$

Now for an arbitrary unit vector \mathbf{s} representing the direction of a translation motion, the corresponding force along the direction is defined as

$$\mathbf{F}_s = -\frac{\partial w(g)}{\partial \mathbf{s}} = -\frac{dw}{dg} \cdot \frac{\partial g}{\partial \mathbf{s}} \quad (45)$$

Since for any vector $\boldsymbol{\tau}$ in the tangential plane $\boldsymbol{\pi}$, the corresponding force must be zero

$$\mathbf{F}_{\boldsymbol{\tau}} = -\frac{dw}{dg} \cdot \frac{\partial g}{\partial \boldsymbol{\tau}} = 0 \quad (46)$$

thus it must have

$$\frac{\partial g}{\partial \boldsymbol{\tau}} = 0 \quad (47)$$

In other words, if the geometric feature remains unchanged when the shape moves along a direction $\boldsymbol{\tau}$, $\boldsymbol{\tau}$ must lie in the tangential plane and the normal direction must be perpendicular to it:

$$\boldsymbol{\tau} \cdot \mathbf{n} = 0$$

For example, consider the contact between a planar face with any other shape. When the face moves horizontally within its plane, the contact region will not change. Thus the contact normal must be parallel to the face normal, as expected.

Now consider a rotation about one axis $\boldsymbol{\alpha}$ with angle α . The corresponding moment about the axis is

$$\mathbf{M}_a = -\frac{\partial w(g)}{\partial \boldsymbol{\alpha}} = -\frac{dw}{dg} \cdot \frac{\partial g}{\partial \alpha} \cdot \boldsymbol{\alpha} \quad (48)$$

If the shape has a rotational symmetry about the axis $\boldsymbol{\alpha}$, the contact region will not change when the shape is rotating about the axis, so $\partial g / \partial \alpha = 0$, and therefore the moment \mathbf{M}_a must be zero. At the same time, assume that the normal force \mathbf{F}_n is applied to the contact point \mathbf{x}_c where $\mathbf{M}_n = 0$. Then, if \mathbf{F}_n is not passing through the rotation axis $\boldsymbol{\alpha}$, it will produce a non-zero moment about the axis: $\mathbf{M} = (\mathbf{p} - \mathbf{x}_c) \times \mathbf{F}_n$, where \mathbf{p} is any point on the axis. Consequently we can conclude that *the normal contact line must pass through the rotation axis of any axi-symmetric shape*, but the location of the intersection is yet to be determined.

For a sphere which has an infinite number of rotational symmetric axes with all passing through the centre, apply the above conclusion to any two different axes. Then it is obvious that the normal force must pass through the centre, as expected.

For two axi-symmetric shapes in contact, the normal force must pass the two axes at the same time if they are not parallel. Furthermore, if a contact point is given, the normal direction \mathbf{n} can be uniquely determined as stated by the following theorem.

Theorem 1 (Contact Normal Theorem for Axi-symmetrical Shapes). *For a given pair of axi-symmetrical shapes, $\mathcal{C}_i : \{\mathbf{n}_i, \mathbf{c}_i\} (i = 1, 2)$ where \mathbf{n}_i are the axes, and \mathbf{c}_i are any two points on the two axes respectively, that are in contact with a contact point at \mathbf{x}_c , and if $\mathbf{n}_1, \mathbf{n}_2$ and \mathbf{x}_c are not co-planar, the corresponding contact normal \mathbf{n} from \mathcal{C}_1 to \mathcal{C}_2 can be uniquely determined by:*

$$\mathbf{n} = (\mathbf{x}_c - \mathbf{c}_1) - \frac{(\mathbf{x}_c - \mathbf{c}_1) \cdot \mathbf{n}_{p_1}}{\mathbf{n}_2 \cdot \mathbf{n}_{p_1}} \mathbf{n}_1 = (\mathbf{c}_2 - \mathbf{p}) - \frac{(\mathbf{c}_2 - \mathbf{p}) \cdot \mathbf{n}_p}{\mathbf{n}_2 \cdot \mathbf{n}_p} \mathbf{n}_2; \quad (49)$$

with $\mathbf{n}_{p_i} = \mathbf{n}_i \times (\mathbf{c}_i - \mathbf{p}) \quad (i = 1, 2)$

The above theorem implies that the same contact normal can be shared by different contact points as long as they are not in co-linear with the contact normal. The following **contact point theorem**, or the converse of the contact normal theorem, confirms this implication.

Theorem 2 (Contact Point Theorem for Axi-symmetrical Shapes). *For a given pair of axi-symmetrical shapes, $\mathcal{C}_i : \{\mathbf{n}_i, \mathbf{c}_i\} (i = 1, 2)$, that are in contact with a known contact normal direction \mathbf{n} , the contact point \mathbf{x}_c satisfies the following condition:*

$$\mathbf{x}_c = \mathbf{c}_1 + t_1 \mathbf{n}_1 + \lambda \mathbf{n} = \mathbf{c}_2 + t_2 \mathbf{n}_2 + \lambda \mathbf{n} \quad (50)$$

where

$$t_i = \frac{\mathbf{n} \cdot [(\mathbf{c}_1 - \mathbf{c}_2) \times \mathbf{n}_i]}{\mathbf{n} \cdot (\mathbf{n}_2 \times \mathbf{n}_1)} \quad (i = 1, 2)$$

and λ is an arbitrary constant, indicating that $\mathbf{x}_c(\lambda)$ forms the normal contact line.

Both theorems are originally proposed in [20], and provide important guidance to enhance both numerical accuracy and stability of contact models for axi-symmetric shapes developed in the traditional approach. However, within the framework of the current energy-conserving contact model, both contact normal and point can be obtained naturally, and automatically satisfy the conditions stated in the above two theorems.

6. Concluding remarks

A unified theoretical framework for developing energy-conserving normal contact models for arbitrarily shaped bodies has been re-established in a more fundamental and rigorous manner in the context of the discrete element method. It is derived based only on the requirement that the contact potential energy must be conserved for an elastic impact of two shapes under any condition. The resulting general energy-conserving contact model states that the normal force as a vector must be the gradient of a contact potential field. When such a contact potential or energy function is specified, a complete normal contact model for a pair of arbitrarily shaped particles, including the contact normal direction, contact point/line and force magnitude, will be automatically determined without introducing any additional assumptions. In this framework, the contact geometry and contact force are indispensably related and are evaluated in a consistent manner. In addition, both single or multiple contact, or any complex contact scenarios can be handled in a unified way, and generally no special treatment is required for multiple contacts. Shapes with symmetric properties that can be used to simplify the contact models have also been discussed.

Due to the importance of the contact energy function in this framework, its fundamental properties have been discussed. These properties will provide some general guidance on how to choose a valid energy function. As mentioned in the introduction, for a given contact energy function, the corresponding normal contact model is viewed as numerical based, and may not be able to accurately capture the true physical behaviour of the contact pair. However, the energy-conserving property of the model ensures that a more robust and stable discrete element modelling will be achieved under complex circumstance.

The developments in this part of the series have laid a solid foundation for establishing specialised energy-conserving normal contacts. Two such models, one based on the contact volume and the other on the overlap, will be presented in later papers of this series [42,43]. The corresponding computational procedures will also be developed and numerical examples covering many contact scenarios will be provided to clearly demonstrate the robustness and applicability of the proposed theory and the models for a wide variety of applications involving arbitrarily shaped particles.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] P.A. Cundall, O.D.L. Strack, A discrete numerical model for granular assemblies, *Geotechnique* 29 (1) (1979) 47–65.
- [2] A. Munjiza, *The Combined Finite-Discrete Element Method*, Wiley & Sons, London, 2004.
- [3] D. Sulsky, S.-J. Zhou, H.L. Schreyer, Application of a particle-in-cell method to solid mechanics, *Comput. Phys. Comm.* 87 (1995) 236–252.
- [4] X. Zhang, *The Material Point Method*, Tsinghua University Press Limited, Beijing, China, 2017.
- [5] H.R. Norouzi, R. Zarghami, R. Sotudeh-Gharebagh, N. Mostoufi, *Coupled CFD-DEM Modeling: Formulation, Implementation and Application to Multiphase Flows*, John Wiley & Sons, 2016.
- [6] L. Jing, O. Stephansson, *Fundamentals of Discrete Element Methods for Rock Engineering: Theory and Applications*, Elsevier, 2007.
- [7] C. O'Sullivan, *Particulate Discrete Element Modelling: A Geomechanics Perspective*, CRC Press, 2011.
- [8] 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.
- [9] S. Ji, L. Lu, *Computational Granular Mechanics and its Engineering Applications*, Springer, ISBN: 978-981-15-3304-4, 2020.
- [10] PFC - Particle Flow Code, Version 6.0, Itasca Consulting Group, Inc. Minneapolis, Itasca, 2018.
- [11] EDEM 2019 User Guide, DEM Solutions Ltd, <https://www.edemsimulation.com/>.
- [12] StarCCM+: <https://www.plm.automation.siemens.com/global/en/products/simcenter/STAR-CCM.html>.
- [13] LIGGGHTS: PUBLIC Documentation, Version 3, DCS Computing GmbH, 2016, <https://www.cfdem.com/media/DEM/docu/Manual.html>.
- [14] Yade Documentation, second ed., 2015, The Yade Project. <http://yade-dem.org/doc/>.
- [15] MatDEM: Fast GPU matrix computation of discrete element method, <http://matdem.com/index.asp?lg=en>.
- [16] J.M. Ting, A robust algorithm for ellipse-based discrete element modelling of granular materials, *Comput. Geotech.* 13 (3) (1992) 175–186.
- [17] T.-T. Ng, Numerical simulations of granular soil using elliptical particles, *Comput. Geotech.* 16 (2) (1994) 153–169.
- [18] X. Lin, T.-T. Ng, Contact detection algorithms for three-dimensional ellipsoids in discrete element modelling, *Int. J. Numer. Anal. Methods Geomech.* 19 (9) (1995) 653–659.
- [19] J.R. Williams, A. Pentland, Superquadrics and model dynamics for discrete elements in interactive design, *Eng. Comput.* 9 (1992) 115–128.
- [20] Y.T. Feng, K. Han, D.R.J. Owen, A generic contact detection framework for cylindrical particles in discrete element modelling, *Comput. Methods Appl. Mech. Engrg.* 315 (2017) 632–651.
- [21] M.A. Hopkins, Discrete element modeling with dilated particles, *Eng. Comput.* 21 (2/3/4) (2004) 422–430.
- [22] L. Liu, S.Y. Ji, A fast detection algorithm based on the envelope function of dilated polyhedron, *Sci. Sin-Phys. Mech. Astron.* 49 (2019) 064601 (in Chinese).
- [23] L. Liu, S.Y. Ji, Bond and fracture model in dilated polyhedral DEM and its application to simulate breakage of brittle materials, *Granul. Matter* 21 (2019) <http://dx.doi.org/10.1007/s10035-019-0896-4>, Article No. 41.
- [24] J. Peters, M. Hopkins, R. Kala, R. Wahl, A poly-ellipsoid particle for non-spherical discrete element method, *Eng. Comput.* 26 (6) (2009) 645–657.
- [25] 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, *Eng. Comput.* 35 (2018) 733–771.
- [26] S. Zhao, J. Zhao, A poly-superellipsoid-based approach on particle morphology for DEM modeling of granular media, *Int. J. Numer. Anal. Methods Geomech.* 43 (2019) 2147–2169.
- [27] Y.T. Feng, D.R.J. Owen, A 2D polygon/polygon contact model: algorithmic aspects, *Eng. Comput.* 21 (2004) 265–277.
- [28] K. Han, Y.T. Feng, D.R.J. Owen, Contact resolution for non-circular discrete objects, *Internat. J. Numer. Methods Engrg.* 66 (3) (2006) 485–501.
- [29] 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, *Int. J. Rock Mech. Min. Sci. Geomech.* 25 (1988) 107–116.
- [30] Y.T. Feng, K. Han, D.R.J. Owen, An energy based polyhedron-to-polyhedron contact model, in: Proceeding of 3rd M.I.T. Conference of Computational Fluid and Solid Mechanics, MIT, USA, 2005, pp. 210–214, June.

- [31] Y.T. Feng, K. Han, D.R.J. Owen, Energy-conserving contact interaction models for arbitrarily shaped discrete elements, *Comput. Methods Appl. Mech. Engrg.* 205–208 (2012) 169–177.
- [32] W. Zhou, G. Ma, X.L. Chang, Y. Duan, Discrete modelling of rockfill materials considering the irregular shaped particles and their crushability, *Eng. Comput.* 32 (4) (2015) 1104–1120.
- [33] R. Kawamoto, E. Andò, G. Viggiani, J. Andrade, Level set discrete element method for three-dimensional computations with triaxial case study, *J. Mech. Phys. Solids* 91 (2016) 1–13.
- [34] Z. Lai, Q. Chen, L. Huang, Fourier series-based discrete element method for computational mechanics of irregular-shaped particles, *Comput. Methods Appl. Mech. Engrg.* 362 (2020) 112873.
- [35] J. Andrade, K. Lim, C. Avila, I. Vlahinić, Granular element method for computational particle mechanics, *Comput. Methods Appl. Mech. Engrg.* 241 (2012) 262–274.
- [36] 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, *Adv. Powder Technol.* 28 (2) (2017) 325–339.
- [37] Z. Zhu, H. Chen, W. Xu, L. Liu, Packing simulation of three-dimensional multi-sized star-shaped particles, *Modelling Simulation Mater. Sci. Eng.* 22 (3) (2014) article id. 035008.
- [38] 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, *Comput. Methods Appl. Mech. Engrg.* 354 (2019) 441–463.
- [39] W. Gao, Y.T. Feng, A coupled 3D discrete elements/isogeometric method for particle/structure interaction problems, *Comput. Part. Mech.* (2019) <http://dx.doi.org/10.1007/s40571-019-00267-8>.
- [40] A. Munjiza, K. Andrews, Penalty function method for combined finite-discrete element systems comprising large number of separate bodies, *Internat. J. Numer. Methods Engrg.* 49 (11) (2000) 1377–1396.
- [41] A. Munjiza, E. Knight, E. Rougier, *Computational Mechanics of Discontinua*, Wiley & Sons, London, 2011.
- [42] Y.T. Feng, Energy-conserving contact interaction models for arbitrarily shaped discrete elements: Contact Volume Based Model and Computational Issues, *Comput. Methods Appl. Mech. Engrg.* (2020) submitted for publication.
- [43] Y.T. Feng, Energy-conserving contact interaction models for arbitrarily shaped discrete elements: Contact Overlap Based Model and Computational Issues, 2020, in preparation.
- [44] P. Wriggers, *Computational Contact Mechanics*, second ed., Springer-Verlag, Berlin, Germany, 2006.
- [45] K.J. Johnson, *Contact Mechanics*, Cambridge University Press, 1985.
- [46] R.D. Mindlin, H. Deresiewicz, Elastic spheres in contact under varying oblique force, *Trans. ASME, J. Appl. Mech.* 20 (1953) 327–344.
- [47] 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, 2017, pp. 29–35.
- [48] J. Marsden, A. Weinstein, *Calculus III*, second ed., Springer, ISBN: 978-0-387-90985-1, 1985.
- [49] O.R. Walton, R.L. Braun, Stress calculations for assemblies of inelastic spheres in uniform shear, *Acta Mech.* 63 (1986) 73–86.