A semi-analytical approach to the elastic loading behaviour of rough surfaces

Xiaogang ZHANG¹, Yali ZHANG^{1,*}, Zhongmin JIN^{1,2,3}

¹ School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

² School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

³ School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, UK

Received: 21 March 2019 / Revised: 10 June 2019 / Accepted: 28 July 2019

© The author(s) 2019.

Abstract: The elastic loading behaviour of rough surfaces is derived based on the physical understanding of the contact phenomena, where the pressure distribution is analytically obtained without any negative values or convergence problems, thus the evolution of the contact behaviour is obtained in a semi-analytical manner. Numerical results obtained by the proposed approach facilitate the understanding of the contact behaviour in the following aspects: 1) the ratio of contact area to load decreases with an increase in real contact area; 2) normal approach-load relationship is approximated by an exponential decay under relatively small loads and a linear decay under relatively large loads; and 3) average gap shows an exponential relationship with load only in moderate load range.

Keywords: surface roughness; real contact area; interfacial stiffness; normal contact stiffness

1 Introduction

Contact of rough surfaces has been a fundamental and challenging problem for many areas. The evolution of the contact behaviour has been considered as critical for understanding the mechanism of many physical phenomena [1] and has profound implications in various engineering applications [2]. For examples, the evolution of real contact area is central to reveal the origin of friction [3], the evolution of deformation and pressure distribution are crucial to understand the mechanism as well as the behaviour of friction [4, 5]. Electrical conduction is derived to be proportional to interfacial stiffness [6, 7]. Similar considerations also apply to heat conduction between rough surfaces, where the interfacial stiffness plays a central role in determining the thermal conduction [8]. Also, the load-displacement relationship is of great importance in determining contact dynamics, as well as the performance and reliability of various engineering apparatus

[9] such as robotic applications and machine tools [10]. Consequently, the accurate understanding of the contact evolution is important for the perspective of revealing the origin and mechanism of friction [11, 12], deriving the electrical and thermal conductance of various technical applications [8], and accurately modelling contact dynamics of machines [10, 11].

Numerous analytical and numerical approaches have been proposed to investigate the contact evolution of rough surfaces [7]. Based on the Hertzian theory, analytical solutions were proposed to study the contact behaviour of rough surfaces, such as the Archard model [12], the Greenwood and Williamson model [13], and the Bush-Gibson-Thomas model [14, 15]. The essence of such approaches is to assume rough surfaces have simplified asperity shapes and height distributions, so the Hertzian solution can be applied to obtain the deformation of asperities, and contact behaviour is therefore derived as a function of surface roughness distribution. However, due to the simplifications on

* Corresponding author: Yali ZHANG, E-mail: zhangyali@swjtu.edu.cn

asperity shapes and height distributions, local asperity interactions are considered in a statistical way, where the interactions of the deformation by different asperities are often neglected [16, 17]. Also, analytical approaches generally cannot give a deterministic distribution of contact pressure and contact spots.

Persson also derived a contact mechanics theory [2]. Its fundamental concept is to solve the contact problem firstly at a coarse scale by neglecting all random roughness, and then include the effects induced by random roughness on mean values or distribution functions by successively adding finer details of the height profiles [18]. The elastic contact of rough surfaces is derived in different length scales by assuming a “transition probability” which specifies the likelihood of the change of pressure as the magnification increases. At low magnification, complete contact happens, and the real contact area is the nominal contact area. As the magnification increases, surface roughness in smaller length scale is detected and the real contact area gets smaller and smaller. However, it is found that Persson’s proof lacks rigour, where the deriving of the diffusion equation assuming full contact and then modelling partial contact by imposing a boundary condition on its solution. Also, comparisons demonstrate that Persson’s solution underestimates the real contact area and therefore overestimates the pressure distribution [19].

With the development of computational resources, numerical approaches were proposed to quantitatively derive the contact behaviour. From a mathematical point view, the contact problem is to solve a system of linear equations under the geometric compliance condition (i.e., the pressure within the contact region is positive and the gap between two surfaces is zero, and the pressure outside contact region is zero and the gap is positive). Since both the pressure distribution and deformation are unknown, a numerical approach generally involves an iteration process to solve the pressure distribution and deformation. A typical widely used approach is the boundary element approach, in combination with the conjugate gradient method (CGM) and Fast Fourier Transform (FFT) speedup strategy [18].

Those approaches have been developed originally to calculate the real contact area and pressure

distribution under a given load. However, when such approaches are adopted to obtain the evolution, it is not effective enough since each step involves an iteration process to get rid of the negative pressure. Also, since numerical approaches are not closed-form solution, where iteration parameters are involved, this wide variety of models not only makes it difficult for experimentalists and engineers to decide which approach to adopt but also throws into question their accuracy and convergence, since predictions from different models do disagree strongly [20].

In the current study, we derive a semi-analytical solution to the contact evolution of rough surfaces based on the physical understanding of the contact process, where the exact evolution of the real contact area, pressure distribution, and deformation of rough surfaces are obtained. This approach was then adopted to investigate the contact evolution of a rigid self-affine surface and an elastic half-space. Contact evolution is characterized in terms of contact area ratio, normal approach and normal stiffness, as well as average gap and interfacial stiffness. Contact behaviour of the elastic self-affine surfaces is further interpreted by comparing the current and existing results.

2 The semi-analytical solution to the elastic loading behaviour

2.1 Physical understanding of the contact process

Inspired by the physical understanding of the contact phenomena, a semi-analytical solution is derived to solve the elastic loading of rough surfaces. Consider the contact of a rough rigid body and an elastic half-space, as shown in Fig. 1(a), where the surface topography of the rigid body is illustrated in Fig. 1(b). When a normal load W is applied on the rigid body, it approaches and penetrates into the elastic half-space, forming a contact region. In the contact region, pressure builds up due to the geometric overlap. The real contact area and the pressure keep increasing until an equilibrium condition is reached, where the normal load is balanced by the pressure integration in the contact region. Consequently, the dry contact problem can be considered as a static equilibrium where a dynamic interaction process takes place before the

Fig. 1 The contact of the rigid rough body and an elastic half-space. (a) Macroscale contact configuration, (b) surface topography of the rigid body, and (c) asperity-scale deformation relationship.

static equilibrium is reached. At each instant of the dynamic interaction, the geometric overlap is determined by the original surface topography. The pressure distribution within the contact region is caused and uniquely determined by the geometric overlap. The deformation of the full domain is determined by the pressure distribution, and the gap of between the deformed elastic half-space and the rigid body determines which node is going into contact next.

2.2 Analytical relationship between deformation and pressure

The contact of a local asperity is analysed to demonstrate the contact relationship, as illustrated in Fig. 1(c). The black dash lines denote their original positions of the rigid body and elastic half-space before the load is applied. At an instant of the dynamic interaction, the rigid asperity moves downwards with a depth of penetration of h_0 , forming a real contact area. The following equations derive the analytical relationships between pressure and deformation. Within the contact area, the deformation v is the sum of a rigid body displacement constant h_0 and the original profile g_0 , as given by Eq. (1).

$$v = g_0 + h_0 \quad (1)$$

Within the contact region, the pressure distribution is caused and uniquely determined by the geometric overlap g_0 . The traditional matrix inverse approach builds up a linear relationship between deformation and pressure, as given by Eq. (2), where C is the deformation matrix.

$$v = C \cdot p \quad (2)$$

The element $C_{ij,kl}$ of the deformation matrix, denoting the deformation coefficient of the node (i, j) by the pressure node (k, l) , is derived based on the Boussinesq solution [21] of a point load on an elastic half-space, as given by Eq. (3), where E^* is the equivalent Young's modulus of the elastic half-space, and Δ is the mesh node size.

$$\frac{\pi E^*}{\Delta} C_{ij,kl} = (|i-k|-0.5) \ln \frac{\sqrt{(|i-k|-0.5)^2 + (|j-l|-0.5)^2} + (|j-l|-0.5)}{\sqrt{(|i-k|-0.5)^2 + (|j-l|+0.5)^2} + (|j-l|+0.5)} + (|j-l|-0.5) \ln \frac{\sqrt{(|i-k|-0.5)^2 + (|j-l|-0.5)^2} + (|i-k|-0.5)}{\sqrt{(|i-k|+0.5)^2 + (|j-l|-0.5)^2} + (|i-k|+0.5)}$$

$$\begin{aligned}
& + (|i-k|+0.5) \ln \frac{\sqrt{(|i-k|+0.5)^2 + (|j-l|+0.5)^2} + (|j-l|+0.5)}{\sqrt{(|i-k|+0.5)^2 + (|j-l|-0.5)^2} + (|j-l|-0.5)} \\
& + (|j-l|+0.5) \ln \frac{\sqrt{(|i-k|+0.5)^2 + (|j-l|+0.5)^2} + (|i-k|+0.5)}{\sqrt{(|i-k|-0.5)^2 + (|j-l|+0.5)^2} + (|i-k|-0.5)}
\end{aligned} \quad (3)$$

Existing contact mechanics approaches often apply iterative processes to get rid of negative pressures induced by the unknown rigid body displacement h_0 . In the current approach, the concept of “boundary node” is introduced, so the pressure distribution is able to be analytically derived without any negative pressure problem. For both the gap and pressure are zero on a boundary node, h_0 can be derived by Eq. (4), where the symbol (N) indexes the element of a vector on the boundary node, y_1 and y_2 are vectors derived by $y_1 = C^{-1}g_0$ and $y_2 = C^{-1}e$, and e is a vector with all entries 1.

$$h_0 = -\frac{y_1(N)}{y_2(N)} \quad (4)$$

Consequently, pressure distribution is obtained as Eq. (5).

$$p = y_1 + h_0 y_2 \quad (5)$$

By applying Eqs. (1)–(5), the pressure distribution within the contact region is analytically obtained. Note, different from existing approaches, here Eqs. (1)–(5) are only formed within the contact region. The deformation in the full domain can also be obtained by applying Eq. (2).

2.3 The semi-analytical solution strategy

A simplified contact configuration, as shown in Fig. 2,

Fig. 2 The contact process of a rigid rough surface on an elastic half-space. (a) Before the load is applied, (b) two nodes in contact, and (c) three nodes in contact.

is adopted to illustrate the numerical implementation strategy, where the rigid body (black solid line) is in contact with the elastic half-space (blue solid line). The contact process is discretised into finite steps. Initially, the node N_1 of the rigid body surface is set as a boundary node, where both the pressure and deformation are zeros. In the next step, the rigid body moves downwards and N_1 penetrates into the elastic half space, as shown in Fig. 2(b). The depth of penetration is particularly chosen to just make N_2 to be a boundary node, thus, the geometric overlap g_0 is analytically determined. By forming the deformation matrix C in the contact region, y_1 and y_2 are obtained and h_0 is determined. Finally, the pressure distribution is uniquely determined by Eq. (5). This step is the iteration core and is repeated by setting another node with the smallest gap into the contact domain. Gradually, more and more nodes are getting into contact, and the pressure gets larger and larger. Finally, an equilibrium between the pressure and the applied load is reached.

The evolution of the contact behaviour is derived by adding nodes one by one into the contact domain in each step. At the n^{th} step, there are n nodes into contact, therefore the deformation matrix C has a dimension of $n \times n$, and the pressure distribution p , as well as vectors y_1 and y_2 have a dimension of n . In each step, the linear equations formulated by the deformation matrix needs to be solved twice for y_1 and y_2 , which is quite expensive for the prospective of computation efficiency and storage cost. The following recursive approach is derived to reduce the computational cost. In the $(n+1)^{\text{th}}$ step, y_1^{n+1} and y_2^{n+1} need to be calculated where the solution of the n^{th} iteration is obtained as $C^n y_1^n = g_0^n$ and $C^n y_2^n = e^n$. The superscript n denotes the n^{th} step, y_1^{n+1} is given by Eq. (6).

$$C^{n+1}y_1^{n+1} = g_0^{n+1} \quad (6)$$

By assuming $y_1^{n+1} = \begin{bmatrix} z_1^n \\ z_1 \end{bmatrix}$, Eq. (6) is rewritten as

Eq. (7), where z_1 is the last element of y_1^{n+1} , c_0 is the last element of the deformation matrix, and c^n is a vector where its elements are obtained from Eq. (3).

$$\begin{bmatrix} C^n & c^n \\ (c^n)^T & c_0 \end{bmatrix} \begin{bmatrix} z_1^n \\ z_1 \end{bmatrix} = \begin{bmatrix} g_0^n \\ g_{n+1} \end{bmatrix} \quad (7)$$

Equation (7) is rewritten as Eqs. (8) and (9).

$$C^n z_1^n + z_1 c^n = g_0^n \quad (8)$$

$$(c^n)^T z_1^n + c_0 z_1 = g_{n+1} \quad (9)$$

Since $C^n z_1^n = y_1^n$, Eq. (8) is rewritten as Eq. (10), where x^n is given by Eq. (11).

$$z_1^n = y_1^n - z_1 x^n \quad (10)$$

$$x^n = (C^n)^{-1} c^n \quad (11)$$

Substitute Eq. (10) into Eq. (9), z_1 is obtained to be $z_1 = \frac{g_{n+1} - (c^n)^T y_1^n}{c_0 - (c^n)^T x^n}$. Finally, y_1^{n+1} is obtained as Eq. (12).

$$y_1^{n+1} = \begin{bmatrix} y_1^n - z_1 x^n \\ z_1 \end{bmatrix} \quad (12)$$

Similarly, y_2^{n+1} is obtained as given by Eq. (13), where

$$z_2 = \frac{e_{n+1} - (c^n)^T y_2^n}{c_0 - (c^n)^T x^n}.$$

$$y_2^{n+1} = \begin{bmatrix} y_2^n - z_2 x^n \\ z_2 \end{bmatrix} \quad (13)$$

Consequently, h_0 is easily obtained by substitute y_1 and y_2 into Eq. (4), and pressure distribution is calculated from Eq. (5). Thus, the computation load is reduced by half due to the reduction of the frequency of solving the linear equations during an iteration.

As can be seen, Eq. (11) now consumes most of the computational source. We may notice that C^n is a symmetric positive-definite matrix. Therefore, the Cholesky factorization can be applied to further reduce computation and memory cost, where the deformation

matrix C^n is rewritten as manipulation of a lower triangle matrix L^n and an upper triangle matrix $(L^n)^T$, and Eq. (11) is rewritten as

$$L^n (L^n)^T x^n = c^n \quad (14)$$

In the $(n+1)^{\text{th}}$ step, known that $C^n = L^n (L^n)^T$, the $(n+1)^{\text{th}}$ Cholesky factorization is $C^{n+1} = L^{n+1} (L^{n+1})^T$, rewritten as Eq. (15), where $[(l^n)^T l_0]$ stands for the last row of L^{n+1} , and l_0 is the last element of the last row.

$$\begin{bmatrix} C^n & c^n \\ (c^n)^T & c_0 \end{bmatrix} = \begin{bmatrix} L^n & 0 \\ (l^n)^T & l_0 \end{bmatrix} \begin{bmatrix} (L^n)^T & l^n \\ 0 & l_0 \end{bmatrix} \quad (15)$$

From Eq. (15), we have

$$L^n l^n = c^n \quad (16)$$

Substitute Eq. (16) into Eq. (14), so Eq. (14) is rewritten as

$$(L^n)^T x^n = l^n \quad (17)$$

Since L^n is a lower triangle, solutions of Eqs. (16) and (17) are written in closed-form, as given by Eqs. (18) and (19).

$$l_m^n = \frac{c_m - \sum_{i=1}^{m-1} L_{m,i}^n l_i^n}{L_{m,m}^n} \quad (18)$$

$$x_m^n = \frac{c_m - \sum_{i=1}^{m-1} L_{i,m}^n x_i^n}{L_{m,m}^n} \quad (19)$$

The calculation of the deformation in the full domain is also time-consuming. Traditionally, the Multi-level Multi-integration (MLMI) method and FFT are used to speed up the calculation of deformation. In the current approach, FFT is adopted to speed up the deformation calculation considering its efficiency [22]. Finally, the solution process is summarised in the following procedure:

- (1) Input loading parameters;
- (2) Obtain initial pressure distribution (two nodes in contact);
- (3) Calculate the deformation outside the contact region by FFT;
- (4) Obtain the gap, determine the next node going to contact;

(5) Obtain the last column of the deformation matrix $\begin{bmatrix} c^n \\ c_0 \end{bmatrix}$;

(6) Calculate l^n by $l_m^n = \frac{c_m - \sum_{i=1}^{m-1} L_{m,i}^n l_i^n}{L_{m,m}^n}$;

(7) Calculate x^n by $x_m^n = \frac{c_m - \sum_{i=1}^{m-1} L_{m,i}^n x_i^n}{L_{m,m}^n}$;

(8) Update L^{n+1} by $L^{n+1} = \begin{bmatrix} L^n & 0 \\ (l^n)^T & \sqrt{(x^n)^T x^n - c_0} \end{bmatrix}$;

(9) Calculate y_1^{n+1} by $y_1^{n+1} = \begin{bmatrix} y_1^n - z_1 x^n \\ z_1 \end{bmatrix}$;

(10) Calculate y_2^{n+1} by $y_2^{n+1} = \begin{bmatrix} y_2^n - z_2 x^n \\ z_2 \end{bmatrix}$;

(11) Calculate h_0 by $h_0 = -\frac{(C^{-1}g_0)_N}{(C^{-1}e)_N}$;

(12) Calculate pressure distribution p by $p = y_1 + h_0 y_2$;

(13) Check if $\iint p \geq W$, or go back to (3).

The above procedure derives the evolution of contact behaviour in a semi-analytical manner. The Cholesky factorization saves half of the storage since only the lower triangle needs to be stored, and the recursion approach saves half of the computation, resulting in a computation load of $O(n_0^2)$ in each iteration. Since the computation load increases with the real contact area, and the major factor determining the computation load is the Cholesky factorization approach solving linear equations (steps (6) and (8)), so the computation load is acceptable under small contact area condition, regardless of the initial configuration. Thus, the computation limitation of the current approach is not the real mesh size of the surface, but the total nodes in the contact area. When the total nodes get large and reach the computation or storage limitation, we may apply the conjugate gradient approach combined with FFT for matrix manipulation to solve Eq. (11) [23]. The current implementation was run on a personal computer, which has an Intel® Core™ i7-7700 CPU and 8 Gb memory. For an elastic contact problem of moderate mesh size as 256×256 , with a contact area of $\pi/16$ percent, the computation time required is about 20 min. Since the real contact area for a real contact is generally in the magnitude of 1%–10%, so

the current approach is practical to run on a personal computer of mesh size from about $(128 \times 128 - 1024 \times 1024)$. Any mesh size larger than 1024×1024 is recommended to use clusters.

The current approach derives the analytical relationship between pressure and deformation, so meshing and calculations are only conducted on the surface of the contact bodies. In contrast, a typical finite element approach calculates the variables based on the stiffness matrix (which is derived on the whole-body domain based on variation principle), so meshing and calculations are conducted on the whole contact body instead of on the surface boundary. Thus, the current approach is much more effective than the finite element approach in terms of the computational load. In addition, by introducing boundary nodes, the evolution of the contact behaviour is able to be semi-analytical derived with a significant improvement in accuracy. Nevertheless, finite element approaches have very good versatility, and can be easily applied to different configurations with materials, structures, and geometries, etc.

The CGM+FFT method has been proven to be the most effective for the elastic contact of rough surfaces [18]. It obtains real contact area and pressure distribution under a given load. In contrast, the current approach derives the evolution of the contact area and pressure distribution based on the analytical relationship, which provides an effective method to accurately calculate the contact stiffness (or interfacial stiffness) for some contact dynamics and electric/thermal contact applications. In addition, although the CGM+FFT approach demonstrates the highest convergence efficiency, its convergence efficiency drops significantly under some extreme conditions such as infinitesimal contact. On the contrary, the current approach produces robust results from infinitesimal to full contact, and it is especially effective for infinitesimal contact. Consequently, the current approach is an important complementary of the existing contact mechanics approaches.

3 Validation of the semi-analytical approach

To validate the current approach, numerical predictions of different geometries were compared with analytical

solutions under different loading conditions. The contact of a rigid cylinder and an elastic half-space, as well as the contact of a rigid sinusoidal surface and an elastic half-space, were considered.

The rigid cylinder has a radius R and a length L . The elastic half space has an equivalent Young's modulus E^* . A load F is applied on the cylinder, forming a contact area of $2a \times L$. Based on the Hertzian theory, the half contact width a is given by $a = \sqrt{\frac{F}{L} \frac{4R}{\pi E^*}}$, and the pressure distribution is obtained as Eq. (20), where p_0 is the Hertzian pressure and $p_0 = \sqrt{\frac{F}{L} \frac{E^*}{\pi R}}$.

$$p(x) = p_0 \sqrt{1 - \left(\frac{x}{a}\right)^2} \quad (20)$$

For comparison, the coordinate x is nondimensionalized by dividing the contact half width, so $\bar{x} = x/a$; and the pressure distribution is nondimensionalized by dividing the Hertzian pressure p_0 , so the nondimensional pressure distribution is obtained $\bar{p} = \sqrt{1 - \bar{x}^2}$.

The rigid sinusoidal surface profile is given by $s(x) = \Delta \cos\left(\frac{2\pi x}{\lambda}\right)$, where Δ is the amplitude, and λ specifies the wavelength. The pressure distribution given by Westergaard [21] under a contact width of $2a$ is given by Eq. (21).

$$p(x) = 2p_0 \frac{\cos\left(\frac{\pi x}{\lambda}\right)}{\sin^2\left(\frac{\pi a}{\lambda}\right)} \sqrt{\sin^2\left(\frac{\pi a}{\lambda}\right) - \sin^2\left(\frac{\pi x}{\lambda}\right)} \quad (21)$$

where p_0 is the average pressure and $p_0 = \frac{\pi E^* \Delta}{\lambda}$. The coordinate x is nondimensionalized as $\bar{x} = 2x/\lambda$. The pressure distribution is nondimensionalized by dividing $p_0/\sin^2\left(\frac{\pi a}{\lambda}\right)$. Finally, the dimensionless pressure distribution is obtained by Eq. (22).

$$\bar{p} = 2\cos\left(\frac{\pi}{2}\bar{x}\right) \sqrt{\sin^2\left(\frac{\pi a}{\lambda}\right) - \sin^2\left(\frac{\pi}{2}\bar{x}\right)} \quad (22)$$

Here we obtained the analytical pressure distribution under three different conditions including full contact ($2a = \lambda$), half contact ($4a = \lambda$), and quarter contact ($8a = \lambda$). Numerical results under the same loading condition were obtained and compared in Fig. 3. Numerical results were obtained under a mesh configuration of $\bar{x} \in (-1, 1)$ and $\Delta\bar{x} = 2/N = 0.0124$. As can be seen, the current numerical results demonstrate good agreement with analytical predictions. Note here, the Westergaard solution derives pressure distribution of sinusoidal surface defined in $\bar{x} \in (-\infty, +\infty)$ with a period of 2, while the results presented here is just in one period where $\bar{x} \in (-1, 1)$. Thus, the solution domain was chosen to be $(-5, 5)$ in the numerical calculation to minimise the discrepancy in pressure caused by nearby waves.

4 Results and discussion

The proposed approach was employed to study the elastic contact of a rough rigid body with an elastic half-space. The equivalent Young's modulus E^* is given by $E^* = E/(1-\nu^2)$, where E is the Young's modulus of and ν is the Poisson's ratio of the elastic half-space. Dimensionless parameters were adopted in the current calculations. Thus, the current results can be easily applied to different conditions by using different units. The nominal contact area (size of the rigid rough body) is assumed to be $A_0 = 10^{-2} \times 10^{-2}$ and has a root-mean-square (RMS) surface roughness of $Sq = 50 \times 10^{-6}$. The normal load is applied up to $W = 0.1 A_0 E^*$ to cover a wide loading range. The numerical approach by Persson

Fig. 3 Validation of the current numerical results with existing analytical results. A: analytical results; N: numerical results; F: full contact; H: half contact; Q: quarter contact.

[24], which generates a self-affine surface similar to experimentally observed topographies, was adopted in the current study. Based on the convolution theorem, the power spectral density can be written as the Fourier transform of the surface geometry. Thus, randomly rough surfaces with any given power spectrum $C(q)$ can be generated using $h(x) = \sum_q B(q)e^{i[q \cdot x + \phi(q)]}$, where q is the roughness wave vector, $\phi(q)$ is independent random variable uniformly distributed in the interval $[0, 2\pi]$, and $B(q) = \frac{2\pi}{\sqrt{A}} [C(q)]^{\frac{1}{2}}$. The power spectrum density with a prescribed Hurst exponent H and cut-offs in the surface spectrum is given by Eq. (23).

$$C(q) = \begin{cases} (2\pi / \lambda_r)^{-2(1+H)}, & 2\pi / \lambda_r > q \\ q^{-2(1+H)}, & 2\pi / \lambda_r > q > 2\pi / \lambda_s \\ 0, & q > 2\pi / \lambda_s \end{cases} \quad (23)$$

The self-affine surface topography, with a mesh size of 512×512 , was generated with a Hurst exponent of 0.8, as presented in Fig. 1(b). The cut-offs in the power spectrum are applied by setting the roll-off wavelength of $\lambda_r = 2.5 \times 10^{-3}$ and a short cut-off wavelength of $\lambda_s = 7.8125 \times 10^{-5}$, where λ_r is a quarter of the width (or length) of the rigid body and λ_s is four times the mesh grid. The RMS slope $\sqrt{\langle |\nabla h|^2 \rangle}$ is obtained to be 0.112. Detailed parameters are presented in Table 1. The evolution of the contact behaviour is investigated and discussed in three aspects: 1) the evolution of real contact area; 2) normal approach and normal contact stiffness; and 3) average gap and interfacial stiffness.

Table 1 Dimensionless contact parameters of a rigid body with an elastic half-space.

Parameter	Value
Equivalent Young's modulus $E^* = E / (1 - \nu^2)$	
Nominal contact area A_0	$10^{-2} \times 10^{-2}$
Load W	$0 \sim 0.1 A_0 E^*$
Length of the rigid body l_1	10^{-2}
Width of the rigid body l_2	10^{-2}
RMS surface roughness of the rigid body Sq	50×10^{-6}
Hurst exponent of the rigid body H	0.8
Roll-off wavelength λ_r	2.5×10^{-3}
Short cut-off wavelength λ_s	7.8125×10^{-5}

4.1 Evolution of real contact area

The evolution and the contact morphologies under different ratios of the contact area are shown in Fig. 4. As can be seen, most of the contact patches are small and sparsely distributed within the nominal contact area, while the majority of the contact points belong to large patches.

Based on existing theoretical and numerical investigations, it is generally accepted that the real contact area demonstrates a linear relationship with the applied load under small loads, and this linear relationship explains Amonton's law of friction. The frictional force is linearly proportional to the real contact area, and the real contact area is proportional to load, thus the frictional force is proportional to load.

To further characterize the effect of load on the real contact area, the real contact area as a function of the applied load was produced and is presented in Fig. 5. To make the current result easier to compare with existing results, the real contact area is nondimensionalized by the nominal contact area A_0 , and the load is nondimensionalized by the product of the nominal contact area and Young's modulus. The current results demonstrate that a linear relationship is a very good approximation for a real contact area of less than 30%, with a Pearson correlation coefficient of $r = 0.999$. For the real contact area larger than 30%, the real contact area to load slope decreases with an increase in load.

To further characterize the linearity of real contact area to load, the contact area ratio k , defined as $k \equiv \left(AE^* \sqrt{\langle |\nabla h|^2 \rangle} \right) / W$, was obtained and compared with existing results, as presented in Fig. 6. Bush et al. [15] obtained the contact area ratio to be $\sqrt{2\pi}$ by their theoretical approach, Persson theory predicted a different k as $\sqrt{8/\pi}$ by his contact mechanics theory [2], and Hyun et al. obtained $k = 2$ by finite element analysis [5]. In the current calculations, k doesn't keep a constant but demonstrates a decreasing trend with an increase in the real contact area, indicating an increasing trend in the average pressure. The different assumptions maybe the main reason of the discrepancies of the above mentioned predictions. In the work by Bush et al. [15], the cap of every single asperity is replaced by a paraboloid with the same height and

Fig. 4 Evolution of real contact area and pressure distribution.

Fig. 5 Real contact area as a function of load.

Fig. 6 Contact area ratio as a function of the real contact area.

principal curvatures as the summit of the asperity. Persson's contact mechanics assumes full contact to derive the diffusion equation and then to model partial contact by imposing a boundary condition on its solution [2]. In the work by Hyun et al. [5], each

asperity is represented by a single node. Based on the current results, the contact area ratio of $k = \sqrt{2\pi}$ by Bush et al. better describes a smaller contact area, while the contact area ratio of $k = \sqrt{8/\pi}$ by Persson better describes a larger contact area. Consequently, a contact area ratio of $k = 2$ is a proper approximation (error less than 10%) for a contact area between 1% and 15%.

4.2 Normal approach and normal stiffness

Normal approach u specifies the displacement of the rigid body under load W . The normal contact stiffness K_n is defined as the negative derivative of load with respect to the normal approach, $K_n = -dW/du$. Normal contact stiffness has an immediate impact on dynamics characteristics and precision of contact parts. In the current work, the normal approach and normal contact stiffness as functions of load were obtained and are presented in Fig. 7, where the normal approach is nondimensionalized by dividing the RMS surface roughness Sq . As can be seen, the normal approach demonstrates a non-linear relationship with applied load, and can be characterized into two regimes. Under a very small load, the normal approach demonstrates an exponential decay with an increase in load, where exponential fitting parameters correspond to $y_0 = -2.56$, $A = 2.26$, and $t = 0.0021$. While under a relatively

large load, the approach-load relationship is approximated by a linear relationship with a slope of -8.93 and an intercept of -0.11 . Obviously, the normal stiffness also demonstrates two regimes, under a relatively small load, the normal contact stiffness grows exponentially with an increase in load, and it quickly tends to be saturated and gradually approaches to 5.8×10^{-3} . Actually, such behaviour is related to the relationship between the real contact area and load, since only the contact patches contribute to the normal stiffness. Under a relatively small load, an increase in load causes a linear increase in the contact area, results in a significant increase in normal contact stiffness. Under relatively large load, however, an increase in load causes in slower increasing rate in the real contact area, resulting in a slower increasing rate in the normal stiffness. Finally, the saturated normal contact stiffness is explained by the saturated real contact area of $A/A_0 = 1$.

4.3 Average gap and interfacial stiffness

Average gap $\bar{g}$ and interfacial stiffness K_i are also important statistic parameters which imply properties of sealing, as well as electrical and thermal conduction. The average gap $\bar{g}$ is defined as the mean separation between the contact solids, and the interfacial stiffness K_i is defined as the negative derivative of load with respect to the average gap, $K_i = -dW/d\bar{g}$. In the current study, the average gap and interfacial stiffness were investigated as functions of the applied load and are presented in Fig. 8. The average gap generally decreases with an increase in load, while the interfacial stiffness increases with an increase load. Existing studies indicate that the average gap (or interfacial stiffness) with load can be approximated by an exponential relationship. However, the current results demonstrate that the average gap fits linearly with the logarithm of dimensionless load only under moderate load, i.e. from 4×10^{-3} to 2×10^{-2} , indicating an exponential decay with an increase in load. Actually, interfacial stiffness physically means how difficult to squeeze the contacting surfaces overlap. Obviously, it is easier under small load since the real contact area is small, and it's getting more and more difficult with an increase in the real contact area, which explains why the contact stiffness demonstrates an increasing growth rate with an increase in load.

Fig. 7 Normal approach as a function of the applied load.

Fig. 8 Average gap as a function of load.

5 Conclusions

The current study derives a new solution to the elastic contact of rough surfaces based on the physical understanding of the contact phenomena. By introducing the boundary node concept, the pressure distribution is analytically obtained without any negative values and convergence problems. The evolution of the contact behaviour including the real contact area, contact spots distribution, and pressure distribution is obtained in a semi-analytical manner. Consequently, numerical results are obtained with the highest accuracy under a given mesh condition. The current approach is an important supplementary of the existing contact mechanics approaches. The contact behaviour of a rigid self-affine surface and an elastic half-space was investigated, and the following findings are obtained.

(1) The linear relationship between the real contact area and applied load is only valid under a small contact area, the contact area ratio continually decreases from 2.5 to 1.0 with an increase in the real contact area. Generally, a contact area ratio of 2 is a

reasonable approximation for the real contact area between 1% and 15%.

(2) The normal approach-load relationship demonstrates a non-linear relationship with applied load, and it is approximated by exponential decay under relatively small load and linear decay under relatively large loads. The normal contact stiffness grows exponentially with an increase in load, and quickly tends to be saturated and gradually approaches to 5.8×10^{-3} in the current calculations.

(3) The average gap only demonstrates an exponential relationship with the applied load under a moderate load range (from 4×10^{-3} to 2×10^{-2}), and the contact stiffness demonstrates an increasing growth rate with an increase in load.

The current approach is derived based on the elastic half-space assumption. Its application is limited to the contact where the contact region is much smaller than the leading dimensions of one or both bodies. Otherwise finite-element method or the boundary element method needs to be used. Also, the current approach investigates the loading behaviour rough surfaces, where the materials are assumed to be elastic and without surface friction. For some materials, however, plastic deformation happens very often at the very tip of asperity and has a significant influence on the pressure distribution and real contact area. In addition, surface friction may also have a significant influence on the pressure distribution and contact behaviour. Thus, future work is expected to extend the current approach by considering surface friction and plastic deformation.

Acknowledgements

The authors acknowledge the financial supports by the National Key Research and Development Program of China (2016YFF0204305), the National Natural Science Foundation of China (51775460 and 51905456), and the China Postdoctoral Science Foundation (2019M653836XB).

Open Access: This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s)

and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- [1] Yastrebov V A, Anciaux G, Molinari J F. Contact between representative rough surfaces. *Phys Rev E* **86**(3): 035601 (2012)
- [2] Persson B N J. Elastoplastic contact between randomly rough surfaces. *Phys Rev Lett* **87**(11): 116101 (2001)
- [3] Ben-David O, Rubinstein S M, Fineberg J. Slip-stick and the evolution of frictional strength. *Nature* **463**(7277): 76–79 (2010)
- [4] Ovcharenko A, Halperin G, Etsion I, Varenberg M. A novel test rig for in situ and real time optical measurement of the contact area evolution during pre-sliding of a spherical contact. *Tribol Lett* **23**(1): 55–63 (2006)
- [5] Hyun S, Pei L, Molinari J F, Robbins M O. Finite-element analysis of contact between elastic self-affine surfaces. *Phys Rev E* **70**(2): 026117 (2004)
- [6] Buczkowski R, Kleiber M, Starzynski G. Normal contact stiffness of fractal rough surfaces. *Arch Mech* **66**(6): 411–428 (2014)
- [7] Pohrt R, Popov V L. Normal contact stiffness of elastic solids with fractal rough surfaces. *Phys Rev Lett* **86**(10): 104301 (2012)
- [8] Barber J R. Bounds on the electrical resistance between contacting elastic rough bodies. *Proc Roy Soc Lond Ser A: Math, Phys Eng Sci* **459**(2029): 53–66 (2003)
- [9] Pan W J, Li X P, Wang L L, Guo N, Mu J X. A normal contact stiffness fractal prediction model of dry-friction rough surface and experimental verification. *Eur J Mech - A/Solids* **66**: 94–102 (2017)
- [10] Shi X, Polycarpou A A. Measurement and modeling of normal contact stiffness and contact damping at the meso scale. *J Vib Acoust* **127**(1): 52–60 (2005)
- [11] Wang R Q, Zhu L D, Zhu C X. Research on fractal model of normal contact stiffness for mechanical joint considering asperity interaction. *Int J Mech Sci* **134**: 357–369 (2017)

- [12] Archard J F. Elastic deformation and the laws of friction. *Proc Roy Soc Lond Ser A: Math, Phys Sci* **243**(1233): 190–205 (1957)
- [13] Greenwood J A, Williamson J B P. Contact of nominally flat surfaces. *Proc Roy Soc Lond Ser A: Math, Phys Sci* **295**(1442): 300–319 (1966)
- [14] Bush A W, Gibson R D, Keogh G P. The limit of elastic deformation in the contact of rough surfaces. *Mech Res Commun* **3**(3): 169–174 (1976)
- [15] Bush A W, Gibson R D, Thomas T R. The elastic contact of a rough surface. *Wear* **35**(1): 87–111 (1975)
- [16] Tian X F, Bhushan B. A numerical three-dimensional model for the contact of rough surfaces by variational principle. *J Tribol* **118**(1): 33–42 (1996)
- [17] Greenwood J A, Putignano C, Ciavarella M. A Greenwood & Williamson theory for line contact. *Wear* **270**(3–4): 332–334 (2011)
- [18] Müser M H, Dapp W B, Bugnicourt R, Sainsot P, Lesaffre N, Lubrecht T A, Persson B N J, Harris K, Bennett A, Schulze K, et al. Meeting the contact-mechanics challenge. *Tribol Lett* **65**(4): 118 (2017)
- [19] Manners W, Greenwood J A. Some observations on Persson's diffusion theory of elastic contact. *Wear* **261**(5–6): 600–610 (2006)
- [20] Carpick R W. The contact sport of rough surfaces. *Science* **359**(6371): 38 (2018)
- [21] Johnson K L. *Contact Mechanics*. Cambridge (UK): Cambridge University Press, 1987.
- [22] Polonsky I A, Keer L M. A fast and accurate method for numerical analysis of elastic layered contacts. *J Tribol* **122**(1): 30–35 (2000)
- [23] Polonsky I A, Keer L M. A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. *Wear* **231**(2): 206–219 (1999)
- [24] Persson B N J. Contact mechanics for randomly rough surfaces. *Surf Sci Rep* **61**(4): 201–227 (2006)

Xiaogang ZHANG. He received his Ph.D. degree in mechanical in 2017 from The University of New South Wales Australia. He joined

the Tribology Research Institute at Southwest Jiaotong University as an assistant researcher in 2018. His current research interests focus on the fundamentals of tribology and biotribology.

Yali ZHANG. She received her Ph.D. degree in mechanical engineering in 2013 from Xi'an Jiaotong University, China. She joined the School of

Mechanical Engineering at Southwest Jiaotong University as a lecturer in 2015. Her current research interests focus on the biotribology of artificial joints.

Zhongmin JIN. He received his Ph.D. degree in mechanical engineering in 1984 from University of Leeds, UK. He joined the School of Mechanical Engineering at Southwest Jiaotong University (SWJTU) from 2015. His

current position is a professor and the dean of the SWJTU-Leeds Joint School. His current research interests include biotribology, biomechanics and medical devices, artificial joint design and manufacturing, and tissue engineering.