The effects of roughness on the area of contact and on the elastostatic friction

FEM simulation of micro-scale rough contact and real world applications

Doctoral Dissertation submitted to the

Faculty of Informatics of the *Università della Svizzera Italiana*

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

presented by

Alessandro Pietro Rigazzi

under the supervision of

Prof. Dr. Rolf Krause

October 2014

Dissertation Committee

Prof. Dr. Illia Horenko	Università della Svizzera Italiana, Switzerland
Prof. Dr. Igor Pivkin	Università della Svizzera Italiana, Switzerland
Prof. Dr. Marco Paggi	Institute for Advanced Studies Lucca, Italy
Prof. Dr. Friedemann Schuricht	Technische Universität Dresden, Germany

Dissertation accepted on 22 October 2014

Prof. Dr. Rolf Krause
Research Advisor
Università della Svizzera Italiana, Switzerland

Prof. Dr. Stephan Wolf
PhD Program Director

I certify that except where due acknowledgement has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; and the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

Alessandro Pietro Rigazzi
Lugano, 22 October 2014

To my late father, for supporting me
in learning numbers before letters.

If numbers aren't beautiful,
I don't know what is.

Paul Erdős

Abstract

Roughness is everywhere. Every object, every surface we touch or look at, is rough. Even when it looks smooth and flat, if analyzed at the proper length scale, it will reveal roughness. Thus, macro-scale, micro-scale, and even nano-scale roughness exist. What is even more fascinating, is that most of the rough structures, which can be observed at a given length scale, repeat themselves at smaller length scales, as in a fractal. The first implication of the rough nature of surfaces is that what we perceive as a full, solid, and smooth contact area, is in reality a collection of fragmented microscopical contact patches, composed of single contact points.

Given its intrinsic complexity, the modeling of the real area of contact has been the subject of a huge amount of studies, which yielded different and contrasting results. As it is easy to imagine, the real area of contact is crucial for many real-world applications, such as the prediction of wear and fretting, charge and heat conduction, and frictional effects. Let alone think of how an acting load is normally believed to be uniformly distributed over the contact surface, and how variable and uneven it must, in reality, appear at microscopic length scales.

The importance of roughness, together with our knowledge in parallel computing and fast solution methods, are the premises of the current work. In this study, we analyze rough contact between realistic surfaces, and we resolve it numerically at micro-scale, to understand its meso- and macro-scale effects. We do this by means of the Finite Element Method, in combination with an optimal multi-grid strategy and a spatial decomposition to perform the computations on highly parallel super-computers. We concentrate on the type of surfaces for which it is believed that molecular and chemical effects can be neglected.

We simulate the contact between an elastic cube and diverse rigid rough surfaces, under different loading conditions, and we derive empirical laws which describe the influence of well known roughness parameters on important features such as contact evolution and static friction production. We also define bounds on the uncertainty of our measurements, to make clear the level up to which our predictions have to be considered reliable and applicable.

Literature on roughness is densely populated by models, approaches, and theoretical predictions about the evolution of the real area of contact. An exhaustive comparison of our results with such corpus of works, articles, books, and theses would be

infeasible. We therefore compare our results on the real area of contact to the predictions of two widely accepted theories (one by B. N. J. Persson, the other by A. W. Bush, R. D. Gibson, and T. R. Thomas), which have often proved to be interpretable as asymptotical bounds, for systems at low pressures. For large pressures, and consequent large areas of contact, we also compare our results to the newly developed and semi-empirical theory by Yastrebov, Molinari, and Anciaux.

Finally, we test our method on the real world problem of tyre-asphalt interactions on wet roads, comparing the results obtained by our method to data from other studies, collected on real highways and runways, and to a theoretical model, which is close, in the assumptions, to our numerical experiments.

Acknowledgements

I wish to thank everyone who supported me during this study: Rolf Krause for being both an academic and a research advisor, providing me with knowledge, motivation, method, and scientific spirit, Illia Horenko and Igor Pivkin for their suggestions throughout the development of this work, Friedemann Schuricht for his lectures and his remarks on elasticity and contact problems, and Marco Paggi for his useful advices about theory, methods, and literature in contact mechanics. I also acknowledge the collaboration of a non-disclosed industrial partner, which provided me with useful insights, based on its long-time experience in this field.

I thank my family and my girlfriend for continuous support, and for coping with me going through the ups and downs every research brings. Thanks to my best friends, the five of them, who kept me laughing, and helped me seeing everything in the right perspective. Thanks to my musical family, for keeping my brain at the right tempo. I also would like to thank Marco Favino for fighting back to back –literally– with me during these years, helping me when I felt stuck, blocked, or inconclusive in this work. Finally, I thank Gioacchino Noris, who was there during my bachelor and master years, and I had the luck to find again on my side, before the final effort which lead to this thesis.

Contents

Figures

Tables

Chapter 1

Introduction: constitutive elements

1.1 Roughness and rough contact

It may seem easy to define *roughness*: everybody, when asked, should be able to indicate, among two or more surfaces, which one is the roughest, after a simple visual or tactile inspection. Nevertheless, when given the task of formally describing roughness, it is difficult to decide on what parameters could characterize it and help classifying surfaces accordingly.

A common –though not unique– choice for such mathematical characterization is based on the Power Spectral Density (PSD) of the surface, which therefore decomposed in wave vectors of different wave lengths, and analyzed in terms of their amplitudes. The PSDs of most real surfaces have been experimentally found to follow a power law (an exponential decay with respect to wave number). And this implies the so-called self-affinity of natural corrugated substrates. Thus, it is possible to prove that, up to a certain level, the surfaces surrounding us in everyday life are similar to fractals, in which geometrical features repeat themselves identically, at different length scales, up to the limits of continuity of matter.

With the sound foundation supplied by this mathematical description, it is possible to develop contact theories, aiming to predict the results of microscopic interactions of surface asperities. There exist several models studying microscopic contact mechanics, developed under different assumptions and with diverse methods, which lead to different results, but share certain common aspects. Their application is of paramount importance to understand problems such as fretting, wear, and friction production in engineering and real world applications.

1.2 Framework of numerical experiments

As it is becoming common for many fields of science, numerical studies have been performed to simulate contact at different length scales, and different methods were

employed, from Particle-Based Methods, to Boundary Elements Methods and Finite Elements Method.

Most numerical simulations in the field are concerned with a flat surface in contact with a self-affine rough solid substrate, with one of the two bodies being rigid, and the other one following the material law generally known as *linear elasticity*. An essential knowledge about material laws is therefore needed to understand the results of numerical simulations of microscopic contact, and what the applicability of their results is.

Theoretical frameworks exist, in which contact problems are solved. One of the classic examples of an analytical solution of a contact is the one derived by Hertz in the nineteenth century, and concerning the contact of two smooth curved elastic bodies. A more general model is the so-called Signorini problem, a boundary value problem equipped with a set of linear constraints, which constitutes the abstract formulation of a contact between two bodies: one of the bodies is modeled as linear elastic, whereas the other one acts as a rigid obstacle.

The Signorini problem is often adopted in numerics. Its weak formulation can efficiently be solved with the Finite Element Method, yielding a solution which has been shown to be a good approximation of reality, when a sufficiently dense discretization of the bodies is used, and the correct material law is imposed.

1.3 Brief introduction to friction

This work is concerned with roughness and its influence on micro- and macroscopic quantities generated by contact. One of the quantities we study is static friction, but since, to the best of our knowledge, the existing analytical models of static friction are not set in the same framework of our numerical experiments and/or do not take the full roughness characterization into account, we just briefly introduce key concepts of friction here, without going into the details of specific models. We instead verify the results we obtain with real-world data, as it is explained below.

Friction is a term indicating the commonly experienced phenomenon of a resistance force opposing to the reciprocal sliding of two surfaces (which do not have to be solid: they can also be fluid). It makes car tyres brake on streets, machinery parts fret and wear, and dissipates energy of any object which is not moving in vacuum. A good understanding of friction and lubricating technologies is the key for energy saving and fault reduction, two factors which drive the modern economy, where the reduction of environmental impact is becoming more and more important. Because of its importance, many publications are concerned with friction, from contact mechanics books (c.f. [Wri06]), to experimental studies ([Lor12]), to dedicated books, such as [Per10]. In this section, we will summarize basic notions of *tribology*, that is, the science devoted to the study of friction. We start this section with some historical remarks, mostly taken from the above mentioned sources, and a more complete treatise

of the subject can be found in [Dow79]. We then illustrate the principal types of friction which are encountered in engineering, and list two of the principal mechanisms which are at their origin. Finally, we will focus on relations which are known –yet not always understood– between roughness and friction.

1.3.1 Historical remarks and basic concepts

The influence of roughness on friction production has been fascinating scientists since the beginning of what we consider the scientific thinking, that is, since the age of Greek philosophers. In fact, the first written pieces of evidence of the notion of friction, are by Aristotle in his *Sense and Senibilia*, c.f. the version annotated by G. R. T. Ross ([Ari06]), and Heron of Alexandria in his tractates on mechanics, c.f. [IMK02]. Both Greek philosphers discussed the importance of roughness in the production of forces opposing to the reciprocal sliding of interfaces. Even before them, the use of lubricants to reduce friction was well known to ancient populations. It is believed that Egyptians experienced friction and mastered lubrication while building the pyramids, and transporting monumental statues, when they learned to pour the correct amount of water on sand, to ease the sliding of wooden sleds transporting blocks, as the famous painting of Figure 1.1 seems to suggest, c.f. [FWP$^+$14]. Leonardo Da Vinci studied friction experimentally and with an engineering appeal, which lead him to the development of rolling elements to reduce frictional forces and wear. He was the first to document the apparent independence of friction on nominal area of contact, and the direct proportionality between load and friction. His results are similar to those found by Charles-Augustin de Coulomb , who formulated a theory of friction, which is condensed in the so-called Coulomb's law (or Amonton-Coulomb law), which relates

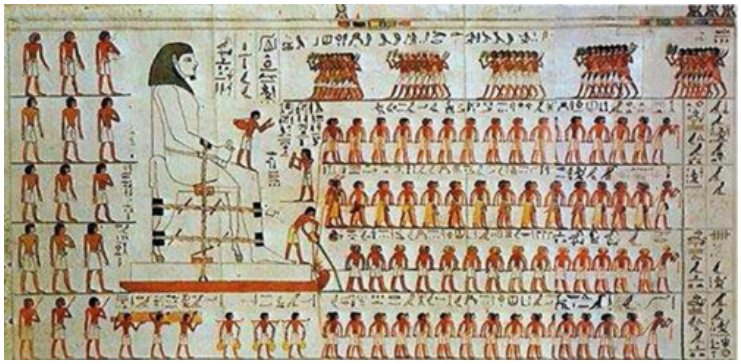

Figure 1.1. Egyptian painting dated 1880 B.C., found in a cave at El-Bershed. The man in front of the statue, seems to be pouring water on sand to ease the sliding of a wooden sled. The painting was first described in [New95].

friction force F_μ and the load acting normally across the surfaces F_N by

$$F_\mu = \mu F_N \tag{1.1}$$

in which the coefficient of friction μ is usually determined experimentally, and depends on the nature of the surfaces in contact. The law, which stood the test of time, is a well-approximating simplification of the frictional behavior, and is the typical form in which friction is introduced to high-school students.

Another concept which is an oversimplification, but also a good engineering approximation, is the distinction between two types of friction, namely *static* and *kinetic*, c.f. [Bla95] and [Sei98]. The best way to understand the difference between static and kinetic friction is by means of the following two definition:

SF If a body is at rest on another surface, to start its sliding motion, a tangential force exceeding the static friction force has to be applied to it.

KF If a body is sliding across a surface, to maintain its velocity, a tangential force in the direction of the motion, and equal in magnitude to the dynamic friction force has to be applied.

Both types of friction are expressed by (1.1), the only difference being the value of the friction coefficient. For most non-viscous materials, it is experimentally found that the static friction coefficient is larger than the kinetic one, and it can be seen by Coulomb's law that kinetic friction is supposed to be independent of velocity of the moving body. This holds generally true for low velocities and non viscous materials. Rubber-like materials, instead, dissipating energy inside the body as heat, are often subject to velocity-dependent kinetic friction, which exceeds static friction. A quite counterintuitive notion, is that for a wheel rolling on a surface, the friction producing the torque needed to make the wheel roll, is of static (and not kinetic) type. This happens because at the contact point between a circle (idealizing the wheel) and a plain (idealizing the surface), the two bodies are actually still, and it is actually the plane which is rotating around the contact point, and not the circle (whose center of mass, moves in a direction parallel to the plane). This obviously only holds in the ideal case, where contact happens at a single point and there is no slip, that is, when the translational velocity v_T of a wheel of radius R and the angular velocity v_A of the same wheel are related by

$$v_T = R v_A \tag{1.2}$$

whereas, in the case of slip, we would have $v_T < R v_A$, and dynamic friction would concur in generating the so-called rolling resistance, c.f. [Per10], [Wri06].

These macroscopic characterizations of friction are what has been used for most engineering applications during the past centuries, but with the availability of new technology and the development of new theories, many steps forward were done in the microscopic description of the phenomenon, aiming to determine what its sources are, and how they can be controlled to reduce its harmful consequences.

1.3.2 The nature of friction

Friction stems from different factors. As a starting point, it is useful to classify them according to the length scale at which they operate. Let us informally define the nano-scale as the scale of geometric features which are in the order of some nanometers, the micro-scale for those which are in the order of some micrometers, and the meso-scale, for millimeters and centimeters. We distinguish between

- Sources of friction acting at the nano-scale.
- Sources of friction acting at the micro- and at the meso-scale.

To the first category belong molecular interactions between boundary layers of atoms of the bodies. Electro-magnetic forces, molecular and atomic bindings, are responsible for the adhesive surface forces which make microscopically flat objects stick together. Such forces are usually measured with atomic force microscopes, as explained in [Per10]. They can yield unexpectedly large friction forces, such in the case of diamond surfaces reciprocally sliding in vacuum, where the boundary Carbon atoms can bind, creating a surprisingly strong force opposed to sliding. The same does not happen if Hydrogen can saturate the Carbon layer, as it happens in a normal setting (not in vacuum). This is a classic example used in many articles, to show the importance of adhesion in certain problems. The original source is [vdOF97], and an follow-up article, relating the force to the sliding direction, is [GCCH07].

The category of forces generated at micro- and meso-scale, is what this work is concerned with, and relates to surface roughness and interactions of substrate asperities. Thus, the fundamental concept, is that the micro- and meso-textures of the surfaces can interact, and the forces which are produced by the contact, when opposed to the movement, are collectively perceived as a macroscopic friction force.

It is clear that both sources of friction can act at the same time, but there are cases in which one of the two prevails, and this is defined in large part by the surface characterization: for very rough surfaces, adhesive forces can be neglected, whereas for microscopically flat surfaces, which are rough only at nano-scale, visco-elastic interactions are irrelevant, c.f. [Per02].

The two sources of friction are studied, also with the intent of finding ways to contrast them, and a classic way of doing so, is lubrication. Lubrication, that is, loosely speaking, the act of putting a liquid substance between the two sliding surfaces, is an enormous field, for which a summary, or a review, would be out of the scope of this work. We point out that most of the lubricating effects act at nano-scale, and are therefore subjects belonging to nano-tribology and tribo-chemistry, and we refer the interested reader to [Per10] for a compact introduction to the field.

1.3.3 Anomalous or not fully understood effects of roughness on friction

Due to its multi-sided nature, friction is still a not completely understood phenomenon. Models exist, to predict frictional behavior of different materials, but laboratory or field experiments are still indispensable to obtain correct results. Though, even in controlled environments, some documented and repeatable results are anomalous and difficult to explain according to classic models. We list two examples which are related to surface roughness, and thus are in principle connected to the present work. One of the two examples turns out to be difficult to prove with our approach, whereas for the other one, the results we obtain are more easily comparable.

The first example of unexpected behavior is the so-called size-effect, c.f. [BLB81]. By performing friction tests on samples of different sizes, it is possible to observe size-dependent friction coefficients. This contradicts the classic theory of friction, for which the area of contact should play no role in building shear resistance. Nevertheless, an interpretation is given in [CP05; CP08], where the phenomenon is put in relation with the fractal nature of natural rough surfaces, and a fractal version of Coulomb's law is derived. This model seems to work independently of the scale at which friction is measured, from sliding experiments performed with small rock samples, to geological fault zones. A reasonable comparison of this model and the numerical results we obtain employing surfaces which are rough on microscopic scale, would require a large effort in computing friction at many different length scales. Though, the accuracy of the numerical results we obtain is strongly affected by the spatial discretization we impose on the elastic domain, and therefore, measures at much larger or much smaller scales would either require an infeasible number of unknowns, or be insufficiently accurate. A different type of approach, with different rough surfaces, could be a suitable solution.

The second phenomenon for which classic friction models are insufficient, is the tyre-asphalt interactions on wet roads. Limited amount of water, such as those present on a street during a normal day of rain, can lead to large loss of traction and braking force, and are responsible for a large increase of car accidents. But the loss of friction which would be predicted by classic models, for such limited presence of water (in the order of one fourth of millimeter), is much smaller than what is measured, and even hydrodynamics effects are not capable of capturing the friction decrease. An interpretation is given in [PTAT05], and is concerned with the reduction of rubber viscosity-related effects, which could happen if water was smoothing the road surfaces only slightly, by filling microscopical pools, where rubber excitation would normally take place. We tested this assumption with our numerical experiments in Section 5.4.2, where we interpret our results according to data collected on real asphalt pavements by different studies.

1.4 Outline of the present work

We aim to quantify the influence of roughness parameters on microscopic contact evolution and related quantities of interest such as real area of contact and static friction force.

This work combines the constitutive elements presented in this brief introduction, and it is organized as follows:

- In Chapter 2, fundamental notions of elasticity and contact problems are given. Constitutive laws of different models of elasticity are derived, and particular cases of contact problems are presented.

- In Chapter 3, we collect state-of-the-art knowledge about self-affine rough surfaces and rough contact theory. We begin with the characterization of a random signal developed by Nayak and apply it to the description of self-affinity features of rough surfaces. We then illustrate different models of rough contact mechanics, briefly comparing their main results.

- In Chapter 4, we give an overview of the techniques and algorithms employed in our numerical experiments. We describe how we efficiently generate surfaces, how we solve the contact problem for a highly irregular obstacle, and we illustrate the basics of the Finite Element Method applied to the problem we are concerned with.

- In Chapter 5, we list the outcomes of several experiments we performed with sets of differently rough surfaces for two types of experiments: normal loading and shear test. We compare our results to experimental data and to one analytical theory about friction on wet roads, a setting which has enough points in common with our assumptions to be approximated accurately by our numerical examples.

Chapter 2

Fundamentals in elasticity and contact theory

In this chapter we give an overview of the theory of elasticity and contact problems, which are constitutive parts of the present work. Section 2.1 is an essential exposition of the derivation of elastic material laws. Section 2.2 focuses on the abstract formulation of a contact problem in elasticity. We conclude the chapter with the analysis of the well known Hertzian contact, in Section 2.3.

Tensor, matrix, and vector quantities are denoted by bold symbols, e.g. $\mathbf{v}$, and their ith component by v_i, whereas the Euclidean norm of a vector is denoted by the name of the vector written in light font, e.g. $\|\mathbf{v}\| = v$. We adopt Einstein's notation: the partial derivative with respect to x_j is abbreviated with the subscript $_{,j}$.

2.1 Elasticity models

In our numerical experiments, we will always employ an elastic domain (a cube), which will be modeled as a continuous solid, that is, its internal structure (which could be at molecular level, but also just at a length scale which is below the observed one) is homogenized, and the material composing our macroscopic object is assumed to respond uniformly in every point. The mathematical study of a body under this assumption is called *continuum solid mechanics*, and it is devoted to the development of models which describe the reaction of the matter to different stimuli, such as deformation, motion, variation of pressure, temperature, and so on. The models are developed starting from classic mechanics principles, that is, conservation of mass, angular and linear momentum, and energy, which are stated as integral equations on arbitrary subparts of the body. The same principles can then be turned in point-wise equations.

In this section, we derive the basis of models of elasticity, that is, the law according to which a body always reaches a configuration in which its stresses are minimized, thus for example, in the absence of constraints and external forces, the body will reach

its nominal initial (stress-free) configuration.

Many text books exist, which treat the subject of elasticity, e. g. [Ogd84], [TT60], or [Cha99]. In this section, we will follow the approach of [Wri06].

2.1.1 Kinematics

We define a body B as an n-dimensional set of points which are in a region of the Euclidean space $\mathbb{E}^n$, with $n = 1, 2, 3$. Such a body, in $\mathbb{E}^3$ is commonly defined as a *simple body*, c.f. [MH94], and we only consider the three-dimensional case for the rest of the section. Every point X of B is called a *particle*, and a *configuration* of B is a one-to-one mapping $\chi : \mathsf{B} \longmapsto \mathbb{E}^3$ which places the particles of B in $\mathbb{E}^3$. Thus, the particle X can be written with respect to the coordinate system of $\mathbb{E}^3$ as $\mathbf{x}$ (the so-called material coordinates), with the relationship $\mathbf{x} = \chi(\mathbf{x})$. The placement, or *deformation* of the body B is the application of the mapping χ to all its points, that is

$$\chi(\mathsf{B}) = \{\chi(X) | X \in \mathsf{B}\}. \tag{2.1}$$

The motion of B is a time-dependent family of deformations $\chi_t : \mathsf{B} \to \mathbb{E}^3$. Thus, the position of the particle X at time $t \in \mathbb{R}$ is

$$\mathbf{x} = \chi_t(X) = \chi(X, t). \tag{2.2}$$

We will refer to $\mathbf{X} = \chi_0(X)$ as the *reference configuration* of B, where $\mathbf{X}$ is the position of particle X in the *current configuration* $\chi_t(X)$. Thus, using (2.2) we can write

$$\mathbf{x} = \chi(\chi_0^{-1}(\mathbf{X}), t) \tag{2.3}$$

where the superscript $^{-1}$ denotes function inversion. To simplify notation, we assume that $\mathbf{X}$ denotes the coordinates of X in the reference configuration (c.f. [TN65]), therefore we can write

$$\mathbf{x} = \chi(\mathbf{X}, t). \tag{2.4}$$

Let the orthogonal Cartesian basis $\mathbf{E}^k$ be the basis of the reference configuration with origin $\mathbf{O}$. Then, for the position vector $\mathbf{X}$, it holds

$$\mathbf{X} = X^k \mathbf{E}^k \tag{2.5}$$

where X^k are the Lagrangian coordinates of $\mathbf{X}$ with respect to $\mathbf{E}^k$. We can now introduce the displacement vector $\mathbf{u}(\mathbf{X})$ as the difference in position vectors of a point in reference and current configurations, that is

$$\mathbf{u}(\mathbf{X}, t) = \chi(\mathbf{X}, t) - \mathbf{X}. \tag{2.6}$$

Figure 2.1 summarizes the quantities we have introduced so far.

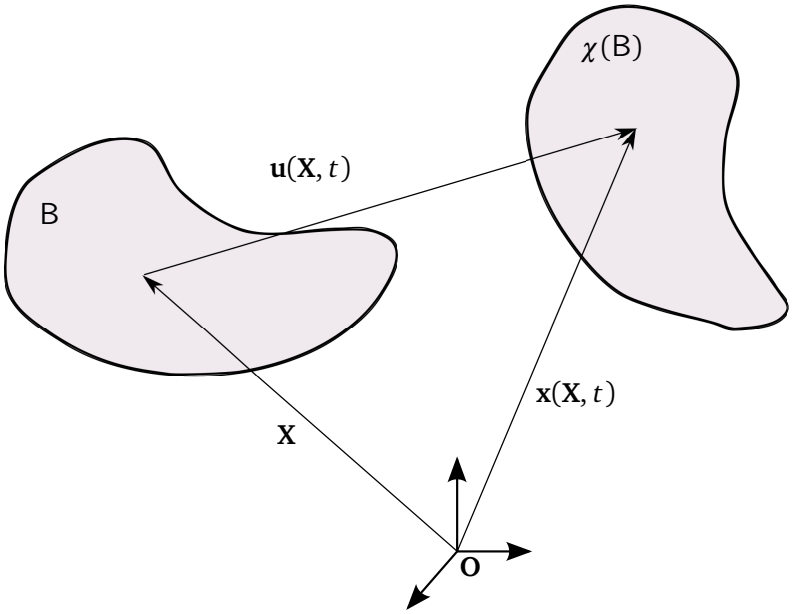

Figure 2.1. Configurations of body B

To describe the deformation process locally, we define the deformation gradient **F** as the gradient of the motion with respect to material coordinates. Thus, **F** is a tensor associating to every material line d**X** in B the line element d**x** in $\chi(\mathsf{B})$, that is:

$$\mathbf{F} := \nabla\chi(\mathbf{X}, t) = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \frac{\partial x_i}{\partial X^k}\mathbf{e}_i \otimes \mathbf{E}^k \tag{2.7}$$

where x_i denotes the ith component of **x**, and $\mathbf{e}_i$ denotes the basis of the spatial or current configuration.

We define

$$J := \det \mathbf{F} \tag{2.8}$$

where J is usually called the deformation determinant, or *Jacobian* determinant. The Jacobian has to be strictly positive, in order to have local admissibility of the deformation. Given these constraints, the deformation gradient is invertible, and it holds

$$\mathrm{d}\mathbf{X} = \mathbf{F}^{-1}\mathrm{d}\mathbf{x} \tag{2.9}$$

To allow measurement of distances in the deformed configuration, with respect to the initial configuration B, we introduce the right Cauchy-Green strain tensor **C**, defined as

$$\mathbf{C} = \mathbf{F}^T\mathbf{F}. \tag{2.10}$$

Since in the initial configuration B, the strain is zero (and therefore $\mathbf{F} = \mathbf{1}$ and $\mathbf{C} = \mathbf{1}$, where **1** denotes the identity tensor), we can introduce the Green-Lagrange strain

tensor $\mathbb{E}$, defined as

$$\mathbf{E} = \frac{1}{2}\left(\mathbf{F}^T\mathbf{F} - \mathbf{1}\right) = \frac{1}{2}(\mathbf{C} - \mathbf{1}) \tag{2.11}$$

It is possible to show (c.f. [Wri06]) that

$$\mathbf{F} = \nabla\mathbf{x} = \nabla\mathbf{X} + \nabla\mathbf{u} = \mathbf{1} + \nabla\mathbf{u} \tag{2.12}$$

and therefore we can rewrite (2.11) as

$$\mathbf{E} = \frac{1}{2}\left[\mathbf{1} + \nabla\mathbf{u} + \nabla\mathbf{u}^T + (\nabla\mathbf{u})(\nabla\mathbf{u})^T - \mathbf{1}\right] = \frac{1}{2}\left[\nabla\mathbf{u} + \nabla\mathbf{u}^T + (\nabla\mathbf{u})(\nabla\mathbf{u})^T\right] \tag{2.13}$$

and approximating this to leading order (disregarding second order terms), we obtain the so-called linearized strain tensor $\boldsymbol{\varepsilon}(\mathbf{u})$ defined as

$$\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T), \tag{2.14}$$

which is used in linear elasticity, under assumption of small deformations.

2.1.2 Balance laws

Balance of mass

The balance (or conservation) of the mass m of a body can be achieved by ensuring that each subpart of the body B has the same mass in reference and current configurations. Let $\tilde{\mathsf{B}} \subset \mathsf{B}$ represent an arbitrary subpart of B, we impose the following condition:

$$m = \int_{\tilde{\mathsf{B}}} \rho_r \mathrm{d}\mathbf{X} = \int_{\chi(\tilde{\mathsf{B}})} \rho \mathrm{d}\mathbf{X}, \quad \forall \tilde{\mathsf{B}} \subset \mathsf{B} \tag{2.15}$$

where ρ_r is the mass density in initial configuration, whereas ρ is the density in current configuration. We can perform the change of variables of the RHS of (2.15), writing the integral in the reference configuration and obtain

$$\int_{\tilde{\mathsf{B}}} \rho_r \mathrm{d}\mathbf{X} = \int_{\tilde{\mathsf{B}}} \rho J \mathrm{d}\mathbf{X}, \quad \forall \tilde{\mathsf{B}} \subset \mathsf{B} \tag{2.16}$$

and therefore, since this condition hold for any arbitrary subpart of B, we can conclude that $\rho_r = J\rho$.

Balance of momentum

The conservation of linear momentum with respect to a volume element in the current configuration $\chi(\mathsf{B})$ can be written as

$$\operatorname{div}\boldsymbol{\sigma} + \rho\bar{\mathbf{b}} = \rho\dot{\mathbf{v}} \tag{2.17}$$

where $\mathbf{v}$ denotes the velocity of a material point in the reference configuration, thus

$$\mathbf{v}(\mathbf{X},t) = \frac{\partial \chi(\mathbf{X},t)}{\partial t} = \dot{\chi}(\mathbf{X},t). \tag{2.18}$$

In (2.17), $\rho\bar{\mathbf{b}}$ represents the sum of volume forces, and $\rho\mathbf{v}$ is the inertia term, which can be neglected for static analysis. The Cauchy stress tensor $\boldsymbol{\sigma}$ relates the stress vector $\mathbf{t}$ to $\mathbf{n}$, which is the vector normal to the surface of $\chi(\mathsf{B})$ through

$$\mathbf{t} = \boldsymbol{\sigma}^T\mathbf{n} \tag{2.19}$$

which is a linear relation proved by the Cauchy theorem. We omit the derivation of angular momentum, and we only report that in the absence of polar stresses, the angular momentum balance is achieved for a symmetric Cauchy stress tensor, thus

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^T \tag{2.20}$$

Equations (2.17) and (2.20) refer to the current configuration $\chi(\mathsf{B})$, but it is also possible to write them with respect to the initial configuration, and this is extremely advantageous for computational reasons, because it makes possible to carry computations always on the same configuration. To obtain such relations, we begin by writing the equivalence of a force in reference and current configurations

$$\int_{\partial\chi(\mathsf{B})} \boldsymbol{\sigma}\mathbf{n}\,\mathrm{d}a = \int_{\partial\mathsf{B}} \boldsymbol{\sigma}J\mathbf{F}^{-T}\mathbf{N}\,\mathrm{d}A = \int_{\partial\mathsf{B}} \mathbf{P}\mathbf{N}\,\mathrm{d}A \tag{2.21}$$

where $\mathbf{N}$ denotes the normal to the surface of B, and $\mathrm{d}a$ and $\mathrm{d}A$ are the area elements of $\chi(\mathsf{B})$ and B, respectively. The tensor $\mathbf{P} = J\boldsymbol{\sigma}\mathbf{F}^{-T}$ is called the first Piola-Kirchhoff stress tensor.

With some algebraic computations, we can write the balance of momentum as

$$\mathrm{div}\,\mathbf{P} + \rho_r\bar{\mathbf{b}} = \rho_r\dot{\mathbf{v}}. \tag{2.22}$$

Notice that using $\mathbf{P}$ to write the conservation of angular momentum, we find that $\mathbf{P}$ is in general not symmetric. Therefore, the second Piola-Kirchhoff stress tensor is defined as $\mathbf{S} = J\mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^{-T}$. The tensor $\mathbf{S}$ is symmetric, and it holds $\mathbf{P} = \mathbf{F}\mathbf{S}$.

Hyperelastic response functions

We finish this section with some brief introductions of hyperelastic constitutive equations, which allow to model the response of materials, by relating the kinematics to the continuity equations. We give, as an example, the definition of a Neo-Hookean material. In our simulations, however, we will employ a simpler linear elastic material law, that is, we will model the response after Hooke's law and we show at the end of

the section, how it is possible to reduce a hyperelastic constitutive equation to a linear elastic law, by using the linearized strain tensor $\boldsymbol{\varepsilon}(\mathbf{u})$ introduced in (2.14).

A material is said to be hyperelastic if the first Piola-Kirchhoff stress tensor $\mathbf{P}$ can be expressed as partial derivative of the strain tensor energy function $\mathscr{W}$ with respect to $\mathbf{F}$, that is

$$\mathbf{P}(\mathbf{F}) = \frac{\partial \mathscr{W}(\mathbf{F}, \mathbf{X})}{\partial \mathbf{F}} \tag{2.23}$$

or, similarly for the second Piola-Kirchhoff stress tensor:

$$\mathbf{S}(\mathbf{C}) = 2\frac{\partial \mathscr{W}(\mathbf{C}, \mathbf{X})}{\partial \mathbf{C}}. \tag{2.24}$$

For a homogeneous material, the neo-Hookean strain energy is defined as

$$\mathscr{W}(\mathscr{C}) = \frac{\mu}{2}[\mathrm{tr}(\mathbf{C}) - 3] - \mu \ln(J) + \frac{\lambda}{2}[\ln(J)]^2. \tag{2.25}$$

where μ and λ are the Lamè constants, which have to be determined by experiments. The formulation of (2.25) we employed is given in [BW08], while in [Wri06], a slightly different, but equivalent, formulation is given. For this energy, we can derive first and second Piola-Kirchhoff tensors as

$$\mathbf{P}(\mathbf{F}) = \frac{\partial \mathscr{W}}{\partial \mathbf{F}} = \mu\left[\mathbf{F} - \mathbf{F}^{-T}\right] + \lambda\left[\ln(J)\right]\mathbf{F}^{-T} \tag{2.26}$$

$$\mathbf{S}(\mathbf{C}) = 2\frac{\partial \mathscr{W}}{\partial \mathbf{C}} = \mu\left[\mathbf{1} - \mathbf{C}^{-1}\right] + \lambda\left[\ln(J)\right]\mathbf{C}^{-1}. \tag{2.27}$$

There exist many other material laws, such as the Fung model [Fun90] (which is usually applied in biology), Mooney-Rivlin model [Moo40; Riv48] (a generalization of the neo-Hookean model), or Blatz-Ko model [FO01], used for compressible rubber materials (c.f. [Wan96]). As mentioned at the beginning of this section, of particular interest for this work there is the linear elastic material law. The model is simple, and the derivation can be found in [TAC+84]. We limit ourselves to reporting the constitutive law, which can be found, e.g. in [Hau02], and reads

$$\boldsymbol{\sigma} = \frac{1}{2\mu}\left\{\boldsymbol{\varepsilon}(\mathbf{u}) + \frac{\nu}{1-2\nu}\,\mathrm{tr}[\boldsymbol{\varepsilon}(\mathbf{u})]\mathbf{1}\right\} \tag{2.28}$$

where $\boldsymbol{\varepsilon}(\mathbf{u})$ has been defined in (2.14), μ denotes the shear modulus and ν denotes the Poisson ratio. This law is also known as Hooke's law, and sometimes it is expressed as function of the Young's modulus E and the bulk modulus K, univocally defined as

$$E = 2\mu(1+\nu) \tag{2.29}$$

$$K = \frac{2\mu(1+\nu)}{3(1-2\nu)}. \tag{2.30}$$

We will describe the elastic body of our experiments via E and ν, to have correspondence with the Hertzian theory of contact. The interpretation of E and ν, is the following. In case of a uniaxial tension (or compression) loading by means of stress σ, for a longitudinal strain ε, for the transverse strain ε_Q it holds

$$\varepsilon_Q = -\nu\varepsilon, \tag{2.31}$$

whereas the stress-strain relation (similar to the simple mono-dimensional Hooke's law for a spring) is

$$\sigma = E\varepsilon. \tag{2.32}$$

Differently stated, E determines the stress (or pressure) needed to compress (or elongate) the continuum, imposing a strain ε, whereas ν determines how much the continuum will expand (or compress) in directions orthogonal to the stress axis, as a result of the uniaxial deformation.

Linear elasticity can also be extended to include viscosity effects. Viscosity can be seen as a variation of the stress tensor according to the time derivative of the strain, thus, loosely speaking, to the speed of the deformation. Rubber, for example, can be orders of magnitude stiffer, when deformations are "fast" compared to when they are "slow", c.f. [Fer61]. Mathematically, the linearization of this property can be modeled by means of the constitutive equation

$$\boldsymbol{\sigma}(t) = \int_{-\infty}^{t} 2\mu(t-\tau)(\dot{\boldsymbol{\varepsilon}}^D(\mathbf{u},\tau)\mathrm{d}\tau + \left\{\int_{-\infty}^{t} K(t-\tau)\left[\operatorname{tr}\dot{\boldsymbol{\varepsilon}}(\mathbf{u},\tau)\right]\mathrm{d}\tau\right\}\mathbf{1} \tag{2.33}$$

where $\dot{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}^D$ denote the derivative with respect to time and the deviator of the linearized strain tensor, respectively, c.f. [Hau02]. The material functions $\mu(\cdot)$ and $K(\cdot)$ are called *shear* and *bulk* functions, respectively. Notice that using constant relaxation functions $\mu(t) = \mu_0$ and $K(t) = K_0$, (2.33) becomes equivalent to the constitutive equation of linear elasticity, (2.28). In the case of non-constant relaxation functions, instead, the stress tensor depends on the so-called *strain history*, that is, the history of all strains which have been applied to the material from $\tau = 0$ to $\tau = t$. The stress tensor can therefore be computed as a sum of increments

$$\Delta_\tau\boldsymbol{\sigma}(t) = 2\mu(t-\tau)\Delta\boldsymbol{\varepsilon}^D(\mathbf{u},t) + K(t-\tau)[\operatorname{tr}\Delta\boldsymbol{\varepsilon}(\mathbf{u},t)]\mathbf{1}. \tag{2.34}$$

Since the stress tensor at time t depends on the strain history, materials which follow this law are considered to have *memory*. We refer the interested reader to [Hau02] for examples of stress tensor increment for different strain histories, and to [WC68] for an adaptation of this model to a Finite Element setting, whereas [Fer61] and [RC07] are two books on viscoelasticity of polymers and present many examples of viscoelastic response functions.

2.2 Contact problems in elasticity

In this section we give the abstract mathematical formulation of the contact problem we are studying in our experiments.

We are interested in a contact problem in the frame of linear elasticity, with mixed Dirichlet and Neumann boundary conditions. In our case, the Boundary Value Problem (BVP) is defined as:

$$
\begin{aligned}
-\nabla\cdot\boldsymbol{\sigma}(\mathbf{u}) &= 0, \quad \text{in } \Omega && (2.35)\\
\mathbf{u} &= \mathbf{d}, \quad \text{on } \Gamma_D && (2.36)\\
\frac{\partial \mathbf{u}}{\partial \mathbf{n}} &= 0, \quad \text{on } \Gamma_N && (2.37)\\
\frac{\partial \mathbf{u}}{\partial \mathbf{n}} &= \mathbf{f}, \quad \text{on } \Gamma_C. && (2.38)
\end{aligned}
$$

We define the elastic domain $\Omega = \mathsf{B} = \chi(\mathsf{B})$ and we model it as linear, homogeneous, and isotropic with strainless reference configuration. Thus, the stress tensor $\boldsymbol{\sigma}(\mathbf{u})$ is defined as in (2.28). The symbol Γ_D denotes the subpart of the domain boundary $\partial\Omega$ on which Dirichlet boundary condition are imposed, and similarly, on $\Gamma_N \subset \partial\Omega$, Neumann conditions are imposed. The subdomain $\Gamma_C \subset \partial\Omega$ is the surface of potential contact, that is, where $\partial\Omega$ can enter in contact with an obstacle $\mathscr{O}$, the force $\mathbf{f}$ is unknown and represents solely the contact forces arising from the elastic interactions of the elastic bulk with the rigid obstacle.

We model our problem as a frictionless, adhesion-free, contact problem, therefore, on Γ_C we introduce the following constraints on $\boldsymbol{\sigma}$:

$$
\begin{aligned}
\sigma_n &\leq 0 && (2.39)\\
\boldsymbol{\sigma}_t &= \mathbf{0} && (2.40)
\end{aligned}
$$

where normal and tangential stress components are denoted respectively by σ_n and $\boldsymbol{\sigma}_t$. Formally $\sigma_n := \sigma_{ij} n_j n_i$ and $((\boldsymbol{\sigma}_t)_i)_{1\leq i\leq 3} := (\sigma_{ij} n_j - \sigma_n n_i)_{1\leq i\leq 3}$. Thus, (2.39) enforces negative normal stresses (the elastic surface can be pushed, but not pulled, since there are no adhesive forces) and (2.40) expresses the absence of point-wise imposed friction, that is, the absence of stress in directions which are tangential to the contact direction (therefore, points slide freely on the obstacle planar surfaces).

The non-penetration condition of the deformed elastic body domain Ω and the obstacle $\mathscr{O}$, can be formalized as the kinematic contact condition, (c.f. [Sch02])

$$\operatorname{int}((\mathbf{x}+\mathbf{u}) \cap \mathscr{O}) = \emptyset. \tag{2.41}$$

In this form, an implementation of the non-penetration condition would be infeasible, therefore we have to introduce $d_{\mathscr{O}} : \mathbb{R}^3 \to \mathbb{R}$, the signed distance function to the obstacle's boundary

$$d_{\mathscr{O}}(\mathbf{x}) = \operatorname{dist}(\mathbf{x}, \mathscr{O}^C) - \operatorname{dist}(\mathbf{x}, \mathscr{O}). \tag{2.42}$$

If we assume that the actual zone of contact is Γ_C, the non-penetration condition Eq. 2.41 is equivalent to

$$d_{\mathscr{O}}(\mathbf{x}+\mathbf{u}) \geq \mathbf{0}, \text{ on } \Gamma_C \tag{2.43}$$

and adding (2.43) to the described set of conditions completes the contact problem. Notice that we define the distance function to be positive outside the obstacle and negative inside.

2.2.1 Signorini problem

An efficient implementation of (2.43), is achieved by linearization. The linearized contact condition

$$\mathbf{u} \cdot \mathbf{n} \leq g(\mathbf{x}+\mathbf{u}) \tag{2.44}$$

makes use of the gap function $g(\cdot)$, that is the distance of a point on the contact boundary to the obstacle. In other words, g is the maximum normal displacement that a point is allowed to assume.

By adding to our set of constraints the complementarity condition

$$\left[\mathbf{u} \cdot \mathbf{n} - g(\mathbf{x}+\mathbf{u})\right] \sigma_n = 0 \tag{2.45}$$

which states that there is no normal stress in the absence of contact, and setting the problem in linear elasticity, we obtain an instance of the well known Signorini problem, which has been the subject of many studies, such as [Kra09] and [KM11].

We will outline our iterative version (designed for geometrically complex obstacles) of such problem in Section 4.2.

2.3 Hertzian contact

The nature of contact problems in elasticity is so complex, that analytical solutions exist only for a small class of examples. Even more difficult are problems where the two contact surfaces are non-conforming, that is, the contact does not happen between two flat regions of the bodies. The possibly best known and most widespread theory of contact is due to Hertz, who derived in [Her82] the analytical solution for the problem of the contact between two curved bodies, as shown in Figure 2.2. The work, which was initially conceived to study the interference caused by the deformations of two lenses in contact, provides a very good prediction of the contact of two non-conforming smooth solids, with a range of validity that we explain in Section 2.3.3. In this section, we recapitulate the results of Hertz theory which are relevant to our study, following the approach of [Joh85], and the interested reader is referred to the same book for other examples, such as contacts of solids with non-smooth boundaries. Other simplified exposures and examples of Hertzian contact can be found, e.g. in [LLP86] and [Pad01]. The results summarized in this section will be of particular interest for

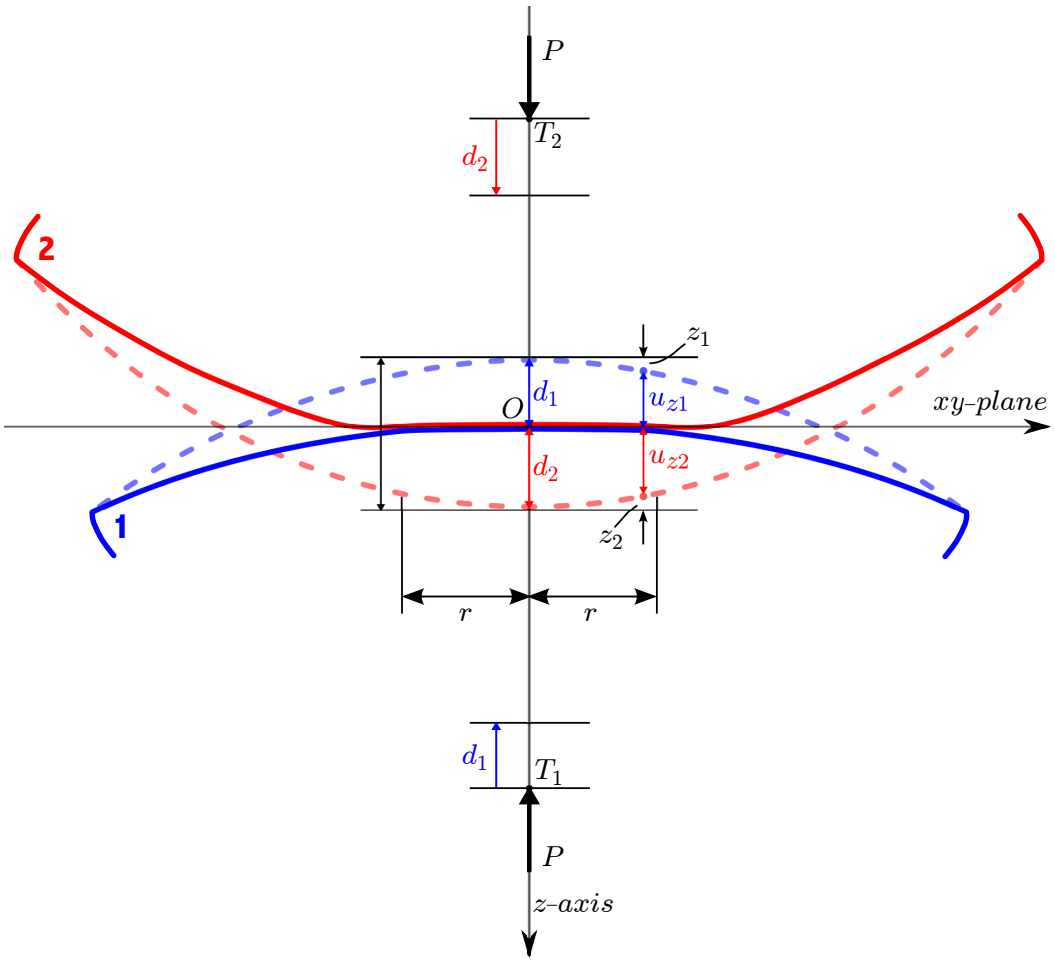

Figure 2.2. Non-conforming smooth convex surfaces in contact.

the theory developed by Bush, Gibson, and Thomas in [BGT75], which we present in Section 3.3.

2.3.1 Geometry of contact region

We are interested in the analysis of the contact developing around a single initial point. Therefore, it is reasonable to identify the origin O of our coordinates system with the first contact point, with the xy-plane being tangent to both surfaces in O, and the z-axis lying along the common normal, pointed towards the inner part of the lower body (see Figure 2.2). We approximate the two contact surfaces with two paraboloids, with major axes aligned to the coordinate system. The equation of the first paraboloid

boundary can be expressed through the equation

$$z_1 = \frac{1}{2R_1'} x_1^2 + \frac{1}{2R_1''} y_1^2 \tag{2.46}$$

where R_1' and R_1'' are the principal radii of curvature of the surface at the origin, aligned to the local coordinate system x_1, y_1, z_1. Similarly we define R_2', R_2'', and x_2, y_2, z_2 for the second paraboloid, and it holds:

$$z_2 = -\left(\frac{1}{2R_2'} x_2^2 + \frac{1}{2R_2''} y_2^2 \right). \tag{2.47}$$

In this description, therefore, convex profiles correspond to positive radii of curvature. We define h as the separation between the two surfaces, therefore $h = z_1 - z_2$, and we transform the two local coordinate system to a global one (x, y, z) in which we can write

$$h = Ax^2 + By^2 + Cxy. \tag{2.48}$$

It is possible to find positive constants A and B such that $C = 0$ and

$$h = Ax^2 + By^2. \tag{2.49}$$

This is proved in detail in Appendix A. For the remainder of this chapter, we will simply define

$$A = \frac{1}{R'}, \quad B = \frac{1}{R''} \tag{2.50}$$

where R' and R'' are the principal relative radii of curvature of the system of interest. The equivalent radius R_e is defined as

$$R_e = \sqrt{R'R''} = \frac{1}{2\sqrt{AB}} \tag{2.51}$$

It can be observed that the contour lines of the separation h given by (2.49) are ellipsoids, with semi-axes which are in the ratio $\sqrt{B}/\sqrt{A} = \sqrt{R'/R''}$.

2.3.2 Point-load, contact area, and pressure distribution

Now we consider the case in which a vertical compressive force P is called a point-load, because it is applied to the points T_1 and T_2 (which have to be sufficiently distant from the contact surfaces, as explained in Section 2.3.3), as depicted in Figure 2.2. The point-load pushes the two points T_1 and T_2 towards O, along the z-axis, and their displacements are d_1 and d_2 respectively. The dashed lines in Figure 2.2 represent how the surfaces would overlap if they did not undergo deformation. The two surface points $S_1(x_0, y_0, z_1)$ and $S_2(x_0, y_0, z_2)$ are aligned vertically and their separation before deformation is described by (2.49). The deformation causes the points S_1 and S_2

to be displaced (with respect to T_1 and T_2) by u_{z1} and u_{z2} respectively. If after the deformation the two surface points coincide, we can write

$$u_{z1} + u_{z2} + h = d_1 + d_2 \tag{2.52}$$

and defining $d = d_1 + d_2$, one can substitute (2.49) in (2.52) and get

$$u_{z1} + u_{z2} = d - Ax_0^2 - By_0^2 \tag{2.53}$$

if the two surface points lie outside the contact area, then

$$u_{z1} + u_{z2} > d - Ax_0^2 - By_0^2 \tag{2.54}$$

We now aim at finding a relation between the point-load P, the parameters A and B, the area of contact, and the pressure distribution internal to the elastic bodies. To do so, since we know that contours of h are elliptic, we assume that the contact region S is an ellipse as well, with semi-axes a and b. Hertz found that the only pressure distribution imposing uniform normal displacement of an elliptic region of the surface is given by

$$\sigma = \sigma_0 \left[1 - \left(\frac{x}{a} \right)^2 - \left(\frac{y}{b} \right)^2 \right]^{\frac{1}{2}}. \tag{2.55}$$

To do so, he compared this problem to the mathematically analogous case of an electrically charged elliptical conductor, with charge intensity varying as the ordinate of a semi-ellipsoid. In such case, the variation of the potential is parabolic, thus, by analogy, in (2.55), σ is a semi-paraboloid and the displacement imposed by such pressure distribution on one body, under the assumption of linear elasticity, is

$$u_z = \frac{1 - \nu^2}{\pi E} \left(L - Mx^2 - Ny^2 \right) \tag{2.56}$$

and therefore, for both bodies

$$u_{z1} + u_{z2} = \left(\frac{1 - \nu_1^2}{\pi E_1} + \frac{1 - \nu_2^2}{\pi E_2} \right) \left(L - Mx^2 - Ny^2 \right) \tag{2.57}$$

and introducing the equivalent elastic modulus E^* defined as

$$\frac{1}{E^*} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \tag{2.58}$$

we can rewrite (2.57) as

$$u_{z1} + u_{z2} = \frac{L - Mx^2 - Ny^2}{\pi E^*} \tag{2.59}$$

which is consistent with (2.53), if

$$
\begin{aligned}
L = d\pi E^* &= \frac{\pi\sigma_0 ab}{2}\int_0^\infty \frac{\mathrm{d}w}{\sqrt{(a^2+w)(b^2+w)w}} = \pi\sigma_0 b\mathbf{K}(e)\\
M = A\pi E^* &= \frac{\pi\sigma_0 ab}{2}\int_0^\infty \frac{\mathrm{d}w}{\sqrt{(a^2+w)^3(b^2+w)w}} = \frac{\pi\sigma_0 b}{e^2a^2}(\mathbf{K}(e)-\mathbf{E}(e))\\
N = B\pi E^* &= \frac{\pi\sigma_0 ab}{2}\int_0^\infty \frac{\mathrm{d}w}{\sqrt{(a^2+w)(b^2+w)^3w}} = \frac{\pi\sigma_0 b}{e^2a^2}\left(\frac{a^2}{b^2}\mathbf{E}(e)-\mathbf{K}(e)\right)
\end{aligned}
\tag{2.60}
$$

where $e=\sqrt{1-b^2/a^2}$ is the eccentricity of the ellipse, with $b<a$, and $\mathbf{K}(e)$ and $\mathbf{E}(e)$ are the complete elliptic integrals of the first and second type, respectively, defined as:

$$\mathbf{K}(e) = \int_0^{\pi/2} \frac{1}{\sqrt{1-e^2\sin^2\theta}}\mathrm{d}\theta \tag{2.61}$$

$$\mathbf{E}(e) = \int_0^{\pi/2} \sqrt{1-e^2\sin^2\theta}\;\mathrm{d}\theta. \tag{2.62}$$

The integrals of (2.60) describe the pressure on the elliptical region, and are derived from potential theory, but we omit the derivation and refer the interested reader to [Her82] or to [Joh85] for a deeper analysis. Since the pressure distribution is ellipsoidal and the total load is P, we can write

$$P = \frac{2}{3}\sigma_0\pi ab \tag{2.63}$$

that is, P is equal to the volume of a semi-ellipsoid with semi-axes a, b, and σ_0.

Finally, to find the ellipse of contact, we write

$$\frac{B}{A} = \frac{N}{M} = \frac{\left(\frac{a}{b}\right)^2\mathbf{E}(e)-\mathbf{K}(e)}{\mathbf{K}(e)-\mathbf{E}(e)} \tag{2.64}$$

and

$$\sqrt{AB} = \frac{\sigma_0 b}{E^* a^2 e^2}\sqrt{\left[\left(\frac{a}{b}\right)^2\mathbf{E}(e)-\mathbf{K}(e)\right][\mathbf{K}(e)-\mathbf{E}(e)]} \tag{2.65}$$

notice that it also holds

$$\frac{B}{A} = \frac{R'}{R''}, \qquad \sqrt{AB} = \frac{1}{2}\sqrt{\frac{1}{R'R''}} = \frac{1}{2R_e}. \tag{2.66}$$

To apply these results to the case of spherical surfaces (surfaces in which R'=R''=R), we introduce the equivalent radius c, defined as the geometric mean of the two semi-axes of the contact ellipse, that is $c=\sqrt{ab}$, which can be computed as

$$c^3 \equiv \sqrt{ab}^3 = \left(\frac{3PR_e}{4E^*}\right)\frac{4}{\pi e^2}\left(\frac{b}{a}\right)^{\frac{3}{2}}\sqrt{\left[\left(\frac{a}{b}\right)^2\mathbf{E}(e)-\mathbf{K}(e)\right][\mathbf{K}(e)-\mathbf{E}(e)]} \tag{2.67}$$

and thus

$$c = \sqrt{ab} = \left(\frac{3PR_e}{4E^*}\right)^{\frac{1}{3}} F_1(e) \tag{2.68}$$

where $F_1(e)$ is a correction factor dependent on the eccentricity e, and tends to 1 as e goes to 0. Thus $F_1(e)$ can be interpreted as a measure of the deviation of c from the radius r of the contact circle resulting from the contact of spherical surfaces under the same loading conditions and with the same R_e (notice that in such contact, $1/R_e = 1/R_1 + 1/R_2$). Thus, in that case

$$r = \left(\frac{3PR_e}{4E^*}\right)^{\frac{1}{3}}. \tag{2.69}$$

The compression d is found from (2.60.1) and (2.63). It holds

$$d = \frac{3P}{2\pi abE^*} b\mathbf{K}(e) = \left(\frac{9P^2}{16E^{*2}R_e}\right)^{\frac{1}{3}} \frac{2}{\pi}\sqrt{\frac{b}{a}}\frac{\mathbf{K}(e)}{F_1(e)} = \left(\frac{9P^2}{16E^{*2}R_e}\right)^{\frac{1}{3}} F_2(e) \tag{2.70}$$

where $F_2(e)$ is again a correction factor going to 1 as e goes to 0, and accounting for the deviation of d from the compression in an equivalent setting, with spherical surfaces. Finally, the maximum pressure is

$$\sigma_0 = \frac{3P}{2\pi ab} = \left(\frac{6PE^{*2}}{\pi^3 R_e^2}\right)^{\frac{1}{3}} F_1(e)^{-2} \tag{2.71}$$

and for the case of spherical surfaces, σ_0 is obtained again setting $F_1(e)$ to 1.

There are other measures derived by Hertz, such as the internal tangential stress, but they are not relevant for the rest of this work, and will therefore be omitted. We reported the results for the area of contact between two paraboloids, but in the literature it is possible to find different examples, such as contact of cylinders, concave surfaces, and contact of a paraboloid and a plane, which will be briefly analyze in Section 3.3.

2.3.3 A remark on validity

Hertz theory is valid only making some assumptions about the area of contact, which must be small, the curvatures, which must be large, and friction, which must be absent. These assumptions, more formally, are:

1 The surfaces are continuous and non conforming, that is, $r \ll R$.

2 The strains are small, so that linear elasticity can be used as a material law, that is, $r \ll R$.

3 Each solid can be considered as an elastic half-space, that is, $a \ll R_{1,2}, a \ll \ell$, where ℓ is the height of the elastic body (the depth of the contact surface with respect to the opposing face of the body).

4 There is no friction acting on the surfaces, that is $\mu = 0$.

These conditions limit the applicability of the theory to real cases. In particular, rubber-like incompressible materials which undergo large deformation and large strain with small pressure, are rarely fitted good by Hertz theory. Materials like rocks and metals (which usually have a Poisson ratio $\nu < 0.3$), instead, are approximated well, as shown in [Joh85].

Chapter 3

State-of-the-art models in rough contact theory

In this chapter we summarize some of the most used and accepted theories of rough contact, starting from the characterization of random rough surfaces defined by Nayak in [Nay71] (Section 3.1), continuing with the contact theories developed by Bush, Gibson, and thomas (Section 3.3), by Persson (Section 3.4), and concluding with the recently published works by Yastrebov, Anciaux, and Molinari (3.5).

3.1 Nayak's characterization of a random process

The characterization of rough (or corrugated) surfaces has always been of large interest for scientists, because in the universe, at some length scale, every surface is rough. Given the complexity and the random nature of such surfaces, a fully deterministic approach to describe them is infeasible, and one has to rely on statistical measures, which will hold on average, for sufficiently large sets of surface specimens. Among the first works which explored the field, there are the ones by M. S. Longuet-Higgins in [LH57a; LH57b; LH62], which are concerned with describing the oceans' moving surfaces, and are a two-dimensional extension of the one-dimensional analysis performed by O. S. Rice in [Ric44; Ric45]. An extended description of random surfaces is introduced by R. Nayak in [Nay71], and is the one used as a starting point by prominent rough contact theories, such as those we will present in the remainder of this chapter. One of the reasons of the success of Nayak's work, a part from the clearness of the argumentation, is the relation he finds between a surface and its one-dimensional profiles, which is of high interest for engineers, since such profiles are easily obtained in laboratories, measuring surfaces with common profilometers. In this section, we will summarize Nayak's theory and highlight those points which are most useful for our work, focussing on isotropic square surfaces, in order to preserve consistency with the numerical experiments described in Chapter 5. Similarly, Nayak's notation will be

adapted to ours.

3.1.1 Characterization of an isotropic random process

We consider a rough surface as a height function $h(x,y)$, assigning to each pair of Cartesian coordinates (x,y) a value of h, that is, the elevation with respect to a reference plane, which, for practical reasons, will be defined as the mean plane of the surface. Thus the mean value of the function $h(x,y)$ over the considered domain $A = [0,L] \times [0,L]$, denoted by $\overline{h}$, will be 0 by construction, and this can be written as

$$\overline{h} = \langle h \rangle = \frac{1}{L^2} \iint_{-L/2}^{+L/2} h(x,y)\,\mathrm{d}x\,\mathrm{d}y = 0 \tag{3.1}$$

where by $\langle \cdots \rangle$ we indicate ensemble averaging. Assuming that the surface is homogeneous, i.e. its description is translationally invariant, its autocorrelation function is defined as

$$R(x,y) = \lim_{L\to\infty} \frac{1}{L^2} \iint_{-L/2}^{+L/2} h(x_1,y_1)h(x_1+x,y_1+y)\ \mathrm{d}x_1\,\mathrm{d}y_1. \tag{3.2}$$

Notice that $R(x,y)$ depends on the in-plane distance vector $\mathbf{x} = (x,y)$ and not on the choice of the point (x_1,y_1), but if the surface is isotropic, then R only depends on the 2-norm of $\mathbf{x}$, denoted by $\|\mathbf{x}\|$.

The Fourier transform of R is possibly the most important way of characterizing surfaces, and it is called the Power Spectral Density (PSD) or simply the power spectrum of the rough surface

$$C(\mathbf{q}) = C(q_x,q_y) = \frac{1}{4\pi^2} \iint_{-\infty}^{+\infty} R(x,y) \exp\left[-i(xq_x + yq_y)\right]\ \mathrm{d}x\ \mathrm{d}y \tag{3.3}$$

where q_x and q_y are the components of the wave-vector $\mathbf{q}$. For isotropic surfaces, $C(\mathbf{q})$ only depends on the 2-norm of $\mathbf{q}$, $\|\mathbf{q}\|$, which will also be denoted simply by q. We can compute the variance of h as $R(0,0)$. In fact, according to (3.2), we obtain

$$\begin{aligned} R(0,0) &= \lim_{L\to\infty} \frac{1}{L^2} \iint_{-L/2}^{+L/2} h(x_1,y_1)h(x_1,y_1)\ \mathrm{d}x_1\,\mathrm{d}y_1 \\ &= \lim_{L\to\infty} \frac{1}{L^2} \iint_{-L/2}^{+L/2} h(x_1,y_1)^2\,\mathrm{d}x_1\,\mathrm{d}y_1 = h_0^2 \end{aligned} \tag{3.4}$$

where h_0 is the root mean square roughness of the surface. We can introduce the inverse Fourier transform of C

$$R(x,y) = \iint_{-\infty}^{+\infty} C(q_x,q_y) \exp\left[i(xq_x + yq_y)\right]\ \mathrm{d}q_x\,\mathrm{d}q_y \tag{3.5}$$

to obtain the relation

$$h_0^2 = R(0,0) = \iint_{-\infty}^{+\infty} C(q_x, q_y)\ \mathrm{d}q_x\ \mathrm{d}q_y \tag{3.6}$$

which indicates that $C(\mathbf{q})$ is a spectral decomposition of h_0^2 into waves with wavenumber $\mathbf{q}$, associated wavelength

$$\lambda = 2\pi / \|\mathbf{q}\| \tag{3.7}$$

and direction along

$$\theta = \arctan\left(\frac{q_x}{q_y}\right). \tag{3.8}$$

3.1.2 Moments of the Power Spectral Density (PSD)

As we have written above, the PSD is one of the most useful and compact ways to characterize a random surface. Many distinctive features of the rough substrate, such as height distribution, peak density, or mean slope, can be deduced from the moments of the PSD. The moment $m_{n_x n_y}$ of C, with $n_x, n_y \in \mathbb{N}$, is defined as follows:

$$m_{n_x n_y} = \iint_{-\infty}^{+\infty} C(q_x, q_y) q_x^{n_x} q_y^{n_y}\ \mathrm{d}q_x\ \mathrm{d}q_y. \tag{3.9}$$

By comparison with (3.6) it can be seen that

$$m_{00} = h_0^2. \tag{3.10}$$

If a profile is taken in the plane $\theta = \theta_0$, the height of the profile is a function only of the distance r from the origin along the profile. For such profile, a definition of the mono-dimensional PSD, denoted by $C_{\theta_0}(q)$, is given by Nayak, but for our purposes, it makes sense only to look at its moments. The n^{th} moment of a profile on the plane θ_0, denoted by $m_{n\theta_0}$, is defined as follows:

$$m_{n\theta_0} = \int_{-\infty}^{+\infty} C_{\theta_0}(q')(q')^n\ \mathrm{d}q' \tag{3.11}$$

and the following relation exists between $m_{n_x n_y}$ and $m_{n\theta_0}$:

$$\begin{aligned} m_{n\theta_0} = {} & m_{n0}\cos^n\theta_0 + C_1^n m_{n-1,1}\cos^{n-1}\theta_0 \sin\theta_0 \\ & + C_2^n m_{n-2,2}\cos^{n-2}\theta_0 \sin^2\theta_0 + \cdots + m_{0n}\sin^n\theta_0 \end{aligned} \tag{3.12}$$

where C_m^n is the binomial coefficient defined as

$$C_m^n = \binom{n}{m} = \frac{n!}{m!(n-m)!}. \tag{3.13}$$

From (3.9) and (3.12), one can derive, for isotropic surfaces, the following identities and relations:

$$m_{20} = m_{02} = m_2 \tag{3.14}$$

$$m_{11} = m_{13} = m_{31} = 0 \tag{3.15}$$

$$m_{00} = m_0 \tag{3.16}$$

$$3m_{22} = m_{40} = m_{04} = m_4 \tag{3.17}$$

where the subscript θ_0 of the one-dimensional moments is dropped because of surface isotropy. Therefore, the only relevant moments for isotropic surfaces are m_0, m_2, and m_4, for which we can derive the following expressions:

$$m_0 = 2\pi \int_0^\infty qC(q)\,\mathrm{d}q \tag{3.18}$$

$$m_2 = \pi \int_0^\infty q^3 C(q)\,\mathrm{d}q \tag{3.19}$$

$$m_4 = \frac{3\pi}{4} \int_0^\infty q^5 C(q)\,\mathrm{d}q. \tag{3.20}$$

Now we assume that $h(x,y)$ can be expressed through an infinite sum of sinusoidal waves (as in a Fourier transformation), and this sum can be written as following:

$$h(x,y) = \sum_n B_n \cos(xq_{xn} + yq_{yn} + \phi_n) \tag{3.21}$$

where ϕ_n is a random phase shift with uniform probability in the range $[0; 2\pi[$. The coefficients B_n can be found with the relation

$$C(q+x, q_y)\,\mathrm{d}q_x\,\mathrm{d}q_y = \frac{1}{2}\sum_{\Delta_q} B_n^2. \tag{3.22}$$

where the summation is performed over all values of n such that (q_{xn}, q_{yn}) lies in the area $\mathrm{d}q_x\,\mathrm{d}q_y$ around (q_x, q_y). From (3.9) and (3.22) we see that

$$m_{n_x n_y} = \frac{1}{2}\sum_n q_{xn}^{n_x} q_{yn}^{n_y} B_n^2. \tag{3.23}$$

3.1.3 Surface statistics

Since the quantities which will be derived in this and the following sections are of statistical nature, it makes sense to analyze their probability densities. To do this, the Central Limit Theorem serves as a basis. Let $\xi_1, \xi_2, \ldots, \xi_n$ be n zero-mean quantities,

each of which is the sum of a sufficiently large number of independent variables. Then, under very general conditions, the joint distribution of ξ_i is Gaussian in n dimensions:

$$p(\xi_1, \xi_2, \ldots, \xi_n) = \frac{1}{\sqrt{(2\pi)^n \Delta}} \exp\left\{-\frac{1}{2} M_{ij} \xi_i \xi_j\right\} \tag{3.24}$$

where M_{ij} is an element of the $n \times n$-matrix $\mathbf{M}$, given by

$$\mathbf{M} = \mathbf{N}^{-1} \tag{3.25}$$

with

$$\mathbf{N} = \begin{pmatrix} \overline{\xi_1^2} & \overline{\xi_1 \xi_2} & \cdots & \overline{\xi_1 \xi_n} \\ \overline{\xi_2 \xi_1} & \overline{\xi_1^2} & \cdots & \overline{\xi_2 \xi_n} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\xi_n \xi_1} & \overline{\xi_n \xi_2} & \cdots & \overline{\xi_n^2} \end{pmatrix}, \tag{3.26}$$

and

$$\Delta = \mathrm{Det}\,(\mathbf{N}). \tag{3.27}$$

Each element of the matrix is defined as

$$\overline{\xi_i \xi_j} = \int \ldots \int_E \xi_i(\phi_1, \ldots, \phi_k)\, \xi_j(\phi_1, \ldots, \phi_k)\, p(\phi_1) \ldots p(\phi_k)\, \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_k \tag{3.28}$$

where E is the probability space of the independent random variables ϕ_j on which the quantities ξ_i depend. As variables ξ_i we can choose

$$\begin{aligned} \xi_1 &= h & \xi_4 &= \partial^2 h / \partial x^2 \\ \xi_2 &= \partial h / \partial x & \xi_5 &= \partial^2 h / \partial x \partial y \\ \xi_3 &= \partial h / \partial y & \xi_6 &= \partial^2 h / \partial y^2 \end{aligned} \tag{3.29}$$

and it is easily observed that, for a sufficiently large amount of surface points, the variables ξ_i satisfy the requirements of the central limit theorem. Therefore, their joint probability density can be expressed through (3.24), and using (3.23), one can derive $\mathbf{N}$. We omit the full derivation of the entries of $\mathbf{N}$, for which an example in given in [Nay71], and we simply it as follows:

$$\mathbf{N} = \begin{pmatrix} m_{00} & 0 & 0 & -m_{20} & -m_{11} & -m_{02} \\ 0 & m_{20} & m_{11} & 0 & 0 & 0 \\ 0 & m_{11} & m_{02} & 0 & 0 & 0 \\ -m_{20} & 0 & 0 & m_{40} & m_{31} & m_{22} \\ -m_{11} & 0 & 0 & m_{31} & m_{22} & m_{13} \\ -m_{02} & 0 & 0 & m_{22} & m_{13} & m_{04} \end{pmatrix} \tag{3.30}$$

and for isotropic surfaces, using the relations given in (3.14), it becomes

$$\mathbf{N} = \begin{pmatrix} m_0 & 0 & 0 & -m_2 & 0 & -m_2 \\ 0 & m_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_2 & 0 & 0 & 0 \\ -m_2 & 0 & 0 & m_4 & 0 & \frac{m_4}{3} \\ 0 & 0 & 0 & 0 & \frac{m_4}{3} & 0 \\ -m_2 & 0 & 0 & \frac{m_4}{3} & 0 & m_4 \end{pmatrix} \tag{3.31}$$

and the inverse is therefore

$$\mathbf{M} = \mathbf{N}^{-1} = \begin{pmatrix} \frac{2m_4}{\Delta_1} & 0 & 0 & \frac{3m_2}{2\Delta_1} & 0 & \frac{3m_2}{2\Delta_1} \\ 0 & \frac{1}{m_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{m_2} & 0 & 0 & 0 \\ \frac{3m_2}{2\Delta_1} & 0 & 0 & \frac{9\Delta_2}{4m_4\Delta_1} & 0 & -\frac{3\Delta_3}{4m_4\Delta_1} \\ 0 & 0 & 0 & 0 & \frac{3}{m_4} & 0 \\ \frac{3m_2}{2\Delta_1} & 0 & 0 & -\frac{3\Delta_3}{4m_4\Delta_1} & 0 & \frac{9\Delta_2}{4m_4\Delta_1} \end{pmatrix} \tag{3.32}$$

with

$$\Delta_1 = 2m_0m_4 - 3m_2^2, \tag{3.33}$$

$$\Delta_2 = m_0m_4 - m_2^2, \tag{3.34}$$

$$\Delta_3 = m_0m_4 - 3m_2^2. \tag{3.35}$$

and (3.27) can be rewritten as

$$\Delta = \frac{4}{27}\left(m_2m_4\right)^2\left(2m_0m_4 - 3m_2^2\right). \tag{3.36}$$

Meanwhile, (3.24) becomes

$$p(\xi_1, \xi_2, \ldots, \xi_6) = \frac{1}{(2\pi)^3\sqrt{\Delta}} \times \exp\left\{-\frac{1}{2}\left[\frac{2m_4\xi_1^2}{\Delta_1} + \frac{9\Delta_2\left(\xi_4^2+\xi_6^2\right)}{4m_4\Delta_1} + \frac{3\xi_5^2}{m_4} + \frac{3m_2\xi_1\left(\xi_4+\xi_6\right)}{\Delta_1} - \frac{3\Delta_3\xi_4\xi_6}{2m_4\Delta_1} + \frac{\xi_2^2+\xi_3^2}{m_2}\right]\right\}. \tag{3.37}$$

This joint probability is the main tool used by Nayak to derive statistics about surface's heights, summits (local maxima), slopes, and curvatures, always in terms of probability distributions.

The probability density $p(\xi_1)$, that is, the distribution of $h(x,y)$, can be derived restricting (3.24) to ξ_1, and obtaining

$$p(\xi_1) = \frac{1}{\sqrt{2\pi \operatorname{Det}(N_{11})}} \exp\left(-\frac{1}{2} M_{11} \xi_1^2\right) \tag{3.38}$$

which can be written noting that

$$M_{11} = \frac{1}{N_{11}} = \left(\overline{\xi_1^2}\right)^{-1} = \frac{1}{h_0^2}. \tag{3.39}$$

Thus, we can rewrite $p(\xi_1)$ in the simple form

$$p(\xi_1) = \frac{1}{\sqrt{2\pi} h_0} \exp\left(-\frac{1}{2}\left(\frac{\xi_1}{h_0}\right)^2\right) \tag{3.40}$$

and it is easily noted that the distribution is, by construction, a zero-mean Gaussian with variance h_0^2.

If we define the norm of the surface gradient as

$$|\nabla h| = \left(\xi_2^2 + \xi_3^2\right)^{\frac{1}{2}} = \left(\frac{\partial h}{\partial x}^2 + \frac{\partial h}{\partial y}^2\right)^{\frac{1}{2}} \tag{3.41}$$

then it is shown in [LH62] that its probability density reads

$$p(|\nabla h|) = \frac{|\nabla h|}{m_2} \exp\left(-\frac{|\nabla h|^2}{2m_2}\right) \tag{3.42}$$

and it is straightforward to find the mean value of $|\nabla h|$, denoted by $\overline{|\nabla h|}$, as

$$\overline{|\nabla h|} = \int_0^\infty |\nabla h|\, p(|\nabla h|)\, \mathrm{d}|\nabla h| = \left(\frac{\pi m_2}{2}\right)^{\frac{1}{2}} \tag{3.43}$$

Nayak stresses the fact that the mean gradient (i.e. the mean value of the norm of the surface gradient) is different from the mean slope measured along a profile line (the intersection of the rough surface with a vertical plane), which for isotropic surfaces is independent of the profile orientation, and therefore, for directions aligned to the axes, it reads

$$\overline{\frac{\partial h}{\partial x}} = \overline{\xi_2} = \overline{\xi_3} = \overline{\frac{\partial h}{\partial y}} = \left(\frac{2m_2}{\pi}\right)^{\frac{1}{2}} \tag{3.44}$$

and this value is a factor $2/\pi$ smaller than the mean gradient found in (3.43). Notice that, from (3.24) it can be seen that

$$p(\xi_1, \xi_2) = p(\xi_1) p(\xi_2) \tag{3.45}$$

and

$$p(\xi_1, \xi_3) = p(\xi_1)p(\xi_3) \tag{3.46}$$

which means that slope and height are statistically independent.

The last distribution we shall report in this summary is that of summit heights. A summit is a local maximum of the corrugated surface, thus, for us, it will represent a possible initial point of contact. To represent a local maximum, the point (x, y) must fulfill the following conditions:

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial y} = 0, \tag{3.47}$$

$$\frac{\partial^2 h}{\partial x^2} < 0, \tag{3.48}$$

$$\frac{\partial^2 h}{\partial y^2} < 0, \tag{3.49}$$

$$\frac{\partial^2 h}{\partial x^2}\frac{\partial^2 h}{\partial y^2} - \left(\frac{\partial^2 h}{\partial x y}\right)^2 \geq 0 \tag{3.50}$$

which can be rewritten according to (3.29) as follows:

$$\xi_2 = \xi_3 = 0, \tag{3.51}$$

$$\xi_4 < 0, \tag{3.52}$$

$$\xi_6 < 0, \tag{3.53}$$

$$\xi_4\xi_6 - \xi_5^2 \geq 0 \tag{3.54}$$

If we look in the infinitesimal area $dA = dxdy$, the probability that the variables ξ_i at (x, y) lie in the range $(\xi_1, \xi_1 + \mathrm{d}\xi_i)$ is $p(\xi_1, \ldots, \xi_6)\,\mathrm{d}\xi_1 \cdots \mathrm{d}\xi_6$. The increments $\mathrm{d}\xi_2$ and $\mathrm{d}\xi_3$ that take place in an area dA are given by

$$\mathrm{d}\xi_2\,\mathrm{d}\xi_3 = J\begin{pmatrix}\xi_2 & \xi_3\\ x & y\end{pmatrix} dA \tag{3.55}$$

where the Jacobian, denoted by $J(\cdot)$, has the value

$$J\begin{pmatrix}\xi_2 & \xi_3\\ x & y\end{pmatrix} = \begin{Vmatrix}\dfrac{\partial \xi_2}{\partial x} & \dfrac{\partial \xi_2}{\partial y}\\[2ex] \dfrac{\partial \xi_3}{\partial x} & \dfrac{\partial \xi_3}{\partial y}\end{Vmatrix} = \begin{Vmatrix}\xi_4 & \xi_5\\ \xi_5 & \xi_6\end{Vmatrix} = \xi_4\xi_6 - \xi_5^2 \tag{3.56}$$

Therefore a point (x, y) will be a summit of height between ξ_1 and $\xi_1 + \mathrm{d}\xi$ if ξ_2 satisfy (3.51) and (3.55), and ξ_i with $i = 4, 5, 6$ take on arbitrary values fulfilling (3.52), (3.53), and (3.54). The probability of the existence of such a point in the area dA is

$$P_{\mathrm{sum}}(\xi_1)dA\,\mathrm{d}\xi_1 = \mathrm{d}\xi_1 \iiint_V p(\xi_1, 0, 0, \xi_4, \xi_5, \xi_6)\,\mathrm{d}\xi_2\,\mathrm{d}\xi_3\,\mathrm{d}\xi_4\,\mathrm{d}\xi_5\,\mathrm{d}\xi_6, \tag{3.57}$$

where P_{sum} is the probability distribution for summits of height ξ_1, and the domain of integration V is defined by (3.52), (3.53), and (3.54). We can rewrite (3.57) using (3.55) and (3.56), substituting for $p(\xi_1,\ldots,\xi_6)$ from (3.37), obtaining

$$P_{\text{sum}}(\xi_1) = \frac{\exp\left(-m_4\xi_1^2/\Delta_1\right)}{(2\pi)^3\sqrt{\Delta}} \iiint_V \left|\xi_4\xi_6 - \xi_5^2\right| \exp(\Xi)\ \mathrm{d}\xi_4\,\mathrm{d}\xi_5\,\mathrm{d}\xi_6 \tag{3.58}$$

with

$$\Xi = -\frac{1}{2}\left[\frac{9\Delta_2(\xi_4^2+\xi_6^2)}{4m_4\Delta_1} + \frac{3\xi_5^2}{m_4} + \frac{3m_2\xi_1(\xi_4+\xi_6)}{\Delta_1} - \frac{3\Delta_3\xi_4\xi_6}{2m_4\Delta_1}\right]. \tag{3.59}$$

Introducing the new variables

$$\begin{aligned} t_1 &= \left(\frac{3}{m_4}\right)^{\frac{1}{2}} \frac{1}{2}(\xi_4+\xi_6), \\ t_2 &= \left(\frac{3}{m_4}\right)^{\frac{1}{2}} \xi_5, \\ t_3 &= \left(\frac{3}{m_4}\right)^{\frac{1}{2}} \frac{1}{2}(\xi_4-\xi_6) \end{aligned} \tag{3.60}$$

and normalizing ξ_1 as follows:

$$\xi^* = \frac{\xi_1}{\sqrt{m_0}} = \frac{\xi_1}{h_0} \tag{3.61}$$

one can write $P_{\text{sum}}(\xi^*)$ as

$$\begin{aligned} P_{\text{sum}}(\xi^*) = {} & \frac{m_4}{m_2}\frac{\sqrt{C_1(\alpha)}}{3(2\pi)^3}\exp(-C_1\xi^{*2})\times \\ & \times \iiint_V{}' \left|t_1^2 - t_2^2 - t_3^2\right| \exp\left[-\frac{1}{2}\left(C_1t_1^2 + t_2^2 + t_3^2 + C_2t_1\xi^*\right)\right]\ \mathrm{d}t_1\,\mathrm{d}t_2\,\mathrm{d}t_3, \end{aligned} \tag{3.62}$$

with

$$\alpha = \frac{m_0m_4}{m_2^2} \tag{3.63}$$

$$C_1 = \frac{\alpha}{2\alpha-3} \tag{3.64}$$

$$C_2 = C_1\sqrt{\frac{12}{\alpha}} \tag{3.65}$$

and the domain of integration V' defined as follows

$$t_1 < 0$$
$$t_2^2 + t_3^2 \leq t_1^2.$$

One can define D_{sum} as the density of summits per unit of area, and it can be computed as

$$\int_{-\infty}^{\infty} P_{\text{sum}}(\xi^*)\, \mathrm{d}\xi^*. \tag{3.66}$$

Notice that, since the scaling of ξ_1 performed to obtain ξ^* only affects heights, it does not influence D_{sum}, which is independent of ξ_1. The integrals of (3.62) and (3.66) can be evaluated analytically obtaining the following closed form for D_{sum}:

$$\frac{1}{6\pi\sqrt{3}} \frac{m_4}{m_2} \tag{3.67}$$

Using $P_{\text{sum}}(\xi^*)$ and D_{sum}, we can define the probability density for summit heights, denoted by $p_{\text{sum}}(\xi^*)$, defined as

$$p_{\text{sum}}(\xi^*) = \frac{P_{\text{sum}}(\xi^*)}{D_{\text{sum}}} \tag{3.68}$$

which can be written as

$$p_{\text{sum}}(\xi^*) = \frac{\sqrt{3}}{2\pi}(\Pi_1 + \Pi_2 + \Pi_3) \tag{3.69}$$

where

$$\Pi_1 = \exp\left(-C_1 \xi^{*2}\right) \left[\frac{3(2\alpha - 3)}{\alpha^2}\right]^{\frac{1}{2}} \xi^*, \tag{3.70}$$

$$\Pi_2 = \frac{3\sqrt{2\pi}}{2\alpha} \exp\left(-\frac{1}{2}\xi^{*2}\right) \left[1 + \operatorname{erf}(\beta)\right] (\xi^{*2} - 1), \tag{3.71}$$

$$\Pi_3 = \sqrt{2\pi} \left[\frac{\alpha}{3(\alpha - 1)}\right]^{\frac{1}{2}} \exp\left\{-\left[\left(\frac{\alpha \xi^{*2}}{2(\alpha - 1)}\right)\right]\right\} \left[1 + \operatorname{erf}(\gamma)\right] \tag{3.72}$$

with

$$\beta = \left[\frac{3}{2(2\alpha - 3)}\right]^{\frac{1}{2}} \xi^* \tag{3.73}$$

$$\gamma = \left[\frac{\alpha}{2(\alpha - 1)(2\alpha - 3)}\right]^{\frac{1}{2}} \xi^*, \tag{3.74}$$

where $\operatorname{erf}(\cdot)$ denotes the error function, defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, \mathrm{d}t. \tag{3.75}$$

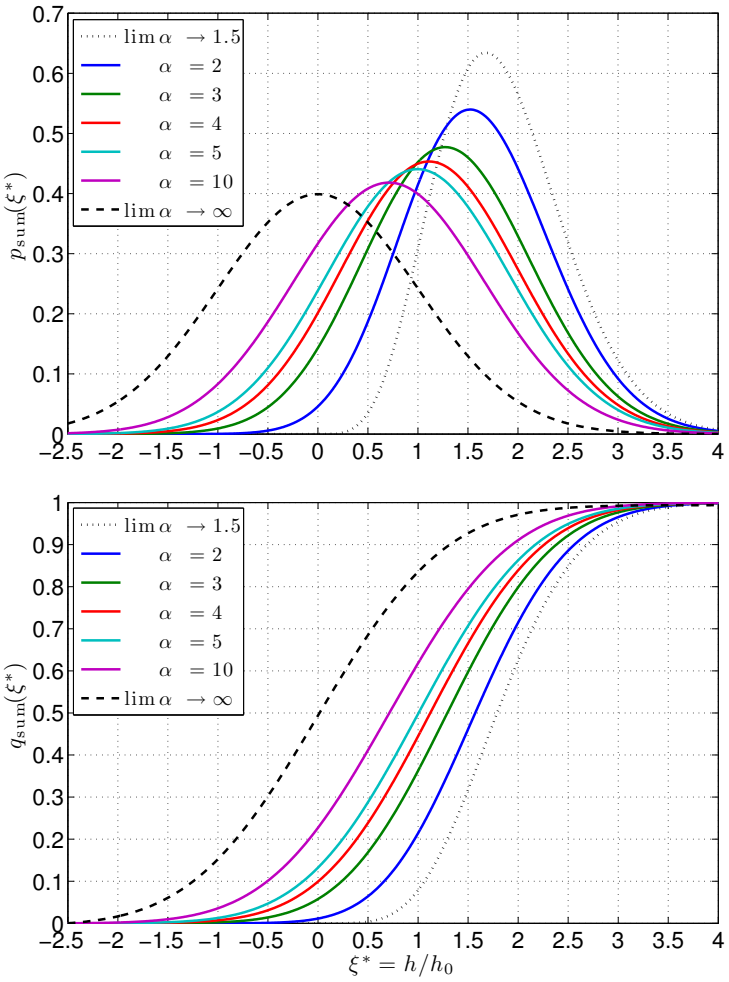

Figure 3.1. Summit distribution for different values of α, with limit cases $\alpha \to 1.5$ and $\alpha \to \infty$. *Top*: Probability density of summit heights $p_{\mathrm{sum}}(\xi^*)$. *Bottom*: Cumlative density of summit heights $q_{\mathrm{sum}}(\xi^*)$.

A remarkable property of $p_{\mathrm{sum}}(\xi^*)$ is its dependence on one only parameter, namely α. This parameter is an indicator of the breadth of the PSD: it is larger for broader spectra, and smaller for narrower ones. Longuet-Higgins has shown in [LH57a] that for any random surface, $\alpha \geq 3/2$, whereas for infinitely broad spectra, α tends to infinity. In Figure 3.1 distributions of summit heights are plotted for different values of α, including the limits $\alpha \to 1.5$ and $\alpha \to \infty$ for which the limits of $p_{\mathrm{sum}}(\xi^*)$ can be evaluated,

obtaining the two following results:

$$\lim_{\alpha\to 1.5} p(\xi^*) = \begin{cases} \frac{2\sqrt{3}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^*\right)\left[\xi^{*2} - 1 + \exp\left(-\xi^{*2}\right)\right], & \xi^* \geq 0 \\ 0, & \xi^* < 0 \end{cases} \tag{3.76}$$

and

$$\lim_{\alpha\to\infty} p(\xi^*) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^{*2}\right) \tag{3.77}$$

It is easy to see that in the case $\alpha \to \infty$, $p_{\text{sum}}(\xi^*)$ follows a perfect normal distribution. In the case $\alpha \to 1.5$, instead, $p_{\text{sum}}(\xi^*)$ is not Gaussian anymore and is biased towards high summits.

The importance of this result is twofold. First, with a single parameter, namely α, the summits of a surface can be characterized, and this parameter, for real isotropic surfaces, can be computed using a surface profile (whereas for anisotropic surfaces, three non parallel profiles are sufficient). Second, for computer generated surfaces, where a desired PSD is imposed (see Section 4.1), the predicted distribution of p_{sum} can be checked, as a measure of quality of the generating algorithm. Moreover, we will see in Section 3.5 that α has a strong influence on the variance of the results which are obtained in rough contact simulations.

It is possible to compute the cumulative density of summits $q_{\text{sum}}(\xi^*)$, defined as

$$q_{\text{sum}}(\xi^*) = \int_{-\infty}^{\xi^*} p_{\text{sum}}(\xi')\,\mathrm{d}\xi'. \tag{3.78}$$

The value of q_{sum} represent the fraction of the total summits for which $h/h_0 \leq \xi^*$. In Figure 3.1 we show $q_{\text{sum}}(\xi^*)$ for the same values of α used for $p_{\text{sum}}(\xi^*)$.

We omit other results derived by Nayak, because they are not relevant for our study. The interested reader is referred to [Nay71] for estimates on mean curvature distribution and techniques to derive α from measured surface profiles.

3.2 Realistic self-affine rough surfaces

As mentioned in Section 1.1, roughness is a natural concept, which anyone is aware of and it is perceived as the local deviation of the surface heights from flatness, c.f. [Per01b]. In other words, the more the points on a surface are distant from the mean plane, the rougher we perceive the surface. But this single measure, which is well represented by the root mean square roughness h_0, defined in 3.1, is not sufficient to mathematically describe roughness. The easiest way to understand the failure of such single measure, is to think of one surface with a known h_0, and the same surface, shrunk with respect to x and y directions, but kept constant in z direction: the resulting surface will be

perceived as rougher (for example after a tactile or visual inspection), but h_0 will actually be the same, as it is only a measure of the standard deviation of $h(\mathbf{x})$ from its mean value, thus, it is only measured with respect to z.

In this section we collect the basic knowledge about the mathematical description of real surfaces at micro-scale, their fractal nature, and the experimental evidence supporting it. The quantity on which we focus is the PSD, which was introduced in Section 3.1, since it is a commonly accepted way to represent rough surfaces in a statistical sense.

3.2.1 Self-affinity and randomness of real surfaces

The best known scientist who strongly claimed that natural structures are fractals, is surely Benoît Mandelbrot, who introduced the concepts of self-similarity and fractional dimension in [Man67], and, later, the term *fractal* itself. As it is now well known, a fractal is a mathematical or geometrical entity showing self-repeated patterns at every scale, and since a full derivation of their mathematical description would be out of the scope of this work, we refer the interested reader to, for example, [Man83], [BL80], or [BS95]. The only quantity of interest for our exposure is the fractal dimension D_f, for which a good definition is given in [RC07]. Let us assume there exists a geometric fractal with an area (or a length, or a volume) a, and the minimal circle (or sphere) circumscribing it has radius r. If we now analyze a subpart of the fractal which can be circumscribed by a smaller circle, with radius $C_r r$, and this subpart has area $C_a a$, then the fractal dimension is defined as

$$D_f = \frac{\log C_a}{\log C_r}. \tag{3.79}$$

This applies to any n-dimensional fractal, and it is straightforward to understand that larger D_f correspond to higher increase of complexity when observing the fractal at a smaller length scale. Fractals are relevant to this work because rough surfaces are, up to a certain degree, self-affine fractals, c.f. [BS95]. For surfaces obtained by fracture (like rocks) or road surfaces, the fractal dimension $D_f \approx 2.2$, c.f. [Sil96], [Bou97], and [Kru97]. Thus, loosely speaking, when observing a rough surface at different length scales (or different *zoom* levels), the structures remain topologically the same and show the same complexity. However, real surfaces are not perfect fractals, and the reason for this is found in the discrete nature of matter: it is quite straightforward to imagine that structures stop repeating themselves, at least at atomic scale. Actually, as argued in [PAT+04], the fractal properties cease to exist at larger length scales than the molecular one, the micro-scale being a good approximation of such limit for surfaces with rms roughness $h_0 > 1$ μm. Besides, self-affine rough surfaces are considered realizations of a random process, c.f. [HPMR04] and [PAT+04], for example. Thus, every surface is classified according to the parameters of the random process generating it, and to statistical quantities which are considered to hold on average

for a sufficiently large set of specimens. In other words, it is assumed that statistical properties of rough surfaces which are realizations of the same random process, for a sufficiently large set of surface specimens, tend to an asymptotic mean value. How fast this limit is approached has been the subject of the investigations of [YAM12], where these quantities are analyzed not only in terms of mean values, but also of variance. In the next section, we shall define the quantities which are commonly used to describe roughness, and their relation to the associated PSD.

3.2.2 Shape of the PSD

The proof of the fractal nature of rough surfaces, has been found experimentally, analyzing their PSD $C(q)$, because self-affinity has a direct influence on its shape. As pointed out in Section 3.1, $C(q)$ is proportional to the Fourier transform of the random surface, c.f. (3.21). If we now assume that surfaces are not continuous, but discrete (again, at least at molecular level), we can reason about the coefficients of the Discrete Fourier Transform (DFT) of such self-affine fractals. Since the coefficients of the DFT are the amplitudes of the waves composing the analyzed discrete signal, it is intuitive to think that geometric structures which are repeating themselves at different length scales imply wave amplitudes which appear scaled by a constant factor at different wave-lengths. Given the logarithmic nature of D_f shown in (3.79), one can deduce that a similar relationship holds for the DFT coefficients, and for the values of $C(q)$ too. Thus, the PSD of a self-affine surface should follow a power law of the form

$$C(q) \sim q^{-K} \tag{3.80}$$

with $K > 0$ depending on D_f.

We omit the derivation of K, which can be found in [BS95], and we simply report the result. To this end, let us introduce the *Hurst exponent* H, also called *roughness exponent*, defined as

$$H = 3 - D_f. \tag{3.81}$$

Using H, for the power spectrum of a perfect self-affine surface it holds

$$C(q) \sim q^{-2(H+1)}. \tag{3.82}$$

Many analyses, e.g. [KP95], [CCI99], and [DAB12], confirm that the PSD of common rough surfaces (like rock or metal surfaces) indeed follow such a power law, at least in the observed domain, that is, in the considered $\mathbf{q}$-space. Since we are concerned with DFT, the length L of the analyzed signal (thus, the length of the rough surface specimen) defines the lower bound of the observable wave-vectors domain. The smallest possible wave-number will be denoted by q_L, to emphasize its connection with L, and its value is

$$q_L = \frac{2\pi}{L}. \tag{3.83}$$

Thus, one must ensure that the size of the specimen is large enough to be able to capture all low frequency waves of the surface. Low frequency waves are important to describe those components of the surface which are large in space: for example, for asphalt-like surfaces, one must ensure that the wave-vectors with smallest spatial frequency can capture objects of the size of the biggest visible rocks, which usually, have diameter no larger than 1.5 cm, thus L should be at least 3 cm, c.f. [PAT$^+$04]. On the other side, the upper bound of the wave-vectors space is denoted by q_1, and is usually referred to as *short-distance cut-off* wave number. The connection with the DFT is again determinant: q_1 depends on the sampling frequency of the signal. Let us assume that the distance between two sampled neighboring points (along x- or y-direction) is δ, then we know that the largest wave number that the DFT can capture is

$$q_1 = \frac{2\pi}{\delta}. \tag{3.84}$$

Thus, q_1 is strictly connected to the spatial resolution of the measuring instrument, which should be of the same order of the diameter of the smallest rough detail of the analyzed surface. For road surfaces, for example, it is argued in [PAT$^+$04] that the radius of the smallest dust particles is in the order of 100 μm. Now that we have defined the boundaries of the q domain, we must mention that the power law is not valid everywhere in the interval $[q_L, q_1]$. There exists a third quantity of interest, namely the *long-distance roll-off* wave vector q_0, with $q_L < q_0 < q_1$. The PSD between

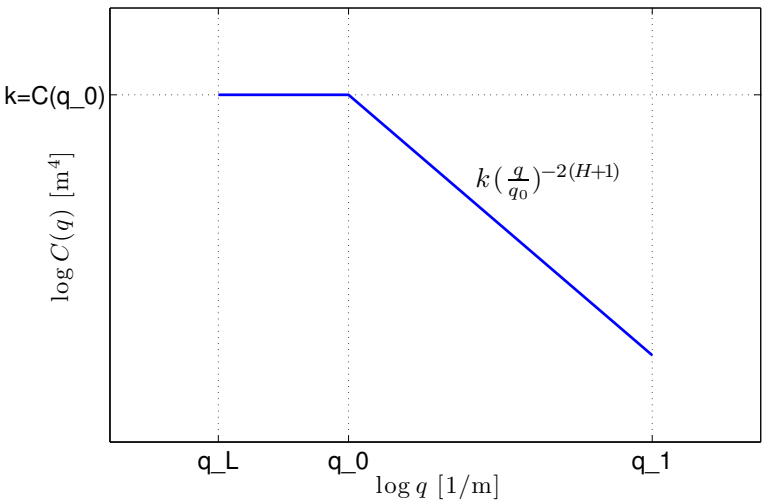

Figure 3.2. Power Spectral Density of a quasi self-affine surface, plotted in full logarithmic scale. In the region between q_L and q_0, $C(q)$ is constant, whereas in the region between q_0 and q_1, it follows an exponential decay.

q_L and q_0 is typically experimentally found to be constant (or only slightly decreasing, but we will consider it constant). The region between q_L and q_0 is often referred to as the (saturation) plateau of the PSD, and the value $C(q_0)$ will be denoted by k. For values of q larger than q_0, C(q) decays exponentially until q_1, as in (3.82).

The above description of the PSD leads to the following function:

$$C(q) = \begin{cases} k, & \text{if } q_L \leq q < q_0, \\ k\left(\dfrac{q}{q_0}\right)^{-2(H+1)}, & \text{if } q_0 \leq q < q_1 \\ 0, & \text{else} \end{cases} \tag{3.85}$$

and its plot is shown in Figure 3.2. Given the relationship between $C(q)$ and the coefficients of the DFT of a given surface, the estimation of the parameters k, H and q_0 is straightforward and in next section, we present the algorithm proposed in [PAT+04], which we implemented to study the surfaces we generate *in silico*.

3.2.3 Experimental determination of $C(q)$

Let us assume we study a surface $h(\mathbf{x})$, extending over the domain $[0,\ L]^2$. The function $h(\mathbf{x})$ is known only point-wise, and we assume that the values we know correspond to the nodes of a square lattice, a grid with uniform spacing δ. Let us further assume we know the coefficients of the DFT of $h(\mathbf{x})$, which can be easily computed by means of a Fast Fourier Transform (FFT), for which efficient parallel implementations are known, c.f. [PA04]. Let $B(\mathbf{q})$ denote the coefficient of the DFT of $h(\mathbf{x})$ associated to the wave-vector $\mathbf{q} = (q_x, q_y)$, then we can write its Fourier series as (see also (3.21))

$$h(x,y) = \sum_{\mathbf{q}} B(\mathbf{q}) \exp\left\{ i \left[q_x x + q_y y + \phi(\mathbf{q}) \right] \right\} \tag{3.86}$$

where i denotes the imaginary unit and $\phi(\mathbf{q})$ is the phase shift angle $\in [0,\ 2\pi[$. Notice that the the components of all possible $\mathbf{q}$-vectors, correspond to the coordinates of another square lattice (in $\mathbf{q}$-space), with lattice constant q_L (the smallest possible wave-number). Thus, all wave-vectors can be written as

$$\mathbf{q} = (mq_L, nq_L) \tag{3.87}$$

where $0 \leq m, n \leq L/\delta$, and $m, n \in \mathbb{N}$, and $B(\mathbf{0}) = 0$. Wave-vectors which have wave-number with $|\mathbf{q}| = q > (L/\delta)q_L = q_1$ are expected to have an associated Fourier coefficient of zero, as they do not belong to the considered wave-vectors domain. Now we recall (3.2) and (3.3) from Section 3.1

$$R(x,y) = \lim_{L\to\infty} \frac{1}{L^2} \int\!\!\int_{-L/2}^{+L/2} h(x_1,y_1)\, h(x_1+x, y_1+y)\, \mathrm{d}x_1 \mathrm{d}y_1$$

$$C(q_x,q_y) = \frac{1}{4\pi^2} \int\!\!\int_{-\infty}^{+\infty} R(x,y) \exp\left[-i(xq_x + yq_y) \right] \mathrm{d}x\mathrm{d}y;$$

Substituting the expression for $R(x, y)$ into the second equation, and then introducing (3.86) we obtain, after some calculations

$$C(\mathbf{q}) = \frac{1}{4\pi^2} \iint \mathrm{d}x\mathrm{d}y \sum_{q'} |B(\mathbf{q})|^2 \exp\left\{ i \left[\left(q_x - q'_x \right) x + \left(q_y - q'_y \right) y \right] \right\} \tag{3.88}$$

and in the limit, for infinite (or sufficiently) many $\mathbf{q}$-vectors, we obtain

$$C(\mathbf{q}) = \frac{L^2}{4\pi^2} |B(\mathbf{q})|^2 . \tag{3.89}$$

Now that we obtained the 2D coefficients of $C(\mathbf{q})$, this can be reduced to the monodimensional function of (3.85) if the surface is isotropic. In such case, we assume that the power spectrum depends only on the wave number q, thus

$$C(\mathbf{q}) = C(\|\mathbf{q}\|) = C(q) \tag{3.90}$$

and we have to radially average the values of $C(\mathbf{q})$. By radial averaging, it is meant that we collect the coefficients associated to any wave vector, and we group them by their norm according to which wave number $q' = mq_L$, with $m \in \{1, 2, \ldots, L/a\}$ which is the closest. Formally, we will create L/a sets S_q, and distribute the coefficients such that:

$$C(\mathbf{q}) \in S_{q'} \Leftrightarrow q' = \arg\min_q \left| \sqrt{q_x^2 + q_y^2} - q \right| . \tag{3.91}$$

Once that every coefficient is assigned to a set, for each value of q' we take the mean value of the coefficients in $S_{q'}$ as the value of $C(q')$. Formally:

$$C(q') = \frac{1}{N_{q'}} \sum_q C(q), \quad \text{with } q \in S_{q'} \tag{3.92}$$

where $N_{q'}$ denotes cardinality of $S_{q'}$.

The resulting PSD could have to be smoothed, since its coefficients could suffer from noise of the measurements, or from the particular geometry of the specimen, c.f. [PAT+04].

Given the computed values of $C(q)$, it is easy to compute k, as the value of the PSD in the saturation plateau and h_0 with (3.10). The short-distance roll-off vector q_0 can be estimated by finding where $C(q)$ starts to decrease exponentially, and the Hurst exponent can then be computed by fitting the power law of (3.85).

Notice that the inverse of the procedure we have just illustrated, can be used to generate rough surfaces starting from a desired PSD, and it is how we generate the datasets employed in our experiments, as we describe in Section 4.1.

3.2.4 Limitations of the description

The description of surfaces with a mono-dimensional power spectral density assumes isotropy of the surfaces themselves. When isotropy does not hold, it is possible to use the two-dimensional PSD, or to compare the PSD's of surface profiles obtained along different directions, as it is described in [Nay71].

Moreover, we remark that the PSD description also implies homogeneity of the surface, that is, the same statistical properties hold in every point $\mathbf{x}$. This is true for many natural surfaces, as for others it is not. Classic examples are wooden surfaces, where depending on the point, that is, on the different cells of the wood which are exposed on the surface, the roughness can be highly different, c.f. [KHB06], [Lei11]. If wooden surfaces have characteristics varying according to $\mathbf{x}$, polishing a surface obtained by fracture will result in a PSD which depends on the height, because polishing will flatten high areas and summits, while leaving lower areas rougher, as indicated in [PAT+04]. In all this cases, different PSD can be adopted to describe the different roughness typologies, but when deriving contact theories from the power spectral information, the inhomogeneity of $C(q)$ is almost never taken into account.

3.3 Bush-Gibson-Thomas's model

A few years after the paper by Nayak summarized in Section 3.1, Bush, Gibson, and Thomas published [BGT75], in which they make use of a part of Nayak's work to derive an *asperity-based* model of rough contact. By "asperity-based" it is meant that the model starts by assuming that surface summits lie on top of geometrically defined asperities. In the case of the Bush-Gibson-Thomas (BGT) model, asperities are modeled as paraboloids, which is a slightly more accurate approximation than the one used by Nayak in [Nay73], which was based on spherical asperities (and because of the similarity with the work presented in this section, we omit it in this work). From the statistical assumptions made on the distribution of summits, points of initial contact are identified, and the contact area is predicted according to the acting load, as in a Hertzian contact (see Section 2.3). The result they found is a linear correlation between load and area of contact, which is still used nowadays as a mean of comparison for more recent theories, such as those proposed in [Per01b] and [YAM14], which we summarize in Section 3.4 and Section 3.5.

3.3.1 Asperity cap modeling

In the BGT model, the cap of each asperity is approximated with a paraboloid having the same height and curvature as the summit of the asperity (for which the distributions can be found in [Nay71]). The paraboloids can be described in implicit form

as:

$$\frac{z_0 - z}{z_0 - h'} = \frac{\left[(x - x_0)\cos\beta + (y - y_0)\right]^2}{a^2} + \frac{\left[(x - x_0)\sin\beta - (y - y_0)\cos\beta\right]^2}{b^2} \tag{3.93}$$

where (x_0, y_0, z_0) are the coordinates of the summit, with $z_0 = h(x_0, y_0)$, β is the angle formed by one of its principal radii of curvature and the x-axis, and a, b, and h' are constants. An example paraboloid is shown in Figure 3.3. From (3.93), one can derive expressions for the second space derivatives of the paraboloids, obtaining

$$\begin{aligned}
\xi_4 &= \frac{\partial^2 h}{\partial x^2} = -2(z_0 - h')\left(\frac{\cos^2\beta}{a^2} + \frac{\sin^2\beta}{b^2}\right) \\
\xi_5 &= \frac{\partial^2 h}{\partial x \partial y} = -2(z_0 - h')\left(\frac{1}{a^2} - \frac{1}{b^2}\right)\sin\beta\cos\beta \\
\xi_6 &= \frac{\partial^2 h}{\partial y^2} = -2(z_0 - h')\left(\frac{\sin^2\beta}{a^2} + \frac{\cos^2\beta}{b^2}\right)
\end{aligned} \tag{3.94}$$

A plane parallel to Oxy with elevation $z = h'$ intersects the paraboloid in an elliptic boundary which possesses the implicit form

$$\frac{\left[(x - x_0)\cos\beta + (y - y_0)\right]^2}{a^2} + \frac{\left[(x - x_0)\sin\beta - (y - y_0)\cos\beta\right]^2}{b^2} = 1 \tag{3.95}$$

and the ellipses has semi-axes of length a and b, forming angles of β and $\beta + \pi/2$ with the x-axis respectively, as shown in Figure 3.3.

3.3.2 Probability distribution of asperity density

We can define $p_h(a, b, z_0)$, that is, the probability distribution of the density (per unit area) of summits with height z_0 intersecting the horizontal plane with $z = h'$ in an ellipse with semi-axes a and b:

$$p_h(a, b, z_0) = \int_0^{\frac{\pi}{2}} p\left(z_1, 0, 0, \xi_4, \xi_5, \xi_6\right) J\begin{pmatrix}\xi_2 & \xi_3 \\ x & y\end{pmatrix} J\begin{pmatrix}\xi_4 & \xi_5 & \xi_6 \\ a & b & \beta\end{pmatrix} d\beta \tag{3.96}$$

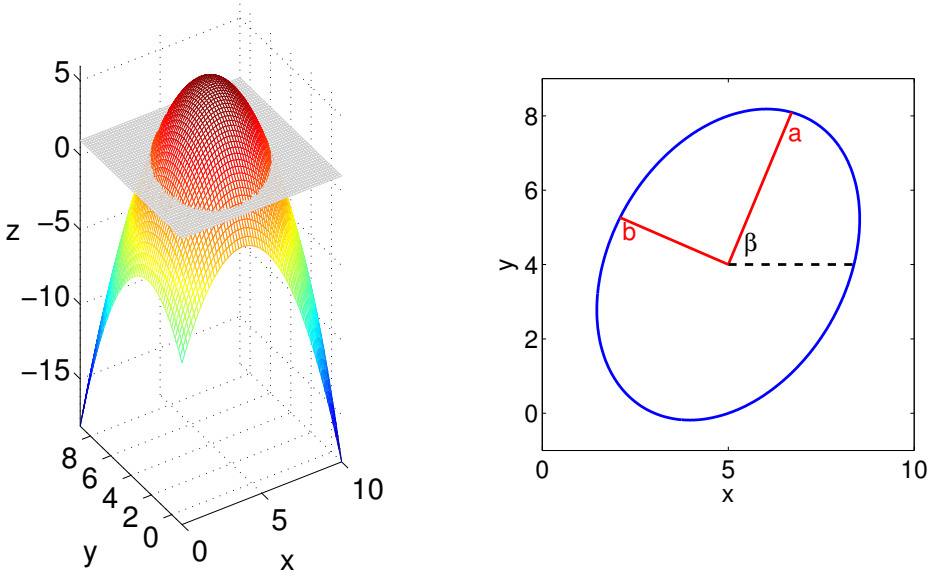

Figure 3.3. Left: A paraboloid with $(x_0, y_0, z_0) = (5, 4, 6)$, $a^2 = 20$, $b^2 = 10$, $h' = 1$ and a horizontal plane with $z = h'$. *Right*: The intersection of the paraboloid with the horizontal plane, with the two semi-axes shown in red.

where the Jacobians can be computed using (3.94), obtaining

$$J\begin{pmatrix}\xi_2 & \xi_3\\ x & y\end{pmatrix} = \begin{Vmatrix}\frac{\partial \xi_2}{\partial x} & \frac{\partial \xi_2}{\partial y}\\ \frac{\partial \xi_3}{\partial x} & \frac{\partial \xi_3}{\partial y}\end{Vmatrix} = \begin{Vmatrix}\xi_4 & \xi_5\\ \xi_5 & \xi_6\end{Vmatrix} = \frac{4(z_0 - h')^2}{a^2 b^2} \tag{3.97}$$

$$J\begin{pmatrix}\xi_4 & \xi_5 & \xi_6\\ a & b & \beta\end{pmatrix} = \begin{Vmatrix}\frac{\partial \xi_4}{\partial a} & \frac{\partial \xi_4}{\partial b} & \frac{\partial \xi_4}{\partial \beta}\\ \frac{\partial \xi_5}{\partial a} & \frac{\partial \xi_5}{\partial b} & \frac{\partial \xi_5}{\partial \beta}\\ \frac{\partial \xi_6}{\partial a} & \frac{\partial \xi_6}{\partial b} & \frac{\partial \xi_6}{\partial \beta}\end{Vmatrix} = \frac{32(z_0 - h')^3 \left|a^2 - b^2\right|}{a^5 b^5} \tag{3.98}$$

and therefore, using (3.24), (3.97), and (3.98), it is possible to analytically solve (3.96), obtaining

$$p_h(a, b, z_0) = \frac{12\sqrt{3}\,(z_0 - h')^5 \left|b^2 - a^2\right|}{m_2^2 m_4 \sqrt{2\alpha - 3}\,\pi^2 a^7 b^7} \exp\left[-\frac{\alpha z_0^2}{(2\alpha - 3)m_0} + \frac{6(z_0 - h')^2}{m_4 a^2 b^2}\right.$$
$$\left.\frac{3z_0(z_0 - h')}{m_2(2\alpha - 3)}\left(\frac{1}{a^2} + \frac{1}{b^2}\right) - \frac{9(\alpha - 1)(z_0 - h')^2}{2m_4(2\alpha - 3)}\left(\frac{1}{a^2} + \frac{1}{b^2}\right)^2\right] \tag{3.99}$$

3.3.3 Bearing area

Now we consider the geometric problem of finding the area of intersection between the rough surface and a plane at an elevation h' (remembering that the mean plane of the rough surface is Oxy). This is equivalent to finding the area fraction $A(h')$ occupied by all points of the surface with $h(x,y) \geq h'$, called cross-sectional area, or *bearing area*. Thus, starting from $q_{\text{sum}}(\xi)$, which is the non-normalized version of (3.78), one can derive

$$A(h') = 1 - q_{\text{sum}}(h') = \int_{h'}^{\infty} p(\xi_1)\,\mathrm{d}\xi_1 = \frac{1}{2}\operatorname{erfc}\left(\frac{h'}{\sqrt{2m_0}}\right) \tag{3.100}$$

where $\operatorname{erfc}(x)$ denotes the complementary error function defined as

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_{x}^{\infty} \exp(-t^2)\,\mathrm{d}t \tag{3.101}$$

It is also possible to compute $A_e(h')$, that is the bearing area fraction for the surface where summits are substituted with paraboloids, which is also called the approximated bearing area. Knowing that the area of an ellipse is πab, and making use of (3.99), one obtains

$$A_e(h') = \int_{h'}^{\infty}\int_{0}^{\infty}\int_{0}^{\infty} \pi ab\, p_h(a,b,z_0)\,\mathrm{d}b\,\mathrm{d}a\,\mathrm{d}z_0 \tag{3.102}$$

which can be rewritten as

$$A_e(h') = \frac{\alpha\sqrt{3}}{2\pi\sqrt{2\alpha-3}}\int_{0}^{\infty}\int_{0}^{\infty} u\exp\left[-\frac{(t+u)^2}{2} - \frac{3\alpha}{2(2\alpha-3)}\left(v - \frac{t+u}{\sqrt{\alpha}}\right)^2\right] f(v)\,\mathrm{d}v\,\mathrm{d}u \tag{3.103}$$

where we introduced

$$f(v) = v - \sqrt{\frac{2}{3}}\exp\left(-\frac{3v^2}{2}\right)\operatorname{Erfi}\left(\sqrt{\frac{3}{2}}v\right) \tag{3.104}$$

with non-dimensional separation t defined as

$$t = \frac{h'}{h_0} \tag{3.105}$$

and imaginary error function (or generalized error function) $\operatorname{Erfi}(x)$

$$\operatorname{Erfi}(x) = \int_{0}^{x} \exp\left(t^2\right)\,\mathrm{d}t. \tag{3.106}$$

It is not possible to derive an analytic solution of (3.103), but for some special cases, a good approximation can be found. For $t \to 0$, the approximation is obtained by

expanding the first three terms of the Taylor series of $A_e(t)$ around $t=0$, which results in

$$\begin{aligned}A_e(t) = &\frac{1}{6\pi}\left[\sqrt{3(2\alpha-3)+2\pi} - 2\tan^{-1}\left(\sqrt{\frac{2\alpha-3}{3}}\right) - \sqrt{\alpha-1}\tanh^{-1}\left(\sqrt{\frac{2\alpha-3}{3(\alpha-1)}}\right)\right]\\ &- \frac{t}{2}\sqrt{\frac{\alpha}{3\pi}}\left\{1+\sqrt{\frac{3}{2\alpha}} - \frac{1}{\sqrt{6}}\left[\log_e\left(\sqrt{3}+\sqrt{2}\right) + \frac{1}{2}\log_2\left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)\right]\right\}\\ &+ \frac{\alpha t^2}{12\pi\sqrt{\alpha-1}}\left\{\frac{\sqrt{3(\alpha-1)(2\alpha-3)}}{\alpha} - \log_e\left[\frac{\sqrt{3(\alpha-1)}+\sqrt{2\alpha-3}}{\sqrt{\alpha}}\right]\right\}\\ &+ \frac{(2\alpha-3)t^3}{36(\alpha-1)\sqrt{2\pi}} + O(t^4). \end{aligned} \tag{3.107}$$

For large t, instead, first one must notice that the integrand of (3.103) is exponentially small apart from the neighborhood of

$$v = \frac{t+u}{\sqrt{\alpha}} \gg 1 \tag{3.108}$$

whereas for large values of v, for (3.104) it holds

$$f(v) = v + O\left(v^{-1}\right) \tag{3.109}$$

therefore, working to leading order of approximation, (3.103) can be rewritten as

$$\begin{aligned}A_e(t) &= \frac{\alpha}{2\pi}\sqrt{\frac{3}{2\alpha-3}}\int_0^\infty u\exp\left[-\frac{(t+u)^2}{2}\right]\int_{-\infty}^\infty v\exp\left[\frac{-3\alpha}{2(2\alpha-3)}\left(v-\frac{t+u}{\sqrt{\alpha}}\right)^2\right]\,\mathrm{d}v\,\mathrm{d}u\\ &= \frac{1}{\sqrt{2\pi}}\int_0^\infty u(t+u)\exp\left[-\frac{(t+u)^2}{2}\right]\,\mathrm{d}u \end{aligned} \tag{3.110}$$

Finally we can apply Laplace's method ([Lap86]) and obtain

$$\frac{1}{\sqrt{2\pi}}\int_0^\infty \exp\left[-\frac{(t+u)^2}{2}\right]\mathrm{d}u = \frac{1}{\sqrt{2\pi}t}\exp\left(-\frac{t^2}{2}\right) \tag{3.111}$$

It is also possible to work to second order of approximation, in that case:

$$A_e(t) = \frac{\exp\left(-\frac{t^2}{2}\right)}{\sqrt{2\pi}t}\left(1-\frac{\alpha+3}{3t^2}+O\left(t^{-4}\right)\right) \tag{3.112}$$

3.3.4 Precision of bearing area fraction $A_e(t)$

The accuracy of the approximation can be studied by means of the ratio $A_e(t)/A(t)$, which tends to 1, for large values of t, where (3.100) becomes

$$A(t) = \frac{\exp\left(-\frac{t^2}{2}\right)}{\sqrt{2\pi}t}\left(1 - t^{-2} + O(t^{-4})\right). \tag{3.113}$$

For large values of α, the model loses validity. In fact, for large values of t, one can estimate the relative error as

$$\frac{A_e(t) - A(t)}{A(t)} = -\frac{\alpha}{3t^2} \tag{3.114}$$

which means that for a desired tolerance ϵ we can find the following bound for t:

$$t > \sqrt{\frac{\alpha}{3\epsilon}} \tag{3.115}$$

Notice that for typical realistic surfaces, $\alpha > 50$, which means that, to reach 10% tolerance, the model would be valid only for the critical separation $t > 12$, which basically means no contact at all, as one can simply compute via $A(t)$. Moreover, for for the range $\alpha \gg 1$ and $t \ll \sqrt{\alpha}$, the area predicted by the model is larger than $A(t)$.

Despite the lack of accuracy at small separations, the model gives an estimate of the area of contact and of its dependence on a load acting normally to push a flat and a corrugated half-space together. To derive this dependency, Hertzian theory (presented in Section 2.3) is used, in combination with the results obtained for $A_e(t)$.

3.3.5 Real area of contact through Hertzian contact

Assuming that every asperity acts independently, each contact with the plane can be modeled as a special case of Hertzian contact.

To be consistent with the notation of Section 2.3, we define the relative radii of curvature of the paraboloid-plane system, and since for the plane it holds

$$R'_{\text{plane}} = R''_{\text{plane}} = \infty \tag{3.116}$$

and the equivalent radii of curvature of this system can easily be derived as done in Appendix A. For the sake of clarity, the semi-axes of the contact ellipse will be denoted by a_c and b_c, keeping a and b as the horizontal semi-axes of the paraboloid replacing the asperity.

We recall (2.53), which is concerned with the displacements of two points in contact S_1 and S_2 belonging to the first (in our case the paraboloid) and second body

(the plane), respectively, and aligned with respect to x and y. For them, knowing the relative radii of curvature, one can write

$$u_{z1} + u_{z2} = d - \frac{x^2}{2R'} + \frac{y^2}{2R''} \tag{3.117}$$

where u_{z1} and u_{z2} are the displacements of S_1 and S_2, and d is the sum of the displacements of the points on which we apply the load (denoted by T_1 and T_2 in Section 2.3).

Following the theory of Hertzian contact, one gets to the system of equations (2.60), which, for the following derivation, has to be written in polar coordinates and becomes

$$\begin{aligned} d &= \frac{\pi\sigma_0 a_c b_c}{E^*} \int_0^{\frac{\pi}{2}} \frac{\mathrm{d}\psi}{\sqrt{a_c^2 \sin^2\psi + b_c^2 \cos^2\psi}} \\ \frac{1}{2R'} &= \frac{\pi\sigma_0 a_c b_c}{E^*} \int_0^{\frac{\pi}{2}} \frac{\sin^2\psi \, \mathrm{d}\psi}{\sqrt{a_c^2 \sin^2\psi + b_c^2 \cos^2\psi}^3} \\ \frac{1}{2R''} &= \frac{\pi\sigma_0 a_c b_c}{E^*} \int_0^{\frac{\pi}{2}} \frac{\cos^2\psi \, \mathrm{d}\psi}{\sqrt{a_c^2 \sin^2\psi + b_c^2 \cos^2\psi}^3}. \end{aligned} \tag{3.118}$$

We recall from (2.63) that the total load P is

$$P = \frac{2}{3}\pi\sigma_0 a_c b_c.$$

whereas for the paraboloid of (3.95), it holds:

$$d = z_1 - h, a^2 = 2\mu R', b^2 = 2\mu R'' \tag{3.119}$$

and therefore, using these results with (3.118), we can write

$$\frac{a^2}{b^2} = \frac{\displaystyle\int_0^{\frac{\pi}{2}} \frac{\cos^2\psi \, \mathrm{d}\psi}{\sqrt{a_c^2 \sin^2\psi + b_c^2 \cos^2\psi}^3}}{\displaystyle\int_0^{\frac{\pi}{2}} \frac{\sin^2\psi \, \mathrm{d}\psi}{\sqrt{a_c^2 \sin^2\psi + b_c^2 \cos^2\psi}^3}} \tag{3.120}$$

and

$$\frac{a_c^2}{a^2} + \frac{b_c^2}{b^2} = 1. \tag{3.121}$$

Now we follow the path used to derive the bearing area fraction $A_e(t)$ and we write, for the total contact area fraction $A_c(t)$

$$A_c(t) = \int_h^\infty \int_0^\infty \int_0^\infty \pi a_c b_c p_h(a, b, z_1) \, \mathrm{d}a \, \mathrm{d}b \, \mathrm{d}z_1 \tag{3.122}$$

that is, the contribution of every contact ellipse is summed to obtain the fraction of the total area which is in contact. Since (3.120) and (3.121) are in implicit form, some changes of variables and some algebraic calculations are needed to obtain an expression of $A_c(t)$ in a form which is suitable for calculations.We omit such derivation for the sake of compactness, and we report the final result, which is:

$$A_c(t) = \tag{3.123}$$

$$\frac{12\alpha}{\pi}\sqrt{\frac{3}{2\alpha-3}}\exp\left(-\frac{\alpha t^2}{2\alpha-3}\right)\int_0^{\frac{\pi}{2}}\int_{\psi=0}^{\frac{\pi}{2}}\int_{\Theta=0}^{\frac{\pi}{4}}\cos\phi\sin^3\phi\, f(\theta,\Theta)\, g(\phi,\Theta)\,\frac{\mathrm{d}\theta}{\mathrm{d}\Theta}\,\mathrm{d}\phi \tag{3.124}$$

where

$$\tan\theta = \frac{a}{b}, \quad \tan\Theta = \frac{a_c}{a_b}, \quad \tan\gamma = \sqrt{\frac{m_0}{m_4}}\sqrt{\frac{1}{a^2}+\frac{1}{b^2}} \tag{3.125}$$

and $f(\theta,\Theta)$, $g(\gamma,\theta)$ are functions which aid in maintaining the expression compact, and can be found in the original paper [BGT75].

Given the already mentioned limitations of the applicability of the theory, which only holds for large separation values, it makes sense to derive an approximation of (3.123) for $t \gg 1$. This can be done again with Laplace's method, obtaining

$$A_c(t) = \frac{\exp\left(\frac{1}{2}t^2\right)}{2\sqrt{2\pi}t} = \frac{1}{2}A(t) \tag{3.126}$$

the mean pressure $P(t)$, instead, can be computed using equations (2.63) and (3.118.1), obtaining

$$P(t) = \frac{2}{3}\pi E^* \int_h^\infty \int_0^\infty \int_0^\infty \frac{(z_1-h)p_h(a,b,z_1)}{\displaystyle\int_0^{\frac{\pi}{2}} \frac{\mathrm{d}\psi}{\sqrt{a_c^2\sin^2\psi + b_c^2\cos^2\psi}}} \mathrm{d}a\,\mathrm{d}b\,\mathrm{d}z_1 \tag{3.127}$$

and, again, the asymptotic limit for large values of t is

$$P(t) = \frac{E^*}{2\pi}\sqrt{\frac{m_2}{2}}\left(\frac{\exp\left(-\frac{1}{2}t^2\right)}{t}\right). \tag{3.128}$$

Therefore, using equations (3.126) and (3.128) we can write

$$\lim_{t\to\infty}\frac{P(t)}{A_c(t)} = \sqrt{\frac{m_2}{\pi}}E^* \tag{3.129}$$

which can be rewritten as

$$\lim_{t\to\infty} A_c(t) = \sqrt{\frac{\pi}{m_2}}\frac{P(t)}{E^*} \tag{3.130}$$

and implies that, for large separations, the area of contact grows linearly with the applied pressure. The coefficient of proportionality, which will be rewritten in a slightly different form when we will compare it to the result obtained by Persson, derived in the next section, is known as the asymptotic limit of the BGT theory, and is commonly used by many works as an upper bound for the expected coefficient of proportionality between load and area of contact. In the remainder of [BGT75], the authors critically compare their results to those of [GW66] and [Mik74], but since those results are derived starting from different assumptions (some of which –like plastic limit– are inapplicable to our experiments), we will not report this comparisons.

Model limitations and extensions

One of the weakest points of the BGT model is the assumptions that asperities act independently. This can hold for large separations, but when the pressure is large enough, the contact islands of different asperities will merge, and the total area of contact will be smaller than the sum of the separate contributions. This phenomenon is often referred to as "long distance interaction" of asperities. We will see in Section 5.2.3 that neglecting long distance interactions result in an overestimation of the coefficient of proportionality between load and area of contact for rough surfaces with small values of root mean square roughness. In [PC10] the BGT model is enhanced to include long distance interactions, thus overcoming its biggest weakness: the coefficient of proportionality is found not to be constant and valid for all surfaces, but strongly influenced by roughness parameters, and from the bandwidth parameter α.

3.4 Persson's theory

In the early 2000s, the rough contact theory proposed by B. N. J. Persson in a series of papers gained remarkable visibility. The theory is based on reasonings about the stress distribution over the area of contact at different *magnification levels*, which we will define formally below, but can be interpreted, loosely speaking, as the "level of zoom" at which the contact is observed. In fact, the basis of the model is the fractal nature of the surfaces, and its consequences on how the asperities of the rough surfaces interact: when observed at low magnification levels, the contact will appear full, whereas decreasing the length scale at which the system is observed, smaller wave length details are taken into account, and the contact area becomes fragmented and smaller. Persson often focuses on polymers interacting with road-like surfaces, for which he derived a complete kinetic friction theory, including effects of rubber viscoelasticity [Per01b; PAT+04] and temperature increase due to polymer excitation [Per06]. But the theory applies equally to different problems, such as the computation of leak rates of rubber seals, [PY08; LP09b; LP10]. Since in this work we only consider static friction, a complete summary of Persson theory for friction is out of scope, and we will

limit ourselves to the derivation of the relationships of load and real area of contact for rough surfaces, c.f. [Per01a; PBC02; Per07; Per08; LP09a], and to a brief exposure of the formula for kinetic friction, c.f. [Per01b]. Since we will employ a fully elastic, adhesion-free model, we will also neglect the parts of his theory regarding plasticity and adhesive forces [MP11; Per02].

3.4.1 Constitutive elements

We consider a rough and rigid half-space in contact with an elastic, flat, square surface, having Young's modulus E and Poisson ratio ν. Thanks to (2.58), we know that, for small deformations, the system is equivalent to one in which both surfaces are elastic, and have Young's moduli E_1, E_2, and Poisson's ratios ν_1, ν_2. Moreover, the system is also equivalent to one with two rough surfaces, if the interfacial gap (the distance between the two undeformed surfaces) is the same, c.f. [Joh85]. Now we observe the system at the length scale $\lambda = L/\zeta$, where ζ is the magnification level, and L is the lateral size of the elastic surface, which is therefore also the lateral size of the nominal contact area A_0 (also known as the apparent area of contact). We define the wave vector $q_L = 2\pi/L$, and write the short-distance cut-off wave vector q_1 as $q_1 = q_L\zeta$. Thus, formally the magnification level is the ratio of the largest to the smallest considered wave vector. Notice that in the remainder of this section, we will always assume that the PSD $C(q)$ is 0 everywhere except in $[q_L, \zeta q_L]$.

3.4.2 Probability distributions $P(\sigma, \zeta)$ and $P(\zeta)$

The fundamental idea of the model is that any contact which is apparently full and smooth, when analyzed at higher magnification levels, appears fragmented. This happens because surfaces are rough at many different length scales, and every increase

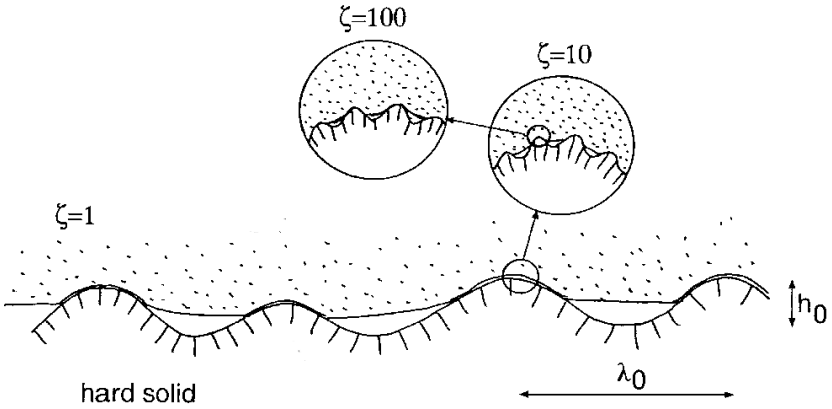

Figure 3.4. Contact depicted at different magnification levels, image from [Per01b].

of the magnification level reveals gaps in areas which are thought to be in complete contact, as it is depicted in Figure 3.4. We now define the stress probability density $P(\sigma,\zeta)$, that is, the function assigning, for a given length scale L/ζ, a probability density to each value of stress intensity σ. Thus, the probability that the stress intensity at a random contact point $\mathbf{x}$ is between σ and $\sigma+\Delta\sigma$ can be expressed by

$$\int_{\sigma}^{\sigma+\Delta\sigma} P(\sigma',\zeta)\mathrm{d}\sigma'. \tag{3.131}$$

We start by assuming that the contact is complete on all length scales, therefore

$$P(\sigma,\zeta) = \left\langle \delta\left(\sigma-\sigma_1(\mathbf{x})\right)\right\rangle. \tag{3.132}$$

Where $\delta(\sigma-\sigma_0)$ is the Dirac delta distribution in σ-space, centered at σ_0, and the notation $\langle\cdot\rangle$ indicates the ensemble averaging, that is, the averaging over different realizations of the height function $h(\mathbf{x})$, which is a random process of the type studied by Nayak (c.f. Section 3.1). Now we assume that increasing the magnification level by $\Delta\zeta$, and the stress becomes $\sigma_1+\Delta\sigma$, therefore we write

$$\begin{aligned} P(\sigma,\zeta+\Delta\zeta) &= \langle\delta(\sigma-\sigma_1-\Delta\sigma)\rangle \\ &= \int_0^\infty \mathrm{d}\sigma'\langle\delta(\sigma'-\Delta\sigma)\,\delta(\sigma-\sigma_1-\sigma')\rangle \\ &= \int_0^\infty \mathrm{d}\sigma'\langle\delta(\sigma'-\Delta\sigma)\rangle P(\sigma-\sigma',\zeta) \end{aligned} \tag{3.133}$$

where in the last step, we exploited the statistical independence of the different regions in ζ over which we average. The ensemble average of the Dirac delta function can be replaced by its Fourier transform

$$\langle\delta(\sigma'-\Delta\sigma)\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}w\left\langle\exp\left(iw(\sigma'-\Delta\sigma)\right)\right\rangle \tag{3.134}$$

and under the assumption that $\Delta\sigma$ is small, we can expand the integrand to second order in $\Delta\sigma$, obtaining

$$\begin{aligned} \langle\delta(\sigma'-\Delta\sigma)\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}w\exp\left(iw\sigma'\right)\left(1-w^2\frac{\left\langle\Delta\sigma^2\right\rangle}{2}\right) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}w\exp\left(iw\sigma'\right) - \int_{-\infty}^{\infty}\mathrm{d}w\exp(-iw\sigma')w^2\frac{\left\langle\Delta\sigma^2\right\rangle}{2} \\ &= \delta\left(\sigma'\right) + \frac{\left\langle\Delta\sigma^2\right\rangle}{2}\frac{\partial^2\delta\left(\sigma'\right)}{\partial\sigma'^2}. \end{aligned} \tag{3.135}$$

Now we substitute (3.135) in (3.133) and we expand the LHS, under the assumption of small $\Delta\zeta$. We write

$$P(\sigma,\zeta)+\frac{\partial P(\sigma,\zeta)}{\partial\zeta}\Delta\zeta=\int \mathrm{d}\sigma' P\left(\sigma-\sigma',\zeta\right)\left[\delta\left(\sigma'\right)+\frac{\left\langle\Delta\sigma^2\right\rangle}{2}\frac{\partial^2\delta\left(\sigma'\right)}{\partial\sigma'^2}\right] \tag{3.136}$$

and by solving the integral applying integration by parts twice, it is obtained

$$\frac{\partial P(\sigma,\zeta)}{\partial\zeta}=\frac{\left\langle\Delta\sigma^2\right\rangle}{2\Delta\zeta}\frac{\partial^2 P(\sigma,\zeta)}{\partial\sigma^2}. \tag{3.137}$$

If we now assume that at the lowest magnification level ($\zeta=1$) the stress over the nominal area of contact is constant and equal to σ_0, we can write

$$P(\sigma,1)=P_0(\sigma)=\delta(\sigma-\sigma_0) \tag{3.138}$$

We can rewrite (3.137) as

$$\frac{\partial P(\sigma,\zeta)}{\partial\zeta}=f(\zeta)\frac{\partial^2 P(\sigma,\zeta)}{\partial\sigma^2}. \tag{3.139}$$

with

$$f(\zeta)=\frac{\left\langle\Delta\sigma^2\right\rangle}{2\Delta\zeta} \tag{3.140}$$

and it can be observed that (3.139) has the form of a one dimensional diffusion Partial Differential Equation (PDE), where density (or temperature) is replaced by $P(\sigma,\zeta)$, the 1D space by the σ-space, the time by the magnification level ζ, and the diffusivity constant is $f(\zeta)$. Thus, as in a normal diffusion, for larger values of time, the density is more diffused over the space (and the integral is constant), in the same manner, for larger magnification level, the stress distribution of (3.139) becomes broader with respect to σ, as shown in Figure 3.5, where larger values of G correspond to larger magnification levels (the function $G(\zeta)$ will be introduced below). The adhesion-free condition can be expressed through the boundary condition

$$P(0,\zeta)=0 \tag{3.141}$$

that means that no contact point can have zero stress, or stated differently, when stress is zero, we observe detachment, and a rigorous proof of this can be found in [PBC02]. We also impose the obvious boundary condition

$$P(\infty,\zeta)=0 \tag{3.142}$$

which will be used to solve (3.139), once we will have found a more tractable form for the diffusivity constant $f(\zeta)$.

We now multiply (3.139) by σ and integrate over σ, and, after some calculations on the RHS, we obtain

$$\frac{\partial}{\partial \zeta}\int_0^\infty \mathrm{d}\sigma \sigma P(\sigma,\zeta) = 0 \tag{3.143}$$

and thus, using the boundary condition (3.138),

$$\int_0^\infty \mathrm{d}\sigma \sigma P(\sigma,\zeta) = \sigma_0. \tag{3.144}$$

If, instead, we integrate both sides of (3.139), we obtain

$$\frac{\partial}{\partial \zeta}\int_0^\infty \mathrm{d}\sigma P(\sigma,\zeta) = -f(\zeta)\frac{\partial P(0,\zeta')}{\partial \sigma} \tag{3.145}$$

hence

$$\int_0^\infty \mathrm{d}\sigma P(\sigma,\zeta) = 1 - \int_1^\zeta \mathrm{d}\zeta' f(\zeta')\frac{\partial P(0,\zeta')}{\partial \sigma}. \tag{3.146}$$

We now want to derive an expression for $P(\zeta)$, which is a function returning the ratio of the real area of contact at magnification level ζ to the nominal area of contact A_0, that is

$$P(\zeta) = \frac{A(\zeta)}{A_0} = \frac{A(\zeta)}{A(1)}. \tag{3.147}$$

Let us start by defining $\langle\sigma\rangle_\zeta$ as the average pressure over $A(\zeta)$ (or, stated differently, over the apparent contact area on the length scale L/ζ). Since the total force acting on the surface must remain constant at every magnification level, we can write

$$\sigma_0 A_0 = \langle\sigma\rangle_\zeta P(\zeta) A_0 \tag{3.148}$$

thus

$$P(\zeta) = \frac{\sigma_0}{\langle\sigma\rangle_\zeta}. \tag{3.149}$$

Now, we simply compute $\langle\sigma\rangle_\zeta$ as

$$\langle\sigma\rangle_\zeta = \frac{\displaystyle\int_0^\infty \mathrm{d}\sigma \sigma P(\sigma,\zeta)}{\displaystyle\int_0^\infty \mathrm{d}\sigma P(\sigma,\zeta)} \tag{3.150}$$

and using (3.149), we can write

$$P(\zeta) = \frac{\sigma_0 \displaystyle\int_0^\infty \mathrm{d}\sigma P(\sigma,\zeta)}{\displaystyle\int_0^\infty \mathrm{d}\sigma \sigma P(\sigma,\zeta)} \tag{3.151}$$

which we can solve using (3.144) and (3.146), obtaining

$$P(\zeta) = 1 - \int_1^{\zeta} \mathrm{d}\zeta' f(\zeta') \frac{\partial P(0,\zeta')}{\partial \sigma} \tag{3.152}$$

In [Per01b], an expression for $f(\zeta)$ is derived. We omit the complete derivation and

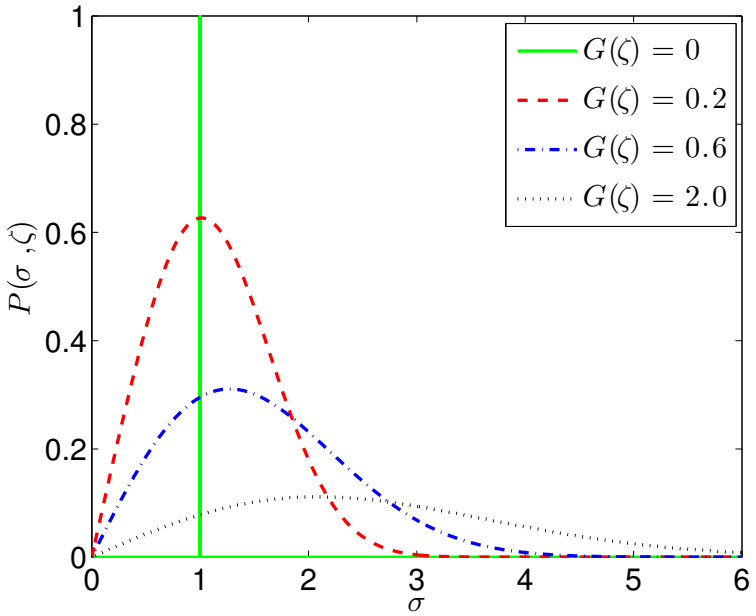

Figure 3.5. Stress distribution at different magnification levels, according to equation (3.157). For $G(\zeta) = 0$, the probability distribution is a Dirac delta distribution centered at σ_0.

we report the two main results. First, a formula for the root mean square normal stress $\langle \sigma_z \rangle$ is found to be[1]

$$\langle \sigma_z \rangle = \frac{\pi}{2} E^{*2} \int_{-\infty}^{\infty} \mathrm{d}q \, q^3 C(q) \tag{3.153}$$

where $C(q)$ denotes as usual the PSD of the rough surface. Since $q = q_L$, and $\Delta\zeta$ is sufficiently small, we can write

$$f(\zeta) = \frac{\langle \sigma^2 \rangle}{2\Delta\zeta} = \frac{\langle \sigma^2 \rangle}{2\Delta q} q_L = \frac{\pi}{4} E^{*2} q_L q^3 C(q) \tag{3.154}$$

[1] In the cited source, the derivation is done for the complex visco-elastic modulus, but here we use the normal elastic modulus, for the sake of consistency.

and for ease of notation we define the function $g(\zeta)$

$$g(\zeta) = \frac{f(\zeta)}{\sigma_0^2} = \frac{\pi}{4}\frac{E^{*2}}{\sigma_0^2} q_L q^3 C(q) \tag{3.155}$$

and can rewrite (3.139) as

$$\frac{\partial P(\sigma,\zeta)}{\partial \zeta} = g(\zeta)\sigma_0^2 \frac{\partial^2 P(\sigma,\zeta)}{\partial \sigma^2}. \tag{3.156}$$

The solution of this type of equation can be found in most text books on PDEs, e.g. [Sal08], and it is shown in [Per01b] to be

$$P(\sigma,\zeta) = \sum_{n=1}^{\infty} \frac{2}{\sigma_Y} \sin\left(\frac{n\pi\sigma_0}{\sigma_Y}\right) \sin\left(\frac{n\pi\sigma}{\sigma_Y}\right) \exp\left[-\left(\frac{n\pi\sigma_0}{\sigma_Y}\right)^2 G(\zeta)\right] \tag{3.157}$$

where σ_Y denotes the yield stress. We show in Figure 3.5, for different values of $G(\zeta)$, the shape of the resulting stress distribution. To generate the plot, we used $n = 10^5$, $\sigma_0 = 1$, and $\sigma_Y = 10^8$. The value of the yield stress is chosen to be very large, in order to simulate the response of a purely elastic material, which is the case which we will consider in the remainder of this section.

Assuming that the flat surfaced body obeys a purely elastic material law (thus with no plasticity, i.e. yield stress $\sigma_Y \to \infty$), we can write

$$P(\zeta) = \int_0^{\infty} \mathrm{d}\sigma P(\sigma,\zeta) \tag{3.158}$$

and employing the boundary conditions (3.141) and (3.142), one can solve (3.156), obtaining

$$P(\zeta) = \frac{2}{\pi}\int_0^{\infty} \mathrm{d}x \frac{\sin x}{x} \exp\left[-x^2 G(\zeta)\right] \tag{3.159}$$

where $G(\zeta)$ is an antiderivative of $g(\zeta)$, which can be written as

$$G(\zeta) = \frac{\pi}{4}\left(\frac{E^*}{\sigma_0}\right)^2 \int_{q_L}^{\zeta q_L} \mathrm{d}q\, q^3 C(q). \tag{3.160}$$

Under the assumption of small loads, $G(\zeta) \gg 1$, it can be noticed that only small values of x will have considerable influence on (3.159), and for such values $x \approx \sin x$, thus

$$P(\zeta) \approx \frac{2}{\pi}\int_0^{\infty} \mathrm{d}x \exp\left[-x^2 G(\zeta)\right] = \frac{1}{\sqrt{\pi G(\zeta)}} \tag{3.161}$$

and since $G(\zeta) \propto \sigma_0$, we can say that the area of contact is proportional to the load, for small loads. The general solution of the PDE is reported in [PAT+04] and is

$$P(\zeta) = \mathrm{erf}\left(\frac{1}{2\sqrt{G(\zeta)}}\right). \tag{3.162}$$

3.4.3 Comparison with the BGT model

Persson derived his model following an approach which is completely different from the one we reported in Section 3.3, and this leads to a generally different result. For small loads, though, it was first noticed in [HPMR04] that the two theories predict a very similar result. One can exploit the definition of m_2 given in (3.19) to rewrite (3.160) as

$$G(\zeta) = \frac{1}{4}\left(\frac{E^*}{\sigma_0}\right)^2 m_2 \tag{3.163}$$

where we recall that we only consider the portion of the PSD bounded by q_L and ζq_L. Thus we can rewrite (3.161) as

$$P(\zeta) = \frac{2}{\sqrt{\pi}} \frac{\sigma_0}{E^* \sqrt{m_2}} \tag{3.164}$$

and assuming that the real area of contact A is equal to $A_0 P(\zeta_{\max})$, it can be written

$$\frac{A}{A_0} = \frac{2}{\sqrt{\pi m_2}} \frac{\sigma_0}{E^*} \tag{3.165}$$

which we can compare with the final asymptotic result of the BGT model, c.f. (3.130), rewritten according to the notation of this chapter as

$$\frac{A}{A_0} = \sqrt{\frac{\pi}{m_2}} \frac{\sigma_0}{E^*}. \tag{3.166}$$

Thus, by comparing (3.165) to (3.166), it is clear that the area predicted by Persson is smaller than the one predicted by the BGT model by a factor equal to

$$\frac{\frac{2}{\sqrt{\pi}}}{\sqrt{\pi}} = \frac{2}{\pi} \tag{3.167}$$

Noting that, for isotropic surfaces, the root mean square slope $\sqrt{\left\langle |\nabla h|^2 \right\rangle}$ is defined in [Nay73]

$$\sqrt{\left\langle |\nabla h|^2 \right\rangle} = \sqrt{2m_2} \tag{3.168}$$

in [HPMR04] a unified formula of the form

$$\frac{A}{A_0} = \kappa \frac{\sigma_0}{E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1} \tag{3.169}$$

is suggested, and κ assumes the value of $\sqrt{8/\pi}$ for Persson's model, and $\sqrt{2\pi}$ for the BGT model.

3.4.4 Extension to kinetic friction computation for elastomers

One of the most known extensions of Persson's theory of contact, is the computation of kinetic friction coefficients for rubber-like materials (which are technically called *elastomers*, a contraction of *elastic polymers*), sliding on hard rough substrates.

The key concept is that for highly viscous materials, the friction force mostly stems from internal dissipation of contact energy. The short wave length asperities of the rough surface, would excite rubber periodically at high frequencies, building up the bulk modulus, that is, the viscous response of the material to variation of strain, which was introduced in (2.33). The classic way of representing bulk (or *loss*) modulus, is to add a complex part to Young's modulus, obtaining the so-called *complex modulus*, c.f. [Fer61], which depends on frequency of the stimulus.

Without going further into details, which can be found in [Per01b], we provide the therein derived equation which relates friction to roughness

$$\mu_k = \frac{1}{2}\int_{q_L}^{\zeta q_L} \mathrm{d}q^3 C(q)P(q) \int_0^{2\pi} \mathrm{d}\theta \cos\theta \,\mathrm{Im}\left(\frac{E^*(qv\cos\theta)}{\sigma_0}\right) \tag{3.170}$$

where $\mathrm{Im}(\cdot)$ denotes the imaginary part of a complex number. The functions $P(q)$ and $G(q)$ have to be re-written for the viscous case, but we omit their expressions here.

Thus, friction is believed to depend only (or mostly) on the loss modulus. A further extension exists, which considers the contribution of temperature (and especially of high temperature spots on the rubber surface), and is given in [Per06; Per14].

3.5 Yastrebov-Anciaux-Molinari's theory

One of the most recent rough contact theories, is the contact evolution law derived by V. Yastrebov, G. Anciaux, and J. F. Molinari. Their model, which we will call the YAM model, pursues the goal of bridging the gap between asymptotic predictions for low and high pressures and is driven mostly by observations of numerical results, obtained using the Boundary Element Method, [YAM12], [YAM14]. In this section, we give an overview of the model assumptions and the derivation of the relationship between load and area of contact, omitting other interesting parts of their works, such as the analysis of the influence of Nayak's parameter α on the variance of the results, or the characterization of the mesh-size related error.

3.5.1 Constitutive elements

In the YAM model, as in the model presented in Sections 3.3 and 3.4, surfaces are considered to be realizations of a random process of the type studied by Longuet-Higgins and Nayak, c.f. Section 3.1. Moreover, surfaces are considered to be self-affine, c.f. Section 3.2, thus as in the previously presented models, the portion of the PSD

which is considered plays a major role in the estimation of the real area of contact. In [YAM12], the results of several simulations with different rough surfaces are studied, initially with a particular focus on small areas of contact (in the order of 10% of the nominal area), in order to prove or disprove linearity between contact area and load at low pressure. By analyzing the derivative of the mean contact pressure with respect to the external pressure, the authors find a power law to be a reasonable fit for the data, and therefore started the derivation of the contact evolution law from the assumption that

$$\frac{\partial \overline{\sigma}}{\partial \sigma_0} = \beta \left(\frac{A_0}{A}\right)^{\mu}, \quad \text{with } \beta > 0,\ 0 < \mu < 1. \tag{3.171}$$

3.5.2 Derivation of the contact law

Starting from (3.171), we can introduce the total force F acting on the two surfaces, and express the mean and the nominal pressure through

$$\overline{\sigma} = \frac{F}{A}, \quad \sigma_0 = \frac{F}{A_0}. \tag{3.172}$$

We can substitute these expressions in the LHS of (3.171) and then divide the equation by A_0 to get

$$\frac{A - F\dfrac{\partial A}{\partial F}}{A^2} = \frac{\beta}{A_0}\left(\frac{A_0}{A}\right)^{\mu} \tag{3.173}$$

which can be multiplied by A^2 and rearranged to obtain

$$\begin{aligned} A - F\frac{\partial A}{\partial F} &= \beta \frac{A^2}{A_0}\left(\frac{A_0}{A}\right)^{\mu} \\ F\frac{\partial A}{\partial F} &= A - \beta A_0 \frac{A^2}{A_0^2}\left(\frac{A_0}{A}\right)^{\mu} \\ F\frac{\partial A}{\partial F} &= A - \beta A_0 \left(\frac{A_0}{A}\right)^{\mu-2}. \end{aligned} \tag{3.174}$$

Separation of variables leads to

$$\int_{A_i/A_0}^{A/A_0} \mathrm{d}\left(\frac{A'}{A_0}\right)\left[\frac{A'}{A_0} - \beta\left(\frac{A'}{A_0}\right)^{2-\mu}\right]^{-1} = \int_{F_i}^{F} \frac{\mathrm{d}F'}{F'}$$

which can be easily solved, obtaining

$$\frac{1}{\mu - 1}\ln\left[\frac{\left(\dfrac{A}{A_0}\right)^{\mu-1} - \beta}{\left(\dfrac{A_i}{A_0}\right)^{\mu-1} - \beta}\right] = \ln\left(\frac{F}{F_i}\right) \tag{3.175}$$

where A_i and F_i can be taken to be the infinitesimal area of contact and force, respectively. Here the YAM model assumes linearity between load and area of contact, and, moreover, it assumes that the unified formula (3.169) holds (that is, it is assumed that the only quantities defining the proportionality factor of the relationship are rms slope and equivalent elastic modulus). Thus

$$\frac{A_i}{A_0} = \kappa \frac{F_i}{A_0 E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1} \tag{3.176}$$

which can be substituted in (3.175) to obtain

$$\frac{1}{\mu - 1} \ln \left[\frac{\left(\frac{A}{A_0}\right)^{\mu-1} - \beta}{\left(\frac{\kappa F_i}{A_0 E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1}\right)^{\mu-1} - \beta} \right] = \ln \left(\frac{F}{F_i}\right) \tag{3.177}$$

since $\mu - 1 < 0$, and

$$\frac{\kappa F_i}{A_0 E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1} \ll 1 \tag{3.178}$$

because F_i is infinitesimal, then β in the denominator is negligible, and taking exponential of both sides of (3.177) we obtain

$$\frac{\left(\frac{A}{A_0}\right)^{\mu-1} - \beta}{\left(\frac{\kappa F_i}{A_0 E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1}\right)^{\mu-1}} = \left(\frac{F}{F_i}\right)^{\mu-1}$$

$$\left(\frac{A}{A_0}\right)^{\mu-1} = \beta + \left(\frac{\kappa F}{A_0 E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1}\right)^{\mu-1}$$

and reintroducing the nominal pressure $\sigma_0 = F/A$, we obtain the final result

$$\frac{A}{A_0} = \left[\beta + \left(\frac{\kappa \sigma_0}{E^*} \sqrt{\left\langle |\nabla h|^2 \right\rangle}^{-1} \right)^{\mu-1} \right]^{\frac{1}{\mu-1}} \tag{3.179}$$

3.5.3 Comparison with BGT and Persson's models

The YAM model manages elegantly to extend to medium pressures the asymptotic results of the BGT and Persson's models for low pressures. The inheritance of the linear relation of area and load for low pressures is still visible, since (3.179) in the limit for $\sigma_0 \to 0$ tends to (3.169). The parameters κ, μ and β have been fitted to the numerical results obtained by means of a FFT-based Boundary Element Method, but the authors stress the difficulties which are encountered in the determination of

such parameters, especially because of the problematic estimation of the contact area for very low pressures. As comparison with the pre-existing models, the computed value of κ found in [YAM12] is $\approx 1.145\sqrt{2\pi}$, whereas in [YAM14], it is said that $\kappa \gtrapprox 1.055\sqrt{2\pi}$. In both cases, thus, an initial slope larger than the one predicted by the other models was found. In the same work, it is analyzed the influence that the bandwidth parameter α of (3.63) has on the fitted parameters, and on the variance of the numerical results. As a strategy for the identification of κ, β, and μ, the authors suggest to fit the data to the derivative of (3.179) with respect to σ_0, first finding the minimum value of κ ensuring that all fitted curves are convex, then performing a non-linear least square to fit β and μ.

3.6 Theories of rough contact, concluding remarks

The overview given in this section, about two prominent and one recent and promising approach to rough contact mechanics, serves the purpose of showing the main accepted ideas about the effects of roughness. All analytical models have limited applicability or reliability, depending on the underlying assumptions. Nevertheless, the asymptotic predictions are good approximations and useful indicators of general trends, which are widely known and agreed on.

As it is common in the scientific community, the debate about validity of assumptions and of models developed on top of them is quite heated. We point out some works which compared the reliability of the different theories, such as [ZBBP04], which performs a full study of many different contact models, and [PDM14], in which important remarks about the impossibility to reach the fractal limit for experimental settings, and interesting relations of H and σ^* to possibly applicable correction factors are derived.

In this work, we omitted the derivation of the well-known Greenwood-Williamson (GW) model, c.f. [GW66], which is considered the first complete treatise of rough contact, but approximates asperities as spherical bumps of equal radius, with their summits following a Gaussian distribution. The model predicts an asymptotic evolution of the area of contact much more similar to the one predicted by Hertz, thus $A \propto \sigma_0^{2/3}$, and it is therefore in contrast with the other more sophisticated models we presented, and with the results of our numerical experiments, see Chapter 5. The interested reader is referred to the original source for an interesting and historically meaningful reading. The model was also extended in [Car09], in which a correction factor is derived, and linear proportionality between load and area of contact is found, whereas in [CDD06], a discrete version of the original GW model is proposed, to improve its accuracy.

Chapter 4

Implementation

In this section, we give a description of the most important numerical techniques we implemented to simulate rough contact. The only pre-existing tool was the contact solver, which is implemented in the software toolbox UG/Obslib++, c.f. [BBJ+97; GK08]. We implemented a parallel self-affine rough surfaces generator, which is described in Section 4.1, we extended the contact solver to be used in an iterative fashion, updating the gap function and computing incremental approximations of the correct solution, for which the algorithm is listed in Section 4.2, and we also developed an optimal complexity algorithm to compute the gap function, working for any obstacle which can be represented as a height function, of which we outline the principal components in Section 4.2.3. In Section 4.3, we introduce the Finite Element Method, and briefly describe the solution method, and the discretization we used in our numerical simulations. Finally, in Section 4.3.4, we analyze the software performance in terms of computing time and its scaling with respect to number of processors.

4.1 Rough surfaces generation

We implemented our surface generator following the algorithm described in [PAT+04]. As already mentioned, the algorithm follows the steps of the surface analysis we presented in Section 3.2.3, thus, we will just outline the procedure, remarking some details which are missing in the original reference. We first illustrate the algorithm and then judge the quality of some resulting surfaces, by comparing their PSD to the prescribed one, and comparing their summit heights distribution to the one predicted by Nayak, see Section 3.1. We point out that different algorithms exist, to generate self-affine surfaces, such as the one based on modified RMD patches described in [PB11], but we preferred to implement our own software, to have better control on its behavior.

4.1.1 Algorithm

The goal of the algorithm is to create a realistic isotropic self-affine rough surface of size $L \times L$, represented by a triangulated square lattice with constant $\delta_{\mathscr{O}}$. On the surface we want to prescribe values for root mean square roughness h_0, long-distance roll-off wave number q_0, and Hurst exponent H. We assume that the smallest wave number q_L and the short-distance cut-off wave number q_1 are defined as:

$$q_L = \frac{2\pi}{L}, \quad q_1 = \frac{2\pi}{\delta_{\mathscr{O}}}. \tag{4.1}$$

We recall that for the height function $h(\mathbf{x})$ representing the surface, it holds, from (3.86), that

$$h(x,y) = \sum_{\mathbf{q}} B(\mathbf{q}) \exp\left\{ i \left[q_x x + q_y y + \phi(\mathbf{q}) \right] \right\},$$

where the phase shift $\phi(\mathbf{q})$ is random and uniformly distributed over $[0, 2\pi[$. Using equation (3.89), we can write

$$B(q) = \frac{2\pi}{L} \sqrt{C(q)} \tag{4.2}$$

where we took advantage of surface isotropy (which translates in radial symmetry of the PSD), to replace $\mathbf{q}$ by its Euclidean norm. Thus, to find the coefficients of the Fourier transform of $h(\mathbf{x})$, all we still need to find is an analytical description of $C(q)$.

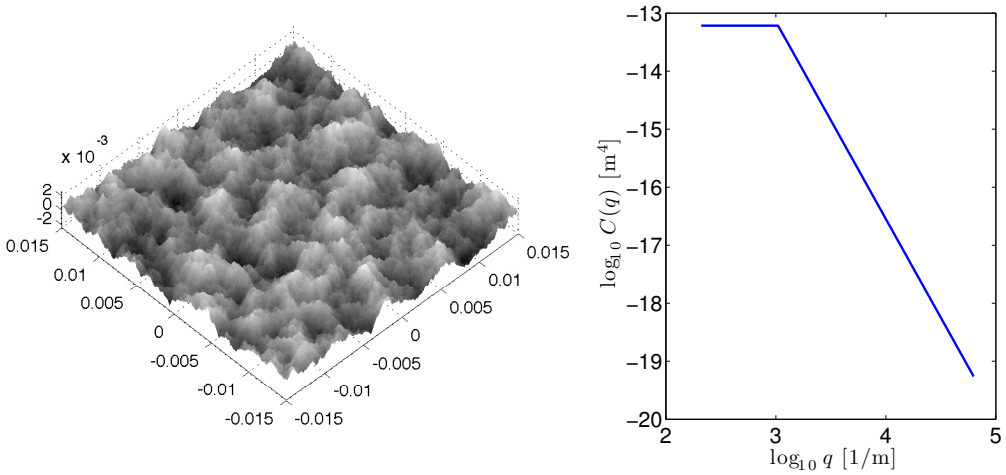

Figure 4.1. Left: A synthetic surface generated for our experiments. It consists of 300×300 points, and the full power spectrum is used. All measures are in meters. *Right*: Logarithmic plot of power spectrum used to generate the surface on the left. In this case, the following parameters are prescribed: $h_0 = 710\,\mu$m, $H = 0.7$, $q_L = 2\pi/0.03$ 1/m, $q_0 = 5q_L$, $q_1 = 2\pi/10^{-4}$ 1/m.

From Section 3.2, we know that the PSD $C(q)$ of an isotropic rough surface must have the form

$$C(q) = \begin{cases} k, & \text{if } q_L \leq q < q_0, \\ k\left(\dfrac{q}{q_0}\right)^{-2(H+1)}, & \text{if } q_0 \leq q < q_1 \\ 0, & \text{else} \end{cases}$$

where q denotes the wave number, which corresponds to the Euclidean norm of the wave vector $\mathbf{q} = (q_x, q_y)$. For ideal self-affine surfaces, it can be assumed that $q_0 = q_L$, but we do not make this assumption, to keep the algorithm more generally applicable. The first step is the definition of the constant k. In [Per01b], it is assumed to be

$$k = \frac{H}{\pi}\left(\frac{h_0}{q_0}\right)^2, \tag{4.3}$$

but we found this to be inaccurate. More precisely, when we used such constant, the surfaces we generated had different h_0 than the one we initially imposed, since we were dropping some of Persson's assumptions. We show in Appendix B the derivation of the correct factor, what approximations can hold for special cases, and what error can be expected, when the wrong approximation is taken. We report here that to obtain the correct value of h_0 is, it must hold:

$$k = \frac{1}{\pi}\left\{\left(1 - \frac{q_L^2}{q_0^2}\right) + \frac{1}{H}\left[1 - \left(\frac{q_0}{q_1}\right)^{2H}\right]\right\}^{-1}\left(\frac{h_0}{q_0}\right)^2. \tag{4.4}$$

Now that we have the analytical description of $C(q)$, we can derive the coefficients of the Fourier transform of $h(\mathbf{x})$, that is, $B(q)$, using (4.2). Thus, we can generate the surface by summing the sinusoidal waves of (3.86), shifting their arguments by the random phase $\phi(\mathbf{q})$. The full discrete spectrum of the surface is composed of wave vectors whose components are multiple of q_L, therefore, if we represent graphically every wave vector in the discrete $\mathbf{q}$-space, we obtain a square lattice, with lattice constant q_L, and since we want to impose that the smallest wave number is q_L, and the short-distance cut-off of the surface is q_1, we select the nodes of the lattice for which $q_L \leq q \leq q_1$, and we call the area they cover the *feasible* $\mathbf{q}$-space. In Figure 4.2, we represent feasible wave vectors (wave vectors belonging to the feasible $\mathbf{q}$-space) as red dots.

Since $h(\mathbf{x})$ is real, it must hold $\phi(-\mathbf{q}) = -\phi(\mathbf{q})$ and $B(-\mathbf{q}) = B(\mathbf{q})$, and these conditions actually make the computations easier, because we can simply sum the contribution of half of the wave vectors, and then multiply this by 2, to obtain the correct height function $h(\mathbf{x})$. In Figure 4.2, we added a blue circle around every wave vector which is used for a surface generation. Notice that wave vectors are selected so that there exist no couple of vectors $(\mathbf{q}', \mathbf{q}'')$ for which $\mathbf{q}' = -\mathbf{q}''$.

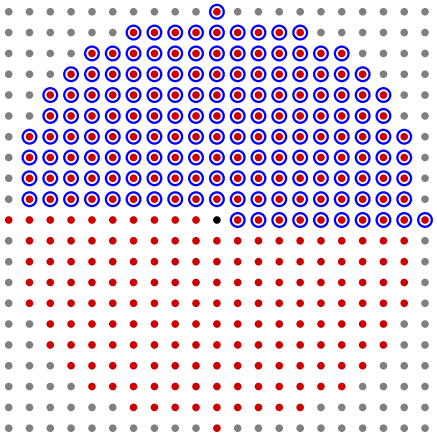

Figure 4.2. The full discrete **q**-space in which we generate the wave vectors. The lattice constant is q_L, and the total amount of points is $[2(L/a)+1]^2$. The red dots are wave vectors for which $q_L \leq \|\mathbf{q}\| \leq q_1$. Blue circles indicate the wave vectors actually used for the computation, which are one half of all the feasible wave vectors.

Summing over all the selected feasible wave vectors, we obtain a self-affine rough surface which has the desired PSD $C(q)$, and, moreover, is fully periodic, because we only used wave vectors for which L is a multiple of their wave length. As a remark, we point out that instead of the summation, once that the coefficients $B(\mathbf{q})$ and the phase shifts $\phi(\mathbf{q})$ are defined, one can simply apply an Inverse Fast Fourier Transform (c.f. [PA04]) to obtain the surface. We prefer the algorithm illustrated, as it is more intuitive and allows for the custom parallelization we describe in next section.

4.1.2 Parallelization in frequency and time

For a realistic surface, the amount of wave vectors to be summed is $O(N^2)$, where $N = L/\delta_{\emptyset}$, and they have to be computed at N^2 points, resulting in a total algorithmic complexity of $O(N^4)$. For example, for one of our surfaces, where $\delta_{\emptyset} = 100$ μm and $L = 3$ cm, the feasible space is composed of roughly 70'000 wave vectors (of which we use 35'000, as explained above), which have to be evaluated in 90'000 points **x**. The computation time of a single surface, for a serial software, is roughly half an hour on a normal desktop computer. Thus, to reduced computation time, we parallelized the code which generates the surfaces. The parallelization is hybrid, employing both MPI ([GLHL+98]) and OpenMP ([DM98]).

We used MPI to subdivide the Cartesian space, that is, to every processor, equally sized subdomains of the surface were given, whereas OpenMP was used to parallelize the summation in frequency space: to every thread of a single processor, a portion

of the **q**-space was assigned. Notice that the algorithm is massively parallel, since both in spatial frequency and in space, the terms of the summations are completely independent: this implies that no communication between threads nor processors has to be executed during the computation. When all threads of a processor have finished their computations, it is sufficient to sum all their contributions to obtain the correct height function of the portion of the subdomain assigned to that processor. When all the processors have finished, the subdomains of the surface are put together, and the complete periodic surface is obtained.

The most common pitfall in this type of parallelization is represented by random variables. In fact, to ensure all processors are using the same phase shifts $\phi(\mathbf{q})$, their values (which are random and uniformly distributed by definition) must be precomputed and shared by all processors. Otherwise, if every processor computes its own phase shifts, the waves will be different from one processor to another, and the different portions of the surface will be inconsistent at the boundaries.

Given the almost complete absence of communication, the parallelization scales almost optimally, and with two processors, each one running eight threads, the computation time is reduced to two minutes, thus, a speedup factor of approximately 15× was achieved.

4.1.3 Quality of the generated surfaces

We measure the quality of the generated surfaces in two ways. The first and trivial one, is checking if the imposed power spectrum can be measured after the generation. The second, and most sophisticated, consists in checking if the summit heights are distributed according to the distribution defined by Nayak, see. (3.69).

To analyze the power spectrum, we followed the procedure that we illustrated in Section 3.2.3, and we can see in Figure 4.3 the comparison between the imposed and the measured power spectral densities. The two PSDs are basically equal, with the computed one slightly larger for the exponential decaying part. We observe the same behavior for all the prescribed PSD.

The summit heights distributions were compared in terms of shape (a simple qualitative inspection), mean value and variance to the distributions predicted by Nayak's theory. To compute mean value $\bar{\xi^*}$ and variance $\sigma^2_{\xi^*}$ of the summit height distribution, we used the formulae derived in [Gre84], which read

$$\bar{\xi^*} = \frac{4}{\pi\alpha}, \quad \sigma^2_{\xi^*} = 1 - \frac{0.8968}{\alpha}. \tag{4.5}$$

We then computed the same parameters for groups of 100 surfaces with same statistical parameters, and obtained the results shown in the bottom plot of Figure 4.4, in which we see that for larger values of α the surfaces we generate tend to have mean values of the summits shifted in positive direction, and this can be observed also in the top plot of Figure 4.4, where the distribution for a group of surfaces with $\alpha = 88.9$ is shown. The

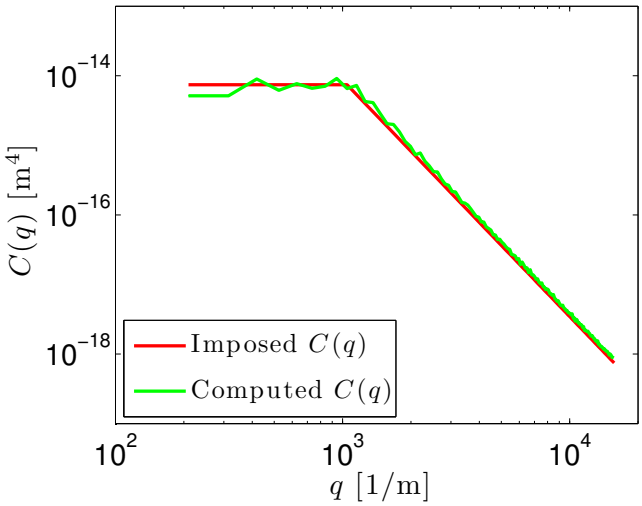

Figure 4.3. Analysis of power spectrum of one generated surface.

shape of the resulting distributions is Gaussian, and variance of the summit heights is close to the theoretical value. Thus, even considering the shifting factor of the summit heights, we judge our surfaces to be accurate enough to be used for numerical experiments, also because the bias is almost constant for large values of α, and it could be easily taken into account to compare our results to those of other studies.

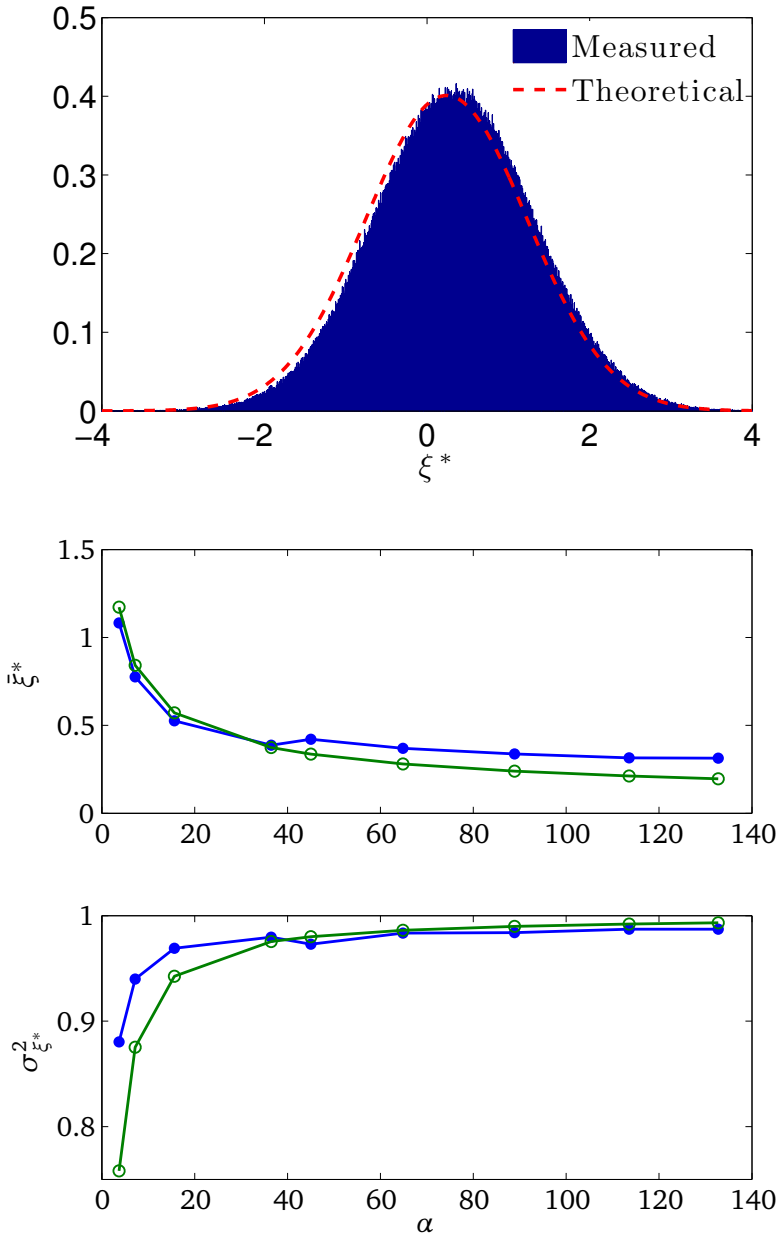

Figure 4.4. Top: Summit height distribution for surfaces with $\alpha = 88.9$. The mean value appears larger than predicted by theory, but the distribution has Gaussian shape, and the variance is preserved. *Bottom*: Mean and variance of summit heights: computed (blue dots) and theoretical (green circles) values.

4.2 Iterative Signorini problem

In this section, we outline the algorithm of what we call an *Iterative Signorini Problem*. Its importance is manly due to its capacity to deal with complex obstacles by means of a sequence of approximate solutions computed with linearized contact conditions.

4.2.1 Generic algorithm

In Section 2.2 we presented a classic Signorini contact problem with linearized contact conditions. The linearized contact condition (2.41) is not directly applicable to large deformation and the reason lies clearly in the linearization itself. Let us indicate the k^{th} node of the mesh by the coordinate vector of its position in reference configuration, i.e. $\mathbf{x}_k$ and the elastic cube's normal pointing outwards from the node as $\mathbf{n}_k$. For each node, the gap function $g(\mathbf{x}_k)$ contains information about the distance from the node to the obstacle surface, this distance being computed along $\mathbf{n}_k$, which for sufficiently small separations, approximates the normal of the closest obstacle point.

From a geometrical standpoint, the normal displacement of the node is bounded by the value of $g(\mathbf{x}_k)$, whereas the tangential displacement is not limited, as if the node's obstacle was a plane, orthogonal to $\mathbf{n}_k$ and at a distance $g(\mathbf{x}_k)$ from the node, as it is depicted in Figure 4.5. Obviously this representation fails to describe a complex obstacle such as a rough surface: when node k does not move only in normal direction, it might slide tangentially on the virtual plane, penetrating the real obstacle and still not violating the linearized constraints, as it can be observed in Figure 4.6. For this reason, we proceed following an iterative approach, and compute a sequence of approximate solutions which converge to a final configuration in which the constraints correctly represent the obstacle locally, and the elastic cube does not penetrate the rough surface.

Algorithm 1 Iterative Contact Problem

given initial geometry, initial gap function $g(\mathbf{x})$, $\alpha_c > 0$
compute $\mathbf{u}^0$
done $\leftarrow$ false
repeat
  compute $g(\mathbf{x}^i + \mathbf{u}^i)$
  compute correction $\mathbf{c}^i$
  $\mathbf{u}^{i+1} \leftarrow \mathbf{u}^i + \alpha_c \mathbf{c}^i$
  if $\dfrac{\|\mathbf{u}^{i+1} - \mathbf{u}^i\|_{\mathbf{M}_\Gamma}}{\|\mathbf{u}^{i+1}\|_{\mathbf{M}_\Gamma}} \leq \text{tol}$ **then**
    $\mathbf{u}^f \leftarrow \mathbf{u}^{i+1}$
    done $\leftarrow$ true
  end if
until not done

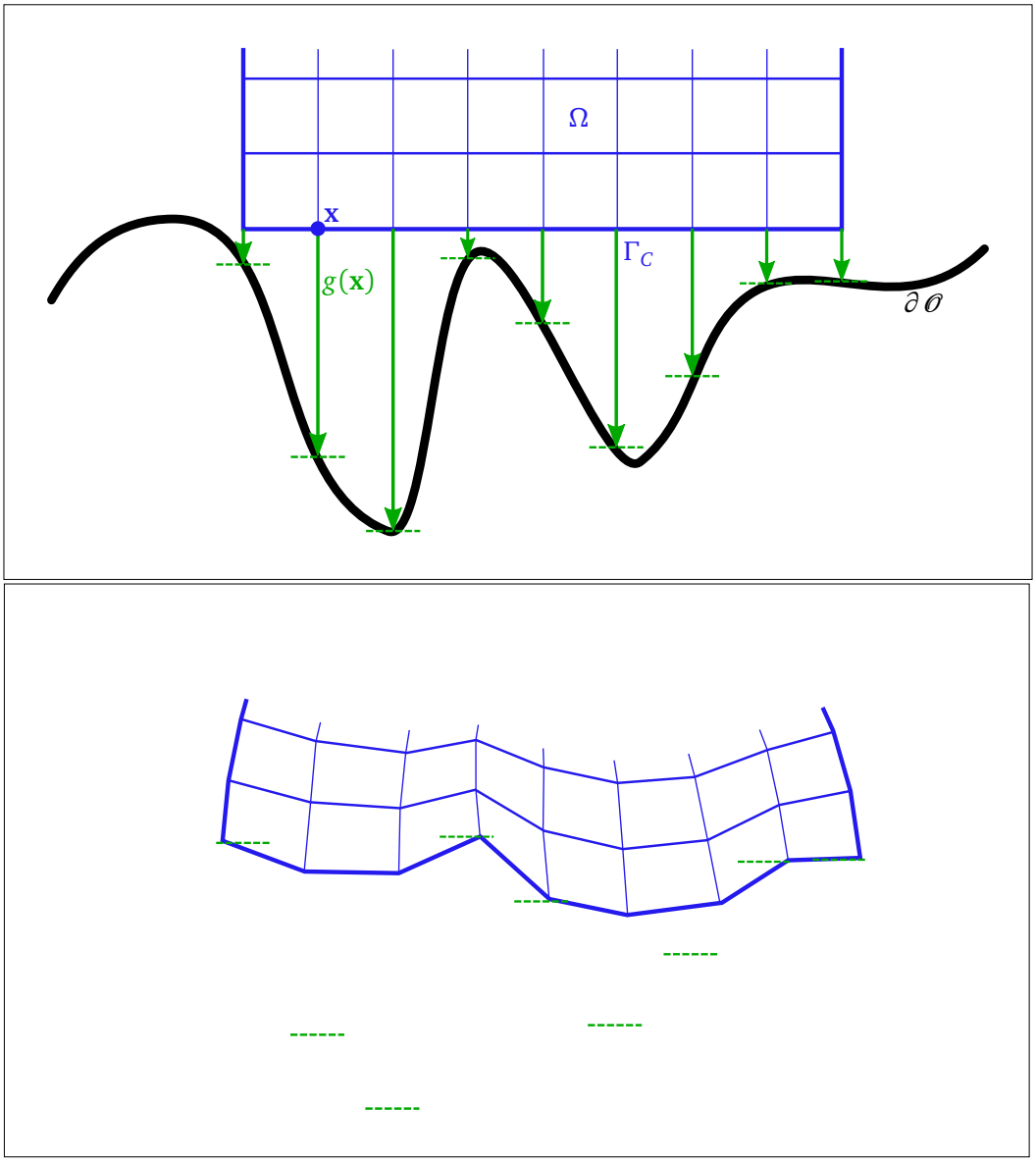

Figure 4.5. Iterative Signorini problem. In blue, the discretized elastic domain Ω (in this case, it is discretized with quadrilaterals, whereas in the real simulations tetrahedra were employed). In black the boundary of the rigid obstacle $\partial\mathcal{O}$ representing the rough surface. In green, the gap function $g(\mathbf{x})$, and the virtual planes which are its only representation in machine memory. *Top*: Initial configuration and gap function. *Bottom*: Hypothetical solution of contact problem. Note that this solution is feasible, since constraints are not violated, but we show in Figure 4.6, that the cube penetrates the original obstacle.

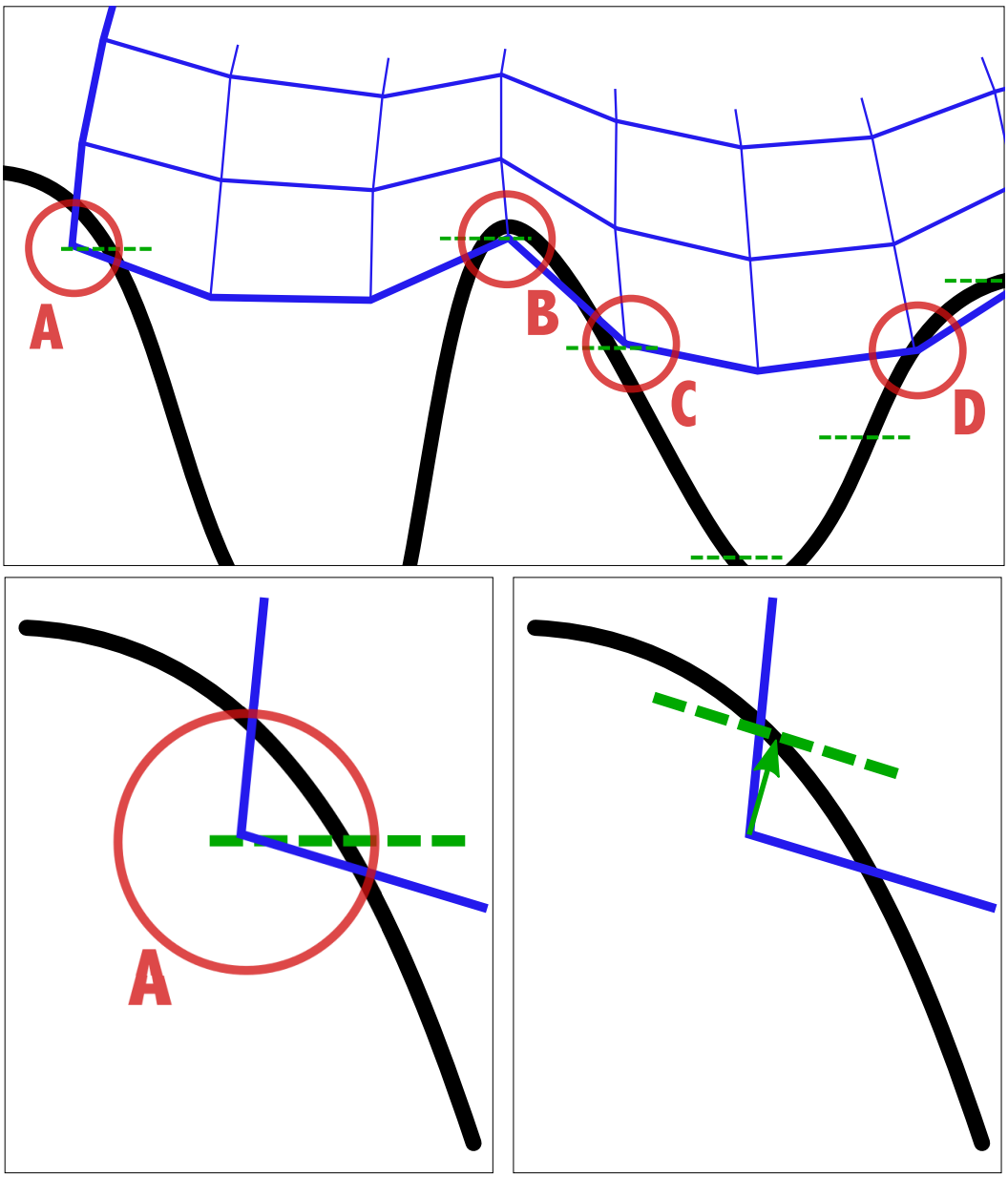

Figure 4.6. Top: possible cases of wrong computed intersections, obtained after the first contact solver step shown in Figure 4.5. Case A, B, and D show nodes which are penetrating the obstacle, even if they do not violate the linear constraint, whereas in case C, a node is blocked by the linear constraint, but it could actually go further downwards. *Bottom*: the gap function, recomputed for case A, which will be projected on the obstacle surface at the next iteration. Note that the intersection is found in negative normal direction.

The iteration scheme is shown in Algorithm 1, and can be summarized as follows: after the i^{th} solution $\mathbf{u}^i$ has been computed, we update the normals $\mathbf{n}$ and the gap function $g(\mathbf{x}^i + \mathbf{u}^i)$ accordingly, as sketched in Figure 4.6. Then we solve the problem with the new constraints, in order to find the correction $\mathbf{c}^i$, which we scale by a factor α_c (a damping factor to avoid oscillations) and we add to our current solution $\mathbf{u}^i$ to obtain $\mathbf{u}^{i+1}$. We repeat this process, until the norm of $\mathbf{c}$ on Γ_C, computed with respect to the boundary mass matrix $\mathbf{M}_\Gamma$, falls below a given tolerance. In our experiments, we set $\alpha_c = 0.8$, which we empirically found to give the fastest convergence of Algorithm 1.

It must be noted that the normals n are recomputed during contact, when the elastic cube is deformed, and are, therefore, generally not parallel to the z axis: this is the reason why we detect forces which have a non-zero projection on the xy plane, even in the frame of linear elasticity, where elastic forces are always orthogonal to the contact plane.

4.2.2 Boundary conditions

In our numerical experiments, we distinguish between two types of problems:

- **Problem A, Normal Load**: Impose vertical displacement on top of the cube, pushing it onto the surface.
- **Problem B, Shear Test**: Solve Problem A, then, on the resulting configuration, impose tangential displacement of the top of the cube.

Problem A serves the purpose of determining characteristic measures of the contact (real area of contact, separation height, vertical force needed to compress the cube), whereas by solving Problem B we can compute the forces which are opposed to the motion, and the elastic energy stored in the deformed elastic cube: the bottom side of the cube, will be pulled by the elastic shear force in the direction of the displacement imposed on the top side, but it will also be pushed in the opposite direction by the shear resistance transferred by the surface asperities.

Formally, the difference between the two problems lies in the values of the Dirichlet Boundary Conditions. In Problem A, (2.36) can be rewritten component-wise as

$$u_x = 0 \tag{4.6}$$
$$u_y = 0 \tag{4.7}$$
$$u_z = d_z < 0, \tag{4.8}$$

whereas in Problem B, it would read

$$u_x = d_x \neq 0 \tag{4.9}$$
$$u_y = 0 \tag{4.10}$$
$$u_z = d_z < 0. \tag{4.11}$$

4.2.3 Distance function

In a contact problem, from a computational standpoint, one of the most time-consuming activity is the identification of the contact itself. In other words, it is not trivial to detect when a full (in the sense of "not hollow") body intersects another one, since they are both discretized with meshes, which are in turn only collections of nodes and edges. Being it a problem of crucial importance, many strategies exist, and we will present typical solutions to compute the gap function $g(\mathbf{x})$ which is used in the iterative scheme described in Section 4.2.

Signed distance functions

Formally, a Signed Distance Function (SDF), is a scalar field of the form

$$\Phi(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R} \tag{4.12}$$

associating to every d-dimensional coordinate vector $\mathbf{x}$ a scalar value which represents the distance of the point with those coordinates from the closest point of a given closed surface (the rigid obstacle, for us), c.f [OF03]. Usually, as convention, points which are outside the surface are assigned a positive distance value, and points inside the surface are assigned a negative value. One of the most important properties of the SDF, is that its gradient $\nabla\Phi(\mathbf{x})$, for regions which are sufficiently close to the surface, points to the closest point on the surface, and, especially, on the surface, it holds

$$\nabla\Phi(\mathbf{x}) = -\mathbf{n}(\mathbf{x}), \tag{4.13}$$

where $\mathbf{n}(\mathbf{x})$ denotes the outward normal of the surface, at point $\mathbf{x}$.

For static obstacles, the SDF can be pre-computed on a regular grid, at a desired step-size, and then, the distance from the obstacle of a given point $\mathbf{x}'$ internal to the grid can be computed by interpolation of the vertices of the grid cell in which $\mathbf{x}'$ is contained. To compute the values of $\Phi(\mathbf{x})$ at each grid node, state-of-the-art fast marching (or sweeping) algorithms exist, such as the one presented in [DE14]. This approximation is fairly reasonable in many applications, for smooth, continuous surfaces, but we tested it for our case, and given the complex nature of the rough surfaces we employ, and the large size of the problem we study, the mesh needed to store sufficient information about the surface was too large to fit into the memory of any machine we used. Therefore, we followed a different approach.

Normal gap and ray-casting

We are interested in finding a good estimate of the *normal gap* between two surfaces. That is, given a point $\mathbf{x}_e$ on the elastic surface $\Gamma = \partial\Omega$, we want to compute which point $\mathbf{x}_{\mathscr{O}}$ on the obstacle surface $\partial\mathscr{O}$ is the closest, at which distance it is, and what is the normal to it. As it is noted in [Yas13], possible inconsistencies arise for complex

geometries, especially when the obstacle is not smooth (because the normal to a non-differentiable point of the surface is not defined), or the obstacle surface is too far from $\mathbf{x}_e$, in which case the distance could be non-unique (for example in case $\mathbf{x}_e$ is in a V-shaped valley of the rough surface, exactly on the symmetry axis).

We decided to adopt an approximation which is precise enough, especially in combination with the iterative scheme we describe in Section 4.2, for sufficiently small separation values. At any point $\mathbf{x}_e$, we compute $\mathbf{n}(\mathbf{x}_e)$, and we compute the distance to $\partial\mathcal{O}$ along that direction. In case of null distance, the node is considered in contact, and as a normal of the contact we use $\mathbf{n}(\mathbf{x}_{\mathcal{O}})$.

Since the obstacle is represented in memory as a height surface, we simply triangulate it (following the connectivity of the lattice on which the nodes lie, and connecting points along the diagonals with $\mathrm{d}x/\mathrm{d}y = 1$). Thus, the problem we solve consists in finding the intersection between the parametric line

$$r : \mathbf{x}_e + \alpha\mathbf{n}(\mathbf{x}) \tag{4.14}$$

and one of the triangles of $\mathcal{O}_\triangle$, which is the triangulated obstacle. This technique, in graphics, is known as *ray-casting*, c.f. [PH04], and the test of intersection between a straight line and a triangle in space can be performed efficiently using the Möller-Trumbore algorithm described in [MT97]. Obviously, checking the intersection between all triangles of $\mathcal{O}_\triangle$ would be computationally infeasible, also because we consider it a periodic obstacle, therefore the amount of triangles would be infinite, but the regularity of the lattice, is helpful. Many spatial structures exist, to efficiently search for objects in space, c.f. [dBCKO08]. In our case, it is sufficient to exploit the lattice regularity of the obstacle to achieve optimal time complexity. To search for possibly intersected triangles, we project r on the xy plane, and only check the intersection between the line and the triangles which belong to cells of the lattice grid which are visited by the line in 2D. To detect which square cells are visited by the line, another algorithm which was invented for graphics purposes is optimal, namely the Bresenham algorithm, c.f. [Bre65]. We refer the interested reader to the original paper for further details.

If the distance traveled by the "ray" is larger than a user-defined threshold, the obstacle is considered to be too far away from the point $\mathbf{x}_e$. Even if we find an intersection, we then search in direction $-\mathbf{n}(\mathbf{x}_e)$, to check if $\mathbf{x}_e$ has penetrated the obstacle, in which case the solver will project it back on $\partial\mathcal{O}$. If the ray intersects the obstacle both in positive and in negative normal direction, than the minimum of the two is chosen, and the state of the node at $\mathbf{x}_e$ (inside or outside the obstacle) is determined accordingly. Notice that the Möller-Trumbore algorithm can also detect if the ray is intersecting the outward or the inward face of the obstacle, thus, it is always possible to understand if the ray was originated inside or outside the obstacle, with respect to the intersected triangle.

As a last remark, we point out that the algorithm we outlined works for any kind

of rigid half-space, that is, for any well defined height function $h(\mathbf{x}) : \mathbb{R}^2 \mapsto \mathbb{R}$. Its complexity in time, for any ray, since it traverses cells linearly, is $O(N^{\frac{1}{2}})$, where N denotes the points of the surface.

4.3 The Finite Element Method

We perform our numerical experiments by means of the Finite Element Method (FEM), which is one of the most used discretization methods for the numerical solution of physical problems formulated as PDEs [Qua09; KO88]. We will describe FEM within the more general class which it belongs to, the class of Galerkin methods.

The basic idea of Galerkin methods is to formulate the variational version of a Boundary Value Problem, and solve the so-called weak problem in a suitable approximation V_h of the space V to which the solution belongs.

We will focus on the Galerkin approximation of the BVP we introduced in Section 2.2. This section is only an introduction to the topic, and a deeper and more general analysis of contact problems, also applied to non-linear cases, can be found in [Sch02] and [KO88].

4.3.1 Galerkin approximation

The problem we are interested in is the BVP defined in (2.35). If (2.35) is satisfied in each point of Ω then the integral of the dot product of this equation with any arbitrary test-function $\mathbf{v}$ (also called virtual displacement, within mechanics context) over the body is zero. The converse statement is also true, c.f. [Yas13], thus

$$\int_\Omega (\nabla \cdot \boldsymbol{\sigma}) \cdot \mathbf{v} \, \mathrm{d}\Omega = 0, \quad \forall \mathbf{v} \in V \tag{4.15}$$

where $V = H^1_{\Gamma_D}(\Omega)$ is a Hilbert space of functions $\mathbf{v} \in L^2(\Omega)$, $\nabla \mathbf{v} \in L^2(\Omega)$, vanishing on Γ_D. If we now integrate (4.15), we obtain

$$\int_\Omega \boldsymbol{\sigma} : (\nabla \mathbf{v}) \, \mathrm{d}\Omega = \int_{\Gamma_D} \boldsymbol{\sigma}\mathbf{n} \cdot \mathbf{v} \, \mathrm{d}\Gamma + \int_{\Gamma_N} \boldsymbol{\sigma}\mathbf{n} \cdot \mathbf{v} \, \mathrm{d}\Gamma + \int_{\Gamma_C} \boldsymbol{\sigma}\mathbf{n} \cdot \mathbf{v} \, \mathrm{d}\Gamma, \quad \forall \mathbf{v} \in V. \tag{4.16}$$

The integral over Γ_D vanishes because of the choice of test functions. And since we impose null force acting on Γ_N, then $\boldsymbol{\sigma}\mathbf{n} = 0$, and the integral on Γ_N vanishes too.

Let us now identify the (linear elastic) stored energy by the function $W : \Omega \times \mathbb{R}^{3\times 3} \to \mathbb{R}$, as done in [Kra09]. Then, the total energy functional $\mathscr{J}(\mathbf{u})$ is composed of the elastic and the potential energy, i.e.

$$\mathscr{J}(\mathbf{u}) = \int_\Omega W(\mathbf{x}, \nabla \mathbf{u}(\mathbf{x})) \, \mathrm{d}\mathbf{x} - \int_{\Gamma_C} \tilde{\mathbf{f}}(\mathbf{x})\mathbf{u}(\mathbf{x}) \, \mathrm{d}\mathbf{a} \tag{4.17}$$

where $\tilde{\mathbf{f}}(\mathbf{x})$ represents the surface force density acting on the contact boundary, as result of the contact deformation. Thus, to solve the problem, one has to minimize $\mathscr{J}$ under the non-penetration constraint introduced in (2.43), that is

$$d_{\mathscr{O}}(\mathbf{u}) \geq \mathbf{0}, \text{ on } \Gamma_C. \tag{4.18}$$

4.3.2 Finite element discretization

We now derive the FEM as an instance of the Galerkin method, in which the space V_h is chosen as a subset of a space of polynomial interpolation.

Let $\mathscr{T}_h$ be a face-to-face *triangulation* of the domain Ω, that is a coverage of Ω with tetrahedral *elements*. The elements are required to be non-overlapping, i.e., given two elements $K, H \in \mathscr{T}_h$, with $K \neq H$, then

$$\dot{K} \cap \dot{H} = \emptyset. \tag{4.19}$$

If the domain is a polytope, the elements cover it completely, i.e.

$$\Omega = \overline{\dot{\bigcup_{K \in \mathscr{T}_h}} K}.$$

We define h_K the diameter of K and $h = \max(h_K)$. We now introduce the interpolation space

$$X_h = \{v_h \in C^0(\overline{\Omega}) : v_{h|K} \in \mathbb{P}_1, \forall K \in \mathscr{T}_h\} \tag{4.20}$$

where $\mathbb{P}_1$ is the space of linear polynomials functions over K. The FE discretization consists in choosing in (4.22) the discrete space

$$V_h = V \cap X_h. \tag{4.21}$$

Thus, we can define the Finite Element Method, as the method to find an approximate solution $u_h \in V_h$ of (4.15), where V_h is a family of subspaces of V that depends on a positive discretization parameter h.

Hence we aim to solve the following Galerkin problem within the Finite Element framework:

find $u_h \in V_h$ such that

$$\mathbf{u}_h = \underset{\mathbf{v}_h}{\operatorname{argmin}}\, \mathscr{J}(\mathbf{v}_h) \tag{4.22}$$

and

$$d_{\mathscr{O}}(\mathbf{u}_h) \geq \mathbf{0}, \text{ on } \Gamma_C. \tag{4.23}$$

Let $\{\varphi_i\}$ denote the set of the Lagrangian basis functions, then the approximate solution can be expressed in the basis $\{\varphi_i\}$ as

$$u_h(\boldsymbol{x}) = \sum_{j=1}^{N_h} u_j \boldsymbol{\varphi}_j(\boldsymbol{x}),$$

where u_j are unknown coefficients, and $N_h = \dim(V_h)$. Thus, the problem defined by (4.22) and (4.23) corresponds to the algebraic problem

$$\mathbf{u} = \underset{\mathbf{v}}{\operatorname{argmin}} \frac{1}{2}\mathbf{v}^T \mathbf{A}\mathbf{v} \tag{4.24}$$

in which the constraint (4.23) must hold. We denote by

- $\mathbf{A}$ the matrix whose entries are $\mathbf{A}_{ij} = a(\boldsymbol{\varphi}_j,\ \boldsymbol{\varphi}_i)$,
- $\mathbf{u}$ the vector whose elements are the coefficients u_j.

The matrix $\mathbf{A}$ is called *stiffness* matrix and in the case of linear elasticity it is symmetric and positive definite. The algebraic formulation offers the possibility to efficiently represent the problem in a numerical software, and several solution methods exist, to find the solution of such algebraic systems.

4.3.3 Solution method and discretization

To compute the solution, we employ our parallel implementation of the conforming non-linear multigrid method as in [Kra09]. The elastic body is distributed among different threads (see Figure 4.7) and the solution is kept consistent over the subdomains boundaries. As a base solver we use an exact LU decomposition and as a smoother we use a non-linear Gauss-Seidel solver.

A complete description of the solution method can be found in the dedicated publication [KRS15]. In Algorithm 2 we report the main steps, mainly to support the reader to grasp the basics of its mechanics, since its application in the field of rough contact is novel. In the algorithm description we introduce ℓ, which is the discretization index of the sequence of finite dimensional subspaces of the appropriate Hilbert space H, $(\mathscr{S}^\ell)_{\ell\geq 0}$, which approximate H. In our case, the spaces $\mathscr{S}^\ell$ are the finite element spaces, which are originating from a family of nested and shape regular meshes, with mes size parameter $h^0 > h^1 > \cdots > h^L$, with $L > 0$ being the index of the finest level under consideration. The functional for which the solution $\mathbf{u}$ is sought is $(J + \varphi^L + \chi_{\mathscr{K}^L})$, where J is a quadratic convex functional and φ^L is assumed to be convex and lower semi-continuous. The third part of the functional, i.e. $\chi_{\mathscr{K}^L}$) is the nonsmooth and nonlinear (but convex) characteristic functional representing the constraints, which, on level ℓ are given by the closed and convex set $\mathscr{D}^\ell_{\tilde{\mathbf{u}}}$. Finally, $\mathscr{X}^\ell_{\tilde{\mathbf{u}}}$ denotes the possibly dependent ansatz space of level ℓ, $\mathscr{Q}_{\tilde{\mathbf{u}}^\ell}$ is a quadratic approximation of $J + \varphi$, and P^L is a nonlinear projection operator which combines the solutions of each one of the m_ℓ subspaces of $(\mathscr{S}^\ell)_{\ell\geq 0}$.

Thus, the algorithm first computes parallel approximate solutions $\mathbf{c}_i^L$, then combines them in $\mathbf{c}^L$ and computes coarse grid corrections in an approximate quadratic space. Finally a sequence of nonlinear but local subproblems is solved along the boundaries, in order to synchronize the solutions.

Algorithm 2 Parallel Nonsmooth Multigrid Method

Initialize $\mathbf{u}^0$.
for $\nu = 0, \ldots, \nu_{\max}$ **do**
  Compute in parallel approximate solutions $\mathbf{c}_i^L$ to the following problems:
    find $\mathbf{w}_i^L \in W_i^L$ such that
    $(J + \varphi^L + \chi_{\mathscr{K}^L})(\mathbf{u}^\nu + \mathbf{w}_i^L) \leq (J + \varphi^L + \chi_{\mathscr{K}^L})(\mathbf{u}^\nu + \mathbf{v}), \quad \forall \nu \in W_i^L$.
  Nonlinear synchronization: $\mathbf{c}^L = P^L(\mathbf{c}_1^L, \ldots, \mathbf{c}_{m_\ell}^L)$.
    Set $\bar{\mathbf{u}} = \mathbf{u}^\nu + \mathbf{c}^L$.
  for $\ell < L$ **do**
    Choose subspace $\mathscr{X}_{\bar{\mathbf{u}}}^\ell$, convex set $\mathscr{D}_{\bar{\mathbf{u}}}^\ell$, $\bar{\mathbf{u}} \in \mathscr{D}_{\bar{\mathbf{u}}}^\ell$ and functional $\mathscr{Q}_{\bar{\mathbf{u}}}^\ell$.
    Parallel coarse grid correction:
      Find $\mathbf{c}^\ell \in \mathscr{D}_{\bar{\mathbf{u}}}^\ell$, such that
      $\mathscr{Q}_{\bar{\mathbf{u}}}^\ell(\bar{\mathbf{u}} + \mathbf{c}^\ell) \leq \mathscr{Q}_{\bar{\mathbf{u}}}^\ell(\bar{\mathbf{u}} + \mathbf{v}), \quad \forall \mathbf{v} \in \mathscr{D}_{\bar{\mathbf{u}}}^\ell$.
    Synchronization: $\mathbf{c}^\ell = P^\ell(\mathbf{c}_1^\ell, \ldots, \mathbf{c}_{m_\ell}^\ell)$.
  end for
  Set $\mathbf{u}^{\nu+1} = \bar{\mathbf{u}} + \sum_{\ell<L} \mathbf{c}^\ell$.
end for

The elastic cube is discretized by means of tetrahedra, the mesh is adaptively refined, therefore it is finer on the contact boundary than in the rest of the domain. A common rule of thumb, when choosing the mesh size of the elastic body at the contact boundary, is to employ a mesh with the lattice constant which is one half of that of the rigid obstacle, i.e., if δ_{Γ_C} and $\delta_{\mathscr{O}}$ denote the lattice constants of the elastic contact boundary and of the rigid obstacle respectively, one would set $\delta_{\Gamma_C} \approx 1/2\,\delta_{\mathscr{O}}$. This choice of δ_{Γ_C} ensures that every rough surface asperity is in contact with at least one complete boundary element of Γ_C. We show in Section 5.3.1 that, even though this choice can lead to a sufficiently good approximation of the contact area for small normal loads, it lacks precision when computing the friction force, for some choices of the roughness parameters. In fact, experiments suggest that the δ_{Γ_C} should be no larger than $1/4\,\delta_{\mathscr{O}}$, for most rough surfaces, to accurately predict friction and area of contact in sheared configurations, but the precision of the results can be greatly improved with finer meshes, especially for some particular roughness parameter.

Since rough surfaces are represented in memory as a height map consisting of 300×300 nodes, it is easy to understand that the amount of nodes needed to discretize the elastic cube is huge, and this, in practice, forces us to employ several processors and massively parallel implementations of the solving routines, to avoid memory issues.

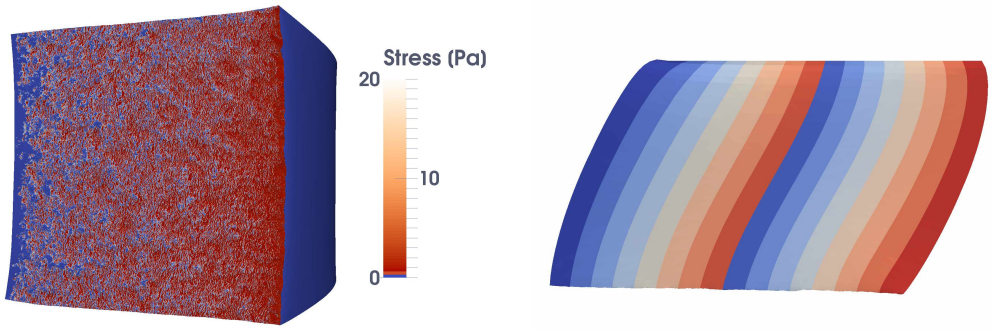

Figure 4.7. Lateral and bottom views of final configuration for one of our experiments. The bottom of the cube remains anchored to the rough surface, due to forces orthogonal to the load produced by the asperities in contact. *Left*: Typical contact stress intensity on the bottom side of the surface, for large loads. *Right*: Subdivision among parallel threads. The load balancer assigns to each thread the same contact area.

4.3.4 Software performance

In this section we analyze the performance of our implementation, in terms of computation time and its scaling with respect to number of processors.

The spatial decomposition was performed subdividing the elastic cube only along x and y directions. This choice does not optimize communication time, because the communication surface grows faster than the number of processors, but it ensures that the contact surface is equally distributed among all processors. The motivation for this choice is straightforward: the computation time is largely dominated by the calculations performed on the contact boundary, whereas solving the linear elastic response occupies only a small fraction of the total time. In Tables 4.1, 4.2 and 4.3, we report the scaling results we obtained employing different mesh sizes to discretize the elastic cube. In all settings, first the cube is discretized uniformly, then the contact surface is refined uniformly five times, leading to a graded adaptive refinement in the elastic body, as it can be observed in Figure 4.8. In the tables we list the total time needed to solve the problem, and we compute the efficiency of the ith parallelization as

$$\mathrm{Eff}_i = \frac{\mathrm{Time}_b \cdot \mathrm{Procs}_b}{\mathrm{Time}_i \cdot \mathrm{Procs}_i}$$

where Procs denotes the number of processors, Time the computation time, and the subscript $_b$ denotes the reference case (the first listed in each table). In the tables, it is shown that the total number of iterations (denoted by *It*) is only mildly influenced by the number of processors. Notice that the total number of iterations includes the amount of iterations needed for the contact boundary stabilization (denoted by CBS in the tables), which shows stability with respect to the number of nodes, too.

Procs	Time	Speedup	Efficiency	It	CBS
32	1824.98	1	1	42	27
64	1018.37	1.79	0.90	45	29
128	563.71	3.24	0.81	48	30
256	339.97	5.37	0.67	52	33

Table 4.1. Scaling results for elastic cube (880k nodes) in contact with rough surface. "It" denotes total number of iterations, and includes iterations needed for contact boundary stabilization (CBS).

Procs	Time	Speedup	Efficiency	It	CBS
128	2094.12 s	–	–	46	35
256	1204.62 s	1.74	0.87	48	37
512	726.51 s	2.88	0.72	53	39
1024	450.98 s	4.64	0.58	57	42

Table 4.2. Scaling results for elastic cube (3.5M nodes) in contact with rough surface. "It" denotes total number of iterations, and includes iterations needed for contact boundary stabilization (CBS).

Procs	Time	Speedup	Efficiency	It	CBS
512	2'219.29 s	–	–	43	33
1'024	1'241.56 s	1.79	0.89	55	38
2'048	864.86 s	2.57	0.64	49	36
4'096	556.92 s	3.98	0.50	54	40

Table 4.3. Scaling results for elastic cube (14.2M nodes) in contact with rough surface. "It" denotes total number of iterations, and includes iterations needed for contact boundary stabilization (CBS).

Procs	Time	Speedup	Efficiency	It	CBS
32	962.15	1	1	35	8
64	480.19	2	1	38	7
128	250.85	3.84	0.96	40	8
256	140.45	6.85	0.86	42	8

Table 4.4. Scaling results for elastic cube (880k nodes) in contact with flat surface. "It" denotes total number of iterations, and includes iterations needed for contact boundary stabilization (CBS).

Procs	Time	Speedup	Efficiency	It	CBS
128	1'179.34	1	1	39	9
256	640.87	1.84	0.92	41	9
512	334.56	3.53	0.88	43	10
1'024	180.88	6.52	0.82	45	11

Table 4.5. Scaling results for elastic cube (3.5M nodes) in contact with flat surface. "It" denotes total number of iterations, and includes iterations needed for contact boundary stabilization (CBS).

Procs	Time	Speedup	Efficiency	It	CBS
512	1'337.41	–	–	39	9
1'024	665.18	2.01	1.01	41	10
2'048	389.33	3.44	0.86	44	11
4'096	229.54	5.83	0.73	47	12

Table 4.6. Scaling results for elastic cube (14.2M nodes) in contact with flat surface. "It" denotes total number of iterations, and includes iterations needed for contact boundary stabilization (CBS).

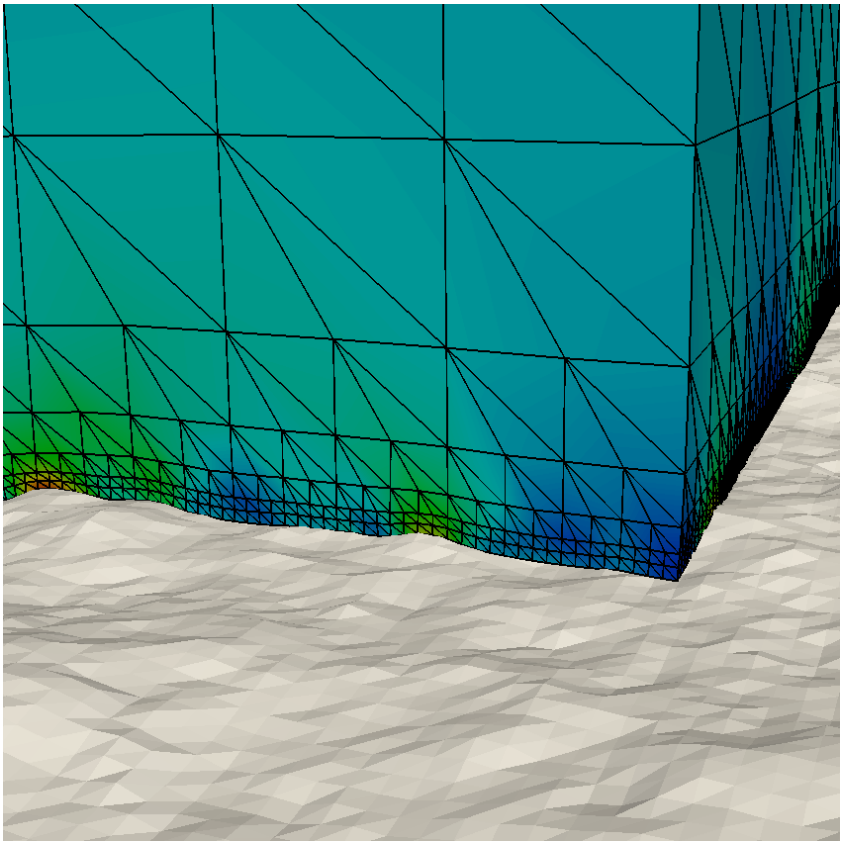

Figure 4.8. Close-up of elastic cube in contact with typical road-like surface for the case $\delta_{\Gamma_C} \approx 1/2\,\delta_{\mathscr{O}}$. Color denotes Von Mises stress.

The identification of the contact nodes for such a complex rough obstacle requires a huge computation time. As means of comparison, we report the results obtained for the same experiments, in which the rough surface is replaced by a plane. We show the scaling results for this case in Tables 4.1 4.5 and 4.6, and it can be observed that the algorithm scales almost perfectly for such simple obstacles.

Chapter 5

Numerical studies of microscopic rough contact

During our study, we performed a large amount of numerical experiments, to understand what roughness parameters influenced real area of contact evolution and friction production. The goal was to exclude parameters with only marginal influence from future studies, to restrict the number of tests one has to perform to have statistically meaningful results. In this section, we illustrate all the results we obtained, and we give remarks about trends and their possible interpretation.

This chapter is subdivided in three main sections. In Section 5.1, we describe the datasets on which we performed the numerical experiments, we give details about surface and cube parameters we varied. Section 5.2 contains the results obtained from experiments performed under vertical (normal) loading conditions, which are compared to the state-of-the-art theories for real area of contact evolution presented in Chapter 3. Section 5.3 concerns the results of our shear tests, in which, after an initial vertical loading, the top of the elastic cube is displaced tangentially, and the shear resistance transferred by the rough surface is measured. The goal is to obtain an estimate of the quantity of static friction force which is produced solely by elastic interactions. Section 5.4.2 illustrates an example of a real-world application of our method: we simulated the effect of water presence rain on a typical road surface, and we computed the difference in friction for different quantities of water, under the assumption of no hydrodynamic interactions, as suggested by [PTAT04; PTAT05]. The results of these specific tests are compared to those of experiments obtained on real roads and runways by other studies.

5.1 Datasets

We begin with a description of the domains on which we performed our experiments. In every experiment, a linear elastic smooth (not rough) cube was used in combination

with a rigid corrugated half-plane obstacle.

The choice of linear elasticity is mostly based on two reasons:

- In the most known rough contact models, such as the GW, BGT, Persson's, and YAM models, second order effects are neglected, c.f. [GW66; BGT75; Per01b; YAM14]. To be able to compare our results to such models, we prefer to use the same assumptions.
- Other papers which study the contact with numerical techniques, such as [KA01; ZBBP04; KSK06; XBR07; HPMR04; PB11; YDPC11], employ linear elasticity. In some works, plasticity is also taken into consideration, but results are anyhow comparable.
- Computation time for linear elastic models is low, compared to other models, which require iterative Newton methods to find a solution. We had to run a total amount of approximately 400 numerical experiments on massively parallel supercomputers, with an estimated total computation time of 500'000 CPU hours. With more sophisticated models, the computation time would have been at least three times as large. We reserve such costly experiments to more focussed (with narrower parameter range) studies.

Moreover, results obtained by the mentioned numerics-based works support the validity of the choice of linear elasticity, since it is shown they are in acceptable agreement with real-world experiments.

5.1.1 Rough surfaces characterization

We performed our experiments on three datasets, and in each dataset we generated surfaces according to a PSD following the definition given in Section 4.1. In each dataset one single parameter was varied, whereas the other were kept identical for every surface. For every value of the parameters we employed, ten statistically identical surfaces were generated. The three datasets are characterized by the following parameters, which we list together with the range in which we chose them (see also Table 5.1):

A Root mean square roughness h_0 ranging from 53 μm to 0.71 mm. Fixed Hurst exponent $H = 0.7$.

B Hurst exponent H ranging from 0.5 to 0.9. Fixed root mean square roughness $h_0 = 0.71$ mm.

C Long-distance cut-off wave number ranging from $2\pi/\delta_{\mathscr{O}}$ to $2\pi/(8\delta_{\mathscr{O}})$, $H = 0.7$, root mean square roughness varying accordingly (see below for a detailed explanation).

The other parameters are the same across all datasets: linear sample size $L = 3$ cm, points vertically aligned to the nodes of a square grid, with lattice constant $\delta_{\mathscr{O}} = 100\ \mu$m, smallest wave number $q_L = 2\pi/L$, long distance roll-off wave number $q_0 = 5q_L$.

The described values of the parameters have theoretical and physical meanings.

- The range of h_0 used in Dataset A covers the rms height values of surfaces like rocks, asphalt, concrete, and also off-road surfaces, as it is described in [PAT$^+$04] and [Lor12], and H was set to a value which is typical for surfaces obtained by fracture, or cracking, as several studies show, see for example [BBCC01], [IIIC03], [CCI99], [DAB12], [Bou97], [Per01b], [PAT$^+$04].

- The root mean square roughness for Dataset B was chosen to maximize the effects of the different fractal dimensions.

- In Dataset C, the long distance cut-off vector was varied to see the effect of low-pass filtering of the PSD on the contact area evolution, and on the variance of the results we obtain: this is intended to be compared to the recent studies published in [YAM12], where similar experiments were performed to understand the importance of the bandwidth parameter α c.f. (3.63). To generate surfaces of Dataset C, we employed the low-pass filtered PSD of a surface with $h_0 = 0.71$ mm. The low pass filter simply cuts high frequencies off completely. The resulting surfaces have, therefore, a slightly diminished rms roughness, which was computed a posteriori, using (3.10). The resulting range for h_0 is shown in Table 5.1. When discussing the results obtained with Dataset C, q_1 will denote the employed short-distance cut-off wave number, whereas by $\hat{q}_1$ will be denoted the theoretical maximum value, therefore $\hat{q}_1 = 2\pi/a$.

We show examples of surfaces belonging to datasets B and C in Figure 5.1, because the influence of H and q_1 is less intuitive than that of h_0 (Dataset A).

Dataset	H [-]	h_0 [m]	$q_1/\hat{q}_1$ [-]
A	0.7	$5.3 \cdot 10^{-5} - 7.1 \cdot 10^{-4}$	1
B	$0.5 - 0.9$	$7.1 \cdot 10^{-4}$	1
C	0.7	$6.9 \cdot 10^{-4} - 7.1 \cdot 10^{-4}$	$\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1$

Table 5.1. Parameters range of generated surfaces. Every surface consists of 300×300 points, spanning 3 cm $\times$ 3 cm, therefore $L = 0.03$ m and $\delta_{\mathscr{O}} = 0.1$mm. The short distance cut-off wavelength $\hat{q}_1 = 2\pi/\delta_{\mathscr{O}}$, and in Dataset C we use different fractions of it. For all datasets, the smallest possible wave-length q_L is $2\pi/L$ and the roll-off wave number q_0 is $5q_L$.

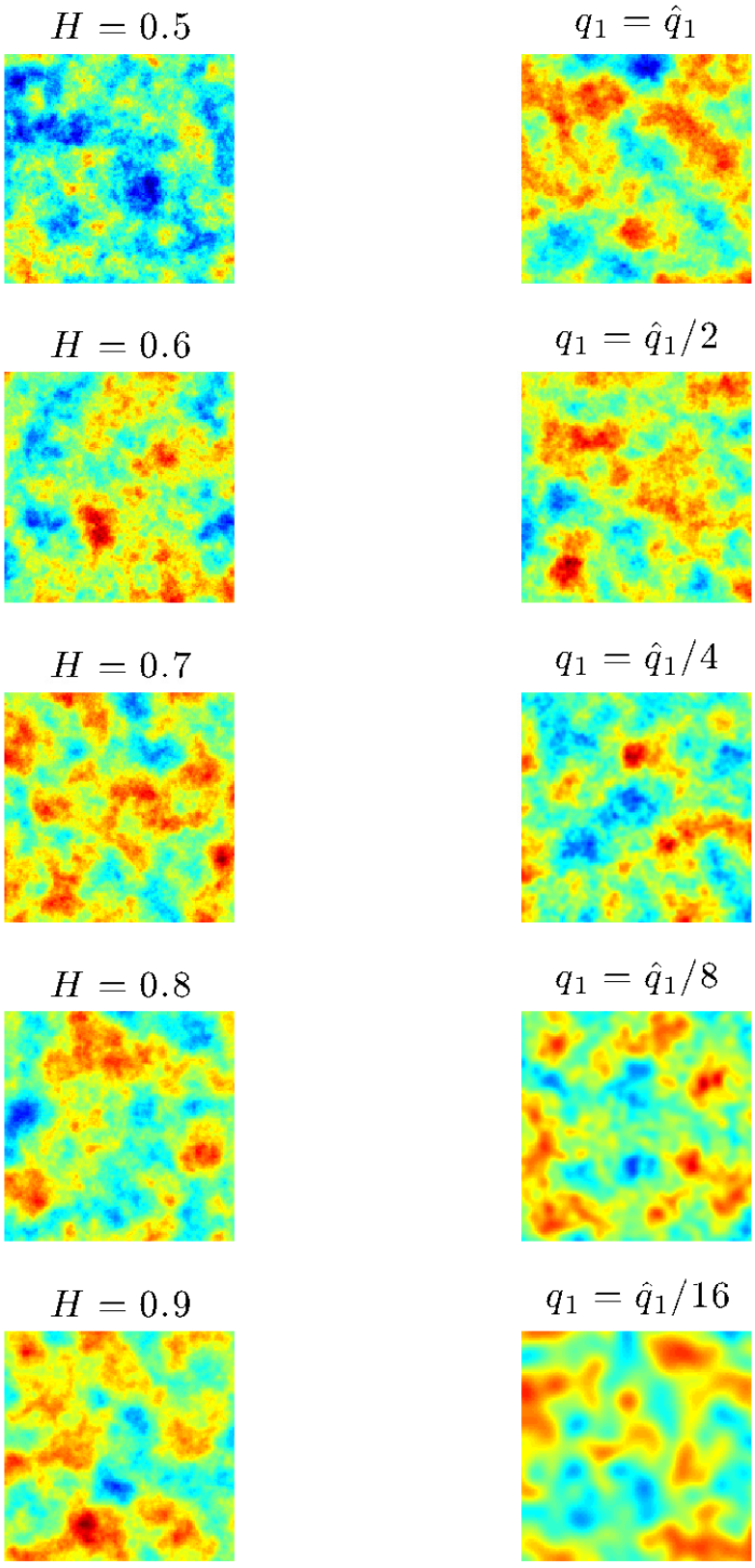

Figure 5.1. Surfaces of datasets B (*Left*) and C (*Right*). Every plot is a one fourth of the original surface it belongs to.

5.1.2 Characterization of the elastic cube

To analyze boundary and size effects, we performed a subset of the normal load experiments on three perfect cubes, differing for their lateral size. The first cube has lateral size $L_c^{\mathrm{M}} = 3$ cm, which is equal to the sample linear size (the smallest period of the rough surface): this ensures that the contact is representative for the surface, that is, every point of the rough surface domain can theoretically get in contact with the elastic cube. Smaller cubes are not considered, as their results could be biased by the local geometry of the corrugated substrate with which they would be in contact. The geometry of a portion of a specimen could be not representative of the whole surface. The second cube has lateral size $L_c^{\mathrm{L}} = 6$ cm, and for the third it holds $L_c^{\mathrm{XL}} = 12$ cm. The superscripts M, L, and XL, will be used to refer to the corresponding elastic cubes. When not specified differently, cube M was employed.

To account for spatial discretization error, on the contact surface we employed four different mesh step-sizes δ_{Γ_C} : 100 μm, 50 μm, 25 μm, and 12.5 μm. Notice that these step-sizes correspond to $\delta_{\mathscr{O}}$, $\delta_{\mathscr{O}}/2$, $\delta_{\mathscr{O}}/4$, and $\delta_{\mathscr{O}}/8$, respectively. When not specified, $\delta_{\Gamma_C} = 50$ μm. For each discretization, it always holds that the tetrahedra which have a side on Γ_C have three edges of length δ_{Γ_C}, oriented along x, y, and z axes. The remaining edges have length $\sqrt{2}\delta_{\Gamma_C}$. As pointed out in Section 4.3.4, since the contact side is refined more than the rest of the cube, the resulting mesh is denser when close to Γ_C and this is shown in Figure 4.8.

The Young modulus E was chosen unitary, since in a linear elasticity static setting with Dirichlet boundary conditions and zero Neumann boundary condition, it plays the only role of linearly scaling all forces.

Finally, to judge the influence of the Poisson's ration ν, introduced in Section 2.1, we performed a set of experiments varying its value from 0.05 (almost purely compressible material) to 0.45 (nearly incompressible material), with an intermediate value of 0.25. When not specified, ν was set to 0.45, as we are more interested in incompressible materials, than in compressible ones.

5.2 Normal load: validation against state-of-the-art models

The first phase of the numerical study, consisted in a series of normal load experiments (denoted as **Problem A** in Section 4.2.1), to study the evolution of the contact area with respect to the applied force, in different settings. There are two motivations for such experiments:

- as presented in Chapter 3, different contact laws exist to predict the results of normal loading, and some of their results have been shown to hold in real-world experiments, under certain conditions;

- the normal load setting is useful to characterize the contact type with known

measures such as separation height and real contact area, and this serves as a preliminary result for shear tests.

We illustrate the strategy we adopted and we start by analyzing contact at low pressures, where asymptotic linear behavior is expected, then we analyze the full normalized pressure range, proposing a polynomial law for which the coefficients show sufficient stability to allow for predictions in an appropriate range of the parameters.

5.2.1 Normal load strategy

The goal of these experiments is to observe the contact evolution for the whole possible pressure range. To do this, we displaced the top of the cube with uniform increments, until the total load $\sigma_0 \approx E$ was reached. We are aware that linear elasticity is not completely reliable at such high pressures, but nevertheless, the results we found provide some useful bounds on the contact evolution. The step-size s of the incremental displacement is proportional to the root mean square roughness of the corrugated obstacle employed in the simulation, and it holds

$$s \approx \frac{L_c^i}{L_c^{\mathrm{M}}} 2h_0. \tag{5.1}$$

We note that we tested different incremental steps, but this did not have any influence on the computed results, and we decided for an increment size which was optimizing computational costs, but still giving a reasonable amount of points to perform model fitting.

5.2.2 Summary of the results

We plot all results of the experiments, to emphasize differences and similarities between the measured responses of the system. Figures 5.2, 5.3, 5.4, and 5.5 show the evolution of the contact area with respect to the normalized pressure, defined as

$$\sigma_0^* = \frac{\sigma_0}{E^*} \left(\sqrt{\left\langle |\nabla h|^2 \right\rangle} \right)^{-1} \tag{5.2}$$

In all plots, two key points result evident:

1 Full contact is not always achievable, especially for very rough surfaces ($h_0 >$ 0.3 mm), for which it is not even possible to arrive to a saturation of the area, because of the limitations of linear elasticity.

2 The evolution of the contact area at large pressures is very difficult to predict by analyzing the low pressure regimes. In fact, initial slopes are similar for the majority of the datasets, even when divergence of the final configurations is observed.

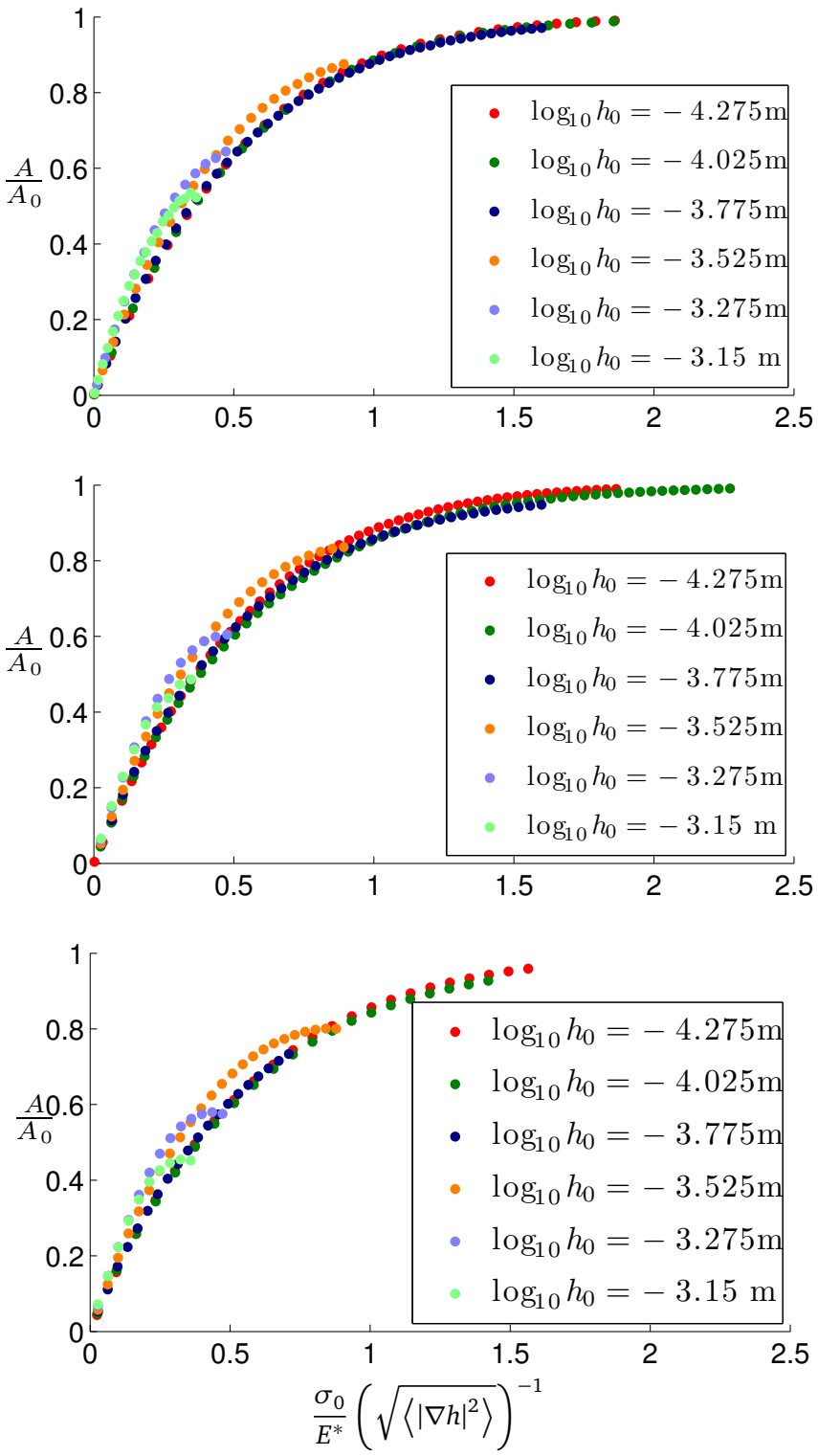

Figure 5.2. Normal load results for Dataset A, with cubes of different sizes. *Top*: cube M. *Middle*: cube L. *Bottom*: cube XL. In the plots, results have been subsampled, for graphical purposes.

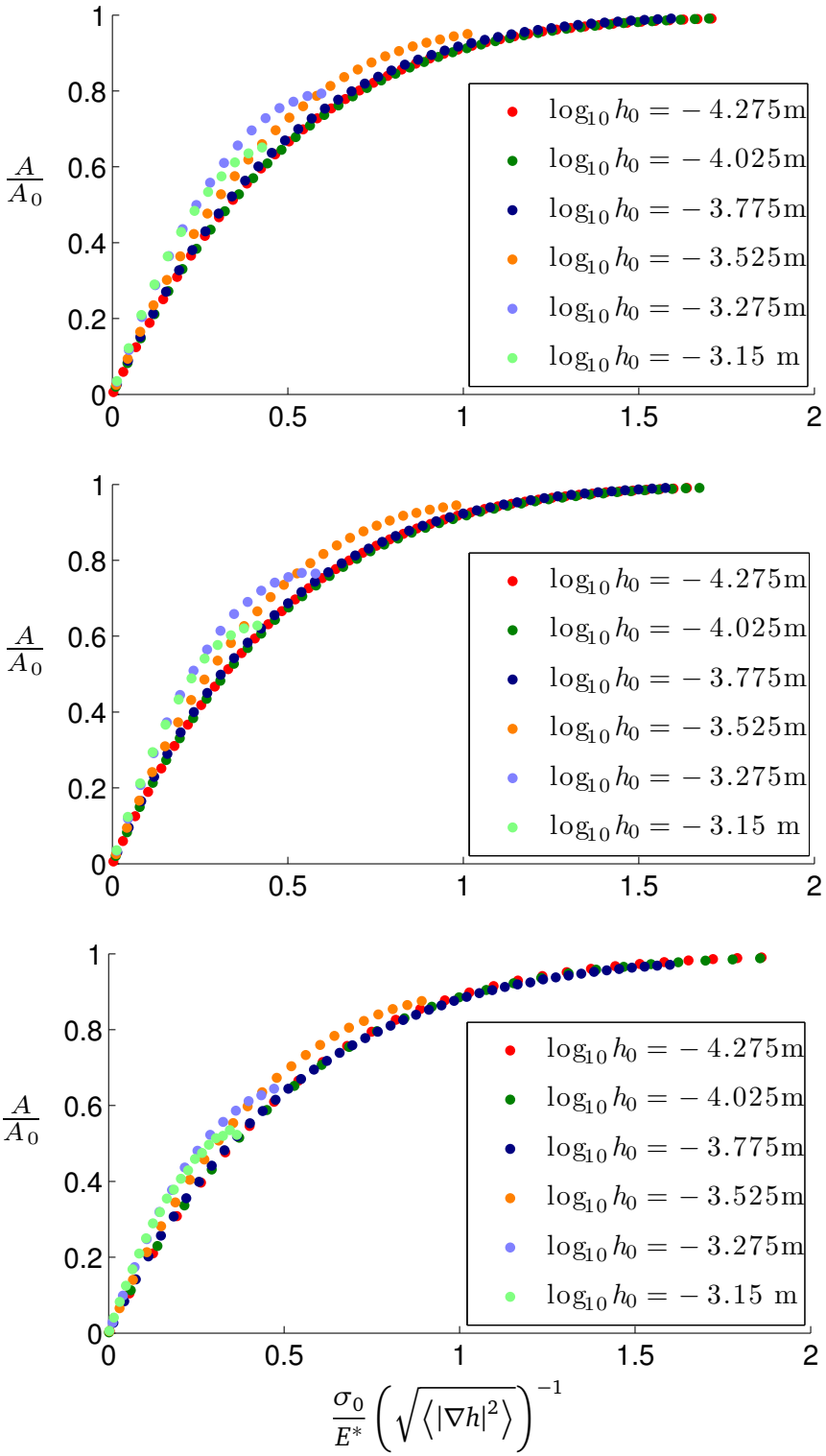

Figure 5.3. Normal load results for Dataset A, with different values of the Poisson ratio ν. *Top*: $\nu = 0.05$. *Middle*: $\nu = 0.25$. *Bottom*: $\nu = 0.45$. In the plots, results have been subsampled, for graphical purposes.

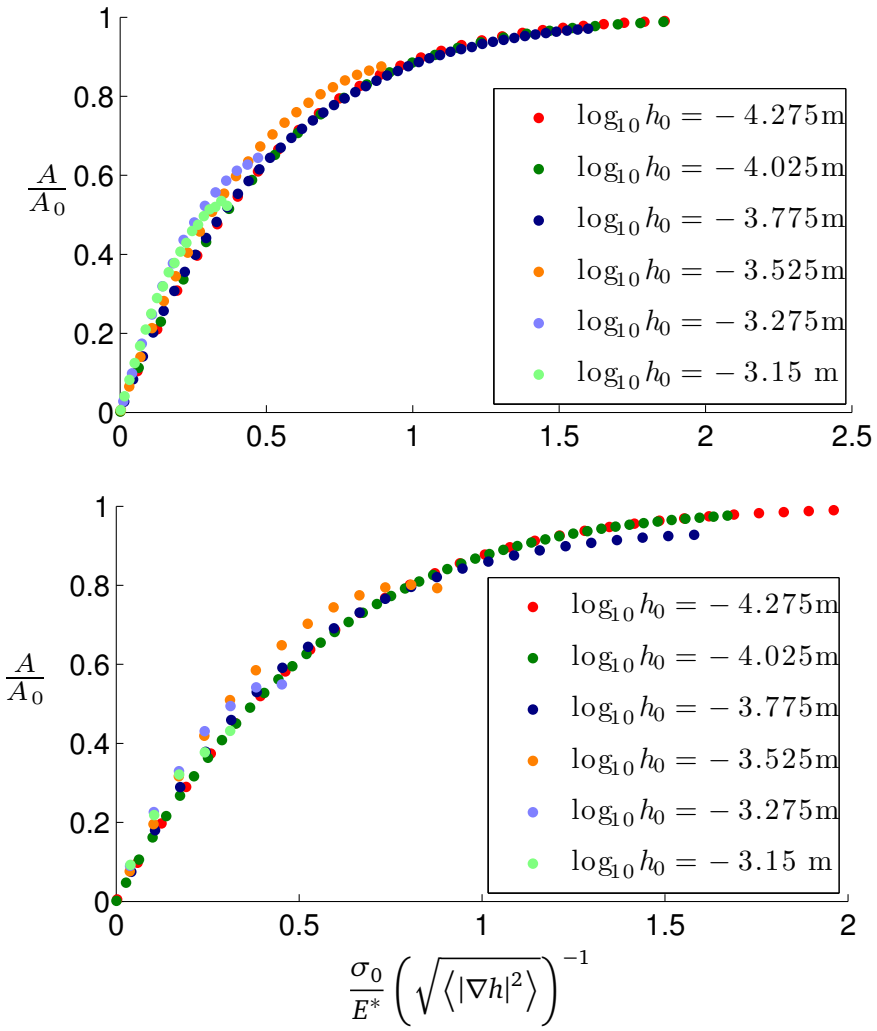

Figure 5.4. Normal load results for Dataset A, with nearly incompressible material and cubes of mesh step sizes. *Top*: $\delta_{\Gamma_C} = 50\ \mu$m. *Bottom*: $\delta_{\Gamma_C} = 25\ \mu$m. In the plots, results have been subsampled, for graphical purposes.

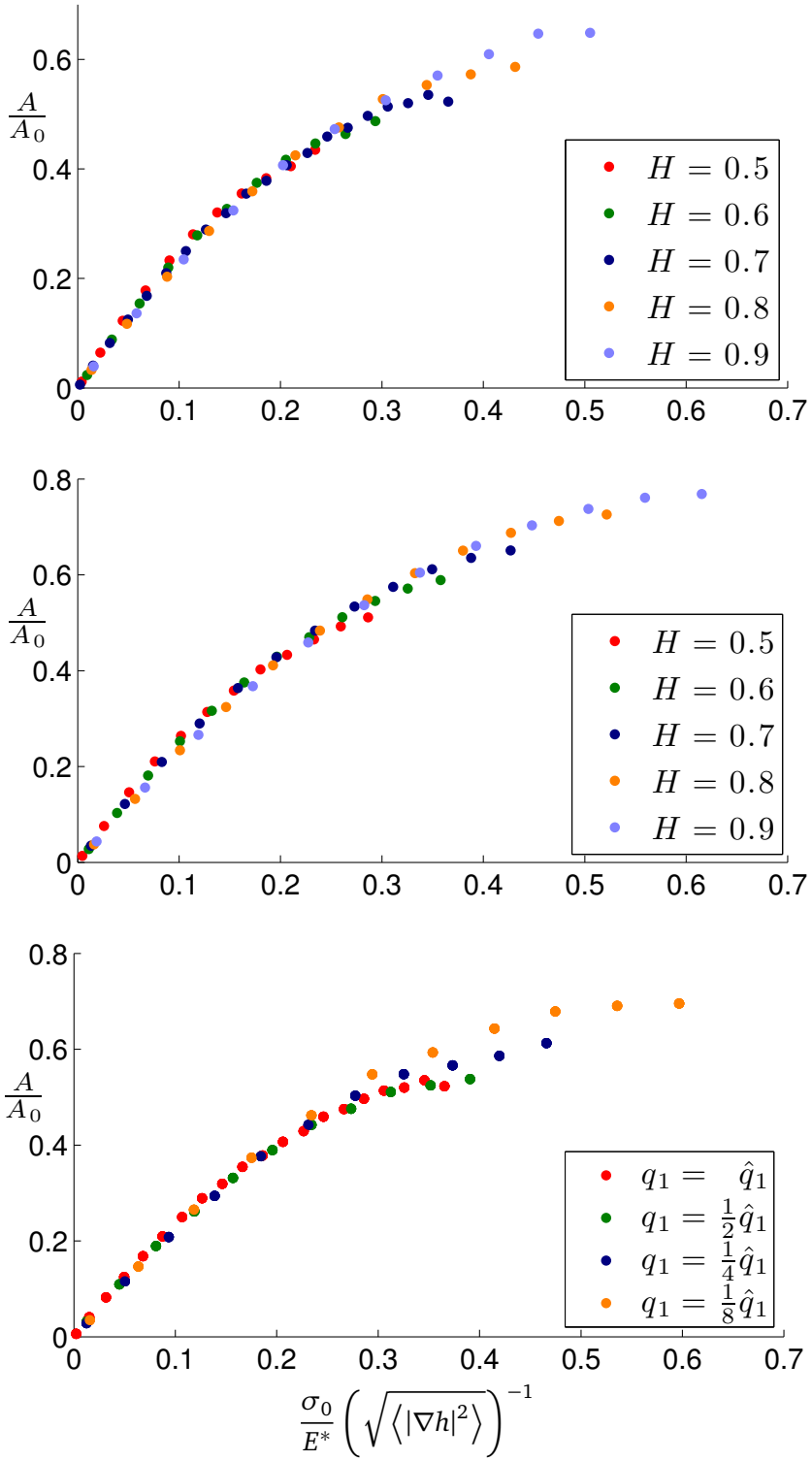

Figure 5.5. Top: Normal load results for Dataset B, and almost compressible cube $\nu = 0.05$. *Middle*: Normal load results for Dataset B and nearly incompressible cube $\nu = 0.45$. *Bottom*: Normal load results for Dataset C, and nearly incompressible cube. In all plots, $\delta_{\Gamma_C} = 50\ \mu$m.

5.2.3 Contact evolution at low pressures

Persson's theory and the BGT model derive a linear law for contact at low pressures, where the real contact area A is below 15% of the nominal contact area A_0. As pointed out by equation (3.169), the two laws can be written as

$$\frac{A}{A_0} = \kappa \frac{\sigma_0}{E^*} \left(\sqrt{\left\langle |\nabla h|^2 \right\rangle} \right)^{-1}$$

where σ_0 is the nominal pressure. The equivalent elastic modulus E^* is defined as $E^* = E/(1-\nu^2)$, with E being Young's modulus and ν Poisson's ratio. By $\sqrt{\left\langle |\nabla h|^2 \right\rangle}$ it is denoted the root mean square slope of the surface. The value of the constant κ is the only difference between the results of the two theories: it is $\sqrt{8/\pi} \approx 1.6$ for Persson, and $\sqrt{2\pi} \approx 2.5$ for the BGT model. We will denote the two limits as κ_{P} and κ_{BGT} respectively.

To study the influence of h_0 on κ, we simulated the contact with Dataset A, and performed a linear regression on the low pressure range data, obtaining the results visible in Figures 5.6 and 5.7. All the results, show a clear positive correlation between h_0 and κ, in other words, the magnitude of κ is proportional to that of h_0, even though a mathematical relation seems difficult to be derived. The values found for κ grow from slightly more than κ_{P} (for small h_0) to slightly less than κ_{BGT} (obtained for large h_0), a part for the cases in which $\nu \leq 0.25$, in which κ_{BGT} is exceeded. These results are in general agreement with [HPMR04], where a similar test for $\nu = 0$ was performed.

We propose an interpretation for this trend, which is based on the underlying assumptions of Persson and BGT theories. The theory suggested by Persson is based on the idea of a homogeneous diffusion of the stresses over the rough surface, driven by the load, which is something happening for small root mean square roughness values, because the summits are small and close, the contact area grows uniformly inside the nominal contact area, and contact patches tend to merge. The BGT model, instead, considers asperities individually and sums their contributions, neglecting interactions of different peaks (so-called *long distance interactions*). Although the premise of the BGT model is not met in general, it can be observed that for very rough surfaces (with large h_0), the asperities are higher and more distant from one another (compared to smoother surfaces), and they therefore act almost independently, indenting the elastic cube: contact regions are isolated and rarely merge. We think that surfaces with values of h_0 contained between the two limits of the studied range correspond to mixed cases, in which both model apply only partially, and this seems to be confirmed by the values of κ we find, which are convex combinations of κ_{P} and κ_{BGT}.

The top plot of Figure 5.6 shows the influence of the size of the elastic cube: larger cubes correspond to smaller relative contact areas. This might happen because the pressure distribution imposed by the Dirichlet conditions is parabolic (as in Hertzian contact), and at the boundaries of the cube, when the size of the cube is sufficiently large, it might become too weak to achieve full contact. The same phenomenon is

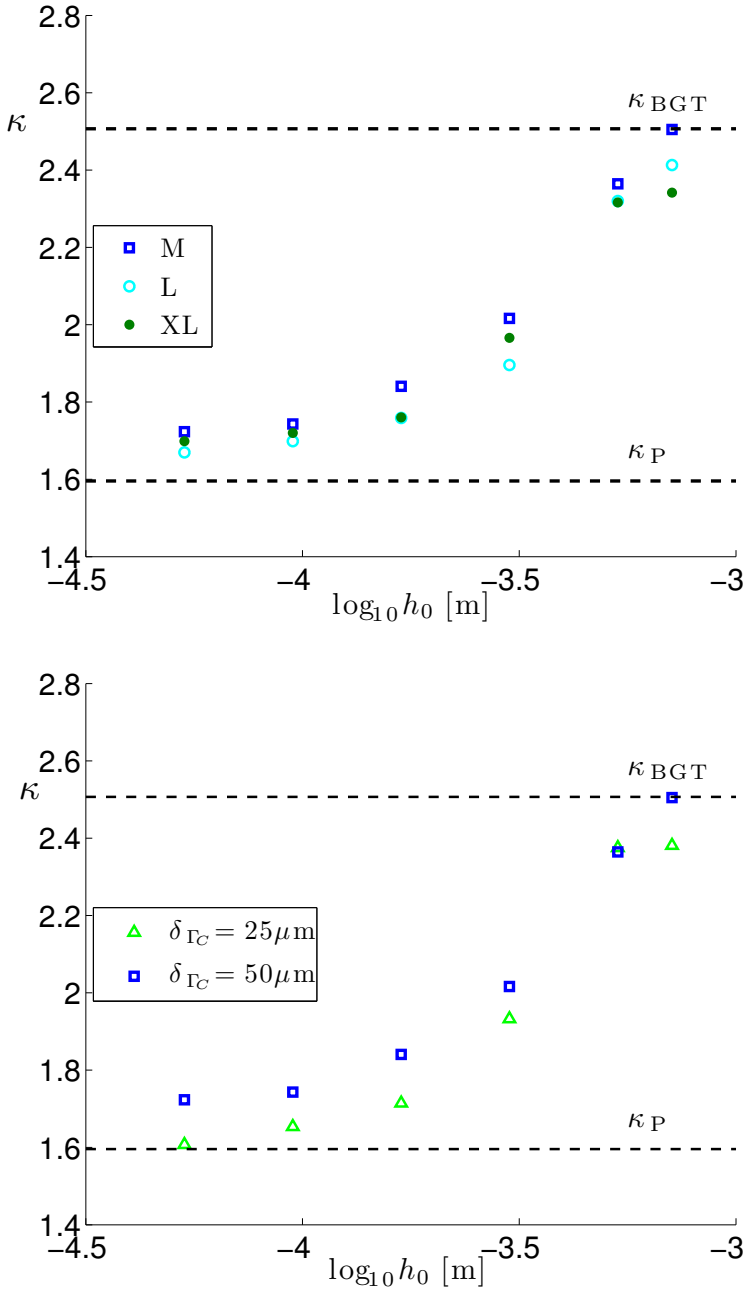

Figure 5.6. Value of κ, obtained by means of linear regression on the results of normal load tests performed on Dataset A at low pressures. *Top*: Influence of cube size with δ_{Γ_C}. *Bottom*: Influence of mesh size on a cube M.

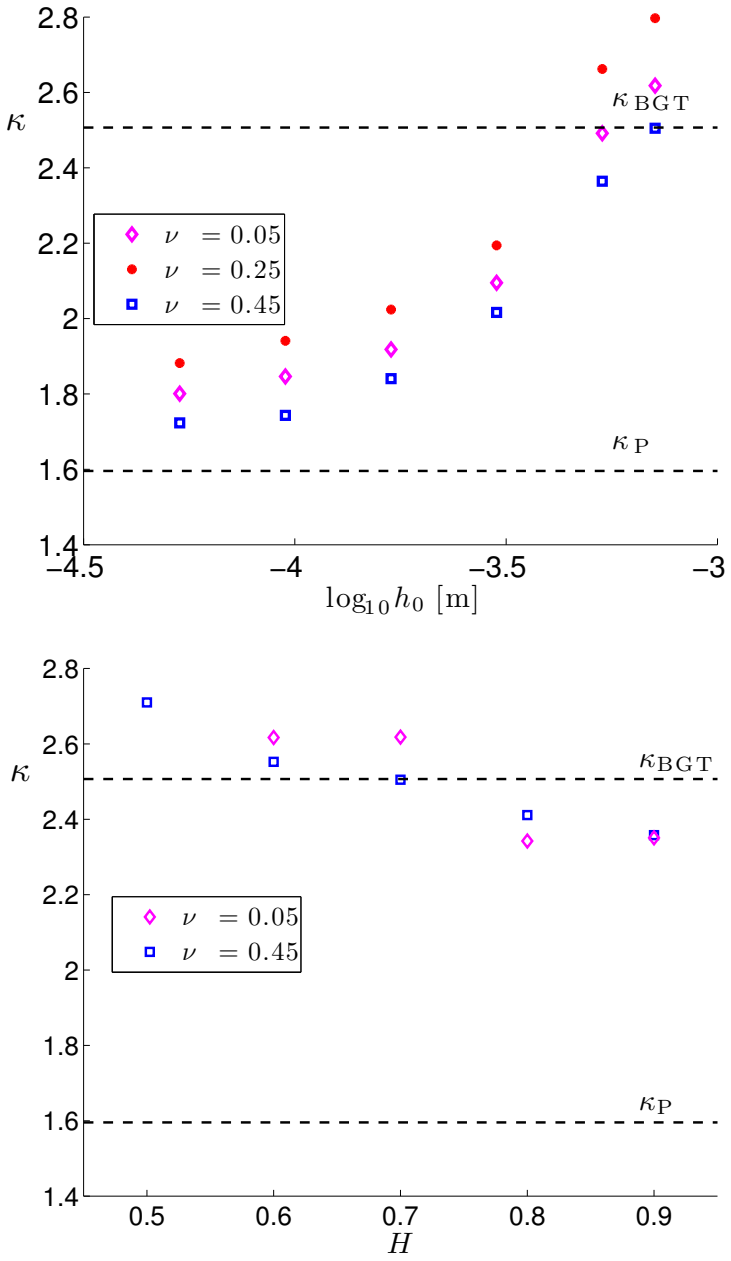

Figure 5.7. Value of κ, obtained by means of linear regression on the results of normal load tests performed on Dataset A (*Top*) and B (*Bottom*) at low pressures. Influence of Poisson's ratio.

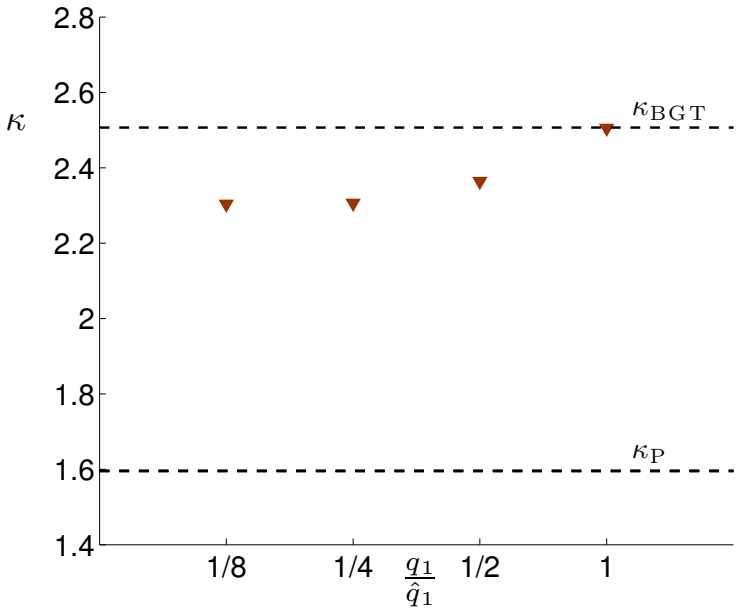

Figure 5.8. Value of κ, obtained by means of linear regression on the results of normal load tests performed on Dataset C at low pressures.

observed for every value of h_0. The bottom plot of Figure 5.6 shows the influence of the mesh size: reducing the mesh step-size, the area of contact results smaller. The same phenomenon is observed for every value of h_0, except for the case in which $h_0 = 0.53$ mm, in which the two values of κ coincide. In our simulations, the area is computed as the ratio between contact and total nodes, thus, when the mesh is finer, every node represents a smaller area, and the approximation of the real area is expected to be more accurate. In Section 5.3 we will face the same effect, and we will use four different mesh sizes, to understand the convergence of the mesh-related error. Finally, Figure 5.7 shows the influence of the Poisson's ratio. It can be observed that the minimum value of κ at each value of h_0 is achieved for $\nu = 0.45$, thus for a nearly incompressible material, whereas the maximum is obtained for $\nu = 0.25$ and the case of an almost compressible material with $\nu = 0.05$ results in an averaged value. This is in contrast with the almost linear positive correlation between area fraction and Poisson's ratio found in [HPMR04], and we suggest a possible interpretation of this difference in Section 5.2.5. We ran experiments on Dataset B and found that H plays a marginal role, as observed in the top plot of Figure 5.7, and this is confirmed in [HPMR04]. From the experiments on Dataset C, it is evident that cutting the last half of the power spectrum has a larger influence than successive reductions of q_1, see Figure 5.8, but the effect seems anyhow to be moderate, as suggested in [YDPC11].

5.2.4 Contact evolution at medium and large pressures

When pressure is increased, and area of contact is larger than 20%, the linear relation is not reliable as an approximation. The result of Persson's theory given in (3.162) can be rewritten in terms of normalized pressure, and it becomes

$$\frac{A}{A_0} = \operatorname{erf}\left(\frac{\sigma_0}{E^*}\sqrt{\frac{2}{\sqrt{\langle|\nabla h|^2\rangle}}}\right). \tag{5.3}$$

For the YAM mode (Section 3.5), contact is described by means of a power-law relationship, which we write again, for the sake of comprehension:

$$\frac{A}{A_0} = \left(\beta + \left(\frac{\sqrt{\langle|\nabla h|^2\rangle}}{\kappa\sigma_0}\right)^{\mu}\right)^{-\frac{1}{\mu}}.$$

We attempted to fit both laws to our results, but the results are not satisfactory. Persson's model seems to be unable to predict the medium and large pressure behavior, also because of its insensitivity to every roughness parameter except for the root mean slope, and for its asymptotic result for infinite pressure, where $A \to A_0$, which is rarely

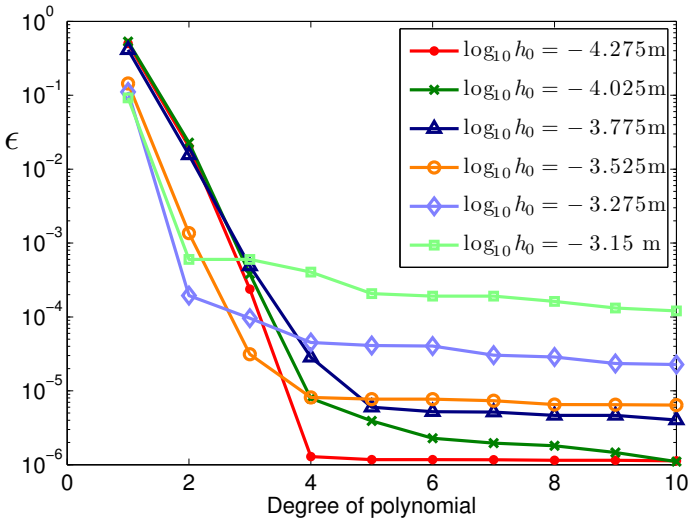

Figure 5.9. Error of interpolating polynomials $\epsilon = 1 - R^2$ plotted against degree of the polynomial, for experiment with nearly incompressible cube M and mesh size 25 μm. Notice that for moderately rough surfaces ($\log_{10} h_0 \leq -3.5$), after fourth degree, the error does not substantially decrease. The quality of the fit depends on the possibility to reach the saturation plateau with linear elasticity.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.06	0.45	-1.32	1.82	0.01
-4.02	-0.07	0.47	-1.34	1.82	0.07
-3.77	-0.14	0.74	-1.65	1.93	0.25
-3.52	-0.32	1.07	-2.02	2.16	0.07
-3.27	4.31	-3.24	-1.87	2.51	0.35
-3.15	-20.68	14.30	-6.34	2.86	3.42

Table 5.2. Coefficients of reference polynomial for Dataset A.

observed in our experiments. The YAM model, instead, can be fitted, and the coefficient of determination R^2 (c.f. [Ric06]) is sufficiently high ($R^2 > 99.9\%$ for most datasets), but for large pressures, the relative area largely exceeds the measured one. Since we are interested in an interpolation of the data, to perform a simplified analysis of the behavior, we tried different interpolating functions (logistic, Gudermannian, different types of sigmoidals), and the best fits were given by polynomials of fourth degree. The constant term of the polynomial was found to be always smaller than the other coefficients by three orders of magnitude, thus, we chose to work with polynomials without constant term. Notice that in most cases, even for experiments with more than 50 results, for the polynomials it holds $R^2 > 99.9995\%$. Adding further orders to the interpolation, does not produce a better approximation, as it can be observed in Figure 5.9, where the error of the fit, computed as $1 - R^2$ is plotted against the order of the polynomial, for different surfaces.

In Figure 5.10 the two interpolations are shown for one dataset, and in the close-ups, it can be observed how the polynomial fitting follows the data more closely, especially for very small and very large pressures. The results are plotted against the normalized pressure σ^* defined in (5.2). We can write our empirical polynomial law as

$$\frac{A}{A_0} = \Pi {\sigma^*}^4 + \Omega {\sigma^*}^3 + \Lambda {\sigma^*}^2 + \kappa \sigma^* \tag{5.4}$$

where we kept κ to denote the coefficient of the linear term, because of the similarity (though, not equality) to the asymptotic coefficient of proportionality of the other models. We find a good stability of the parameters for moderately rough surfaces, except for the cube XL, where boundary effects seem lead to instability of the parameters, as shown in the top plot of Figure 5.13. With respect to ν and to δ_{Γ_C}, the parameters show a consistent and predictable behavior, as it can be observed in Figures 5.12, 5.13, and Figure 5.11, which we show at the end of this section.

As a reference, in table Table 5.2, we list the values of the coefficients we find for a cube M with $\nu = 0.45$, $\delta_{\Gamma_C} = 50\ \mu$m, and the complete list of the coefficients we found for the different settings is given in Appendix C, together with the error $\epsilon = 1 - R^2$ for

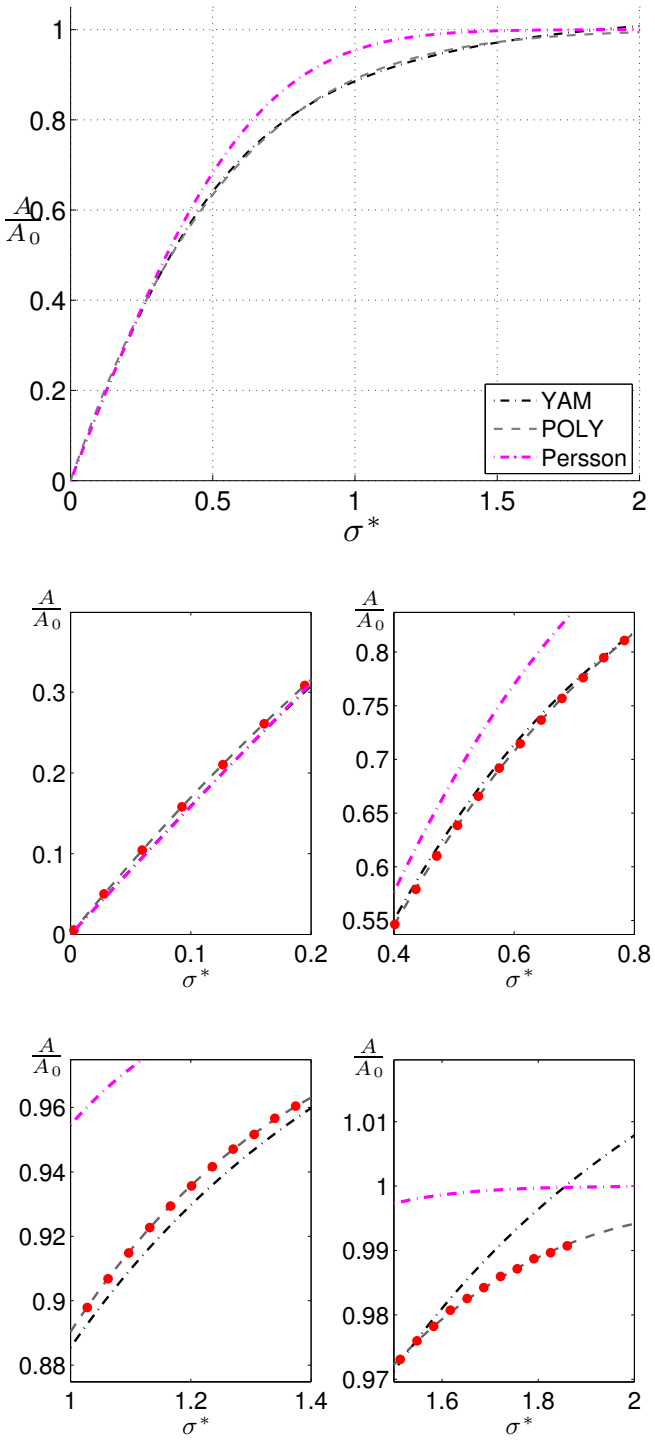

Figure 5.10. Comparison of YAM contact law, Persson's model and polynomial fit. Relative area of contact plotted against normalized pressure σ^*. *Top*: the full pressure range. *Bottom*: close-ups of specific regions of the pressure range, where the fits diverge (notice that in the top left plot, Persson's model overlaps with the polynomial interpolation). Experimental results are displayed as red dots.

every fit. In the appendix, it can be noted that the precision of the polynomial fit of problems of the same size and spatial discretization increases for

- lower values of h_0
- lower values of q_1
- lower values of ν.

Thus, the quality of the polynomial interpolation is inversely proportional to the perceived roughness of the surface, and to the Poisson's ratio (therefore to the transversal expansion of the cube).

We remark that the polynomials we find are a good interpolation of experimental results, but they are generally not good for extrapolation for larger pressures than those for which data points are measured. Anyhow, for moderately rough surfaces, where the plateau region can be reached, all the polynomials possess a maximum, which is located at a pressure which is slightly larger than the limit one we used in our experiments. We think that the value of σ^* where the area of contact would reach the maximum could be a good estimate of the end of the validity range of the polynomial, and after that value, the area can be considered constant, according to our results.

5.2.5 Differences from other studies

Even though at low pressures, our results are compatible with the asymptotic predictions of other models, and with numerical results of other studies, when going to larger pressures, we observe a slightly different behavior. A possible source of difference, from works like [HPMR04], resides in the behavior at the boundaries of the contact area.

In our experiments, given the null Neumann conditions of (2.35), the cube can expand in x and y directions. On distant (expanded) boundaries, the vertical pressure is less than in the center of the area of contact (since, macroscopically, the pressure distribution is parabolic), and this could result in a different contact pressure, which in turn influences the maximum achievable area of contact.

We also notice that we mostly studied nearly incompressible solids, whereas it seems that in other studies, e.g. [YAM12], a low value of ν is usually employed, to limit transversal expansion. We remark that large values of ν make the numerical problem harder to solve, which is a well known issue of FEM simulations: for example, in a typical run, the time to compute the correct (iterated) solution of one single step, can double, when going from $\nu = 0.25$ to $\nu = 0.45$.

The boundary effects caused by transversal expansion are not included in any analytical theory that we have presented. At low pressures, for small values of ν, they can be neglected, but for materials like rubber, which undergo large strains, it might be an inaccurate approximation.

5.2.6 Periodic boundary conditions

To mitigate the effect of transversal expansion, it was suggested to apply periodic boundary conditions on the elastic cube, in x and y direction. This could in principle work for normal load experiments, but there exists the possibility it gives rise to numerical instabilities, because of the rough nature of the obstacle. In fact, let $\mathbf{x}'$ be a point on Γ_C and let it further be on one lateral side of the cube, for example on the side facing negative x direction, so that $\mathbf{x}' = (0, y', z')$, in a periodic setting. Now let us assume $\mathbf{x}'$ is in contact with the obstacle $\mathcal{O}$. Then, at every solution step, the point $\mathbf{x}'$ is in contact with point on $\partial\mathcal{O}$, but since we are enforcing periodicity, $\mathbf{x}'$ also corresponds to the point on the other side of the cube $\mathbf{x}'' = (L, y', z')$. Thus, only if the periods of the cube and of the rough surface correspond *exactly*, it is theoretically possible that the contact is solved correctly, because in that case, the points on $\partial\mathcal{O}$ which are in contact with $\mathbf{x}'$ and $\mathbf{x}''$ are exactly at the same height (being them at distance L), and thus the contact is solved correctly. But as soon as machine error is present, the coordinates of the contact points could differ slightly, and cause instabilities in the solver. Moreover, in the shear tests we present in Section 5.3, we observe compression of the cube with respect to x, thus the two points of contact would not be at distance L anymore, resulting in inconsistency of closest obstacle point, and impossibility of convergence.

5.2.7 Conclusive remarks

In this section, we measured the influence of roughness parameters both on the low and on the high pressure ranges. We showed that full contact is not always reachable, especially for incompressible solids, at least using linear elasticity. We also showed that low pressure predictions are less influenced by roughness characteristics, thus, linear regressions show stable values of the parameter κ.

5.2.8 Plots of polynomial coefficients for different numerical settings

We provide plots of the polynomial coefficients we found as fits of the different numerical results. When possible, we plot ranges in which the parameters show stability, or predictable and consistent behavior. A complete list of the parameters is to be found in Appendix C. We recall that the we employed a fourth order polynomial without constant term, of the form $\Pi\sigma^{*4} + \Omega\sigma^{*3} + \Lambda\sigma^{*2} + \kappa\sigma^{*}$. An error estimation of each polynomial regression can be found in Appendix C, where it is scaled by 10^4, to represent its first significant digits.

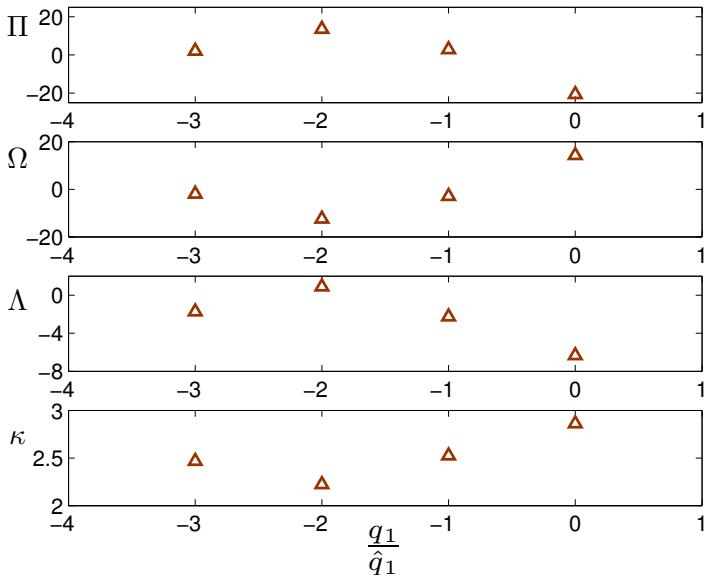

Figure 5.11. Value of the polynomial coefficients obtained by fitting the fourth order polynomial to the results of Dataset C, for a cube with $\delta_{\Gamma_C} = 0.5$ μm and $\nu = 0.45$.

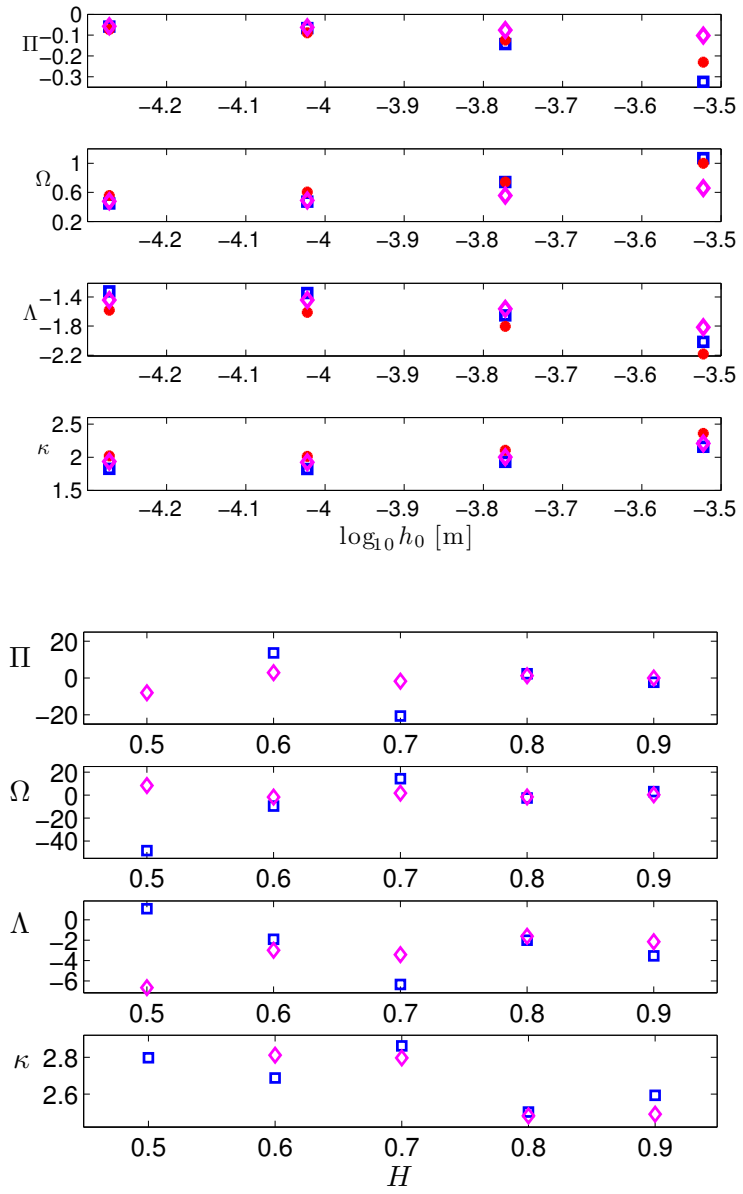

Figure 5.12. Value of the polynomial coefficients obtained by fitting the fourth order polynomial to the results of Dataset A. Influence of Poisson's ratio: magenta diamonds represent results obtained for $\nu = 0.05$, red dots for $\nu = 0.25$, and blue squares for $\nu = 0.45$. For $H = 0.5$, some coefficients are out of the plotted range.

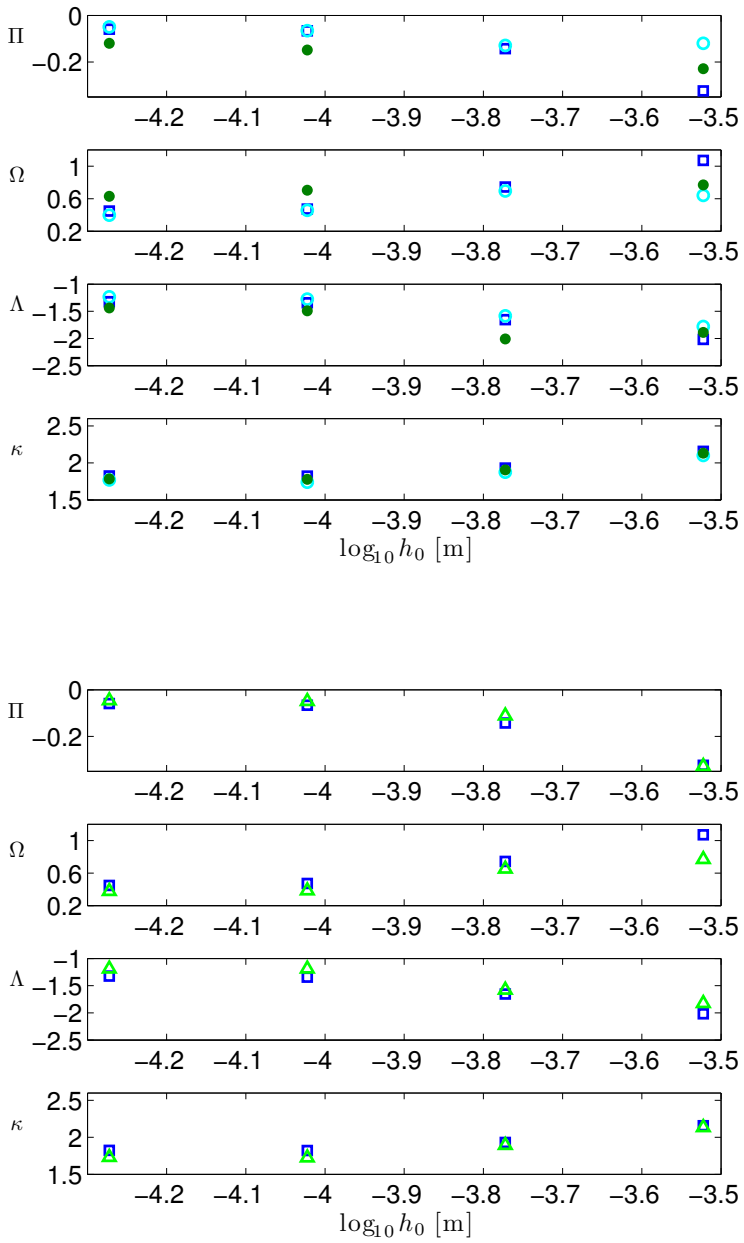

Figure 5.13. Value of the polynomial coefficients obtained by fitting the fourth order polynomial to the results of Dataset A. *Top*: Influence of cube size: blue squares represent results for cube M, cyan circles for cube L, dark green dots for cube XL. For the cube XL, some coefficients are out of the plotting range, because of boundary effects. *Bottom*: Influence of mesh size: blue squares represent results obtained for $\delta_{\Gamma_c} = 0.5$ μm, light green triangles for $\delta_{\Gamma_c} = 0.25$ μm.

5.3 Computation of elastostatic friction by means of shear tests

As mentioned in Section 1.3, friction is a very complicated phenomenon. It is still not fully understood, and its effects are sometimes unpredictable. Many factors concur to build friction, and we recall that, among these, those which are believed to be the most influent are:

F1 Viscoelastic response of the materials in contact.

F2 Macro- and micro- texturing of the surfaces.

F3 Chemical, or atomic-scale physical interactions between boundary atoms.

A full study of such a complicated phenomenon is too ambitious for the present work. We therefore focused on one particular instance of friction, in which less factors intervene, setting the problem in a simpler framework, which is suitable for our FE approach.

Thus, in this section, we aim to numerically measure elastostatic friction, between moderately rough and highly rough surfaces. In the quasi-static setting, **F1** can be neglected, as we are studying a system evolving at infinitesimal speed (or, equivalently, we assume infinite relaxation time), thus the constitutive law for linear viscoelastic behavior (2.33) becomes identical to that of linear elasticity, (2.28), and viscous effects do not show up. Using large values of h_0, we limit ourselves to micro-scale roughness, disregarding roughness at the nano-scale, and we are in a setting in which adhesive forces should be negligible, as stated in [Per01b], thus, **F3** can also be neglected, and we will mathematically characterize **F2** and quantify its influence.

5.3.1 Experimental setting and results

In the unconstrained case, a prescribed displacement of the top of the cube results in the same displacement of the bottom face (as in a rigid body translation), and we observe the same behavior if we use a flat surface as obstacle. Due to the contact constraints and to the surface roughness, the solution of Problem B is a configuration of the cube in which the elastic shear force induced by the displaced top is balanced by the sum of the forces transferred by the rigid asperities on the bottom side of the cube.

Our experiments are subject to some feasibility limits. As explained in Section 5.1, we investigated the influence of the mesh size on the experimental results. We believed this mesh size had to be related only to the lattice constant of our rigid obstacles, but we found that there is a strong correlation between the convergence of our results and the rms roughness h_0. As it turns out, for datasets with $h_0 < 0.3$ mm, the needed mesh size of the elastic cube is one eighth of the obstacle lattice constant, and this implies that, for obstacles which are represented by 300×300 points, we need to

solve a non-smooth contact problem with approximately 45M unknowns. We note that such computations are possible only thanks to our parallel implementation of an optimal multi-grid solver (c.f. [Kra09]), but one single complete shear test, performed on 2048 processors, can last up to 24 hours. For this reason, we decided to perform a limited amount of simulations for such finer meshes, and studies of variance for intra-set measurements are performed only for coarser meshes.

Another limitation to our approach comes from the fact that we employ a linear elastic material, but to obtain measurable friction forces on almost smooth surfaces ($h_0 \approx 0.0001$ m $\approx \delta_\varnothing$), we need to apply a large load on top of the cube. Thus, we decided to impose a fixed displacement for all experiments, and based on the experience we earned on the normal load experiments, we set the normal displacement to 9 mm, which results in an area of contact larger than 20-30% for all datasets. We stress again that we are using a purely elastic material law, with infinite yield stress: thus no viscosity or plasticity effect will be produced.

The surface datasets are those described in Section 5.1, and the Poisson's ratio for all tests is 0.45.

General trends

We will treat every dataset separately and describe its results in a separate section, but some trends are common to all samples.

The mesh size has a larger influence on the results of the shear tests, than on those of the normal load experiments. This is due to the fact that the final configuration in the shear tests highly depends on the frictional response, because for sufficiently large friction forces, the trailing part of the bottom face of the cube (the so-called "tail" of the footprint) tends to detach from the obstacle: larger friction forces correspond to bigger tails, thus to smaller areas of contact. This also means that the area in which indenting forces are applied is smaller and, locally, much larger micro-forces (which can cause fretting and cracks) can be predicted. These effects are relevant especially for surfaces with small h_0, as the one used for the experiments shown in Figure 5.15

As it can be expected and is well-known from practice, in shear tests the stress concentrates in the front part of the footprint. Locally, around each asperity, the front part of the contact is weaker, because the elastic surface is pulled away from the rigid obstacle, whereas in the rear part, stress is larger, because the two surfaces are pushed together. These effects are relevant especially for surfaces with large h_0, as the one used for the experiments shown in Figure 5.14.

Dataset A and the influence of root mean square roughness

As already stated, Dataset A serves the purpose of determining the influence of the root mean square roughness h_0 on the contact area and on the friction force. It also shows clearly that the discretization size of the elastic cube must be chosen considering h_0.

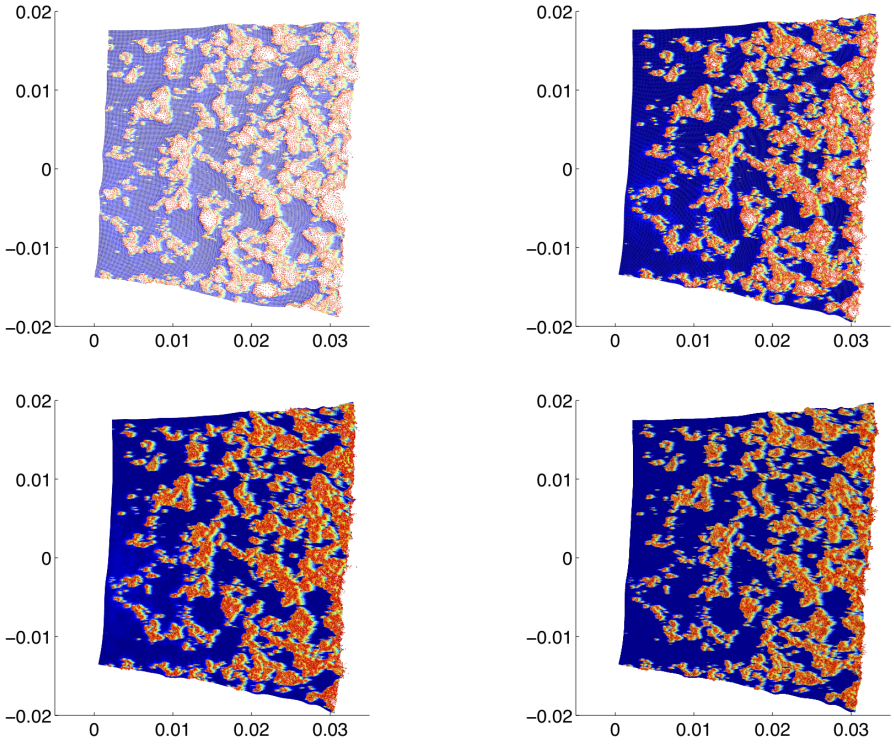

Figure 5.14. Bottom side of the elastic cube at the end of shear tests performed with different meshes, for a surface with $h_0 = 710$ μm. Even with $\delta_{\Gamma_C} = \delta_{\mathscr{O}}$ (top left image), the asperities of the rough surface block the cube, which is deformed. By increasing the mesh density by two (top right), we predict a final configuration which is already qualitatively identical to the one obtained with $\delta_{\Gamma_C} = 1/4\,\delta_{\mathscr{O}}$ (bottom left) and $\delta_{\Gamma_C} = 1/8\,\delta_{\mathscr{O}}$ (bottom right). Blue areas are not in contact with the rough surface.

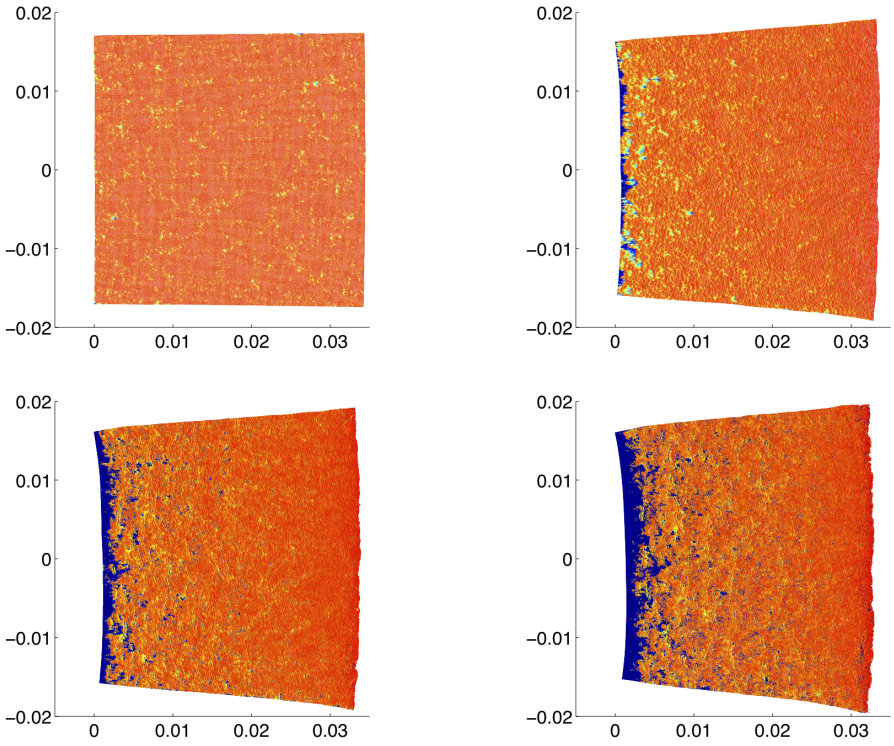

Figure 5.15. Bottom side of the elastic cube at the end of shear tests performed with different meshes, for a surface with $h_0 = 53$ μm. With $\delta_{\Gamma_C} = \delta_{\mathscr{O}}$ (top left image), the cube slides freely, undergoing only small deformation. With $\delta_{\Gamma_C} = 1/2\,\delta_{\mathscr{O}}$ (top right), a detached region (in blue) is visible. Increasing the mesh density, friction becomes larger and the detachment of the cube from the obstacle grows bigger, because of the increased deformation (bottom left image is obtained with $\delta_{\Gamma_C} = 1/4\,\delta_{\mathscr{O}}$, and bottom right with $\delta_{\Gamma_C} = 1/8\,\delta_{\mathscr{O}}$).

Quantitative Analysis Two phenomena can be observed, performing the experiments on Dataset A. First, the area of contact at the end of the normal test, shown in the top plot of Figure 5.16, shows a clear trend: it decreases monotonically with respect to h_0. Shear tests, instead, require sufficiently fine meshes, because for very small values of h_0, results are misleading when δ_{Γ_C} is $1/2\,\delta_{\mathscr{O}}$ or larger, and this can be observed in Figures 5.16 and 5.17. We studied the full range of h_0 for a mesh size of $1/4\,\delta_{\mathscr{O}}$. To interpolate the predicted value of the friction coefficient with respect to h_0 we employed MATLAB's function `polyfit` and fitted a quadratic polynomial, which attains a maximum friction value at $h_0 = 2.36 \cdot 10^{-4}$ m.

For the feasibility limits described in Section 4.3.3, we studied only a subset of the dataset with a mesh size of $1/8\,\delta_{\mathscr{O}}$. Quantitatively, results differ more relevantly for smaller than for larger values of h_0 and this can be seen in Figures 5.16 and 5.17. Based on the previous observations, we again fit a quadratic polynomial to the resulting friction and observe larger values for small values of h_0. The maximum value of this polynomial is attained at $h_0 = 2.07 \cdot 10^{-4}$ m. We speculate that for values of h_0 which are smaller than those we studied, a much finer mesh would have to be employed, but we recall that with that type of roughness, adhesion would not be negligible anymore. We stress the fact that larger friction coefficients obtained on finer meshes influence the area of contact of the sheared configuration, since it can make the detached footprint tail larger, as observed in Figure 5.15.

The contact type can vary significantly when we solve problems for obstacles spanning more than one order of magnitude for h_0. A surface with $h_0 \approx 10^{-4}$ m is smooth, compared to one for which $h_0 \approx 10^{-3}$ m. With the same imposed displacement on top of the cube, the contact for the former is almost complete (neglecting the detached tail), whereas for the latter it happens in small, fragmented areas. The vertical reaction force opposed by smooth surfaces is large, because of wide almost flat areas, and the horizontal force produced by the asperities will be small, compared to the normal load applied on top of the cube. Nevertheless, the frictional response is significantly different from zero even for surfaces with $h_0 \approx \delta_{\mathscr{O}}$.

Statistical Analysis For the computed friction coefficients and areas of contact, we studied the coefficients of variation (computed as the ratios of the standard deviation to the mean value of a distribution) and we found it to be less than 3% across the whole range of h_0. These results are listed in Table D.1 of Appendix D. We recall that the variance tests were done on the cube with mesh size $1/2\,\delta_{\mathscr{O}}$: ten surfaces with statistically equivalent roughness parameters were created and the same experiment was performed on each of them.

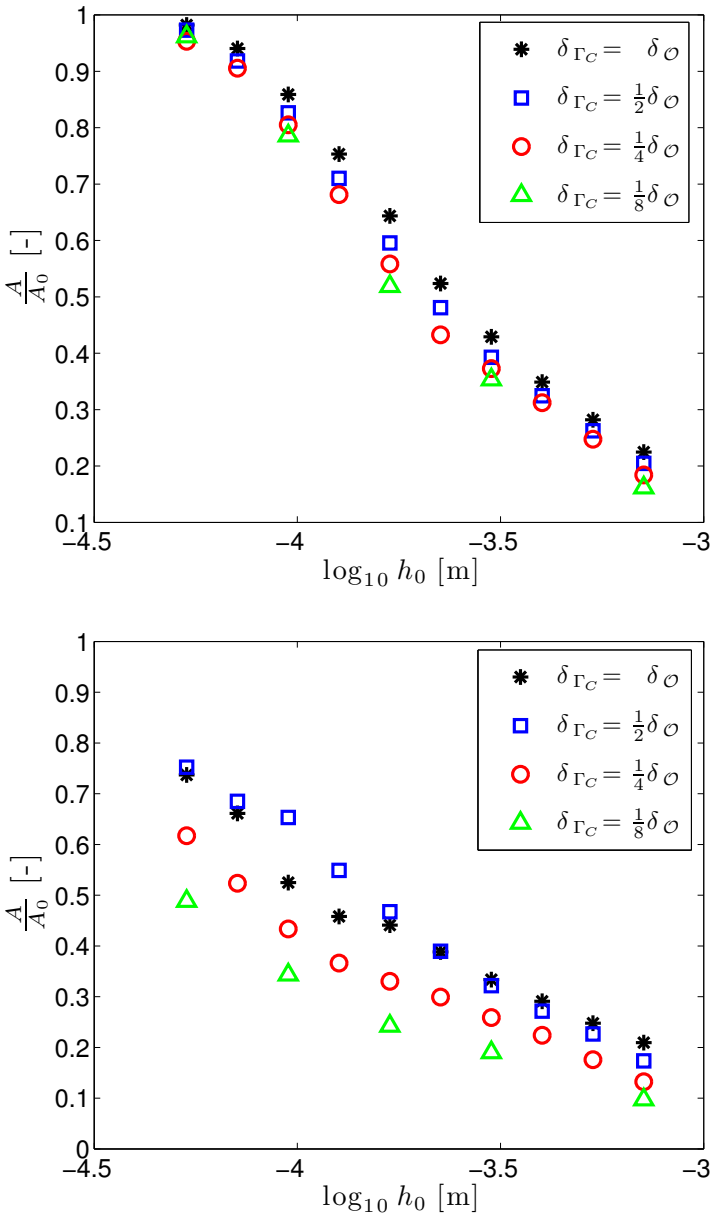

Figure 5.16. Top: Real area of contact for normal load test preceding shear phase, showing a clear trend for all refinement levels. For small values of h_0, the contact is almost complete, whereas for large values it reaches 15% of the nominal contact area. *Bottom*: Real area of contact after shearing test. For finer meshes, the computed frictional resistance force is larger, and this results in a larger detached footprint tail, which diminishes the total area of contact.

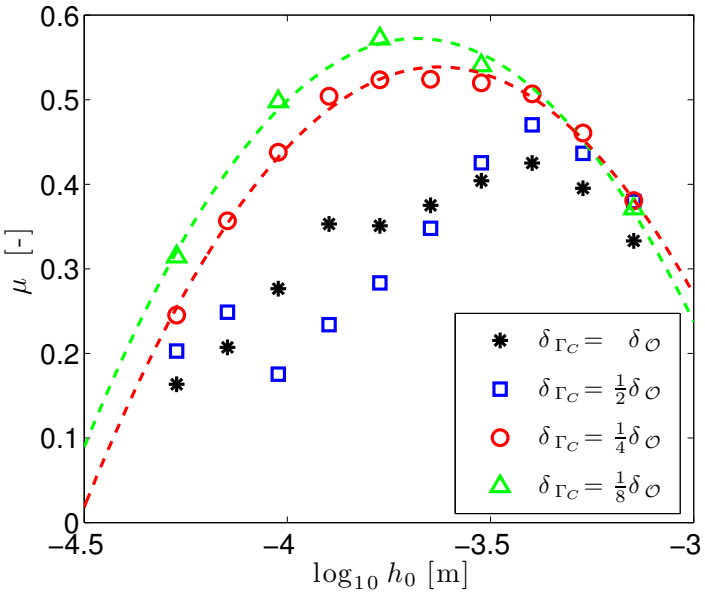

Figure 5.17. Friction computed for shear tests, varying the root mean square roughness of the obstacle and the mesh size of the elastic cube.

Dataset B and the influence of Hurst exponent

As described in Section 5.1, all surfaces belonging to Dataset B have the same statistical characteristics, except for the Hurst exponent H, which assumes values ranging from 0.5 to 0.9.

Quantitative Analysis With respect to Dataset A, we observe that refining the mesh does not change the predicted behavior: trends are correctly identified already for the mesh size of $1/2\,\delta_{\mathcal{O}}$, and results only suffer a minor discretization error, when compared to those obtained with a mesh size of $1/4\,\delta_{\mathcal{O}}$. This can be observed in Figure 5.18, where we took one surface for each Hurst exponent and compared the results obtained with two differently mesh step-sizes.

We find friction to be monotonically increasing with respect to H, as it can be observed in Figure 5.19. This effect seems to be a direct consequence of different sizes of the area of contact, shown in the top plot of Figure 5.19: for smaller H (larger D_f), the amount of microscopic asperities on the surface is larger, thus the contact is fragmented into small areas, which act with weaker intensity on the cube's elastic surface. When we fit a quadratic polynomial through the mean values of μ of every subgroup we obtain an approximation, which is visible in Figures 5.18 and 5.19.

Statistical Analysis To study the uncertainty, we again employ ten different realizations of the random surfaces for each parameter setting and the mesh size of the elastic cube is $1/2\,\delta_{\mathcal{O}}$, and we obtained coefficient of variations which are less or equal to 3% for both friction and area of contact, these results are listed in Table D.2 of Appendix D.

Dataset C and the influence of the short distance cut-off wave vector

We recall that in Dataset C, we reduce the short-distance cut-off vector. To do so, we simply filter out high frequencies of the PSD. As a consequence, h_0 is decreased. The resulting range of the root mean square roughness is listed in Table 5.1: the highest deviation we obtain from the original h_0 is less than 2%, and it is obtained for $q_1 = \hat{q}_1/8$.

Quantitative Analysis The low-pass filter influences both μ (bottom plot of 5.20) and A (top plot of Figure 5.20). By cutting high frequencies off, we smoothen the surfaces, filtering out small asperities. This reduces the rms roughness and enlarges the area of contact, as visible in Figure 5.20, because the fragmentation due to micro-asperities is eliminated, and contact happens in big smooth islands, or "patches". Friction coefficient is affected by the filtering too, and for narrower powers spectral densities, we

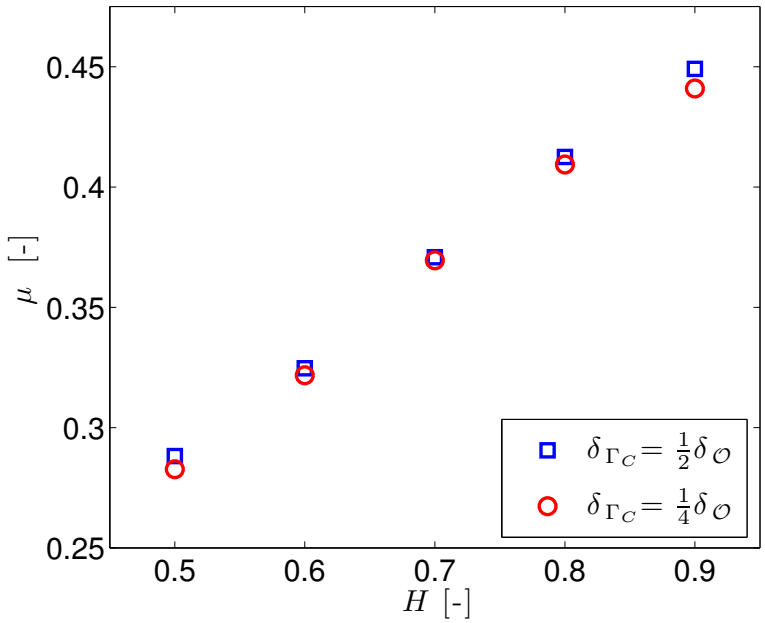

Figure 5.18. Friction computed for shear tests on Dataset B, varying the mesh size of the elastic cube.

obtain larger friction coefficients, as seen in Figure 5.20.

Statistical Analysis Results collected for the subgroups with the smallest cut-off wave vectors show larger variances and coefficients of variations, as listed in Table D.3 of Appendix D, and this can be observed for both measured quantities, friction and real area of contact. We give an interpretation of this phenomenon according to the theory of random, gaussian signals.

The results of Dataset C show that low-pass filtering has a relevant influence on

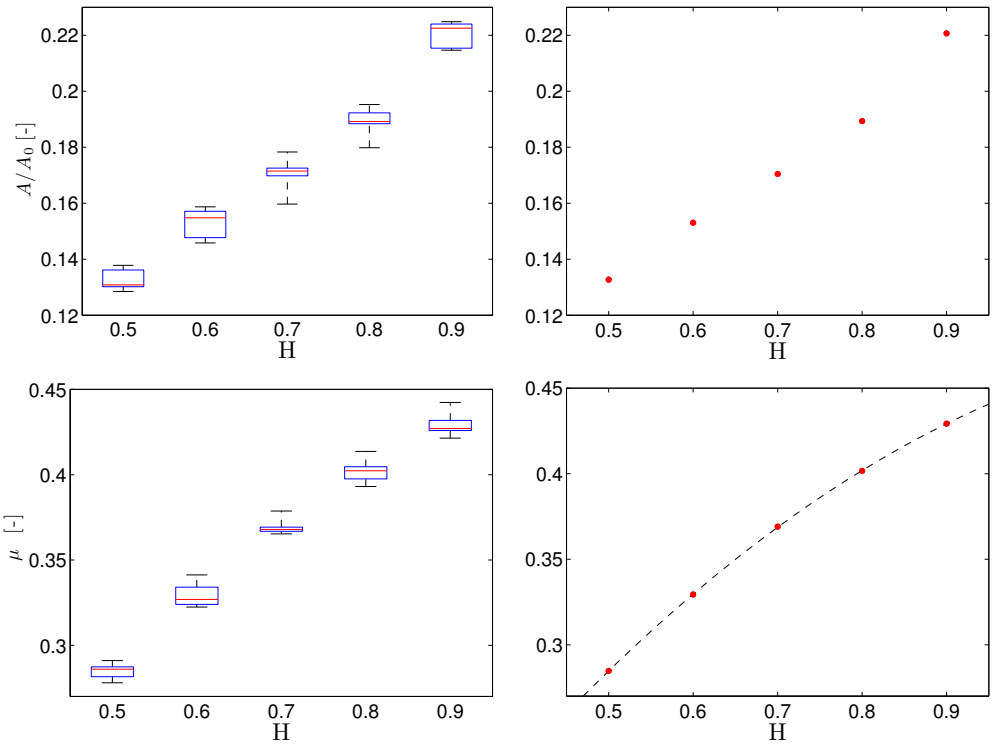

Figure 5.19. Top: Measured real area of contact for Dataset B. *Left*: The real area of contact fraction for the final (sheared) configuration of our numerical experiments. Each box represents a subgroup of ten specimens with same statistical measures. *Right*: The average area of contact for each subgroup. *Bottom*: Friction results for Dataset B. *Left*: Box plot of the computed friction coefficient μ, obtained as the ratio between horizontal and vertical forces induced by the displacement of the top of the cube. Each box represents a subgroup of ten specimens with same statistical measures. *Right*: Average friction coefficient μ for the different subgroups and quadratic fit.

the simulations, and the differences in the results of the four subgroups can not be explained in terms of the reduced h_0, for which the variation is less than 2%. We suggest an interpretation according to the theory of Nayak, presented in Section 3.1, and to some observations about variance of the numerical results, made in [YAM12]. We computed the value of Nayak's parameter $\alpha = m_0 m_4 / m_2^2$: for every subgroup of Dataset C, and the results are listed in Table 5.3. We computed the summit heights probability densities for the four subgroups and obtained the results visible in Figure 5.21, where limit distributions for $\alpha \to \infty$ and $\alpha \to 1.5$ can also be observed. We recall that for $\alpha \to \infty$, that is, for an infinite bandwidth, the summit heights would be distributed according to a normal distribution and that for an isotropic random surface $\alpha \geq 1.5$. The case of $\alpha \to 1.5$ is the one giving rise to the distribution which deviates the most

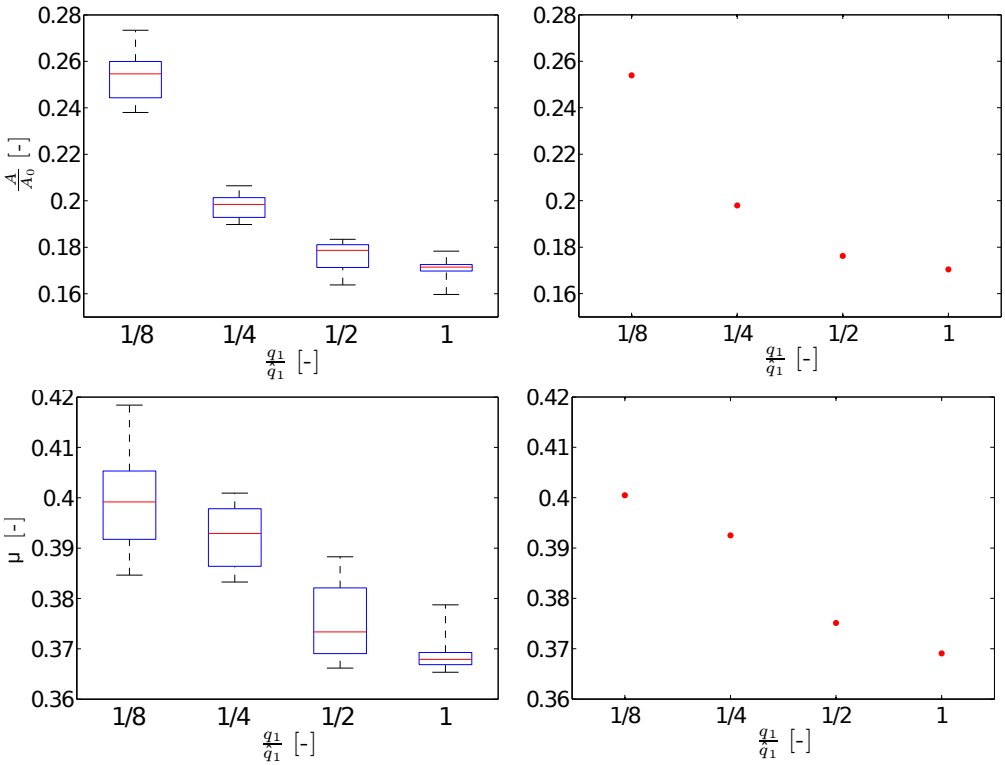

Figure 5.20. Top: Computed real area of contact for Dataset C. *Left*: The real area of contact fraction for the final (sheared) configuration of our numerical experiments. Each box represents a subgroup of ten specimens with same statistical measures. *Right*: The average area of contact for each one of the four subgroups in a semilogarithmic plot. *Bottom*: Friction results for Dataset C. *Left*: The friction coefficient μ for different short distance cut-off wave vectors q_1. Each box represents a subgroup of ten specimens with same statistical measures. *Right*: The mean values of the friction coefficients for the different subgroups.

$q_1/\hat{q}_1$	m_0	m_2	m_4	α
1	$5.001 \cdot 10^{-7}$	4.146	$3.055 \cdot 10^{9}$	88.895
1/2	$4.984 \cdot 10^{-7}$	2.624	$5.039 \cdot 10^{9}$	36.469
1/4	$4.939 \cdot 10^{-7}$	1.620	$8.308 \cdot 10^{8}$	15.627
1/8	$4.831 \cdot 10^{-7}$	0.958	$1.367 \cdot 10^{7}$	7.1916

Table 5.3. Principal power spectrum moments and bandwidth parameter α defined in (3.63), for the four subgroups of Dataset C.

from the normal distribution. In Figure 5.21 is observable how the low-pass filtering of the power spectrum increases the probabilities of high peaks of the corresponding surface.

We suppose that the presence of very high smooth summits makes long-distance interactions between summits more relevant, since such interactions depend on the reciprocal positions of the peaks, which is random and varies strongly in the different realizations of the random process generating surfaces. These variations are a source of uncertainty in our global measures (area of contact and shear resistance). Further

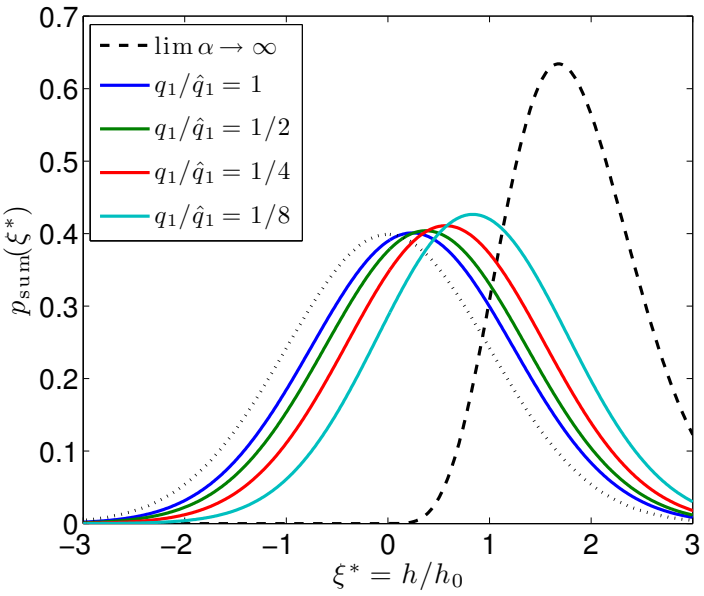

Figure 5.21. Probability density function of summit relative heights ($p_{\mathrm{sum}}(\xi^*)$) for surfaces of Dataset C, compared to the perfect Gaussian obtained for $\alpha \to \infty$ and the limit distribution obtained for $\alpha \to 1.5$.

evidence of the importance of long distance interactions is provided in [YAM12], where a deeper analysis of cut-offs is also performed, and the resulting surfaces are classified as *representative* or *non-representative*, based on how close their height distribution is to a Gaussian probability distribution. The representativity of surfaces which are distant from a Gaussian distribution is low, meaning that many specimens are needed to compute a reasonable mean value for rough contact quantities, and this is exactly what happens for surfaces with narrow bandwidths, that is small ratios of q_1 to q_0.

5.3.2 Conclusive remarks

In this section, we showed that all roughness parameters have a significant influence on area of contact and friction for shear tests. This should be kept in mind, when performing real experiments, and all roughness parameters should always be listed, or at least, computable (by means of Fourier analysis on a digitalized sample of the real surface, or with other measures). From a numerical standpoint, much attention has to be paid to the choice of the mesh step size, which has to be sufficiently small, for small values of h_0. We also proved that wave vectors with high frequencies have an importance for both the mean and the variance of the computed quantities, thus, in simulations and in practical experiments, the full spectrum should be employed, or measured.

A comparison of our results to existing analytical models of static friction (e.g. [CEB88], [MF94], and [KE04; LET10]), or to analytical models of kinetic friction (such Persson's one, see (3.170)), is hard, as the premises of other models are different (such as plastic deformations of asperities, or viscoleastic materials) and a full study of the influence of each roughness parameter, to the best of our knowledge, has not been done by other studies.

5.4 Verification

The classic paradigm of computational science is composed of three phases:

CS1 Development and implementation of a model.

CS2 Validation of results against known theory.

CS3 Verification of numerical predictions with real world data.

Phase **CS1** was done in the very beginning of this work, and it is obviously the basis from which the results could be obtained. Phase **CS2** was mostly performed for the normal load experiments, where theories which are set in frameworks comparable to the one within which we performed our experiments are present and well established. For phase **CS3**, we needed to find large datasets of results, obtained with real experiments, in labs, or controlled settings. For the same results, we also needed to have access to a good characterization of the rough surfaces, in terms of PSD. Since one of the fields in which friction is more relevant is the automotive industry, a natural choice seemed to be the tyre-road interactions. We know that we are not in the perfect setting for studying elastomeric response to excitation, since we are neglecting temperature and viscosity effects, but at the same time, we think it is interesting to understand if at least a part of the tyre-road friction can be explained by the purely elastic interactions of rubber with the hard corrugated substrate, which is supposed to undergo negligible plastic deformations, as it is harder than low excited rubber by three orders of magnitude, c.f. [TZU10]. As we have pointed out, our samples have statistical characteristics similar to road-like surfaces, especially those surfaces which are in the middle of the h_0 range of Dataset A. Thus, we searched for some valuable, unbiased and freely accessible data of friction on roads.

5.4.1 A real world example: simplified tyre-road dynamics on wet asphalt

We found suitable studies in [HPB$^+$00], [Ame08], and [Mee09], which are publications related to studies commissioned by different American departments of transportation. In [Mee09], the skid number SN40R, which is a measure of the friction coefficient on a wet road, was computed. As we will explain better in Section 5.4.2, wetting of surfaces is possibly accountable for the loss of hysteretic response of rubber sliding on asphalt, thus, a slightly wet surface is somehow more similar to our setting, since it reduces the bulk of the tyre treads, but keeps the elastic interactions with large asperities unaffected. Therefore, we compared the results we obtained for surfaces which are similar to the ones used in [Mee09] (we use the surfaces of Dataset A), and obtain the result shown in Figure 5.22. Even though there are some differences, our results are compatible with measured data, with a tendency to give higher values of friction, but a recognizable trend. Notice also that the maximum friction measured on

roads, is located close to where we found it, near a mean profile depth of 0.8 mm, or a root mean square roughness of 0.2 mm.

Obviously, information about the road surfaces is incomplete, and the origin of the discrepancy between the values could stem from this factor, from a different loading condition, or from the difference between the values of static and dynamic friction coefficients, because in the absence of viscosity, static friction coefficient is usually experimentally found to be larger than the kinetic one. However, considering that we have not taken viscous effects into account, but a part of them should be present in data collected in real experiments, it is possible that the amount of friction due to elastic interactions that we compute overestimates the real value. We anyhow consider the results encouraging, and we believe they demonstrate that our approach is indeed reasonable and can predict, with acceptable limitations, realistic results. We remark that in all publications mentioned above, the friction coefficient on wet roads is measured between 0.4 and 0.65. For this reason, in next section we compare our assumptions and our results to the theory developed in [PTAT04; PTAT05], for wet tyre-road interactions.

For a deep analysis of the influence of friction on tyre-road dynamics, the interested reader is referred to [Pac06] and [SG04] for two models which are used as reference by the automotive industry.

5.4.2 Comparison with analytical theory

The loose term *aquaplaning* (or hydroplaning), refers to the loss of control experienced while driving a vehicle on a very wet road. The most common and important effects of aquaplaning (c.f. [Ame08]) are

E1 severe reduction of braking force, with consequent increase of braking distance;

E2 loss of traction power, with consequent impossibility of steering, and controlling the vehicle.

Aquaplaning starts when the tire grooves can not channel the water present on the road out of the contact area, and three major factors concur in triggering this situation, c.f. [HPB$^+$00]:

F1 the size and the composition of the road texture;

F2 the height of the water present on the road;

F3 the rolling speed of the wheel.

Usually, the phenomenon is analyzed empirically, based on road tests, and the purposes of the studies are usually to design better road surfaces (understanding the best texture and composition to use for road pavement), better tyres (maximizing the

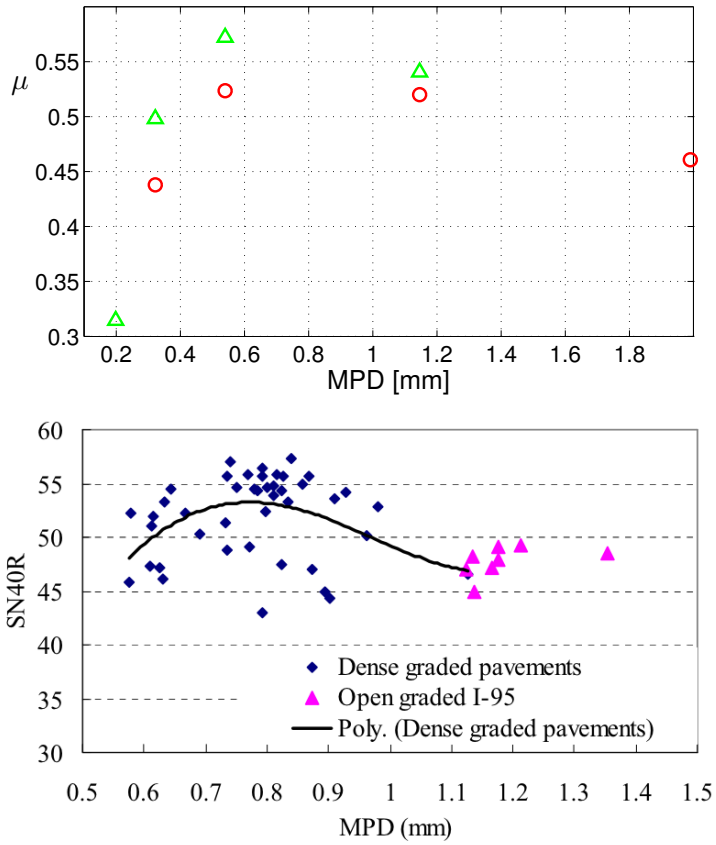

Figure 5.22. Top: Results obtained on Dataset A, plotted against Mean Profile Depth of surfaces. *Bottom*: Measurements obtained on two different types of roads in New Jersey, image taken from [Mee09]. The skid number SN40R is obtained multiplying μ_k by 100. Note the different ranges for the MPD between numerical and experimental surfaces.

channelling capacity of the grooves), and to give advices to drivers (in terms of limit speeds where aquaplaning begins, and of practices to avoid when asphalt is wet).

Analytical models describing aquaplaning are, to the best of our knowledge, rare, and consider hydrodynamical effects which are beyond the scope of this work, but the interested reader can find a very good reference in [CYA+01].

We will focus on a simplified model which links roughness to the wet road dynamics, but in conditions where aquaplaning does not start, and **E1** and **E2** are limited in magnitude. The model is described in [PTAT04], [PAT+04], and [PTAT05]. It is based on Persson's theory for friction of elastomers on rough surfaces, and takes into account the power spectrum of typical road surfaces. Note that in terms of friction reduction,

it is commonly accepted that at slow velocities, the drop in the friction coefficient for a wet surface is approximately 20-30%, c.f. [MW83] and [Hei03].

The model is based on three assumptions:

A1 the power spectrum of the road surface is decreased over the **q**-domain, because tyre rubber seals water "pools" present on the surface (preventing water from leaking out), and short wavelength details are smoothened out;

A2 the friction reduction experienced on wet surfaces is due to reduction of the bulk modulus, and this reduction is caused by the smoothing of short wavelength asperities, which usually excite rubber and dissipate energy internally to it, creating most of the tyre-road friction;

A3 hydrodynamic behavior of water can be neglected, because of the speed at which the contact happens: water acts like a rigid substrate, when the pools containing it are sealed by the tyre rubber.

It is interesting to notice that the study we used to validate our results in Section 5.4, is based on data collected on wet surfaces, in a situation similar to the one assumed by this very model.

Even though our numerical experiments do not account for viscosity, we can partially test the assumptions **A1-3**, by means of a simple experiment: we can simulate the presence of various water coverages on road-like surfaces and use the wet surface, in which water acts as a solid wall, to perform our shear tests. In this way, we will be able to understand if the presence of water has an influence on the static friction measured in the shear tests.

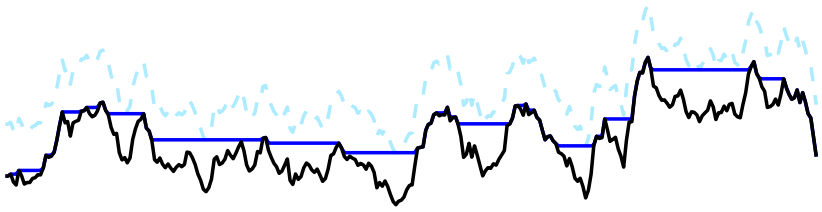

Figure 5.23. Two-dimensional example of water filling. The rough profile is represented by the black line, the light blue dashed line is the initial water with constant depth, the solid blue line is the resulting water configuration. Surface pools are filled with water up to the maximum depth, or to the maximum level of water prescribed by the user.

To simulate the presence of water on the surfaces, we implemented an algorithm which fills all pools which have a depth d smaller than a user-defined value $d_{\max}$. Notice that not all water surfaces will have the same height, as there will be water pools standing on top of "hills", and other ones will be at the very bottom of the rough surface. The algorithm simply starts assuming that there is an equal amount of water $d_{\max}$ over all the surface, and then removes water which is higher than the boundaries of the pools in which it is contained. Figure 5.23 clarifies the algorithm for a simple two-dimensional example. In Figure 5.24 it is shown the limit water distribution we get for a typical road-like surface: in such conditions, the total water coverage of the surface is around 40%, and any added water will flow out of the surface. To simulate water flowing out of the surface, but still be able to consider the complete specimen wet, one can not simply impose that the water level at the boundaries is zero, because this will dry pools on the boundaries, which in a periodic setting, should just be filled like other pools. Thus, to remove excess water, we simulated the water configuration on a much larger surface, in which we used our specimens as tiles, and built a 3×3-tiles surface, on which we imposed zero-boundary conditions. Then, we took as final result the water configuration on the central tile, and replicated it periodically to obtain the obstacle used in our simulations.

A typical result of the algorithm can be observed in Figure 5.24, and in [PTAT04], a result qualitatively equal to ours is shown for a scan of a real road surface, filled numerically with water.

We simulated contact with a surface with parameters close to the ones measured in [Lor12] for asphalt roads, thus we took a surface from Dataset A, with $h_0 \approx 0.3$ mm, for which, in our shear tests, the static friction coefficient $\mu = 0.54$. We will employ the same load used in the shear tests of Section 5.3, hence, the displacement on top of the cube will be 9 mm.

With the water limit distribution, we found the friction coefficient to be equal to the dry case. This lead us to try to add more water to the system, simulating standing water, as it can be encountered on very wet roads, or where puddles caused by pavement depressions are present. To do this, we added periodic boundary conditions to our water-road system. We simulated coverages of up to 80% of the projected area, and the results for the static friction coefficients are shown in Figure 5.25. It can be observed that the static friction coefficients starts to decrease, only when the covered surface is more than 40%.

In a study of airplane accidents due to wet runways ([CAHS09]), commissioned by the British Air Accidents Investigation Branch, the normal slipperiness due to wet asphalt (the one we are concerned with in these experiments) is called *viscous hydroplaning*, and it is said that it can occur for water coverages with depths of less than 0.25 mm, and this is the mean value of the water pool depth we observe for coverages around 60%. For deeper pools, the real *dynamic hydroplaning* is supposed to start. We note that these two intervals roughly correspond to the intervals in which the

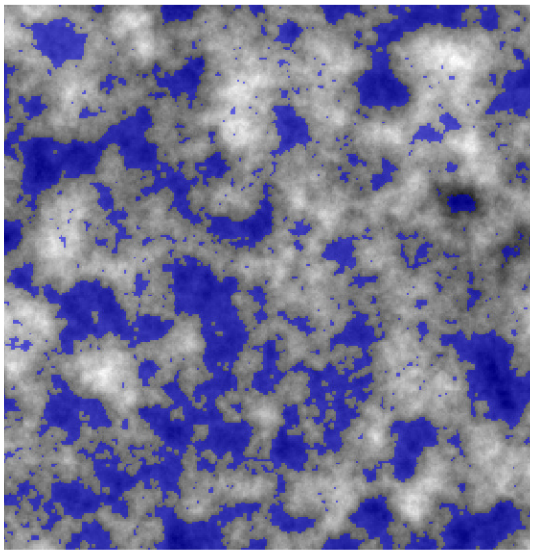

Figure 5.24. Realistic self-affine rough surface covered with water, imposing zero-boundary conditions on a larger domain. The total water coverage in this case is 40% of the surface projected area.

computed elastic static frictional forces are or are not appreciably affected.

We propose an interpretation of these results. We know from the results of our experiments, that the initial area of contact for the shear test, is $\approx 40\%$, thus, when water covers a smaller fraction of the total rough surface, the wet areas do not enter in contact with the elastic solid, a part for the smaller higher pools, which anyhow do not prevent the interaction with large asperities. Only when the fraction of surface covered by water is overlapping with the original area of contact, friction begins to decrease.

Thus, we can say that the purely elastic component of friction, for roads with no water stagnation, is not influenced, and can not be accounted responsible for viscous hydroplaning (normal slip due to wet roads). In contrast, when water is so much that large asperities are covered (like on a flooded runway), it is possible that hydrodynamics effects (which are not simulated here), can prevent elastic interactions, and lead to total loss of traction and braking force.

If we now compare our results to those obtained in [PTAT04], it can be assumed that the hysteretic, bulk, friction is responsible for most of the tyre-road resistance (the friction coefficient for rubber sliding on dry asphalt is usually measured around unity), and it could effectively be reduced by the water smoothing effect.

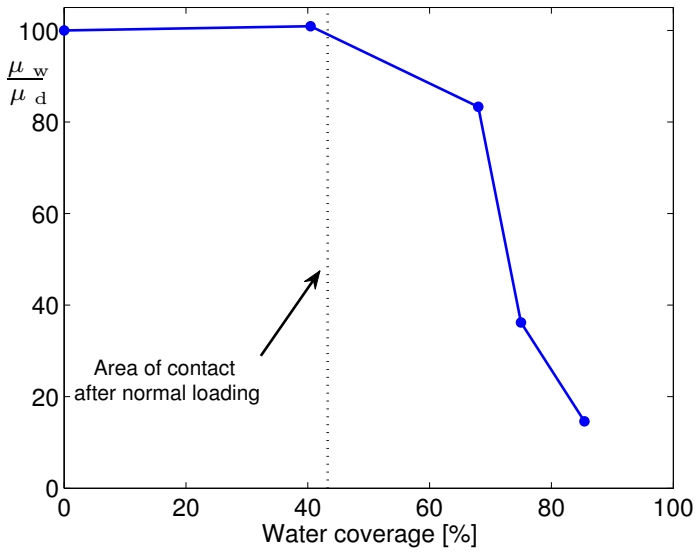

Figure 5.25. Ratio of elastostatic friction coefficient on wet surface to coefficient on dry surface, expressed as percentage, against water coverage. The limit coverage for zero-boundary conditions is 40.45%. Larger coverages are obtained by stagnation (periodic boundary conditions).

As a remark, we would like to compare the PSDs of the surfaces for different water coverages, as it is done in [PTAT04]. If we measure the PSD of a rough surface which is covered with water, we obviously observe a decrease over the whole **q**-space, as it can be seen in Figure 5.26. This is clearly due to the presence of flat water surfaces. Though, the differences in the PSD are not sufficient to explain the variation in the friction coefficients: for example, for a 70% coverage, the friction coefficient is one third of that measured for the dry surface, but the root mean square roughness is 0.1 mm, and with such a surface, according to our calculations, the friction coefficient should be reduced by at most 10%. This inconsistency is caused by the PSD failing in measuring the real effects of roughness, for surfaces which are not homogeneous.

5.4.3 Conclusive remarks

In this section, we compared our friction prediction to real-world data and to an analytical model of tyre-road interaction in wet conditions, a setting in which water is supposed to reduce the viscous response of rubber, but not to prevent contact of tyre treads and asphalt.

The comparison of our results with the real world data seems to confirm that the presence of small amounts of water on surfaces might drastically reduce hysteretic

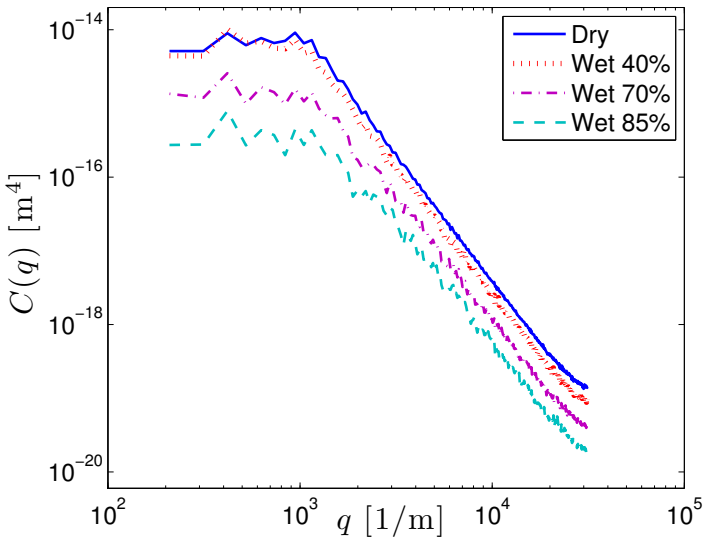

Figure 5.26. PSD of wet surfaces with different water coverages, notice that increasing the coverage, the PSD decreases over all the q-space.

rubber effects, and that our experiments possibly overestimates slightly the amount of friction produced by elastic forces.

Though, it seems that the simple measure of the PSD of the wet surface is not sufficient to predict the new friction coefficient, as it homogenizes the description of a surface in which roughness varies locally.

We remark that rubber sliding on dry asphalt has a friction coefficient of 0.9-1 (c.f. [Com11], [Pac06], or any mechanical engineering, or automotive engineering book), and the difference in magnitude between this value and the one which we compute can therefore partially be explained in terms of viscous energy dissipation, a phenomenon we do not include in our experiments by design.

Chapter 6

Conclusions

In this work, we showed that a part of static friction can be numerically computed by means of a Finite Element Simulation of an iterative version of the classic Signorini contact problem.

We studied the influence that different roughness parameters have on macroscopic contact features such as friction and real area of contact, and we empirically computed bounds on the uncertainty of the values we obtained.

We studied the effects of different numerical choices, such as the mesh step size and the spatial dimension of the problem, as well as the influence of the chosen Poisson's modulus on the results.

We found good agreement of our results with state-of-the-art theories for rough contact at low pressures, whereas for large pressures, our results tend to diverge from analytical models.

We gave our interpretation of observed phenomena, and described the results by means of the coefficient of a polynomial function, which is capable of interpolating results of normal loading experiments with a sufficient accuracy, especially for certain acceptably large ranges of roughness, domain size, and elasticity parameters.

We compared the results obtained in our experiments to real world data of tyre-road interactions, obtained for settings in which our assumptions are partially valid, that is, where viscosity and plasticity have smaller influence. Our results slightly overestimate measured data, but the trends are clearly recognizable, even if information about the real experiments is not complete. We find a reasonable approximation of the limit quantity of water needed to start the phenomenon of aquaplaning, which finds confirmation in reports and measurements on the phenomenon.

We compared the same results to an analytical theory of tyre slipping on wet roads, and found that the validities of our approach and of the analytical theory rely on a similar set of assumptions, and thus, our approach can quantify the amount of friction produced by purely elastic interactions with a reasonable accuracy.

Given the promising results, many possibilities are open for future development of

this theory. Surely, several other numerical experiments will be performed, employing different material laws, and experimental settings, but further development of our model will be possible only with experimental data collected in controlled environment, with information about the surfaces and the materials involved.

Appendix A

Relative radii of curvature R' and R''

In this section we derive closed forms for the relative radii of curvature R' and R'' used in Section 2.3. We proceed step by step, following [Her82], [Joh85], and [LLP86]. For the sake of completeness, we shall add the steps of the proof which are not present in any of the references we cited.

Let us again recall we are concerned with two smooth curved paraboloids in contact, with equations

$$\begin{aligned} z_1 &= \frac{1}{2R_1'}x_1^2 + \frac{1}{2R_1''}y_1^2 \\ z_2 &= -\left(\frac{1}{2R_2'}x_2^2 + \frac{1}{2R_2''}y_2^2\right) \end{aligned}$$

We define h as the separation between the two surfaces, therefore $h = z_1 - z_2$. Notice that x_1 and y_1 form an angle ϑ with x_2 and y_2 respectively, whereas $z_1 \equiv z_2$. We can now transform the coordinates in which we express h to a system x, y, z, where x form angles α and β with x_1 and x_2 respectively. This corresponds to rotating the local coordinate systems, and the transformations are

$$\begin{cases} x_1 = x\cos\alpha + y\sin\alpha \\ y_1 = -x\sin\alpha + y\cos\alpha \end{cases} \qquad \begin{cases} x_2 = x\cos\beta + y\sin\beta \\ y_2 = -x\sin\beta + y\cos\beta \end{cases} \tag{A.1}$$

which allow us to rewrite z_1 and z_2 as

$$\begin{aligned} z_1 = &\\ &\frac{1}{2R_1'}\left(x^2\cos^2\alpha + xy\cos\alpha\sin\alpha + y^2\sin^2\alpha\right) \\ &-\frac{1}{2R_1''}\left(x^2\sin^2\alpha - xy\cos\alpha\sin\alpha + y^2\cos^2\alpha\right) \end{aligned} \tag{A.2}$$

and

$$z_2 = \frac{1}{2R_2'}\left(x^2\cos^2\beta + xy\cos\beta\sin\beta + y^2\sin^2\beta\right) - \frac{1}{2R_2''}\left(x^2\sin^2\beta - xy\cos\beta\sin\beta + y^2\cos^2\beta\right). \tag{A.3}$$

We can therefore write h in the form

$$h = z_1 - z_2 = Ax^2 + By^2 + Cxy \tag{A.4}$$

with

$$\begin{aligned} A &= \frac{1}{2}\left(\frac{1}{R_1'}\cos^2\alpha - \frac{1}{R_1''}\sin^2\alpha\right) - \frac{1}{2}\left(\frac{1}{R_2'}\cos^2\beta - \frac{1}{R_2''}\sin^2\beta\right) \\ B &= \frac{1}{2}\left(\frac{1}{R_1'}\sin^2\alpha - \frac{1}{R_1''}\cos^2\alpha\right) - \frac{1}{2}\left(\frac{1}{R_2'}\sin^2\beta - \frac{1}{R_2''}\cos^2\beta\right) \\ C &= \frac{1}{2}\left(\frac{1}{R_1'}\sin\alpha\cos\alpha - \frac{1}{R_1''}\sin\alpha\cos\alpha\right) - \frac{1}{2}\left(\frac{1}{R_2'}\sin\beta\cos\beta - \frac{1}{R_2''}\sin\beta\cos\beta\right) \end{aligned} \tag{A.5}$$

which, with the help of trigonometric transformations, become

$$\begin{aligned} A &= \frac{1}{4}\left[\frac{1}{R_1'} + \frac{1}{R_2'} + \frac{1}{R_1''} + \frac{1}{R_2''} + \left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)\cos 2\alpha + \left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)\cos 2\beta\right] \\ B &= \frac{1}{4}\left[\frac{1}{R_1'} + \frac{1}{R_2'} + \frac{1}{R_1''} + \frac{1}{R_2''} - \left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)\cos 2\alpha - \left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)\cos 2\beta\right] \\ C &= \frac{1}{2}\left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)\sin 2\beta - \frac{1}{2}\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)\sin 2\alpha \end{aligned} \tag{A.6}$$

The only condition we impose, is that $C = 0$, so that (A.4) can be rewritten as

$$h = Ax^2 + By^2 \tag{A.7}$$

It can be observed in (A.6.3) that

$$C = 0 \quad \Leftrightarrow \quad \frac{\frac{1}{2}\left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)}{\sin 2\alpha} = \frac{\frac{1}{2}\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)}{\sin 2\beta} \tag{A.8}$$

and this condition is satisfied by the triangle shown in Figure A.1, for which we can compute the length of the side w as

$$w = v\cos 2\alpha + u\cos 2\beta = \frac{1}{2}\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)\cos 2\alpha + \frac{1}{2}\left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)\cos 2\beta = A - B, \tag{A.9}$$

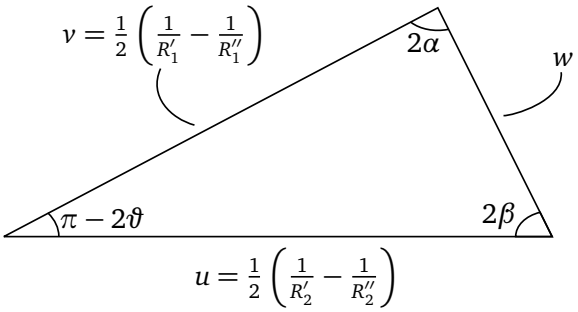

Figure A.1. The triangle satisfying (A.8).

but w must also obey the law of cosines for the triangle it belongs to, that is:

$$w = \frac{1}{2}\left[\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)^2 + \left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)^2 - 2\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)\left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)\cos(\pi - 2\vartheta)\right]^{\frac{1}{2}} \tag{A.10}$$

therefore, using (A.9) and exploiting properties of cosine, we can state the condition

$$A - B = \frac{1}{2}\left[\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)^2 + \left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)^2 + 2\left(\frac{1}{R_1'} - \frac{1}{R_1''}\right)\left(\frac{1}{R_2'} - \frac{1}{R_2''}\right)\cos(2\vartheta)\right]^{\frac{1}{2}} \tag{A.11}$$

and, it is easy to see that

$$A + B = \frac{1}{2}\left(\frac{1}{R_1'} + \frac{1}{R_2'} + \frac{1}{R_1''} + \frac{1}{R_2''}\right). \tag{A.12}$$

Finally we can define an angle τ, such that

$$\cos\tau = \frac{A - B}{A + B} \tag{A.13}$$

and we get the closed forms

$$A = \frac{\left(\frac{1}{R_1'} + \frac{1}{R_2'} + \frac{1}{R_1''} + \frac{1}{R_2''}\right)\cos^2\left(\frac{\tau}{2}\right)}{2}, \quad B = \frac{\left(\frac{1}{R_1'} + \frac{1}{R_2'} + \frac{1}{R_1''} + \frac{1}{R_2''}\right)\sin^2\left(\frac{\tau}{2}\right)}{2}. \tag{A.14}$$

The principal relative radii of curvature are then defined as

$$\frac{1}{2R'} = A, \qquad \frac{1}{2R''} = B, \tag{A.15}$$

hence

$$h = \frac{1}{2R'}x^2 + \frac{1}{2R''}y^2. \tag{A.16}$$

Appendix B

Factor k for rough surface generation

We want to investigate the validity of the Persson's formula [1] for the scaling factor k of the isotropic power spectral density which is stated in [Per01b] as

$$C(q) = k\left(\frac{q}{q_0}\right)^{-2(H+1)}, \text{ if } q_0 < q < q_1, \tag{B.1}$$

where

$$k = \frac{H}{\pi}\left(\frac{h_0}{q_0}\right)^2. \tag{B.2}$$

Our considerations will show that this formula for k is a limit case of a refined, more general one taking into account the roll-off effect at q_0, as well as the upper (q_1) and lower (q_L) cut-off wave vectors.

As introduced in Section 3.2, it holds

$$C(q) = \begin{cases} k & \text{, if } q_L \le q < q_0, \\ k\left(\frac{q}{q_0}\right)^{-2(H+1)} & \text{, if } q_0 \le q < q_1 \\ 0 & \text{, else} \end{cases}.$$

From Section 3.1 we know, by (3.10) that

$$h_0^2 = 2\pi \int_0^\infty qC(q)\,dq$$

[1] We note that Persson's definition of h_0 in [Per01b] is different from ours by a factor of 2, since he defines $h_0^2/2 = \langle h^2 \rangle$. For the sake of consistency, in this Appendix we will conform to our notation, where $h_0 = \langle h^2 \rangle$.

Directly from definition we have

$$\begin{aligned}\int h(x)^2\,\mathrm{d}x &= 2\pi\int_{q_L}^{q_0} kq\,\mathrm{d}q + 2\pi\int_{q_0}^{q_1} k\frac{q^{-2H-1}}{q_0^{-2(H+1)}}\,\mathrm{d}q \\ &= 2\pi\left[\frac{k}{2}(q_0^2-q_L^2)+kq_0^{2H+2}\frac{1}{-2H}\left(q_1^{-2H}-q_0^{-2H}\right)\right] \\ &= \pi k\left\{\left(q_0^2-q_L^2\right)+\frac{q_0^2}{H}\left[1-\left(\frac{q_0}{q_1}\right)^{2H}\right]\right\}. \end{aligned} \tag{B.3}$$

Substituting this result in the RHS of (3.10), we get

$$h_0^2 = \pi k q_0^2\left\{\left(1-\frac{q_L^2}{q_0^2}\right)+\frac{1}{H}\left[1-\left(\frac{q_0}{q_1}\right)^{2H}\right]\right\}. \tag{B.4}$$

Herein we see that as long as $q_L \neq q_0$ there is a permanent influence of the plateau region on h_0. Now we can find the value of k from (B.4).

$$k = \frac{1}{\pi}\left\{\left(1-\frac{q_L^2}{q_0^2}\right)+\frac{1}{H}\left[1-\left(\frac{q_0}{q_1}\right)^{2H}\right]\right\}^{-1}\left(\frac{h_0}{q_0}\right)^2. \tag{B.5}$$

which is the value ensuring that all surfaces possess the prescribed h_0.

If we approximate (B.4) for the case $q_1 \gg q_0$, we get

$$h_0^2 \approx \pi k q_0^2\left[\left(1-\frac{q_L^2}{q_0^2}\right)+\frac{1}{H}\right]. \tag{B.6}$$

and the resulting solution for k is

$$k \approx \frac{H}{\pi}\left[\left(1-\frac{q_L^2}{q_0^2}\right)H+1\right]^{-1}\left(\frac{h_0}{q_0}\right)^2. \tag{B.7}$$

If q_L and q_0 are close, the influence of the plateau vanishes. In case $q_L = q_0$, from (B.4) the mean square roughness of the surface becomes

$$h_0^2 = \frac{\pi k q_0^2}{H}\left[1-\left(\frac{q_0}{q_1}\right)^{2H}\right] \tag{B.8}$$

If we now also add $q_1 \gg q_0$ we get approximately

$$h_0^2 \approx \frac{2\pi k q_0^2}{H}. \tag{B.9}$$

Consider now (B.5) in the case $q_L = q_0$ where the power spectral density below q_0 does not influence h_0 (as assumed in [PAT+04]). Then we get

$$k = \frac{H}{\pi}\left[1-\left(\frac{q_0}{q_1}\right)^{2H}\right]^{-1}\left(\frac{h_0}{q_0}\right)^2. \tag{B.10}$$

Assuming additionally that $q_1 \gg q_0$ we get the approximated value of k suggested in [Per01b]

$$k \approx \frac{H}{\pi}\left(\frac{h_0}{q_0}\right)^2. \tag{B.11}$$

To compute the error induced by the approximated value of k, we computed $\sqrt{m_0}$ for the PSD of surfaces with different Hurst exponents, keeping the other parameters identical, as in Dataset B (c. f. Section 5.3). In Figure B.1 one can see that with the k proposed by Persson, h_0 is different from the one we prescribed, and this difference is clearly correlated to H. Relative deviation can be as high as 100%, as $H \to 1$. If we use our corrected factor k, the h_0 value we get is exactly the same we prescribed. This deviation can be explained by the fact that in the surfaces we generate $q_L \neq q_0$.

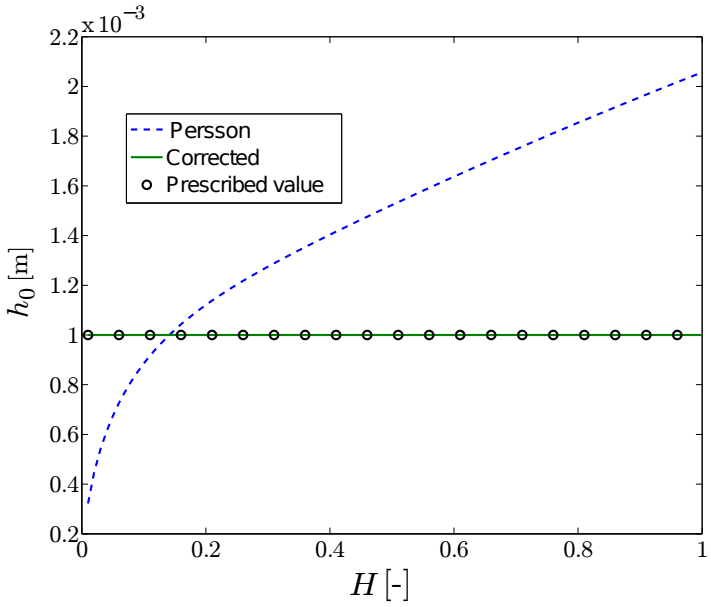

Figure B.1. Comparison of the computed h_0 value on numerically produced surfaces with different Hurst exponents, for corrected and uncorrected (Persson's approximation) factor k.

Appendix C

Computed coefficients of the interpolation polynomials

In this appendix, we list all the parameters we found for the polynomial interpolation given in (5.4), which relates real contact area fraction and normalized pressure for normal load experiments, along with the error $\epsilon = 1 - R^2$ scaled by a factor 10^4, to display its first significant digits. In each captions, the numerical setting is reported.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.06	0.45	-1.32	1.82	0.01
-4.02	-0.07	0.47	-1.34	1.82	0.07
-3.77	-0.14	0.74	-1.65	1.93	0.25
-3.52	-0.32	1.07	-2.02	2.16	0.07
-3.27	4.31	-3.24	-1.87	2.51	0.35
-3.15	-20.68	14.30	-6.34	2.86	3.42

Table C.1. Dataset A. Cube M, $\delta_{\Gamma_c} = 50$ μm, $\nu = 0.45$.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.05	0.38	-1.19	1.73	0.01
-4.02	-0.05	0.38	-1.19	1.73	0.06
-3.77	-0.11	0.65	-1.58	1.89	0.14
-3.52	-0.33	0.77	-1.82	2.13	0.30
-3.27	-9.78	8.66	-4.97	2.60	2.25
-3.15	50.24	-25.47	-0.08	2.36	1.75

Table C.2. Dataset A. Cube M, $\delta_{\Gamma_c} = 25$ μm, $\nu = 0.45$.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.05	0.40	-1.23	1.77	0.01
-4.02	-0.07	0.46	-1.28	1.74	1.01
-3.77	-0.13	0.69	-1.58	1.88	0.18
-3.52	-0.12	0.64	-1.78	2.10	0.37
-3.27	0.64	-1.22	-1.99	2.42	0.22
-3.15	10.02	-6.24	-2.32	2.54	4.53

Table C.3. Dataset A. Cube L, $\delta_{\Gamma_C} = 50$ μm, $\nu = 0.45$.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.12	0.63	-1.44	1.78	0.21
-4.02	-0.15	0.70	-1.49	1.78	0.45
-3.77	-0.72	1.60	-2.01	1.91	0.10
-3.52	-0.23	0.77	-1.89	2.13	0.10
-3.27	4.26	-5.37	-0.83	2.36	0.26
-3.15	9.80	-9.36	-1.21	2.44	1.21

Table C.4. Dataset A. Cube XL, $\delta_{\Gamma_C} = 50$ μm, $\nu = 0.45$.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.06	0.48	-1.44	1.93	0.04
-4.02	-0.06	0.49	-1.44	1.92	0.05
-3.77	-0.08	0.56	-1.56	2.00	0.06
-3.52	-0.10	0.66	-1.82	2.21	0.03
-3.27	0.16	-0.10	-2.13	2.60	0.07
-3.15	-1.70	1.76	-3.42	2.80	0.24

Table C.5. Dataset A. Cube M, $\delta_{\Gamma_C} = 50$ μm, $\nu = 0.05$.

h_0	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
-4.27	-0.07	0.56	-1.58	2.02	0.05
-4.02	-0.09	0.60	-1.61	2.01	0.06
-3.77	-0.12	0.74	-1.80	2.11	0.10
-3.52	-0.23	1.00	-2.18	2.36	0.07
-3.27	0.90	-0.65	-2.45	2.79	0.21
-3.15	-1.57	0.79	-3.50	2.94	0.32

Table C.6. Dataset A. Cube M, $\delta_{\Gamma_C} = 50$ μm, $\nu = 0.25$.

H	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
0.50	-8.03	8.38	-6.65	3.19	0.76
0.60	2.89	-1.75	-2.99	2.81	0.52
0.70	-1.70	1.76	-3.42	2.80	0.24
0.80	1.20	-1.54	-1.61	2.48	0.30
0.90	-0.03	0.23	-2.15	2.49	0.10

Table C.7. Dataset B. Cube M, $\delta_{\Gamma_c} = 50$ μm, $\nu = 0.05$.

H	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
0.50	112.91	-48.32	1.09	2.80	0.75
0.60	13.62	-9.45	-1.92	2.69	1.70
0.70	-20.68	14.30	-6.34	2.86	3.42
0.80	2.34	-2.52	-2.01	2.50	0.98
0.90	-2.33	3.07	-3.53	2.59	3.11

Table C.8. Dataset B. Cube M, $\delta_{\Gamma_c} = 50$ μm, $\nu = 0.45$.

$q_1/\hat{q}_1$	Π	Ω	Λ	κ	$10^4 \cdot \epsilon$
1.00	-20.68	14.30	-6.34	2.86	3.42
0.50	2.89	-2.90	-2.25	2.52	1.00
0.25	13.58	-12.43	0.88	2.22	2.80
0.12	2.04	-1.97	-1.73	2.47	2.17

Table C.9. Dataset C. Cube M, $\delta_{\Gamma_c} = 50$ μm, $\nu = 0.45$.

Appendix D

Statistical analysis of shear tests

In this section, we give details about the statistical analysis performed on the results of the shear tests presented in Section 5.3, to compute static friction. For the listed quantities (friction coefficient μ and relative area of contact $A/A_0 = A'$), mean values (indicated by the bar above the relative symbol), standard deviations (denoted by s) and coefficients of variation (s^*) are shown. All experiments were performed with an elastic cube of edge length L and $\delta_{\Gamma_C} = 1/2\,\delta_{\mathcal{O}}$ (50 μm), hence, mean coefficient of friction and area of contact are not accurate for small values of h_0.

$\log h_0$	-4.14	-4.02	-3.89	-3.77	-3.64	-3.52	-3.39	-3.27	-3.14
$\bar{\mu}$	0.2508	0.1737	0.2312	0.2919	0.3534	0.4282	0.4634	0.4401	0.3275
s_μ	0.0024	0.0043	0.0024	0.0059	0.0029	0.0042	0.0070	0.0054	0.0056
s^*_μ	0.0096	0.0248	0.0102	0.0202	0.0083	0.0097	0.0150	0.0124	0.0172
$\bar{A'}$	0.6837	0.6535	0.5490	0.4577	0.3858	0.3180	0.2689	0.2253	0.1535
$s_{A'}$	0.0018	0.0040	0.0033	0.0054	0.0027	0.0043	0.0039	0.0029	0.0044
$s^*_{A'}$	0.0026	0.0061	0.0061	0.0119	0.0069	0.0134	0.0147	0.0129	0.0289

Table D.1. Results obtained with experiments performed on Dataset A.

H	0.5	0.6	0.7	0.8	0.9
$\bar{\mu}$	0.2848	0.3275	0.3692	0.4020	0.4291
s_μ	0.0048	0.0056	0.0037	0.0064	0.0067
s^*_μ	0.0169	0.0172	0.0101	0.0158	0.0157
$\bar{A'}$	0.1327	0.1535	0.1707	0.1891	0.2210
$s_{A'}$	0.0039	0.0044	0.0052	0.0044	0.0042
$s^*_{A'}$	0.0291	0.0289	0.0308	0.0232	0.0189

Table D.2. Results obtained with experiments performed on Dataset B.

$q_1/\hat{q}_1$	1/8	1/4	1/2	1
$\bar{\mu}$	0.4005	0.3925	0.3751	0.3692
s_μ	0.0110	0.0063	0.0073	0.0037
s^*_μ	0.0274	0.0162	0.0195	0.0101
$\bar{A'}$	0.2539	0.1980	0.1763	0.1707
$s_{A'}$	0.0110	0.0057	0.0068	0.0052
$s^*_{A'}$	0.0432	0.0290	0.0386	0.0308

Table D.3. Results obtained with experiments performed on Dataset C.

Bibliography

[Ame08] AASHTO, American Association of State Highway and Transportation Officials. *Guide for Pavement Friction*. American Association of State Highway and Transportation Officials, 2008.

[Ari06] Aristotle. *De Sensu and De Memoria*. New York, Arno Press, 1906.

[BBCC01] M. Borri-Brunetto, B. Chiaia, and M. Ciavarella. Incipient sliding of rough surfaces in contact: a multiscale numerical analysis. *Computer Methods in Applied Mechanics and Engineering*, 190(46-47):6053–6073, 2001.

[BBJ+97] P. Bastian, K. Birken, K. Johannsen, S.Lang, N. Neuß, H. Rentz-Reichert, and C.Wieners. UG – a flexible software toolbox for solving partial differential equations. *Computing and Visualization in Science*, 1:27–40, 1997.

[BGT75] A.W. Bush, R.D. Gibson, and T.R. Thomas. The elastic contact of a rough surface. *Wear*, 35(1):87 – 111, 1975.

[BL80] M. V. Berry and Z. V. Lewis. On the weierstrass-mandelbrot fractal function. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 370(1743):459–484, 1980.

[Bla95] P. J. Blau. *Friction Science and Technology*. Dekker Mechanical Engineering. Taylor & Francis, 1995.

[BLB81] S. Bandis, A.C. Lumsden, and N.R. Barton. Experimental studies of scale effects on the shear behaviour of rock joints. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts*, 18(1):1 – 21, 1981.

[Bou97] E. Bouchaud. Scaling properties of cracks. *Journal of Physics: Condensed Matter*, 9(21):4319, 1997.

[Bre65] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.

[BS95] A.-L. Barabási and H. E. Stanley. *Fractal Concepts in Surface Growth*. Cambridge University Press, Cambridge, UK, 1995.

[BW08] J. Bonet and D. Wood. *Nonlinear continuum mechanics for finite element analysis*. Cambridge, 2 edition, 2008.

[CAHS09] K. Conradi, T. J. Atkinson, S. J. Hawkins, and C. J. Scott. Report 1/2009 G-XLAC G-BWDA G-EMBO. Technical report, Air Accidents Investigation Branch, 2009.

[Car09] G. Carbone. A slightly corrected greenwood and williamson model predicts asymptotic linearity between contact area and load. *Journal of the Mechanics and Physics of Solids*, 57(7):1093 – 1102, 2009.

[CCI99] A. Carpinteri, B. Chiaia, and S. Invernizzi. Three-dimensional fractal analysis of concrete fracture at the meso-level. *Theoretical and Applied Fracture Mechanics*, 31(3):163 – 172, 1999.

[CDD06] M. Ciavarella, V. Delfine, and G. Demelio. A "re-vitalized" greenwood and williamson model of elastic contact between fractal surfaces. *Journal of the Mechanics and Physics of Solids*, 54(12):2569 – 2591, 2006.

[CEB88] W. R. Chang, I. Etsion, and D. B. Bogy. Static friction coefficient model for metallic rough surfaces. *Journal of Tribology*, 110(1):57–63, 01 1988.

[Cha99] P. Chadwick. *Continuum mechanics*. Dover Publications, Inc., 1999.

[Com11] Commissioni romande di matematica, di fisica e di chimica. *Formulari e tavole: matematica, fisica, chimica*. Éditions G d'Encre, 2011.

[CP05] Alberto Carpinteri and Marco Paggi. Size-scale effects on the friction coefficient. *International Journal of Solids and Structures*, 42(9-10):2901 – 2910, 2005.

[CP08] A. Carpinteri and M. Paggi. Size-Scale Effects on Strength, Friction and Fracture Energy of Faults: A Unified Interpretation According to Fractal Geometry. *Rock Mechanics and Rock Engineering*, 41(5):735–746, 2008.

[CYA+01] X. Claeys, Jingang Y., L. Alvarez, R. Horowitz, C. Canudas de Wit, and L. Richard. Tire friction modeling under wet road conditions. In *American Control Conference, 2001. Proceedings of the 2001*, volume 3, pages 1794–1799 vol.3, 2001.

[DAB12] H. Dyer, D. Amitrano, and A.-M. Boullier. Scaling properties of fault rocks. *Journal of Structural Geology*, 45(0):125 – 136, 2012. Fault zone structure, mechanics and evolution in nature and experiment.

[dBCKO08] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd ed. edition, 2008.

[DE14] F. Dang and N. Emad. Fast iterative method in solving eikonal equations: A multi-level parallel approach. *Procedia Computer Science*, 29(0):1859 – 1869, 2014. 2014 International Conference on Computational Science.

[DM98] L. Dagum and R. Menon. Openmp: An industry-standard api for shared-memory programming. *IEEE Comput. Sci. Eng.*, 5(1):46–55, January 1998.

[Dow79] D. Dowson. *History of Tribology*. Longman, 1979.

[Fer61] J.D. Ferry. *Viscoelastic properties of polymers*. Wiley, 1961.

[FO01] Y. B. Fu and R.W. Ogden. *Nonlinear Elasticity: Theory and Applications*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2001.

[Fun90] Y. C. Fung. *Biomechanics: Motion, Flow, Stress, and Growth*. Springer-Verlag, 1 edition, 1990.

[FWP+14] A. Fall, B. Weber, M. Pakpour, N. Lenoir, N. Shahidzadeh, J. Fiscina, C. Wagner, and D. Bonn. Sliding friction on wet and dry sand. *Phys. Rev. Lett.*, 112:175502, Apr 2014.

[GCCH07] G. Gao, R. J. Cannara, R. W. Carpick, and J. A. Harrison. Atomic-scale friction on diamond: A comparison of different sliding directions on (001) and (111) surfaces using md and afm. *Langmuir*, 23(10):5394–5405, 2007. PMID: 17407330.

[GK08] C. Groß and R. Krause. Import of geometries and extended informations into obslib++ using the exodus ii and exodus parameter file formats. Technical Report 712, Institute for Numerical Simulation, University of Bonn, Germany, January 2008.

[GLHL+98] W. D. Gropp, A. Lumsdaine, S. Huss-Lederman, B. Nitzberg, and E. Lusk. *MPI*. The MIT Press, 1998.

[Gre84] J. A. Greenwood. A unified theory of surface roughness. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 393(1804):133–157, 1984.

[GW66] J. A. Greenwood and J. B. P. Williamson. Contact of Nominally Flat Surfaces. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 295(1442):300–319, 1966.

[Hau02] P. Haupt. *Continuum Mechanics and Theory of Materials*. Advanced Texts in Physics. Springer, 2002.

[Hei03] G Heinrich. Elastomerreibung und kontactmechanik, (deutsches institute fü kautschuktechnologie, 2003.

[Her82] H. Hertz. Über die berührung fester elastischer körper. *Journal für die reine und angewandte Mathematik*, 92:156–171, 1882.

[HPB+00] J. J. Henry, National Cooperative Highway Research Program, National Research Council (U.S.). Transportation Research Board, American Association of State Highway, and Transportation Officials. *Evaluation of Pavement Friction Characteristics*. NCHRP synthesis. National Academy Press, 2000.

[HPMR04] S. Hyun, L. Pei, J.-F. Molinari, and M. O. Robbins. Finite-element analysis of contact between elastic self-affine surfaces. *Phys. Rev. E*, 70:026117, Aug 2004.

[IIIC03] Mohsen A. Issa, Mahmoud A. Issa, M. S. Islam, and A. Chudnovsky. Fractal dimension – a measure of fracture roughness and toughness of concrete. *Engineering Fracture Mechanics*, 70(1):125 – 137, 2003.

[IMK02] G. L. Irby-Massie and P.T. Keyser. *Greek Science of the Hellenistic Era: A Sourcebook*. Routledge, 2002.

[Joh85] K. L. Johnson. *Contact Mechanics*. Cambridge University Press, 1985.

[KA01] Y. A. Karpenko and A. Akay. A numerical model of friction between rough surfaces. *Tribology International*, 34(8):531 – 545, 2001.

[KE04] L. Kogut and I. Etsion. A static friction model for elastic-plastic contacting rough surfaces. *Journal of Tribology*, 126(1):34–40, 01 2004.

[KHB06] M. Kilic, S. Hiziroglu, and E. Burdurlu. Effect of machining on surface roughness of wood. *Building and Environment*, 41(8):1074 – 1078, 2006.

[KM11] R. Krause and C. Mohr. Level set based multi-scale methods for large deformation contact problems. *Appl. Numer. Math.*, 61(4):428–442, April 2011.

[KO88] N. Kikuchi and J.T. Oden. *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1988.

[KP95] J. Krim and G. Palasantzas. Experimental observations of self-affine scaling and kinetic roughening at sub-micron lengthscales. *International Journal of Modern Physics B*, 09(06):599–632, 1995.

[Kra09] R. Krause. A nonsmooth multiscale method for solving frictional two-body contact problems in 2d and 3d with multigrid efficiency. *SIAM Journal of Scientific Computing*, 31(2):1399–1423, January 2009.

[KRS15] R. Krause, A. Rigazzi, and J. Steiner. A parallel multigrid method for constrained minimization problems and its application to friction, contact, and obstacle problems. *Computing and Visualization in Science*, 2015. (to appear).

[Kru97] J. Krug. Origins of scale invariance in growth processes. *Advances in Physics*, 46(2):139–282, 1997.

[KSK06] A. Kumar, F. Sadeghi, and C. M. Krousgrill. Effect of surface roughness on normal contact compression response. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 220(2):65–77, 2006.

[Lap86] P. S. Laplace. Memoir on the probability of the causes of events (english translation). *Statistical Science*, 1(3):364–378, 08 1986.

[Lei11] C. Leising. Wood surface roughness using 3d profilometry. Technical report, Nanovea, 2011.

[LET10] L. Li, I. Etsion, and F. E. Talke. Contact area and static friction of rough surfaces with high plasticity index. *Journal of Tribology*, 132(3):031401–031401, 06 2010.

[LH57a] M. S. Longuet-Higgins. The statistical analysis of a random, moving surface. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 249(966):321–387, 1957.

[LH57b] M. S. Longuet-Higgins. Statistical Properties of an Isotropic Random Surface. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 250(975):157–174, October 1957.

[LH62] M. S. Longuet-Higgins. The statistical geometry of random surfaces. In *Proceedings of the 13th Symposium in Applied Mechanics of the American Mathematical Society: Vol. 13: Hydrodynamic Instability*, pages 105–142. AMS, American Mathematical Society, 1962.

[LLP86] L. D. Landau, E. M. Lifchits, and L. P. Pitaevski. *Course of theoretical physics: theory of elasticity*. Butterworth-Heinemann, 1986.

[Lor12] B. Lorenz. *Contact Mechanics and Friction of Elastic Solids on Hard and Rough Substrates*. Schriften des Forschungszentrums Jülich / Reihe Schlüsseltechnologien: Reihe Schlüsseltechnologien. Hochschulbibliothek der Rheinisch-Westfälischen Technischen Hochschule Aachen, 2012.

[LP09a] B. Lorenz and B. N. J. Persson. Interfacial separation between elastic solids with randomly rough surfaces: comparison of experiment with theory. *Journal of Physics: Condensed Matter*, 21(1):015003, 2009.

[LP09b] B. Lorenz and B. N. J. Persson. Leak rate of seals: Comparison of theory with experiment. *EPL (Europhysics Letters)*, 86(4):44006, 2009.

[LP10] B. Lorenz and B. N. J. Persson. Leak rate of seals: Effective-medium theory and comparison with experiment. *The European physical journal / E*, 31, 2010. Record converted from VDB: 12.11.2012.

[Man67] B. Mandelbrot. How long is the coast of britain? statistical self-similarity and fractional dimension. *Science*, 156(3775):636–638, 1967.

[Man83] B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freedman and Co., New York, 1983.

[Mee09] J. N. Meegoda. Non-contact skid resistance measurement. Technical report, New Jersey Institute of Technology, 2009.

[MF94] H. Matsukawa and H. Fukuyama. Theoretical study of friction: One-dimensional clean surfaces. *Phys. Rev. B*, 49:17286–17292, Jun 1994.

[MH94] J. E. Marsden and T. J. R. Hughes. *Mathematical Foundation of Elasticity*. Dover editions, 1 edition, 1994.

[Mik74] B. B. Mikić. Thermal contact conductance; theoretical considerations. *International Journal of Heat and Mass Transfer*, 17(2):205 – 214, 1974.

[Moo40] M. Mooney. A Theory of Large Elastic Deformation. *Journal of Applied Physics*, 11:582–592, September 1940.

[MP11] N. Mulakaluri and B. N. J. Persson. Adhesion between elastic solids with randomly rough surfaces: Comparison of analytical theory with molecular-dynamics simulations. *EPL*, 96(6):66003, 2011.

[MT97] T. Möller and B. Trumbore. Fast, minimum storage ray-triangle intersection. *J. Graph. Tools*, 2(1):21–28, October 1997.

[MW83] W. E. Meyer and J. D. Walter. *Frictional Interaction of Tire and Pavement 85*. American Society for Testing and Materials, 1983.

[Nay71] P. Ranganath Nayak. Random process model of rough surfaces. *Journal of Tribology*, 93(3):398–407, 07 1971.

[Nay73] P. R. Nayak. Random process model of rough surfaces in plastic contact. *Wear*, 26(3):305 – 333, 1973.

[New95] Percy E. Newberry. *The tomb of Tehuti-Hetep*. Number [3] in Memoir / Archaeological Survey of Egypt ; [3] ; Archaeological Survey of Egypt: Memoir. Egypt Exploration Society, London, 1895.

[OF03] S. Osher and R. P. Fedkiw. *Level set methods and dynamic implicit surfaces*. Applied mathematical science. Springer, New York, N.Y., 2003.

[Ogd84] R. W. Ogden. *Non-Linear Elastic Deformations*. Dover Publications, 1984.

[PA04] W.P. Petersen and P. Arbenz. *Introduction to Parallel Computing*. Oxford texts in applied and engineering mathematics. Oxford University Press, 2004.

[Pac06] H.B. Pacejka. *Tyre and Vehicle Dynamics*. Automotive engineering. Butterworth-Heinemann, 2006.

[Pad01] N. P. Padture. Hertzian contacts. In K.H. J. Buschow P. Veyssière, R. W. Cahn, M. C. Flemings, B. Ilschner, E. J. Kramer, and S. Mahajan, editors, *Encyclopedia of Materials: Science and Technology (Second Edition)*, pages 3750 – 3752. Elsevier, Oxford, second edition edition, 2001.

[PAT+04] B. N. J. Persson, O. Albohr, U. Tartaglino, A. I. Volokitin, and E. Tosatti. On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *Journal of Physics: Condensed Matter*, 17(1):R1–R62, 2004.

[PB11] M. Paggi and J. R. Barber. Contact conductance of rough surfaces composed of modified RMD patches. *Int. J. Heat Mass Transfer*, 54:4664–4672, 2011.

[PBC02] B. N. J. Persson, F. Bucher, and B. Chiaia. Elastic contact between randomly rough surfaces: Comparison of theory with numerical results. *Phys. Rev. B*, 65:184106, Apr 2002.

[PC10] Marco Paggi and Michele Ciavarella. The coefficient of proportionality κ between real contact area and load, with new asperity models. *Wear*, 268(7-8):1020 – 1029, 2010.

[PDM14] N. Prodanov, W. B. Dapp, and M. H. Müser. On the contact area and mean gap of rough, elastic contacts: Dimensional analysis, numerical corrections, and reference data. *Tribology Letters*, 53(2):433–448, 2014.

[Per01a] B. N. J. Persson. Elastoplastic contact between randomly rough surfaces. *Phys. Rev. Lett.*, 87:116101, Aug 2001.

[Per01b] B. N. J. Persson. Theory of rubber friction and contact mechanics. *Journal of Chemical Physics*, 115(8):3840–3861, 2001.

[Per02] B. N. J. Persson. Adhesion between elastic bodies with randomly rough surfaces. *Phys. Rev. Lett.*, 89:245502, Nov 2002.

[Per06] B. N. J. Persson. Rubber friction: role of the flash temperature. *Journal of Physics: Condensed Matter*, 18(32):7789–7823, 2006.

[Per07] B. N. J. Persson. Relation between interfacial separation and load: A general theory of contact mechanics. *Phys. Rev. Lett.*, 99:125502, Sep 2007.

[Per08] B. N. J. Persson. On the elastic energy and stress correlation in the contact between elastic solids with randomly rough surfaces. *Journal of Physics: Condensed Matter*, 20(31):312001, 2008.

[Per10] B. N. J. Persson. *Sliding Friction: Physical Principles and Applications*. NanoScience and Technology. Springer Berlin Heidelberg, 2nd edition, 2010.

[Per14] B.N.J. Persson. Role of frictional heating in rubber friction. *Tribology Letters*, 56(1):77–92, 2014.

[PH04] M. Pharr and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[PTAT04] B. N. J. Persson, U. Tartaglino, O. Albohr, and E. Tosatti. Sealing is at the origin of rubber slipping on wet roads. *Nat Mater*, 3(12):882–885, 12 2004.

[PTAT05] B. N. J. Persson, U. Tartaglino, O. Albohr, and E. Tosatti. Rubber friction on wet and dry road surfaces: The sealing effect. *Phys. Rev. B*, 71(3):035428, January 2005.

[PY08] B. N. J. Persson and C. Yang. Theory of the leak-rate of seals. *Journal of Physics: Condensed Matter*, 20(31):315011, 2008.

[Qua09] A. Quarteroni. *Numerical modeling for differential problems*. Springer Milan, 4 edition, 2009.

[RC07] M. Rubinstein and R. H. Colby. *Polymer Physics*. Oxford University Press, 3rd edition, 2007.

[Ric44] S. O. Rice. Mathematical analysis of random noise. *Bell System Technical Journal*, 23(3):282–332, 1944.

[Ric45] S. O. Rice. Mathematical analysis of random noise. *Bell System Technical Journal*, 24(1):46–156, 1945.

[Ric06] J. Rice. *Mathematical Statistics and Data Analysis*. Number p. 3 in Advanced series. Cengage Learning, 2006.

[Riv48] R. S. Rivlin. Large elastic deformations of isotropic materials. iv. further developments of the general theory. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 241(835):379–397, 1948.

[Sal08] S. Salsa. *Partial differential equations in action*. Springer, 1 edition, 2008.

[Sch02] F. Schuricht. Variational approach to contact problems in nonlinear elasticity. *Calculus of Variations and Partial Differential Equations*, 15(4):433–449, 2002.

[Sei98] S. Seireg. *Friction and Lubrication in Mechanical Design*. Mechanical Engineering. Taylor & Francis, 1998.

[SG04] J. Svendenius and M. Gäfvert. *A Semi-empirical Tire-model Including the Effects of Camber*. Institutionen för reglerteknik, Lunds tekniska högskola. Lund University, 2004.

[Sil96] V. Silberschmidt. Fractal and multifractal characteristics of propagating cracks. *J. Phys. IV France*, 06(C6):C6–287–C6–294, 1996.

[TAC+84] C. Truesdell, S.S. Antman, D.E. Carlson, G. Fichera, E. Gurtin, and P.M. Naghdi. *Mechanics of Solids: Volume II: Linear Theories of Elasticity and Thermoelasticity, Linear and Nonlinear Theories of Rods, Plates, and Shells*. Handbuch der Physik. Springer Berlin Heidelberg, 1984.

[TN65] C. Truesdell and W. Noll. *The non-linear field theory of mechanics*. Springer, 3 edition, 1965.

[TT60] C. Truesdell and R. Toupin. *The classical field theories*. Springer, 1960.

[TZU10] R. Tarefder, A. Zaman, and W. Uddin. Determining hardness and elastic modulus of asphalt by nanoindentation. *International Journal of Geomechanics*, 10(3):106–116, 2010.

[vdOF97] R. J. A. van den Oetelaar and C.F.J. Flipse. Atomic-scale friction on diamond(111) studied by ultra-high vacuum atomic force microscopy. *Surface Science*, 384(1-3):L828 – L835, 1997.

[Wan96] Y. Wang. *Finite Deformations of a Generalized Blatz-Ko Material*. University of Plymouth, 1996.

[WC68] J. L. White and Boeing Company. *Finite Elements in Linear Viscoelasticity*. Boeing Company, 1968.

[Wri06] P. Wriggers. *Computational Contact Mechanics*. Springer, 2006.

[XBR07] L. Xiao, S. Bjorklund, and B.G. Rosén. The influence of surface roughness and the contact pressure distribution on friction in rolling/sliding contacts. *Tribology International*, 40(4):694 – 698, 2007. NORDTRIB 2004.

[YAM12] V. A. Yastrebov, G. Anciaux, and J.-F. Molinari. Contact between representative rough surfaces. *Phys. Rev. E*, 86:035601, Sep 2012.

[YAM14] V. A. Yastrebov, G. Anciaux, and J.-F. Molinari. From infinitesimal to full contact between rough surfaces: evolution of the contact area. Technical Report arXiv:1401.3800, CERN, Jan 2014.

[Yas13] V. A. Yastrebov. *Numerical Methods in Contact Mechanics*. ISTE. Wiley, 2013.

[YDPC11] V. A. Yastrebov, J. Durand, H. Proudhon, and G. Cailletaud. Rough surface contact analysis by means of the Finite Element Method and of a new reduced model. *Comptes Rendus Mécanique*, 339:473–490, 2011.

[ZBBP04] G. Zavarise, M. Borri-Brunetto, and M. Paggi. On the reliability of microscopical contact models. *Wear*, 257(3-4):229 – 245, 2004.