Forward and inverse problems in mechanics: from a single to thousands of interacting bodies

by

Saviz Mowlavi

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author .....
Department of Mechanical Engineering
July 31, 2022

Certified by .....
Ken Kamrin
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by .....
Nicolas Hadjiconstantinou
Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Theses

Forward and inverse problems in mechanics: from a single to thousands of interacting bodies

by

Saviz Mowlavi

Submitted to the Department of Mechanical Engineering
on July 31, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Mechanics is the branch of physics that characterizes how bodies deform in response to forces, which can involve two categories of problems. In forward problems, one seeks to predict the response of the system given full knowledge of its physical and geometric properties. In inverse problems, some of the physical or geometric properties of the system are unknown and are either to be identified through experiments, or to be designed to optimize a desired objective. Although the physical laws governing the deformation of single elastic bodies have been known for over a century, forward problems involving thousands of interacting elastic bodies still elude simple and accurate models, while inverse problems involving even a single elastic body lack effective solution methods.

In this thesis, we investigate forward and inverse problems in systems ranging from a single elastic body to thousands of interacting ones. In a first part, we derive analytically a model for the contact force between elastically anisotropic bodies. We then implement this contact model into a computational framework for the forward dynamics of systems composed of hundreds of interacting bodies, which we leverage to showcase examples where the elastic anisotropy of each body affects the macroscopic behavior of the system. In a second part, we derive a homogenized continuum model to predict the forward dynamics of granular materials consisting of millions of interacting elastic particles, such as sand, with a particular focus on the accurate description of the onset and arrest of flow in response to external loading variations. Besides its predictive abilities, this model also sheds light on the physical mechanisms responsible for various unique features of avalanches and landslides such as their large initial acceleration. In a last part, we propose a topology optimization framework for the inverse problem of identifying hidden voids or rigid inclusions in an elastic body using measurements of the surface deformation in response to a prescribed surface loading. This framework combines recent advances in machine learning with level-set methods and the equations governing the deformation of single elastic bodies. We demonstrate the effectiveness of our method in identifying the number, locations, and shapes of hidden voids and rigid inclusions in elastic and hyperelastic materials.

Thesis Supervisor: Ken Kamrin
Title: Professor of Mechanical Engineering

Acknowledgments

First and foremost, I would like to extend my deepest gratitude towards my advisor, Prof. Ken Kamrin. Ken's level of involvement in my research during the past four years, combined with his ever-burning passion for mechanics and tackling hard problems, have made my PhD experience extremely enjoyable. He is that rare professor who loves taking a deep dive into his students' research problems, always enthusiastically discussing technical details at any level, and of any sort. These discussions have been extremely helpful in tackling the manifold challenges that inevitably arise over the course of a PhD, and have had a major impact in shaping the present thesis. Just as important, Ken has managed to strike a delicate balance between giving me excellent advice while leaving enough freedom for me to grow my skills and confidence as an independent researcher. Finally, I also thank Ken for placing his trust in me by accepting to take me into his group when I was already a third-year graduate student at MIT looking for a new research direction.

I am also grateful to my thesis committee members, Prof. Gareth McKinley and Prof. Jörn Dunkel, for their time and their interest in my work. Through much helpful advice and encouraging comments, they have played a very supportive role and I am honored that they were part of my committee. I am indebted to Prof. Themis Sapsis for welcoming me in his group during my first three years at MIT, and thank him for being very understanding of my wishes. Amongst many professors in the MechE department with whom I have had a chance to interact over the years, I would like to express my appreciation towards Prof. Irmgard Bischofberger and Prof. Wim van Rees for a wonderful experience TAing 2.005, when they taught back it in 2018. I am also particularly grateful to Irmgard for acting like a family member ever since and for giving me the chance to interact with many invited speakers over the years, both on campus and at the dinner table of fancy restaurants! Finally, I also thank Leslie Regan, Saana McDaniel, Barbara Smith, and Christina Spinelli, who strive so hard every day to make sure that every student within the department feels welcome and like at home.

Outside of MIT, I would first like to express heartfelt gratitude towards Dr. Saleh Nabi and Dr. Mouhacine Benosman from Mitsubishi Electric Research Laboratories (MERL), with whom I had a fantastic internship experience during summer 2021. Not only did this experience help me acquire expertise that became critical to enable the work presented in the fourth chapter of the present thesis, it also showed me how enjoyable academic research can be in an industrial research lab thanks to the many daily interactions between its research staff. I am also greatly indebted to them for succeeding in bringing me back to MERL as a Visiting Research Scientist, position that I am super excited to start next month. Next, I would like to acknowledge a number of academic mentors who have each contributed to shaping me as a researcher through the last decade. I thank Prof. François Gallaire for introducing me to the world of research and for placing so much faith in me, as well as Prof. Cristóbal Arratia who, together with François, expertly supervised my first research project. I thank Prof. PT Brun who later joined this research collaboration, kindly hosted me at Princeton for a week back in 2018, never missed a chance to support me through

words of encouragement and other actions, and with whom I shared an unforgettable 3-day bike & beach trip from Boston to Provincetown (remember the calzone?). I thank Prof. Marco Ramaioli for mentoring my first and, to date, only experimental research experience during an internship at Nestlé Research, of which I keep very fond memories. I thank Dr. Edouard Boujo for taking interest in one of my previous works and many ensuing passionate scientific exchanges. I thank Prof. Mattia Serra and Prof. L Mahadevan for a very fun and fruitful research side project that came to form during a class that I took at Harvard. Finally, I thank Prof. Steven Johnson and Prof. J-C Nave for inviting me to contribute to a project that was based on my earlier work with François and PT, and which gradually turned into a friendly regular chat filled with Ukulele demos and random conversations.

Of course, half (if not more!) of the fun during a PhD comes from life outside work, for which I am immensely grateful to a number of close friends whom I have had the chance to meet during my time in Cambridge. To the members of our Dynamically Qualified squad – Boyu, Cam, Essie, Rohit – I will forever cherish our countless meals at school and outside, our game nights, our hiking, surfing, skiing, and Six Flags trips, and our COVID zooms. Boyu, I also remember you being the first fluid mechanician I met at MIT, and I keep fond memories of our many fluid-filled discussions. Rohit, I could not possibly have hoped for a better roommate at 25 Andrew St and will never forget our many chats, runs and bike rides, in addition to all of the above. Kasturi, thank you for always providing such nice company; I am already sad at the thought that you will soon leave Cambridge, and I will always miss our dinners with Rohit and Cindy at 25 Andrew. Ishita, thank you for being one of Cindy's closest friends and such a great roommate. Shashank, I vaguely remember seeing you chatting with Ken outside 1-314 long before I joined the group; little did I know that we would later become so close, to the point that it's pointless for me enumerating everything we've done together – now let me work! Maytee, I am so glad that I got to know you by joining the group; you are such a pleasant person to hang out with and I look forward to many more activities together! Aaron, thank you for being such a great labmate and friend; I will never forget our morning run together in Vermont and our little excursion in the river in the White Mountains. Florian, my life during the first two years at MIT would have been different without you – we simply had way too much fun with all our swimming sessions, Indian buffets and dim sum, bike rides, BSO outings, math exchanges, potlucks, and so much more. Salman, thank you for following me (wink) since high school all the way to Cambridge to provide such great company and host delightful Nowruz dinners. Arko and Katie, thanks for all the good times at our respective homes, in restaurants, at the pool, in apple orchards, and whatever we'll do next! Abhinav, Jvo and Corbin, thank you for making me feel part of your MSEAS family; I have greatly enjoyed your friendship throughout the years – Abhinav, looking forward to the next breakfast at Café Madeleine! Xiaoyu, thank you for being such a close friend; I will never forget our countless meals at school and at dim sum, our many volunteer outings to the Food Bank and elsewhere, our hikes, and your crazy house parties. Lup Wai, you were this friend that I knew I could always count on – I miss all our lunches, our free food scavenging, our Ashdown and SP brunches, our volunteering outings,

our scientific discussions, and our potlucks. Jérome, it was a lot of fun hanging out with you during potlucks, movie nights, and working on the European Career Fair together! Antoine, we started out as labmates, then you became a great research collaborator, and finally you became a great friend; I will always treasure our lengthy lunches and dinners (in pure French tradition!) and our BSO outings. Nisha, I keep heartwarming memories of the time we were acting together as Service Chairs, and of our subsequent friendship with all its profound discussions. Chun Man and Manho, thank you for all the time spent eating, volunteering, roaming the White Mountains, and having fun together. James, thank you for all the fun, laughter, and teasing during my first year at MIT. David, I am eternally grateful to serendipity for making us bump into each other in Ashdown, marking the beginning of a long friendship that has endured your departure from Cambridge. Chinmay, thank you for the bike rides and for bringing me on your first flight after getting the license. Simon, thank you for all the long walks and the crazy conversations. Alex and Yuki, thank you for your company at MERL last summer and your friendship – I can't wait to see you guys again next month! Isha, it has been a real honor having you as both a scientific collaborator and a close friend. Finally, thank you, John, Daniel, Mukund, Ravi, Mo, Sid, Cherry, Jerry, Rachel, Wael, Alexis, Steven, Dixia, Michel, Karine, Will, Qiong, João, Juncal, Victor, for your friendship over the years.

I also wish to thank the local Bahá'í communities of Cambridge, Somerville and Boston, which have been like a home outside home. Amongst many people, I wish in particular to express my gratitude towards Zarrin, Brian, Donna, and Arman for treating MIT Bahá'í students like their own family, Cristina, Clarissa and Shoghi for our friendship and collaboration on the social justice devotionals, Shabnam for your company in the MechE department, and Sujay and Yukie for the good times and fun conversations.

Finally, I simply would not be in this place, at this time, without the uncountable efforts and sacrifices of my parents. Papa and 媽媽, a huge thank you for everything you have done, and for always encouraging us to follow our own passions, whichever they might be. 弟弟, thank you for being such a great brother, so fun to talk to, and always present when needed. Sára'ng, you are always looking after me, and I look forward to getting to know you one day. I am also deeply grateful to my wife Cindy's parents for their care and treating me like their child. Last but not least, Cindy, there are no words to express my heartfelt gratitude for your unwavering love, support, and – as is crucially important for a graduate student's motivation – for always showing such great interest in my work!

Contents

Chapter 1

Introduction

1.1 Motivation

Mechanics is the branch of physics that characterizes how bodies deform in response to forces. In this thesis, we focus on elastic bodies, which are characterized by the property that they do not dissipate energy when deforming under load [73]. Intuitively, an elastic body returns to its original size and shape once the load is removed. The physical laws governing the deformation of single elastic bodies are well-known; they consist of mass and momentum conservation laws together with a constitutive law specifying the relationship between the stress and the strain at every point in the body [111]. Denoting by $\mathbf{u}(\mathbf{x}, t)$ the displacement field that characterizes the motion at time t of every initial point $\mathbf{x}$ in a single body, these physical laws can be expressed as a governing equation of the form

$$\mathbf{f}(\mathbf{u}; \boldsymbol{\theta}) = \mathbf{0}, \tag{1.1}$$

where $\boldsymbol{\theta}$ comprises all the physical and geometric parameters that describe the problem: the elastic constants defining the constitutive law (which depend on the material that the body is made of), the geometry of the object, the applied loading, etc. When $\boldsymbol{\theta}$ is fully specified, solving (1.1) for $\mathbf{u}$ yields the displacement field that describes the deformation of the body. This is known as a *forward problem*, where one seeks to predict the response of the system given full knowledge of its physical and geometric properties.

Consider now a general situation involving N interacting elastic bodies, whose motion will be described by the combined displacement fields $\mathbf{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ within each body. In this case, one can use the governing equation (1.1) for the deformation of each body to obtain a general governing equation for the entire system, which we denote

$$\mathbf{F}(\mathbf{U}; \boldsymbol{\Theta}) = \mathbf{0}, \tag{1.2}$$

where $\boldsymbol{\Theta}$ now denotes the physical and geometric parameters describing the entire problem consisting of all N bodies: the elastic constants and geometry of each body, and the loading applied to the whole system. As in the single-body case, one can

use (1.2) to solve the forward problem of finding the combined displacement fields $\mathbf{U}$ given full knowledge of $\boldsymbol{\Theta}$.

Since the forward solution of (1.1) or (1.2) is generally intractable analytically except in very simple cases, one resorts to computational methods such as the finite-element method in which the continuous field $\mathbf{u}(\mathbf{x}, t)$ within each body is replaced by discrete values $\mathbf{u}_i^j$ defined at discrete times t_j and initial locations $\mathbf{x}_i$ on a mesh. The governing equations (1.1) or (1.2) are then discretized in space and time, yielding a numerical solution for all discrete values $\mathbf{u}_i^j$. Although this procedure is generally tractable for problems involving a single body, its computational cost rapidly becomes prohibitive as the number of objects involved increases. This calls for the development of reduced models that can replace (1.2) to describe the dynamics of the system in a computationally more efficient fashion while retaining enough accuracy.

In many practical applications, such as when the number of bodies involved is very large, one is not concerned with the deformation field within each body. This motivates a first class of reduced models to replace (1.2), called the discrete-element method (DEM) [201]. In the DEM, the deformation of every body is lumped into the motion of its center of mass, which can be computed efficiently using Newton's equations of motion. This method requires a *contact force law* that specifies the force generated between two contacting particles as a function of their overlap distance. As such, the accuracy of any model based on the DEM is tied to the accuracy of the contact force law being used. So far, an exact contact force law only exists for the case of elastically isotropic smooth particles, which implies that systems consisting of any other type of particle (rough, elastically anisotropic, plastic, etc) will incur errors when modeled with the DEM.

For granular materials such as sand which contain thousands to billions of particles, methods based on the DEM are still too computationally expensive. In that case, since one is not usually concerned with the individual motion of every particle, an other alternative to replace (1.2) is a second class of reduced models based on homogenization and continuum mechanics [3]. The idea is to average out, or homogenize, the motion and the forces of every constituent particle so that the discrete system of many bodies can be described by an equivalent, 'effective' continuum material. This effective continuum material is described by the same mass and momentum conservation laws contained in the governing equation (1.1) for the deformation within a single elastic body. However, since the stress and strain within this effective continuum material correspond to averages over discrete particles, the constitutive relationship that relates them will be very different from the case of a single elastic material. In fact, the right form for this constitutive relationship – that yields identical behavior between the effective continuum material and the discrete system – has proven elusive over decades and is still an active subject of research today. Challenges stem from the complex behavior of granular materials, including their ability to behave like a solid, a fluid, and a gas, their stochasticity, and their nonlocality.

Besides forward problems, another class of problems relevant to engineering applications is that of *inverse problems*. In inverse problems, some of the physical or geometric properties contained in $\boldsymbol{\theta}$ or $\boldsymbol{\Theta}$ are unknown; they are instead to be found so that the corresponding solution $\mathbf{u}$ of (1.1), or $\mathbf{U}$ of (1.2), minimizes a user-defined

objective $J(\mathbf{u})$ or $J(\mathbf{U})$. This can correspond either to identifying unobserved properties of an actual system given measurement data (in which case J measures the discrepancy between predictions and measurements), or to designing properties so that a hypothetical system achieves a target functionality (in which case J is a design objective that is minimum when the functionality is achieved). Inverse problems are solved using iterative methods, where a starting guess for the unknown properties is iteratively refined until the corresponding state $\mathbf{u}$ or $\mathbf{U}$ minimizes J. Even for the case of a single elastic body, inverse problems are notoriously difficult to solve when the number of properties to be found is large, since searching through a high-dimensional space requires many solutions of (1.1) and adjoint-based methods to guide the search direction. An additional difficulty concerns the case where geometric properties are to be found; in that case it is desirable to avoid restricting the topology of the solution, which calls for a solution methodology that can handle merging and splitting of shapes.

1.2 Outline of the thesis

In this thesis, we address several forward and inverse problems in systems ranging from a single elastic body to thousands of interacting ones, related to the themes presented above.

In Chapter 2, we develop an elastically anisotropic DEM model for the accurate modeling of systems consisting of hundreds to thousands of elastically-anisotropic particles, whose constitutive relationship depends on the orientation of the crystalline atomic structure within each particle. Using the governing equation (1.1) for the elastic deformation within each particle, we derive analytically an exact model for the force generated between two contacting elastically anisotropic bodies. We then implement a simplified yet accurate version of this contact model into a custom-developed DEM code, which we leverage to showcase examples where the elastic anisotropy of each body affects the macroscopic behavior of the system.

In Chapter 3, we extend an existing homogenized continuum model to predict the forward dynamics of granular materials consisting of millions of interacting elastic particles, such as sand, with a particular focus on the accurate description of the onset and arrest of flow in response to external loading variations. Through simulations of the model in an idealized geometry, we shed light on the physical mechanisms responsible for various unique features of avalanches and landslides such as their large initial acceleration or their spontaneous arrest. Further, we evaluate the predictive abilities of the model by comparing its predictions with DEM simulations in a couple of different geometries.

In Chapter 4, we propose a topology optimization framework for solving the inverse problem of identifying hidden voids or rigid inclusions in a single elastic body, using measurements of the surface deformation in response to a prescribed surface loading. This framework combines recent advances in machine learning with the known equations governing the deformation of single elastic bodies, and hinges on the introduction of a novel eikonal regularization of the material density field parametrizing

the geometry. Using examples involving a variety of geometries, we demonstrate the effectiveness of our proposed method to identify the topology, location and shapes of hidden structures in both linear elastic and nonlinear hyperelastic materials.

1.3 Contributions

All three main chapters of this thesis constitute original work jointly carried out by myself and my advisor, Ken Kamrin. Chapters 2 and 3 were published in the following two articles, while Chapter 4 forms the basis of a manuscript in preparation.

- S. Mowlavi and K. Kamrin. Contact model for elastically anisotropic bodies and efficient implementation into the discrete element method. *Granular Matter*, 23(2):1-29, 2021.
- S. Mowlavi and K. Kamrin. Interplay between hysteresis and nonlocality during onset and arrest of flow in granular materials. *Soft Matter*, 17(31):7359-7375, 2021.

Chapter 2

Contact model for elastically anisotropic bodies and efficient implementation into the discrete element method

2.1 Introduction

Beginning with the seminal paper of Cundall and Strack [41], the Discrete Element Method (DEM) has rapidly established itself as a method of choice for simulating the behavior of granular materials in a wide range of situations [201, 123, 72]. In this approach, Newton's equations of motion are integrated individually for every particle in the system, taking into account body forces as well as surface forces that arise from the interactions of contacting particles. Contact force laws dictate the magnitude of these surface forces as a function of the overlap between adjacent particles. As such, they are an essential ingredient of any DEM simulation, and a multitude of contact laws of various complexities have been formulated to account for effects as varied as friction [41, 189, 180], damping [25], torsion [52], cohesion [171, 116, 126], plasticity [181], and so forth.

Contact force models may be divided into two broad classes [200]. The first concerns models that are formulated based on an exact or approximate solution of the physics governing the contact problem at the scale of the individual grains. The most prominent example is Hertz' contact law [83], which gives an expression for the normal force generated by the elastic deformation of two contacting spheres. Hertz' contact law, which is based on the exact solution of the continuum elasticity equations for this problem, takes a remarkably simple form wherein the force is dependent on the three halves power of the overlap distance between the particles [92]. Contact force models belonging to the second class are formulated empirically, balancing ease of implementation and computational cost with accuracy of the results. Cundall and Strack's linear spring-dashpot model falls under this second category.

Although the second class of methods is particularly useful when one wants to

incorporate physical mechanisms that elude simple analytical solutions, the first class is preferable when one is concerned with the precise quantification of the forces in a granular medium. For instance, numerous studies [139, 86, 35] have investigated the distribution and properties of interparticle forces in granular materials and their connection with the external loading characteristics. In the case of elastically isotropic bodies, for which the contact force is independent of the direction of contact, Hertz' contact law will return the exact forces as long as the deformation of the bodies is small and contact points on the same particle are not too close. Most materials in nature, however, are elastically anisotropic. Whenever the size of individual particles becomes small enough, crystalline grains become apparent relative to the particle size [56], and the contact force between particles will be direction-dependent as a result of elastic anisotropy. Myriad engineering processes such as additive manufacturing [130] or ceramic packings [76, 40] involve powders of fine particles, and the accurate quantification of interparticle forces in these cases calls for a contact force law that can take elastic anisotropy into account. Clearly, one expects van der Waals forces to play an important role at these small scales, but the relative strength of such attractive forces decreases with increasing load and their modeling has already been treated previously [93, 10]. Furthermore, monocrystalline granular particles of a larger size do exist, both naturally [104] and artificially [49]. Such particles notably play a key role in novel experimental methods for inferring particle-wise strain tensors in opaque packings by exploiting X-ray diffraction [75, 87].

In this chapter, we derive a contact law for the normal elastic force that is generated between two elastically anisotropic bodies of arbitrary geometry, as long as the surfaces are smooth and frictionless. Our approach begins with the formulation of a numerical procedure for the exact analytical solution to the continuum elasticity problem, which builds on more than fifty years of research in the contact mechanics literature [9]. In particular, several authors have sought to extract a relationship between indentation force, depth, and contact area during the unloading branch of an indentation test, wherein an axisymmetric rigid indentor is pressed against an elastically anisotropic half space [184, 185, 178, 183, 48]. The exact solution procedure that we present here extends the scope of these studies to the case of two contacting, elastically anisotropic bodies with smooth and non-spherical geometry, which lacks a detailed treatment in the previous endeavors.

We then simplify the exact solution into a readily implementable anisotropic contact force law, which in the particular case of spherical contacting bodies B_1 and B_2 of radii R^{B_1} and R^{B_2} takes the form

$$F = \frac{4}{3}\tilde{E}^c_*(\alpha^{B_1}, \beta^{B_1}, \alpha^{B_2}, \beta^{B_2})R^{1/2}\delta^{3/2}, \tag{2.1}$$

where F is the normal force and δ the overlap between the two bodies, $R = (1/R^{B_1} + 1/R^{B_2})^{-1}$ is the composite radius, $\tilde{E}^c_*$ is a material-specific composite modulus depending on two sets of Euler angles (α^B, β^B) describing the orientation of the contact normal direction with respect to the internal axes of bodies $B = B_1$ and B_2. The *only* difference between the simplified anisotropic contact law (2.1) and Hertz' famil-

iar contact law for isotropic bodies lies in the orientation-dependence of the composite modulus $\tilde{E}^c_*$, which calls upon the entire set of elastic constants for the material comprising each body. This similarity between the isotropic and simplified anisotropic contact laws extends to smooth particles of abritrary shape, as we show later in the chapter.

The simplification utilizes Vlassak *et al.*'s [183] idea of truncating the Fourier series expansion of the surface Green's function to its constant term, which was shown in [183] to result in accurate predictions of the force generated by a rigid spherical indentor on a half space made of sapphire. We demonstrate that this accuracy is retained in our contact law for generic smooth contacting particles over a wide range of materials, with the simplified anisotropic contact law differing from the exact solution by less than 1% in all considered cases. We then show how to efficiently implement these formulas in a DEM scheme, taking advantage of the offline precomputation of the material-specific contact modulus over all possible orientations. Finally, we present two examples involving assemblies of single-crystal zirconia particles that display how anisotropy at the particle level alters macroscopic behavior and can be exploited in applications.

The chapter is organized as follows. We begin by formulating the contact problem and describe the solution methodology in Section 2.2, based on which an exact anisotropic contact force law is then derived in Section 2.3. Through successive simplifications of the exact solution, we then propose in Section 2.4 two simplified contact force laws, which we compare with the exact solution in Section 2.5 after validating the latter against finite-element simulations. An implementation of the first simplified contact law into a DEM code is then presented in Section 2.6, along with two example applications featuring elastically anisotropic particles. Conclusions close the chapter in Section 2.7. Finally, we invite the reader interested in the implementation details to consult the appendices.

2.2 Problem statement and solution methodology

2.2.1 Definition of the contact problem

We consider two elastically anisotropic bodies B_1 and B_2, comprised of materials having elasticity tensors $\mathbb{C}^{B_1}$ and $\mathbb{C}^{B_2}$. Throughout the chapter, quantities with a superscript B_1 and B_2 will refer to body B_1 and body B_2, respectively, and quantities with a superscript B will refer to either body interchangeably. In the reference unstressed configuration, the two bodies are contacting at a single point and are separated by a common tangent contact plane, as pictured in Figure 4-1(a). Let the contact point O be the origin of a cartesian coordinate system (x, y, z), where the x-y plane is the common tangent plane and the z-axis is directed along the inward normal of body B_1. The initial gap $g_0(x, y)$ measures the gap between the undeformed bodies, and is given to lowest order by

$$g_0(x, y) = Mx^2 + Ny^2, \tag{2.2}$$

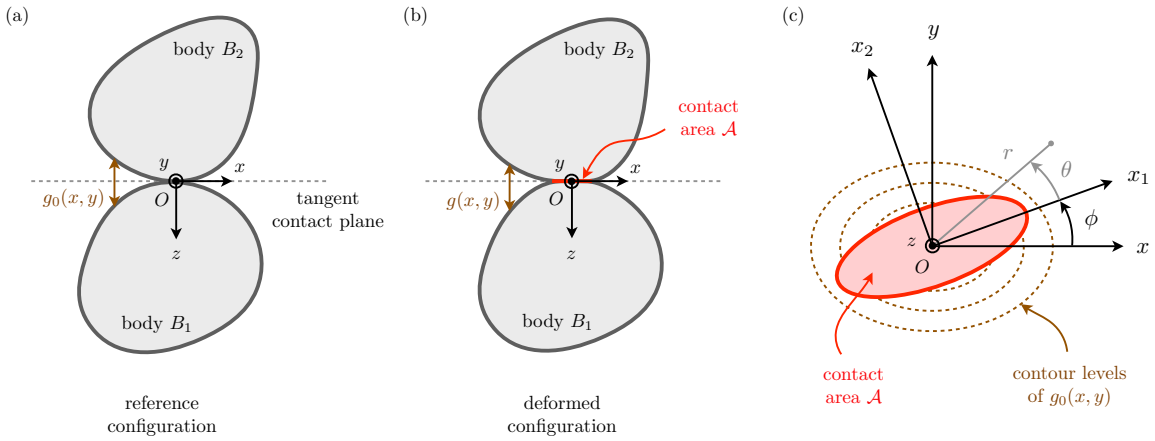

Figure 2-1: Geometry of the contact problem. (a) In the reference configuration, the two bodies are contacting at a single point O and separated by a common tangent plane. The coordinate system (x, y, z) is defined in such a way that the (x, y)-axes, spanning the tangent plane, are aligned with the principal axes of the contour levels of the initial gap function $g_0(x, y)$. (b) In the deformed configuration, the two bodies are pressed against each other with a normal force F, resulting in a relative displacement normal to the tangent plane as well as the establishment of a finite contact region $\mathcal{A}$. (c) In the tangent contact plane, the contact area takes the shape of an ellipse whose major and minor axes (x_1, x_2) are rotated by an angle ϕ with respect to the (x, y) axes. The set of polar coordinates (r, θ) used in Section 2.3.2 is defined with respect to the (x_1, x_2) axes.

where the x- and y-axes have been chosen so that they align with the principal axes of the contour levels of $g_0(x, y)$, and $N \geq M$ by convention. In this work, we only consider bodies with a smooth and convex surface, for which the first-order terms of $g_0(x, y)$ are zero and M, N are both positive. While outside the scope of this chapter, we mention that formulae to obtain M and N from the principal radii of curvature of bodies B_1 and B_2 at the contact point are given in the books of Johnson [92] and Barber [12]. In the specific case of ellipsoidal bodies, the calculation of the principal radii of curvature knowing the contact point and orientations of B_1 and B_2 is nontrivial and explained in the appendix of [199].

The bodies are then pressed against each other with a force F directed along the normal to the contact plane, which results in a relative displacement δ and the establishment of a finite contact region $\mathcal{A}$, as pictured in Figure 4-1(b). We denote the vertical surface displacement generated in each body along the z-axis by $w^{B_1}(x, y)$ and $w^{B_2}(x, y)$, both measured positive into the respective body. The final gap $g(x, y)$ is then given by

$$g(x, y) = g_0(x, y) - \delta + w^{B_1}(x, y) + w^{B_2}(x, y). \tag{2.3}$$

Inside the contact region $\mathcal{A}$, the gap $g(x, y)$ must vanish, which implies

$$w^{B_1}(x, y) + w^{B_2}(x, y) = \delta - g_0(x, y), \quad (x, y) \in \mathcal{A}. \tag{2.4}$$

Outside the contact region, the gap $g(x, y)$ must be positive, which translates as

$$w^{B_1}(x, y) + w^{B_2}(x, y) > \delta - g_0(x, y), \quad (x, y) \notin \mathcal{A}. \tag{2.5}$$

We assume that the surfaces are frictionless, so that there is only a normal traction (that is, a pressure) $p(x, y)$ between the bodies, which resultant over the contact area $\mathcal{A}$ is equal to F. The boundary conditions (2.4) and (2.5) are supplemented by the condition that $p(x, y) > 0$ for $(x, y) \in \mathcal{A}$, and $p(x, y) = 0$ for $(x, y) \notin \mathcal{A}$.

The problem, therefore, is to find the contact area $\mathcal{A}$ and pressure distribution $p(x, y)$ such that the resulting surface displacements satisfy the boundary conditions (2.4) and (2.5). In this way, the normal force F between the two bodies can be related with their relative displacement δ. This elasticity problem was first solved analytically by Hertz [83] for elastically isotropic bodies, leading to the well-known Hertz contact law. The solution process is, however, much more cumbersome for elastically anisotropic bodies. While integral expressions have been derived and solution strategies have been suggested by various authors using a range of mathematical techniques [191, 11, 178, 183, 65, 10], an exact step-by-step solution scheme for generally-shaped contacting surfaces, including the non-circular case $M \neq N$, is still missing.

2.2.2 Solution methodology

Our general solution strategy for the contact problem is based on Hertz's derivation of the elastically isotropic case [83, 92], and proceeds in a similar way for both isotropic and anisotropic bodies. First, one introduces the simplification that the contact region

$\mathcal{A}$ is flat, and that the surface displacements generated by the pressure distribution $p(x, y)$ are equal to those that would be produced in equivalent semi-infinite bodies (i.e., elastic half-spaces) loaded with the same surface pressure distribution over the same contact region. In order for this simplification to hold, the size of the contact area must be small with respect to the dimensions of each body as well as their principal radii of curvature at the contact point. This simplification, first introduced by Hertz, enables one to express the combined surface displacements as the convolution

$$w^{B_1}(x, y) + w^{B_2}(x, y) = \sum_{B \in \{B_1, B_2\}} \iint_{\mathcal{A}} \hat{w}^B(x - x', y - y') p(x', y') dx' dy', \tag{2.6}$$

where $\hat{w}^B(x - x', y - y')$ is the vertical surface displacement at (x, y) produced by a unit concentrated normal load at (x', y') on the surface of an elastic half-space. As we will see later, the Green's function $\hat{w}^B(x, y)$ is a known quantity that depends on the elasticity tensor $\mathbb{C}^B$ of body B together with, for anisotropic bodies, its orientation with respect to the contact plane.

The next step is to find the shape of the contact region and distribution of pressure such that the combined surface displacements predicted by (2.6) agree with the boundary conditions (2.4) and (2.5). Consider a flat elliptical[1] contact area with semi-axes lengths a_1 and a_2,

$$\mathcal{A} = \left\{ (x_1, x_2) : \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} < 1 \right\}, \tag{2.7}$$

where $a_2 \leq a_1$ by convention, and the (x_1, x_2) coordinates are rotated by some yet-unknown angle ϕ about the (x, y) coordinates, as shown in Figure 4-1(c). In addition, consider a pressure distribution of the form

$$p(x_1, x_2) = p_0 \left(1 - \frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} \right)^{\xi}, \quad (x_1, x_2) \in \mathcal{A}, \tag{2.8}$$

where the exponent ξ is unknown[2] in advance. For the respective cases of isotropic and anisotropic bodies, Hertz [83] and Willis [191] showed that when $\xi = 1/2$ (and only then), the postulated contact area (2.7) and pressure distribution (2.8) produce combined surface displacements (2.6) that are compatible with the conditions (2.4) and (2.5), thereby validating the functional forms (2.7) and (2.8). In fact, in the isotropic case, it can be immediately shown that (2.7) and (2.8) solve (2.4-2.6) by appealing to a known analogous result from potential theory (see [110] for details).

The problem is now reduced to finding the scalar parameters a_1, a_2, ϕ, and p_0, given M, N, δ, as well as the orientation and elastic constants of the contacting

[1] Hertz was guided by his observations of elliptic optical interference fringes between two contacting glass lenses, which is the very problem that motivated his subsequent analysis of the contact deformation [92].

[2] Asymptotic arguments, however, require that for smooth contacting bodies the contact pressure tend to zero at the boundary of the contact area [12], which implies that ξ is positive.

bodies. Once this is done by equating the coefficients in (2.4) and (2.6), the contact law for the force F can be obtained through the relation

$$F = \iint_{\mathcal{A}} p(x_1', x_2')dx_1'dx_2' = \frac{2}{3}\pi p_0 a_1 a_2. \tag{2.9}$$

The following section goes through this process in detail, beginning with elastically isotropic bodies in Section 2.3.1 and followed by anisotropic bodies in Section 2.3.2. In each case, the Green's function is first presented (equations (2.10) and (2.22) for isotropic and anisotropic bodies, respectively) and inserted in the convolution integral (2.6) to obtain the surface displacements (equations (2.12) and (2.28) for isotropic and anisotropic bodies, respectively). The unknown scalar parameters are then found by enforcing the boundary condition (2.4), eventually leading to a relation between the contact force F and the relative displacement δ (equations (2.18) and (2.34) for isotropic and anisotropic bodies, respectively).

2.3 Derivation of the exact contact force

2.3.1 Isotropic bodies

Green's function and surface displacements

We begin with a review of the solution for elastically isotropic bodies, which we will later refer to when developing a simplified anisotropic solution. In the isotropic case, the Green's function $\hat{w}^B(x_1, x_2)$ is axisymmetric and given in closed form as

$$\hat{w}^B(x_1, x_2) = \frac{1}{\pi E_*^B (x_1^2 + x_2^2)^{1/2}}, \tag{2.10}$$

where E_*^B is the plane strain modulus of body B, defined from its Young's modulus E^B and Poisson's ratio ν^B as

$$E_*^B = \frac{E^B}{1 - (\nu^B)^2}. \tag{2.11}$$

Inserting (2.10) into (2.6) and using (2.9), we find that the combined surface displacement within the contact area $\mathcal{A}$ caused by the pressure distribution (2.8) is [12]

$$w^{B_1}(x_1, x_2) + w^{B_2}(x_1, x_2) = \frac{3F}{4\pi a_1 E_*^c}\left(I_0(e) - \frac{x_1^2}{a_1^2}I_1(e) - \frac{x_2^2}{a_1^2}I_2(e)\right), \tag{2.12}$$

where E_*^c is the composite plane strain modulus,

$$E_*^c = \left(\frac{1}{E_*^{B_1}} + \frac{1}{E_*^{B_2}}\right)^{-1}, \tag{2.13}$$

e is the eccentricity of the contact area,

$$e = \sqrt{1 - \left(\frac{a_2}{a_1}\right)^2}, \tag{2.14}$$

and $I_0(e)$, $I_1(e)$, and $I_2(e)$ are integrals defined as

$$I_0(e) = \int_0^\pi \frac{d\theta}{(1 - e^2 \cos^2\theta)^{1/2}}, \tag{2.15a}$$

$$I_1(e) = \int_0^\pi \frac{\sin^2\theta d\theta}{(1 - e^2 \cos^2\theta)^{3/2}}, \tag{2.15b}$$

$$I_2(e) = \int_0^\pi \frac{\cos^2\theta d\theta}{(1 - e^2 \cos^2\theta)^{3/2}}. \tag{2.15c}$$

It now remains to identify the surface displacements (2.12) with the boundary condition (2.4) in order to solve for the unknowns a_1, e, ϕ, and F. This last step of the solution process is described hereafter.

Contact force solution

The solution procedure presented here is similar to that given in Barber [12], with the exception that the latter reference uses complete elliptic integrals of the first and second kind instead of (2.15). This leads to a numerically ill-posed problem when e vanishes, which we avoid by working with expressions (2.15).

Equating the surface displacements (2.12) with the boundary condition (2.4), we find that the pressure distribution (2.8) gives the correct surface displacements provided that the (x_1, x_2) axes coincide with (x, y) (that is, $\phi = 0$, which means that the major and minor axes of the pressure distribution are aligned with those of the initial gap function). In addition, this yields the relations

$$\frac{3FI_0(e)}{4\pi a_1 E_*^c} = \delta, \tag{2.16a}$$

$$\frac{3FI_1(e)}{4\pi a_1^3 E_*^c} = M, \tag{2.16b}$$

$$\frac{3FI_2(e)}{4\pi a_1^3 E_*^c} = N. \tag{2.16c}$$

By combining (2.16b) and (2.16c), we obtain a simple nonlinear equation for the eccentricity,

$$\frac{I_2(e)}{I_1(e)} - \frac{N}{M} = 0. \tag{2.17}$$

The contact force F then follows from (2.16a) and (2.16b) as

$$F = \frac{4\pi}{3} E_*^c \frac{[I_1(e)]^{1/2}}{[I_0(e)]^{3/2}} M^{-1/2} \delta^{3/2}, \tag{2.18}$$

where the material parameter E_*^c is defined in (2.13). In summary, the isotropic contact force law requires the solution of equation (2.17) for e, after which F can be obtained with (2.18).

Spherical case

We conclude our review of isotropic materials with a discussion on the form of the Hertzian solution for the limiting case of spherical contacting bodies, which results in the celebrated expression commonly referred to as the Hertz contact law [200, 107]. Consider two contacting spheres of radii R^{B_1} and R^{B_2}. To lowest order, the gap between the undeformed bodies is given by

$$g_0(x, y) = \frac{x^2}{2R} + \frac{y^2}{2R}, \tag{2.19}$$

where $1/R = 1/R^{B_1} + 1/R^{B_2}$. Therefore $M = N = 1/2R$, in which case (2.17) gives $e = 0$, and (2.18) reduces to the Hertz contact law,

$$F = \frac{4}{3} E_*^c R^{1/2} \delta^{3/2}. \tag{2.20}$$

2.3.2 Anisotropic bodies

Green's function and surface displacements

For anisotropic bodies, there is no direct algebraic expression for the Green's function $\hat{w}^B(x_1, x_2)$. Various integral expressions have been derived by different authors, starting with Willis [191] who performed a Fourier transform in the x-y plane and solved implicitly the resulting equations. Willis' expression, however, requires the simultaneous solution of multiple nonlinear integral equations, making it challenging to work with in practice. Instead, we utilize in this chapter a direct integral expression for the Green's function derived by Barnett and Lothe [13], obtained by solving the Fourier-transformed equations using a formalism due to Stroh [176].

First, let the coordinates (X_1^B, X_2^B, X_3^B) represent a basis that is preferentially oriented for the material structure in body B, and with respect to which the components of the elasticity tensor are $\mathbb{C}_{ijkm}^B$. In body B, the stress and strain are therefore everywhere related as

$$\epsilon_{ij}^B = \mathbb{C}_{ijkm}^B \sigma_{km}^B, \tag{2.21}$$

where σ_{km}^B and ϵ_{ij}^B are, respectively, the components of the local stress and strain tensors in the (X_1^B, X_2^B, X_3^B) basis. Figure 2-2(a) depicts this body-centric basis for the same bodies B_1 and B_2 introduced in Figure 4-1, but here viewed from the global reference frame $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$. (The latter is introduced for future reference and will not be referred to in this section.) We introduce the unit normal $\mathbf{n}$ to the contact plane, which is directed from body B_1 to body B_2, i.e. along the negative z-direction. As pictured in Figures 2-2(b) and 2-2(c), we denote by (n_1^B, n_2^B, n_3^B) the components of $\mathbf{n}$ in the (X_1^B, X_2^B, X_3^B) basis of each body. Then, Barnett and Lothe's expression

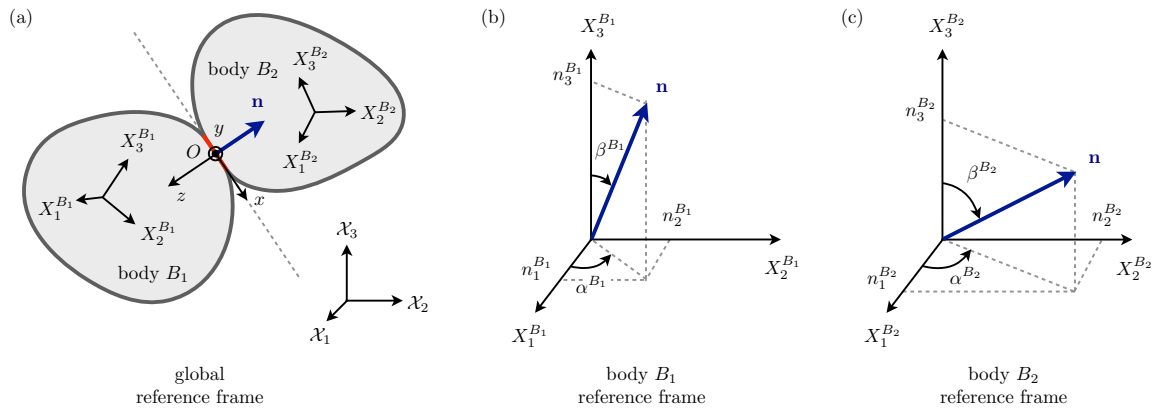

Figure 2-2: Further details on the geometry of the problem. (a) The two contacting bodies depicted in Figure 4-1 are here viewed from the global reference frame, defined by the coordinates $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$. We represent the local coordinates bases $(X_1^{B_1}, X_2^{B_1}, X_3^{B_1})$ and $(X_1^{B_2}, X_2^{B_2}, X_3^{B_2})$ of bodies B_1 and B_2, the contact normal and tangent plane directions (x, y, z), as well as the unit normal $\mathbf{n}$ to the tangent contacting plane. (b,c) In the reference frames $(X_1^{B_1}, X_2^{B_1}, X_3^{B_1})$ and $(X_1^{B_2}, X_2^{B_2}, X_3^{B_2})$ of body B_1 and B_2, respectively, the unit-length contact normal $\mathbf{n}$ can be parameterized either by its coordinates (n_1^B, n_2^B, n_3^B), or by the two Euler angles (α^B, β^B).

for the vertical displacement at a point P in the x-y plane due to a concentrated unit vertical load at the origin reads[3] (see the appendix of [185])

$$\hat{w}^B(\mathbf{x}) = \frac{1}{|\mathbf{x}|}\left[n_k^B G_{km}^{-1}\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right) n_m^B\right], \tag{2.22}$$

where $\mathbf{x}$ is the position vector of P. The matrix $[\mathbf{G}]$ in the above equation is defined as

$$G_{ij}(\mathbf{t}) = \int_0^{2\pi} \left(\{\mathbf{rr}\}_{ij} - \{\mathbf{rs}\}_{ik}\{\mathbf{ss}\}_{kr}^{-1}\{\mathbf{sr}\}_{rj}\right) d\gamma, \tag{2.23}$$

where $\mathbf{r}, \mathbf{s}, \mathbf{t}$ are unit vectors such that $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ forms a right-hand Cartesian system, γ is the angle between $\mathbf{r}$ and some fixed point in the plane perpendicular to $\mathbf{t}$, and the matrices $(\mathbf{ab})$ are defined as

$$\{\mathbf{ab}\}_{jk} = a_i \mathbb{C}_{ijkm}^B b_m, \tag{2.24}$$

with (a_1, a_2, a_3) and (b_1, b_2, b_3) denoting the components of vectors $\mathbf{a}$ and $\mathbf{b}$ in the (X_1^B, X_2^B, X_3^B) basis. It now remains to substitute (2.22) into (2.6) and solve the resulting integral. This is no easy task, but Barber and Ciavarella [10] have suggested

[3] Although the unit normal $\mathbf{n}$ is shared between bodies B_1 and B_2 and hence points in opposite directions with respect to each body's surface, expression (2.22) is valid for both bodies since it is quadratic in the components of $\mathbf{n}$.

an efficient strategy for doing so, which we formalize here.

We define the set of polar coordinates (r, θ) as $(x_1, x_2) = (r\cos\theta, r\sin\theta)$, as shown in Figure 4-1(c). The angle θ is measured with respect to the (x_1, x_2) axes, which are rotated by an as-yet-unknown angle ϕ with respect to the (x, y) axes. Since the orientation of the latter with respect to the (X_1^B, X_2^B, X_3^B) basis is known, we write the Green's function (2.22) in the 'rotated' polar coordinates (r, θ) as

$$\hat{w}^B(r, \theta; \phi) = \frac{1}{r}\left[n_k^B G_{km}^{-1}(\theta; \phi) n_m^B\right] = \frac{h^B(\theta; \phi)}{r}, \tag{2.25}$$

where the presence of ϕ emphasizes the dependence of the polar Green's function on the orientation ϕ of the (x_1, x_2) basis. For completeness, we shall mention that $h^B(\theta; \phi)$ is also a function of the material parameters as well as the (known) orientation of the (x, y, z) basis with respect to the (X_1^B, X_2^B, X_3^B) basis, which can be characterized by a rotation matrix as described in Appendix A.1. In an effort to preserve clarity of exposure, however, we have omitted this dependence in our notation.

As a consequence of Maxwell's reciprocal theorem (see [12]), the function $h^B(\theta; \phi)$ satisfies the relation $h^B(\theta; \phi) = h^B(\theta + \pi; \phi)$ and therefore admits the Fourier expansion

$$h^B(\theta; \phi) = \sum_{m=0}^{\infty} a_m^B(\phi)\cos 2m\theta + \sum_{m=1}^{\infty} b_m^B(\phi)\sin 2m\theta. \tag{2.26}$$

Due to the way that the angles θ and ϕ are defined in Figure 4-1(c), we necessarily have $h^B(\theta; \phi) = h^B(\theta + \phi; 0)$. As a consequence, the Fourier coefficients $a_m^B(\phi)$ and $b_m^B(\phi)$ can be expressed as

$$a_m^B(\phi) = a_m^B(0)\cos 2m\phi + b_m^B(0)\sin 2m\phi, \tag{2.27a}$$
$$b_m^B(\phi) = -a_m^B(0)\sin 2m\phi + b_m^B(0)\cos 2m\phi. \tag{2.27b}$$

Therefore, the knowledge of $h^B(\theta; 0)$ suffices to calculate the Fourier coefficients $a_m^B(\phi)$ and $b_m^B(\phi)$. Given the elasticity tensor $\mathbb{C}^B$ as well as the orientation of the (x, y, z) basis with respect to the (X_1^B, X_2^B, X_3^B) basis, we present in Appendix A.2 an algorithm for computing $h^B(\theta; 0)$. In practice, the Fourier coefficients $a_m^B(0)$ and $b_m^B(0)$ decay very quickly with m, and we have found that truncating the Fourier series at $m = 5$ is perfectly adequate.

As shown in Barber and Ciavarella [10], the integral (2.6) can then be solved in polar coordinates using (2.26), leading to the combined surface displacement

$$w^{B_1}(x_1, x_2) + w^{B_2}(x_1, x_2) = \frac{3F}{4a_1}\Bigg\{\sum_{m=0}^{\infty} a_m(\phi)\left[I_{0,m}(e) - \frac{x_1^2}{a_1^2}I_{1,m}(e) - \frac{x_2^2}{a_1^2}I_{2,m}(e)\right] + \frac{x_1 x_2}{a_1^2}\sum_{m=1}^{\infty} b_m(\phi) I_{3,m}(e)\Bigg\}, \tag{2.28}$$

where $a_m(\phi) = a_m^{B_1}(\phi) + a_m^{B_2}(\phi)$, $b_m(\phi) = b_m^{B_1}(\phi) + b_m^{B_2}(\phi)$, and the integrals $I_{0,m}(e)$,

$I_{1,m}(e)$, $I_{2,m}(e)$, and $I_{3,m}(e)$ are defined as

$$I_{0,m}(e) = \int_0^\pi \frac{\cos(2m\theta)d\theta}{(1-e^2\cos^2\theta)^{1/2}}, \tag{2.29a}$$

$$I_{1,m}(e) = \int_0^\pi \frac{\sin^2\theta\cos(2m\theta)d\theta}{(1-e^2\cos^2\theta)^{3/2}}, \tag{2.29b}$$

$$I_{2,m}(e) = \int_0^\pi \frac{\cos^2\theta\cos(2m\theta)d\theta}{(1-e^2\cos^2\theta)^{3/2}}, \tag{2.29c}$$

$$I_{3,m}(e) = \int_0^\pi \frac{\sin(2\theta)\sin(2m\theta)d\theta}{(1-e^2\cos^2\theta)^{3/2}}. \tag{2.29d}$$

Note that these integrals relate to the ones defined in (2.15) for isotropic bodies as $I_{0,0}(e) = I_0(e), I_{1,0}(e) = I_1(e)$, and $I_{2,0}(e) = I_2(e)$. Finally, identifying the surface displacement (2.28) with the boundary condition (2.4), one can solve for a_1, e, ϕ, and F. This requires an iterative approach which we describe hereafter.

Contact force solution

We now present a solution procedure that goes beyond the solutions detailed in [183] and [10], which are restricted to the specific case $M = N$. First, we express the initial gap function g_0 in the (x_1, x_2) coordinates; see Figure 4-1(c). This can be done by substituting the coordinate transformation relations

$$x = x_1\cos\phi - x_2\sin\phi, \tag{2.30a}$$

$$y = x_2\cos\phi + x_1\sin\phi, \tag{2.30b}$$

into (2.2), leading to

$$\begin{aligned} g_0 = {} & x_1^2(M\cos^2\phi + N\sin^2\phi) + x_2^2(M\sin^2\phi + N\cos^2\phi) \\ & + x_1x_2(N-M)\sin 2\phi. \end{aligned} \tag{2.31}$$

Equating the surface displacements (2.28) with the boundary condition (2.4) in the (x_1, x_2) coordinates, we obtain the relations

$$\frac{3F}{4a_1}\sum_{m=0}^{\infty} a_m(\phi)I_{0,m}(e) = \delta, \tag{2.32a}$$

$$\frac{3F}{4a_1^3}\sum_{m=0}^{\infty} a_m(\phi)I_{1,m}(e) = M\cos^2\phi + N\sin^2\phi, \tag{2.32b}$$

$$\frac{3F}{4a_1^3}\sum_{m=0}^{\infty} a_m(\phi)I_{2,m}(e) = M\sin^2\phi + N\cos^2\phi, \tag{2.32c}$$

$$\frac{3F}{4a_1^3}\sum_{m=1}^{\infty} b_m(\phi)I_{3,m}(e) = (M-N)\sin 2\phi. \tag{2.32d}$$

We recast (2.32b) to (2.32d) into two equations for ϕ and e:

$$(M-N)\sin 2\phi \sum_{m=0}^{\infty} a_m(\phi) I_{2,m}(e) - (M\sin^2\phi + N\cos^2\phi) \sum_{m=1}^{\infty} b_m(\phi) I_{3,m}(e) = 0, \tag{2.33a}$$

$$(M\cos^2\phi + N\sin^2\phi) \sum_{m=0}^{\infty} a_m(\phi) I_{2,m}(e) - (M\sin^2\phi + N\cos^2\phi) \sum_{m=0}^{\infty} a_m(\phi) I_{1,m}(e) = 0. \tag{2.33b}$$

Together, (2.33a) and (2.33b) form a nonlinear system of equations for e and ϕ that can be solved numerically according to the procedure described in Appendix A.3, after which the only remaining unknowns are a_1 and F. Combining (2.32a) and (2.32b), we find that F is given by

$$F = \frac{4}{3} \frac{[\sum_{m=0}^{\infty} a_m(\phi) I_{1,m}(e)]^{1/2}}{[\sum_{m=0}^{\infty} a_m(\phi) I_{0,m}(e)]^{3/2}} (M\cos^2\phi + N\sin^2\phi)^{-1/2} \delta^{3/2}. \tag{2.34}$$

In summary, the anisotropic contact force law requires the calculation of $h^B(\theta;0)$ from equation (2.25), after which the Fourier coefficients $a_m^B(\phi)$ and $b_m^B(\phi)$ can be found using (2.26) and (2.27). These can then be substituted into equations (2.33a) and (2.33b) to calculate e and and ϕ, before finally obtaining F through equation (2.34). Observe that the anisotropic solution retains the power 3/2 dependence of F on δ from the isotropic solution (2.18). Moreover, in the limiting case of isotropic materials, one obtains $a_0^B(\phi) = 1/\pi E_*^B$ and $a_m^B(\phi) = b_m^B(\phi) = 0$ for all $m > 0$, and this solution procedure appropriately reduces to the isotropic one given in Section (2.3.1).

Limitations for an implementation in DEM

We end this section with a discussion on issues of computational cost. While the solution procedure presented in this section is reasonably fast so long as one is merely interested in computing the force between two bodies under a few different situations, it is nevertheless too expensive for direct implementation into a DEM code. Indeed, the latter case requires a calculation of the force at every contact and at every time step, in which case the solution scheme quickly becomes prohibitively expensive. An alternative option is to precompute, for a given material, a look-up table of stored solution values for e and ϕ that would then be accessed during the course of the DEM simulation, with only the force F remaining to compute from (2.34). However, such a table would have to be four-dimensional – three parameters to describe the orientation of the (x, y, z) basis with respect to the (X_1^B, X_2^B, X_3^B) basis, and one for the ratio N/M – due to the coupling between equations (2.33) for e and ϕ, and the Fourier coefficients of the Green's function. In practice, this is not possible from a storage requirement standpoint, which essentially precludes the applicability of the exact contact force law (2.34) to the DEM. In order to circumvent this issue, we

discuss in the following section two possible simplification strategies, which both rely on shortening the form of the Green's function (2.25) appearing in the exact solution. We also propose an efficient implementation of the simplified solutions into DEM simulations.

2.4 Simplifications of the anisotropic contact force

2.4.1 Isotropic truncation of the Green's function

This approximation follows the exact anisotropic solution detailed in Section 2.3.2, with the crucial difference that the Fourier expansion (2.26) of the Green's function (2.25) is truncated after the constant term $a_0^B(\phi)$, so that $a_m^B(\phi) = b_m^B(\phi) = 0$ for all $m > 0$. This idea of truncating the Green's function was introduced by Vlassak *et al.* [183] in the context of a rigid indentor pressing against an anisotropic half space.

Setting $m = 0$ in (2.27) reveals that a_0^B is not a function of ϕ, as expected since the constant term is equal to the average of $h^B(\theta;\phi)$ over all θ. In contrast to the other Fourier coefficients, it follows that a_0^B no longer depends on the full orientation of the (x, y, z) basis with respect to the (X_1^B, X_2^B, X_3^B) basis attached to body B, but only on the orientation of the unit contact normal $\mathbf{n}$ with respect to (X_1^B, X_2^B, X_3^B). As shown in Figures 2-2(b) and 2-2(c), this relative orientation can be parameterized either by the components (n_1^B, n_2^B, n_3^B) of $\mathbf{n}$ or by the two Euler angles (α^B, β^B), both measured with respect to the local (X_1^B, X_2^B, X_3^B) basis. The two representations are related as

$$(n_1^B, n_2^B, n_3^B) = (\cos\alpha^B\sqrt{1-\cos^2\beta^B}, \sin\alpha^B\sqrt{1-\cos^2\beta^B}, \cos\beta^B), \tag{2.35a}$$

$$(\alpha^B, \beta^B) = (\operatorname{arctan2}(n_2^B, n_1^B), \arccos n_3^B), \tag{2.35b}$$

where $\operatorname{arctan2}(\cdot,\cdot)$ denotes the four-quadrant inverse tangent. From here on, we will indicate the contact normal direction with respect to body B in terms of the Euler angles (α^B, β^B), and the dependence of a_0^B on the latter will be denoted explicitly.

After truncation of the Fourier series, the Green's function (2.25) reduces to the same form as that for isotropic bodies,

$$\hat{w}^B(r) = \frac{a_0^B(\alpha^B, \beta^B)}{r} = \frac{1}{\pi\tilde{E}_*^B(\alpha^B, \beta^B)r}, \tag{2.36}$$

where $\tilde{E}_*^B(\alpha^B, \beta^B)$ is the plane strain modulus of the equivalent isotropic body, defined by Vlassak *et al.* [183] as

$$\tilde{E}_*^B(\alpha^B, \beta^B) = \frac{1}{\pi a_0^B(\alpha^B, \beta^B)}. \tag{2.37}$$

In (2.36) and (2.37), the superscript B attached to a_0^B and $\tilde{E}_*^B$ indicates a dependence of these quantities on the elasticity tensor $\mathbb{C}^B$ of body B, which may differ between bodies B_1 and B_2.

By virtue of the similarity between the truncated Green's function (2.36) and its isotropic counterpart (2.10), the rest of our solution proceeds in an analogous way to isotropic materials and is considerably simpler than the full anisotropic solution. Similar to the isotropic case detailed in Section 2.3.1, the phase angle $\phi = 0$ and the eccentricity e satisfies

$$\frac{I_2(e)}{I_1(e)} - \frac{N}{M} = 0. \tag{2.38}$$

Hence, the normal force F is expressed as

$$F = \frac{4\pi}{3}\tilde{E}^c_*(\alpha^{B_1}, \beta^{B_1}, \alpha^{B_2}, \beta^{B_2})\frac{[I_1(e)]^{1/2}}{[I_0(e)]^{3/2}}M^{-1/2}\delta^{3/2}, \tag{2.39}$$

where $\tilde{E}^c_*$, the composite plain strain modulus of the equivalent isotropic bodies, is given by

$$\tilde{E}^c_*(\alpha^{B_1}, \beta^{B_1}, \alpha^{B_2}, \beta^{B_2}) = \left(\frac{1}{\tilde{E}^{B_1}_*(\alpha^{B_1}, \beta^{B_1})} + \frac{1}{\tilde{E}^{B_2}_*(\alpha^{B_2}, \beta^{B_2})}\right)^{-1}. \tag{2.40}$$

Spherical case

Similarly to isotropic materials, the particular case of spherical bodies lends itself to further simplification. As discussed in Section 2.3.1, the gap function coefficients for two contacting spheres of radii R^{B_1} and R^{B_2} are given by $M = N = 1/2R$, with $1/R = 1/R^{B_1} + 1/R^{B_2}$. It then follows that the eccentricity $e = 0$, and the normal force F reduces to

$$F = \frac{4}{3}\tilde{E}^c_*(\alpha^{B_1}, \beta^{B_1}, \alpha^{B_2}, \beta^{B_2})R^{1/2}\delta^{3/2}. \tag{2.41}$$

Efficient implementation in DEM through a look-up table

The simplified solutions (2.39) and (2.41) obtained from the truncation of the Green's function assume the same form as the exact isotropic solutions (2.18) and (2.20), with the exception of $\tilde{E}^c_*$, the composite plain strain modulus (2.40). In the anisotropic solution, the latter depends on the relative orientation of the contact normal with respect to the two bodies through the equivalent plane strain modulus $\tilde{E}^B_*(\alpha^B, \beta^B)$ defined in (2.37). The computation of $\tilde{E}^B_*(\alpha^B, \beta^B)$ through the truncation of the Green's function (2.25) is rather demanding, which prevents its online integration into a DEM code. Nevertheless, we may leverage the fact that besides the angles α^B and β^B, the quantity $\tilde{E}^B_*$ solely depends on the elasticity tensor $\mathbb{C}^B$ of body B.

An effective remedy to the computational cost issue is thus to create, for every different material $\mathbb{C}^B$ present in the simulation, a table $[\tilde{E}_*](\cdot, \cdot\,; \mathbb{C}^B)$ of values of the equivalent plane strain modulus spanning all contact normal directions $\alpha^B \in [0, 2\pi]$ and $\beta^B \in [0, \pi]$. These two-dimensional look-up tables are to be precomputed offline and their values interpolated online according to the instantaneous values of α^B and β^B when (2.40) is called during the course of the DEM simulation. In this way, the simplified anisotropic contact laws (2.39) and (2.41) are equally fast to compute as

their isotropic counterparts, save for the interpolation of the look-up tables. Given a material, we describe in Appendix A.4 an algorithm for the calculation of such a look-up table — this table is then *shared among all bodies made of the same material*. The value of the composite plain strain modulus $\tilde{E}_*^c$ corresponding to two contacting bodies B_1 and B_2 can then be retrieved from two (or one, if $\mathbb{C}^{B_1} = \mathbb{C}^{B_2}$) precomputed tables $[\tilde{E}_*](\cdot,\cdot\,;\mathbb{C}^{B_1})$ and $[\tilde{E}_*](\cdot,\cdot\,;\mathbb{C}^{B_2})$ according to the algorithm presented in Appendix A.5.

Lastly, the solution to (2.38) for the eccentricity e of the contact area, which is required for non-spherical particles in both the isotropic and anisotropic contact laws, takes just a few Newton-Raphson iterations to converge[4] and can either be directly implemented into a DEM code, or stored in another one-dimensional look-up table as a function of the ratio M/N.

With $\tilde{E}_*^c$ and e in hand, the contact force can be readily calculated from (2.39). A Python implementation of the computational approach described in this section, including the computation of the look-up table, has been shared in an online repository at `https://github.com/smowlavi/AnisotropicGrains.git`.

2.4.2 Ad hoc computation of the plane strain modulus

As we have noted above, the main issue with the first simplification strategy lies in the need to compute the anisotropic Green's function (2.25) in order to obtain the equivalent plane strain modulus $\tilde{E}_*^B(\alpha^B,\beta^B)$ defined in (2.37). In this section, we present an alternative, ad hoc approach to obtain $\tilde{E}_*^B$ that is much faster to compute, yet retains directional information and makes full use of all elastic constants of the material. Recall that for isotropic materials, the plain strain modulus is given by

$$E_*^B = \frac{E^B}{1-(\nu^B)^2}, \tag{2.42}$$

where E^B and ν^B are respectively the Young's modulus and Poisson's ratio of body B. Returning to anisotropic materials, we may define an ad hoc equivalent plain strain modulus $\tilde{E}_*^B(\alpha^B,\beta^B)$ through a direct generalization of the above expression. We substitute E^B and ν^B with the *effective* Young's modulus $E_{\mathbf{n}}^B(\alpha^B,\beta^B)$ and *effective* Poisson's ratio $\nu_{\mathbf{n}}^B(\alpha^B,\beta^B)$ along the contact normal direction $\mathbf{n}$, giving

$$\tilde{E}_*^B(\alpha^B,\beta^B) = \frac{E_{\mathbf{n}}^B(\alpha^B,\beta^B)}{1-(\nu_{\mathbf{n}}^B(\alpha^B,\beta^B))^2}. \tag{2.43}$$

The effective material quantities $E_{\mathbf{n}}^B$ and $\nu_{\mathbf{n}}^B$ are defined the same way as for isotropic materials, with the exception that they now depend on the relative orientation (α^B,β^B) of the unit normal $\mathbf{n}$ with respect to the body. First, consider a state of uniform uni-

[4] To speed up convergence, one may start the iterations from $e = 2e_g/\sqrt{3}$ with $e_g = \sqrt{1-M/N}$, which provides an excellent approximation to the solution in the range $0 < e_g < 0.4$ and remains reasonably accurate up to $e_g \simeq 0.8$ (see [12], Section 3.3.1).

axial stress along $\mathbf{n}$,

$$\boldsymbol{\sigma} = \sigma \mathbf{n} \otimes \mathbf{n}, \tag{2.44}$$

which induces a strain $\boldsymbol{\epsilon}^B = \mathbb{S}^B \boldsymbol{\sigma}$, with $\mathbb{S}^B$ the compliance tensor of particle B. The resulting normal strain along the contact normal $\mathbf{n}$ is then given by

$$\epsilon^B_{\mathbf{n}} = \mathbf{n} \cdot \boldsymbol{\epsilon}^B \mathbf{n} = \mathbf{n} \cdot (\mathbb{S}^B \boldsymbol{\sigma}) \mathbf{n}, \tag{2.45}$$

and the resulting normal strain in the transverse direction is given by

$$\epsilon^B_{\mathbf{t}} = \frac{1}{2\pi} \int_0^{2\pi} (\mathbf{t} \cdot \boldsymbol{\epsilon}^B \mathbf{t}) d\gamma = \frac{1}{2\pi} \int_0^{2\pi} (\mathbf{t} \cdot (\mathbb{S}^B \boldsymbol{\sigma}) \mathbf{t}) d\gamma, \tag{2.46}$$

where $\mathbf{t}$ is a unit vector orthogonal to $\mathbf{n}$, and γ is the angle between $\mathbf{t}$ and an arbitrary fixed point in the plane perpendicular to $\mathbf{n}$. Denoting $\mathbf{u}$, $\mathbf{v}$ a fixed orthogonal basis within that plane, the substitution $\mathbf{t} = \cos\gamma\, \mathbf{u} + \sin\gamma\, \mathbf{v}$ enables the explicit calculation of the above integral, leading to

$$\epsilon^B_{\mathbf{t}} = \frac{1}{2} (\mathbf{u} \cdot (\mathbb{S}^B \boldsymbol{\sigma}) \mathbf{u}) + \frac{1}{2} (\mathbf{v} \cdot (\mathbb{S}^B \boldsymbol{\sigma}) \mathbf{v}). \tag{2.47}$$

The effective Young's modulus and Poisson's ratio along $\mathbf{n}$ are thus

$$E^B_{\mathbf{n}}(\alpha^B, \beta^B) = \frac{\sigma}{\epsilon^B_{\mathbf{n}}}, \qquad \nu^B_{\mathbf{n}}(\alpha^B, \beta^B) = -\frac{\epsilon^B_{\mathbf{t}}}{\epsilon^B_{\mathbf{n}}}. \tag{2.48}$$

Finally, we insert the above quantities back into the ad hoc definition (2.43) of the equivalent plane strain modulus, and we use (2.38) and (2.39) to find the resulting normal force. We note that while the computation of the plane strain modulus using the ad hoc approach described here is much faster than the Green's function approach described in Section 2.4.1, it is still more demanding than simply retrieving a precomputed value from a look-up table. Therefore, it is also advantageous to use the latter approach in this case, creating a table of values $[\tilde{E}_*](\cdot, \cdot\,; \mathbb{C}^B)$ of the ad hoc plain strain modulus as a function of α^B and β^B, for every material present in the simulation.

2.4.3 Summary of the exact and simplified laws

For the convenience of the reader, we provide in Table 2.1 a summary of the exact and simplified anisotropic contact force laws that we have presented in Sections 2.3.2, 2.4.1, and 2.4.2. We display separately the general case of an elliptic gap function (i.e. $M \neq N$) and the limiting case of a circular gap function (i.e. $M = N$), for which the simplified contact laws assume an even cleaner form[5]. Note that in Section 2.4.1,

[5]Willis [191] demonstrated that the contact area remains elliptic in the exact solution for a circular gap function and general anisotropic media. Thus, the exact contact law still requires the coupled solution of e and ϕ through (2.33), while its simplified counterparts simply return a circular contact area as described in Section 2.4.1.

Contact force law	Elliptic g_0 ($M \neq N$)				Circular g_0 ($M = N$)			
	F	$\tilde{E}_*^B$	e	ϕ	F	$\tilde{E}_*^B$	e	ϕ
Exact (Section 2.3.2)	(2.34)	–	(2.33)	(2.33)	(2.34)	–	(2.33)	(2.33)
Truncated (Section 2.4.1)	(2.39)	(2.37)	(2.38)	0	(2.41)	(2.37)	0	0
Ad hoc (Section 2.4.2)	(2.39)	(2.43)	(2.38)	0	(2.41)	(2.43)	0	0

Table 2.1: Summary of the exact and simplified anisotropic contact laws.

we have described the circular limit in the context of two spherical contacting bodies of radii R^{B_1} and R^{B_2}, in which case $M = N = 1/2R$ with $1/R = 1/R^{B_1} + 1/R^{B_2}$. From here on, we will refer to the simplified laws described in Sections 2.4.1 and 2.4.2 as the *truncated* and *ad hoc* contact laws, respectively.

2.5 Validation of the contact force laws

In this section, we first validate the accuracy of the exact contact force law against finite-element method (FEM) simulations, before comparing the accuracy of the two simplified contact force laws with respect to their exact counterpart. To this effect, we perform numerical calculations of the force experienced by a flat rigid plate (body B_2) pressing against a smooth elastic body made of a given material (body B_1), as pictured in Figure 2-3(a). Following our previous convention, we parameterize the direction of the unit normal $\mathbf{n}$ to the contact plane with respect to the local coordinate basis $(X_1^{B_1}, X_2^{B_1}, X_3^{B_1})$ of body B_1 by the two Euler angles α^{B_1} and β^{B_1} depicted in Figure 2-3(a).

We consider a number of possible scenarios by changing (*i*) the material, (*ii*) the orientation of B_1, represented by α^{B_1} and β^{B_1}, as well as (*iii*) the geometry of the smooth surface, defined by the gap function (2.2). In terms of materials, we selected three different crystals spanning a wide range of degrees of symmetry. First is iron (Fe), which has a cubic crystalline structure described by three independent elastic constants $C_{11} = 231$, $C_{44} = 116$, and $C_{12} = 135$ GPa, as determined in [158]. Second is quartz (SrO_2), which has a trigonal crystalline structure described by 6 independent elastic constants, measured by [84] as $C_{11} = 87.2$, $C_{33} = 106$, $C_{44} = 57.2$, $C_{12} = 6.57$, $C_{13} = 12.0$, and $C_{14} = -17.2$ GPa. Finally, third is zirconia (ZrO_2), which has a monoclinic crystalline structure described by 13 independent elastic constants, which were characterized by [30] as $C_{11} = 361$, $C_{22} = 408$, $C_{33} = 258$, $C_{44} = 99.9$, $C_{55} = 81.2$, $C_{66} = 126$, $C_{12} = 142$, $C_{13} = 55.0$, $C_{15} = -21.3$, $C_{23} = 196$, $C_{25} = 31.2$, $C_{35} = -18.2$, and $C_{46} = -22.7$ GPa.

2.5.1 Validation of the exact contact force law

We begin by comparing predictions of the exact contact force law with results from FEM simulations. The 3D setup presented in Figure 2-3(a) is implemented in ABAQUS

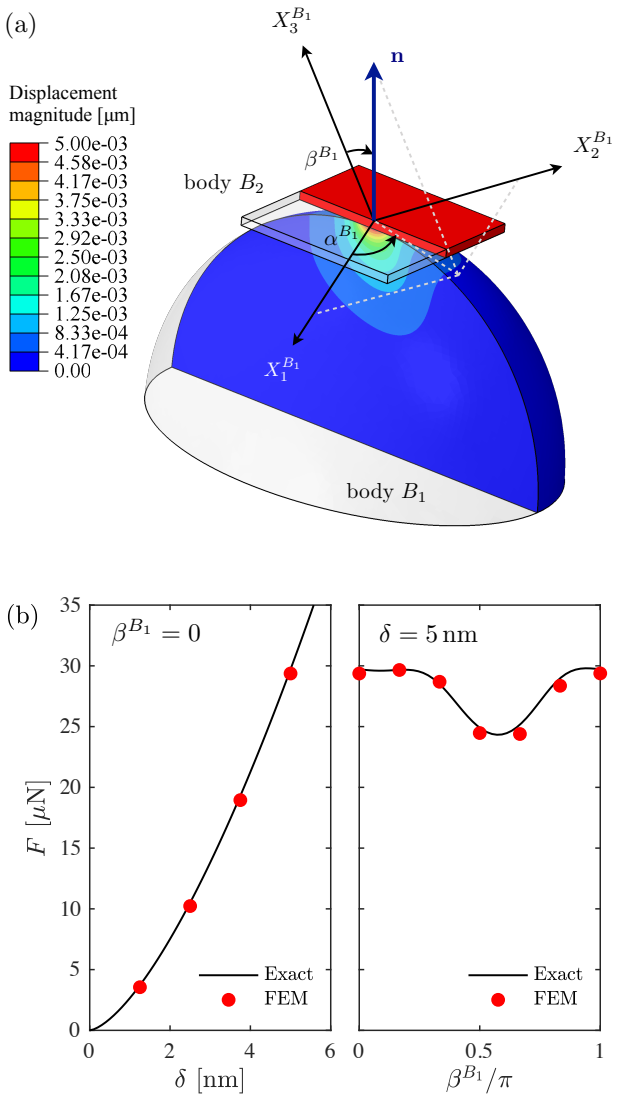

Figure 2-3: (a) A flat rigid plate B_2 is assigned a vertical displacement δ into a smooth elastic body B_1 with possibly unequal principal radii of curvature. The orientation of the local coordinate basis of B_1 with respect to the contact normal direction $\mathbf{n}$ is parameterized by the Euler angles α^{B_1} and β^{B_1}. The contour levels on the vertical cut display the magnitude of the elastic displacement generated by an indentation depth $\delta = 5\,\mathrm{nm}$, as computed in FEM. (b) Contact force predicted by the exact contact law (solid line) and the FEM simulations (red dots) for $\alpha^{B_1} = \pi/2$ and $\beta^{B_1} = 0$ as a function of δ (left pane), and for $\alpha^{B_1} = \pi/2$ and $\delta = 5\,\mathrm{nm}$ as a function of β^{B_1} (right pane).

(2017). The elastic body B_1 is designed with principal radii of curvature at the contact point of $R_1 = 0.5\,\mu$m and $R_2 = 0.25\,\mu$m, corresponding to gap function coefficients $M = 1/2R_1 = 1\mu\text{m}^{-1}$ and $N = 1/2R_2 = 2\mu\text{m}^{-1}$ [12]. The contact interaction between B_1 and B_2 is modeled following the 'surface-to-surface' formulation, with hard contact in the normal direction and no friction in the tangential direction. We assign the elastic properties of quartz (SrO_2) to B_1 by specifying its full elasticity tensor and rotating the corresponding material directions $(X_1^{B_1}, X_2^{B_1}, X_3^{B_1})$ according to the desired values of α^{B_1} and β^{B_1}. Body B_1 is discretized using 112752 quadratic tetrahedral elements (C3D10M), and its base is pinned in the vertical direction. An incremental vertical displacement directed into B_1 is prescribed to the flat rigid plate B_2, which is defined as an 'analytical rigid surface'. The analysis is carried out using the explicit solver by moving the rigid plate at a rate slow enough to ensure that the deformation proceeds in a quasi-static manner, as verified by the fact that (*i*) the total kinetic energy never exceeds 0.03% of the total internal energy, and (*ii*) the contact force measured at the plate is within 1% equal to the sum of the vertical reaction forces at the basal nodes of B_1.

The contour levels on the vertical cut of body B_1 in Figure 2-3(a) display the magnitude of the elastic displacements induced by an indentation depth $\delta = 5\,$nm, as computed in FEM, for material orientation $\alpha^{B_1} = \beta^{B_1} = \pi/2$. Interestingly, the anisotropy of the constitutive relation is reflected in the absence of axisymmetry (with respect to the contact normal direction) of the elastic displacement field. For a more quantitative analysis, the red dots in the left pane of Figure 2-3(b) depict the FEM contact force for material orientation $\alpha^{B_1} = \pi/2$, $\beta^{B_1} = 0$ and four different values of the vertical displacement δ of B_2 into B_1. These FEM results are in excellent agreement with the corresponding predictions from the exact contact force law shown by the solid line; the power 3/2-dependence of F on δ is also clearly visible. Conversely, the red dots in the right pane of Figure 2-3(b) display the FEM contact force for $\delta = 5\,$nm and different material orientations defined by $\alpha^{B_1} = \pi/2$ and varying values of β^{B_1}. Once again, the FEM results agree well with the exact contact force law shown by the solid line, with the difference between the two not exceeding 3.4%.

The slight discrepancy between FEM and theoretical results observed in Figure 2-3(b) can be attributed to various reasons. On the one hand, the FEM solution is dependent on the resolution of the mesh in the vicinity of the contact region, and further refinement of the mesh would reduce errors arising from the numerical discretization. On the other hand, the exact contact force law relies on the assumptions that the size of the contact area is small with respect to the dimensions of B_1 as well as its radii of curvature at the contact point. Such assumptions are never satisfied exactly, thus invariably lead to small errors when the contact law is applied to a real-case scenario. Notwithstanding, the comparisons displayed in Figure 2-3(b) exhibit a sufficient level of agreement to validate both the accuracy and the implementation of the exact contact force law.

2.5.2 Accuracy of the simplified contact force laws

Polar visualizations

We now proceed with the comparison of the two simplified contact force laws with respect to their exact counterpart. As in the previous section, we calculate the contact force experienced in the setup pictured in Figure 2-3(a), this time using a wider range of materials, orientations and surface geometries. We first show polar visualizations of the directional dependence of the force predicted by the exact solution and its two simplifications for an indentation depth (overlap) $\delta = 100\,\mathrm{nm}$. Due to the symmetry exhibited by the Green's function (2.22) with respect to the sign of the unit normal $\mathbf{n}$, the behavior of the force is completely specified for all materials by the hemisphere $\alpha^{B_1} \in [0, 2\pi]$, $\beta^{B_1} \in [0, \pi/2]$. Thus, we visualize the directional dependence of the contact force by projecting each direction point on the hemisphere to a plane through stereographic projection, in such a way that the data pertaining to the orientation $(\alpha^{B_1}, \beta^{B_1})$ will be displayed at the location $(\tan(\beta^{B_1}/2)\cos\alpha^{B_1}, \tan(\beta^{B_1}/2)\sin\alpha^{B_1})$ in a disk of unit radius.

Figures 2-4 and 2-5 display such polar visualizations of the exact contact force law and its two simplifications for a circular gap function ($M = N = 1\,\mu\mathrm{m}^{-1}$) in Figure 2-4 and an elliptic gap function ($M = 1\,\mu\mathrm{m}^{-1}$, $N = 2\,\mu\mathrm{m}^{-1}$) in Figure 2-5. In the elliptic case, we have chosen to orient the principal axes (x, y) of the gap function along the polar (α, β) directions. For both figures, (a,b,c) correspond to iron, (d,e,f) to quartz, and (g,h,i) to zirconia. Surprisingly, we notice that the truncated force law is remarkably close to the exact solution for all materials, contact directions and shapes of the gap function. This result is extremely promising for DEM applications since the truncated law can return a near-exact contact force at a very reasonable cost (presuming that one uses a look-up table approach as described in Section 2.4.1 and Appendix A.5). The ad hoc approximation, on the other hand, deviates further away from the exact solution. The accuracy with which it predicts the shape of the contour levels of the force depends on the degree of symmetry of the material – it performs very well in this regard for iron, reasonably well for quartz, and more poorly for zirconia. More importantly, it fails to correctly predict the extrema of the force and displays a much stronger dependence of the latter on the contact normal direction, as compared with the other solutions.

The relative dependence of the force on the contact normal direction is not strongly influenced by the geometry of the gap function, as one observes by comparing Figures 2-4 and 2-5. The most apparent difference between the circular and elliptic gap functions under a given overlap distance is that the force is lower in the elliptic case for all contact normal orientations, which is expected since the case $M = 1\,\mu\mathrm{m}^{-1}$, $N = 2\,\mu\mathrm{m}^{-1}$ has a higher mean curvature at the contact point than the case $M = N = 1\,\mu\mathrm{m}^{-1}$. Curiously, the other geometrical features of the exact solution – namely the eccentricity e and orientation ϕ of the contact area – are more sensitive to the geometry of the gap function, as shown in Appendix A.6.

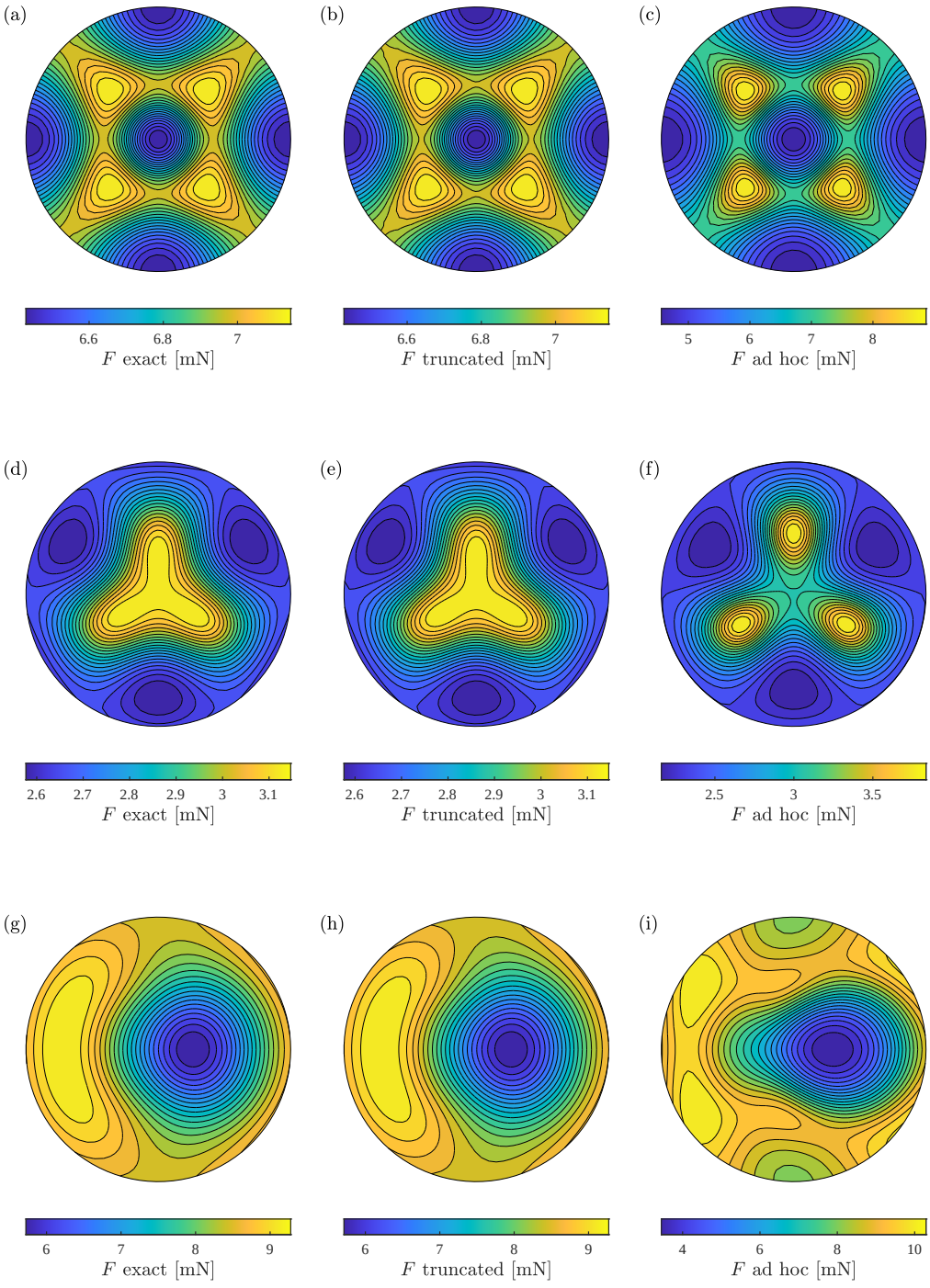

Figure 2-4: Polar visualizations of the normal force F predicted by the exact contact force law and its two simplifications for iron (a,b,c), quartz (d,e,f), and zirconia (g,h,i), under indentation depth $\delta = 100\,\text{nm}$ and gap function coefficients $M = N = 1\,\mu\text{m}^{-1}$.

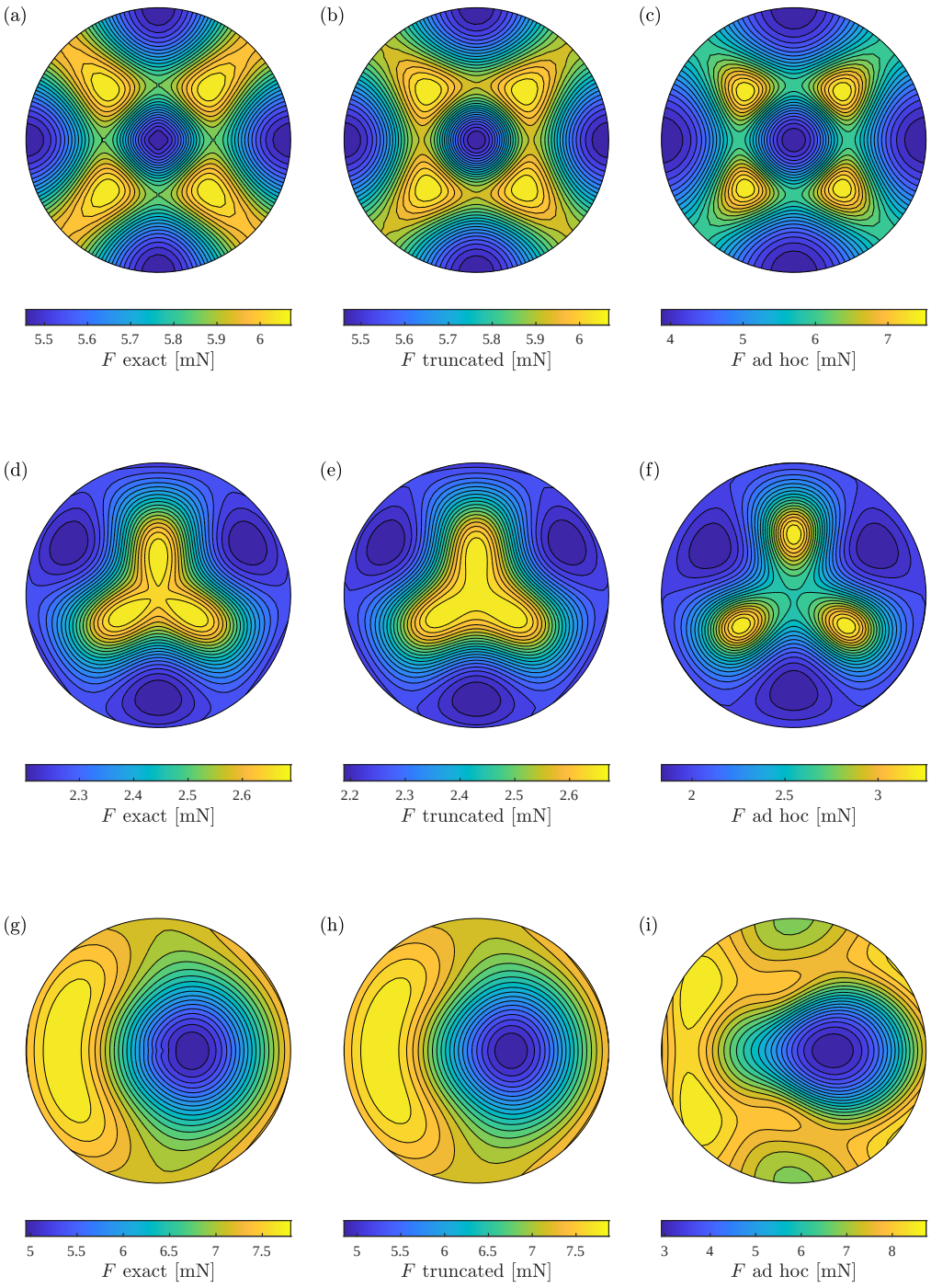

Figure 2-5: Polar visualizations of the normal force F predicted by the exact contact force law and its two simplifications for iron (a,b,c), quartz (d,e,f), and zirconia (g,h,i), under indentation depth $\delta = 100\,\text{nm}$ and gap function coefficients $M = 1\,\mu\text{m}^{-1}$ and $N = 2\,\mu\text{m}^{-1}$.

Error analysis

Next, we perform a quantitative analysis of the accuracy of the two simplified contact force laws with respect to their exact counterpart. Let us first define e_g, the eccentricity of the gap function, as

$$e_g = \sqrt{1 - \frac{M}{N}}. \tag{2.49}$$

The quantity e_g measures the eccentricity of the contour levels of the gap function $g_0(x, y)$, in the same way that e quantifies the eccentricity of the boundary of the contact area. A circular gap function corresponds to $e_g = 0$. We first quantify the sensitivity of the various contact laws with respect to the contact normal direction for different values of e_g, using the same overlap $\delta = 100\,\text{nm}$ as prescribed before. To this effect, Figure 2-6 shows the mean and extrema values, over all contact normal directions, of the normal force predicted by the exact and simplified contact laws versus e_g, for iron (a), quartz (b) and zirconia (c). We again observe that the truncated law is very accurate, while the ad hoc law exaggerates the dependence of the force on the orientation. For a more quantitative comparison, we define, for a given value of e_g and a given material, the relative error

$$\mathcal{E} = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi/2} \frac{|F_\text{s}(\alpha, \beta) - F_\text{e}(\alpha, \beta)|}{F_\text{e}(\alpha, \beta)} \sin\beta d\alpha d\beta, \tag{2.50}$$

where F_e and F_s refer, respectively, to the exact and simplified solutions. Thus, (2.50) returns the mean relative error over all contact normal orientations. Figure 2-7 shows the error as a function of the eccentricity of the gap function e_g for the three materials considered previously. The accuracy of the truncated law is remarkable for small values of the gap function eccentricity e_g, and remains very good as e_g increases, with the relative error $\mathcal{E}$ remaining near or under 1%. The ad hoc law, on the other hand, behaves more poorly with the error being on the order of 10% for the three materials.

We now study the behavior of the exact and simplified contact laws as the constitutive relation approaches the isotropic limit. For this purpose, we construct an arbitrary cubic material of varying anisotropy ratio AR defined by [196] as

$$\text{AR} = \frac{2C_{44}}{C_{11} - C_{12}}, \tag{2.51}$$

with the particular case AR = 1 corresponding to an isotropic material, and we pick the same values for C_{11} and C_{12} as for iron. (For reference, iron then corresponds to the case AR = 2.41.) Figure 2-8 shows the mean and extrema values, over all contact normal directions, of the normal force predicted by the exact and simplified contact laws for this arbitrary material and a circular gap function, that is, $e_g = 0$. As expected, the two simplified contact laws degenerate to the exact solution in the limiting case AR = 1 of an isotropic material. This remains true for a finite value

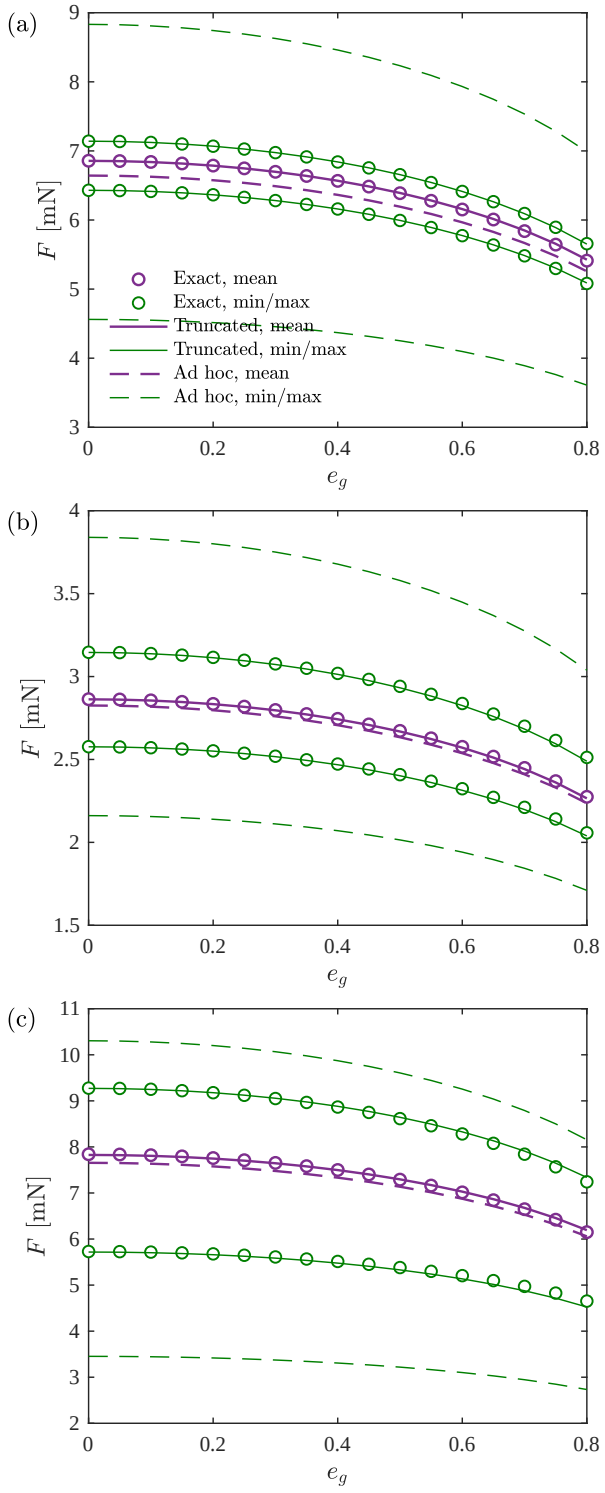

Figure 2-6: Mean, maximum and minimum values of the normal force F, over all contact normal directions, predicted by the exact and simplified contact laws as a function of the eccentricity e_g of the gap function, for (a) iron, (b) quartz, and (c) zirconia.

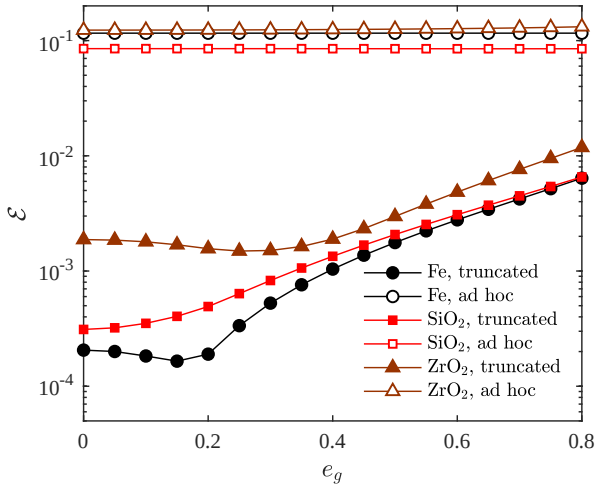

Figure 2-7: Mean relative error $\mathcal{E}$ over all contact normal directions of the force F predicted by the two simplified contact laws as a function of the eccentricity e_g of the gap function, for iron, quartz, and zirconia.

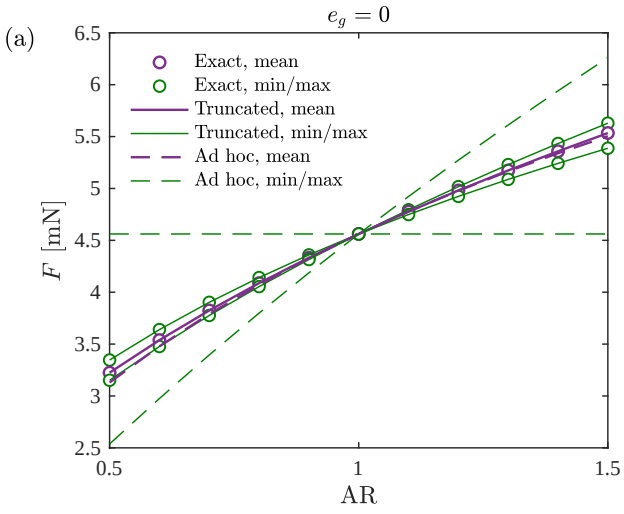

Figure 2-8: (a) Mean, maximum and minimum values of the normal force F, over all contact normal directions, predicted by the exact and simplified contact laws for an arbitrary cubic material with varying anisotropy ratio AR and $e_g = 0$.

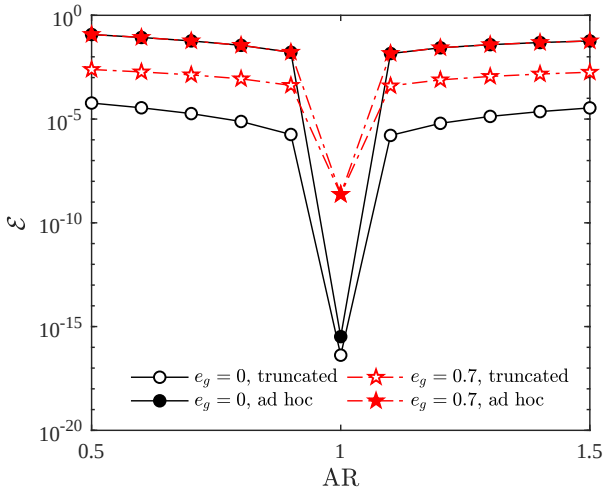

Figure 2-9: Mean relative error $\mathcal{E}$ over all contact normal directions of the force F predicted by the two simplified contact laws as a function of the anisotropy ratio AR of an arbitrary cubic material.

of e_g, as displayed in Figure 2-9 for $e_g = 0.7$ in terms of the mean relative error ϵ defined in (2.50). Here again, the truncated contact law is remarkably close to the exact solution for all values of anisotropy ratio AR and gap function eccentricity e_g, with the relative error $\mathcal{E}$ remaining under 1%, while the second approximation returns larger relative errors on the order of 10%.

Going forward, we select the truncated simplification as the contact force law of choice for implementation into a DEM code. Implemented with a look-up table approach, the computation of this force law is equally fast as the ad hoc simplification, yet returns results that are accurate to within 1% for ellipsoids and within 0.1% for spherical particles.

2.6 Applications

We now implement the truncated anisotropic contact law described in Section 2.4.1 into a custom DEM code, which enables the simulation of granular materials composed of elastically anisotropic particles. Even though the contact laws that we have derived are applicable to arbitrarily-shaped bodies as long as their surfaces are smooth and convex, we will here restrict ourselves to spherical particles. We will show two examples, one static and one dynamic, where the anisotropy of the constitutive relation induces changes in the macroscopic properties of the system.

2.6.1 Equations of motion

Consider a system of N spherical particles $i = 1, \dots, N$ with (possibly distinct) elasticity tensors $\mathbb{C}_i$. In this section, we adopt a slight change of notation and indicate

quantities pertaining to body i with a subscript i, in line with conventions from the DEM literature. The positions and orientations of the particles are described by a set of generalized coordinates $\{\mathbf{q}_i\} = (\{\mathbf{r}_i\}, \{\boldsymbol{\epsilon}_i\})$, where $\{\mathbf{r}_i\} \in \mathbb{R}^3$ denotes the position of the center of mass of body i, and $\{\boldsymbol{\epsilon}_i\} \in \mathbb{R}^4$ is a set of Euler parameters (unit quaternions) that characterizes the orientation of body i, both in the global reference frame $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$ shown in Figure 2-2(a). The linear and angular velocities of the particles are described by generalized velocities $\{\mathbf{v}_i\} = (\{\dot{\mathbf{r}}_i\}, \{\boldsymbol{\omega}_i\})$, where $\{\boldsymbol{\omega}_i\} \in \mathbb{R}^3$ is the angular velocity of body i in the global frame and relates to the time derivative of the Euler parameters $\{\boldsymbol{\epsilon}_i\}$ as (see [61, 58] or equation (9.3.37) in [78])

$$\{\dot{\boldsymbol{\epsilon}}_i\} = \frac{1}{2}[\mathbf{A}(\boldsymbol{\epsilon}_i)]\{\boldsymbol{\omega}_i\}, \tag{2.52}$$

with the matrix $[\mathbf{A}(\boldsymbol{\epsilon}_i)] \in \mathbb{R}^{4\times 3}$ defined as

$$[\mathbf{A}(\boldsymbol{\epsilon}_i)] = \begin{bmatrix} -\epsilon_{i,1} & -\epsilon_{i,2} & -\epsilon_{i,3} \\ \epsilon_{i,0} & \epsilon_{i,3} & -\epsilon_{i,2} \\ -\epsilon_{i,3} & \epsilon_{i,0} & \epsilon_{i,1} \\ \epsilon_{i,2} & -\epsilon_{i,1} & \epsilon_{i,0} \end{bmatrix}. \tag{2.53}$$

Two bodies i and j, with diameters d_i and d_j, interact when their signed overlap function,

$$\delta_{ij} = \frac{d_i + d_j}{2} - |\mathbf{r}_i - \mathbf{r}_j|, \tag{2.54}$$

is positive. Denoting by $c_i = \{j : \delta_{ij} \geq 0\}$ the set of particles that are in contact with body i, the generalized velocities can be integrated in time using Newton's equations of motion,

$$m_i\{\ddot{\mathbf{r}}_i\} = \sum_{j\in c_i}\{\mathbf{F}_{ij}\} + m_i\{\mathbf{g}\}, \tag{2.55a}$$

$$I_i\{\dot{\boldsymbol{\omega}}_i\} = \sum_{j\in c_i}(a_{ij}\{\mathbf{n}_{ij}\} \times \{\mathbf{F}_{ij}\}), \tag{2.55b}$$

where m_i and I_i denote respectively the mass and moment of inertia of particle i. At each contact, $a_{ij} = (d_i - \delta_{ij})/2$ denotes the distance from the center of mass of particle i to its contact point with particle j, the unit normal vector $\mathbf{n}_{ij} = (\mathbf{r}_j - \mathbf{r}_i)/|\mathbf{r}_j - \mathbf{r}_i|$ is directed from i to j, and the force $\{\mathbf{F}_{ij}\}$ consists of normal and tangential components,

$$\{\mathbf{F}_{ij}\} = F_{ij}^n\{\mathbf{n}_{ij}\} + F_{ij}^t\{\mathbf{t}_{ij}\}, \tag{2.56}$$

where the tangent unit vector $\mathbf{t}_{ij}$ belongs to the contact plane and depends on the history of relative tangential velocities of i and j at the contact point. In this chapter, we consider frictionless[6] bodies so that $F_{ij}^t = 0$. The normal force F_{ij}^n comprises an

[6] A direct consequence of this assumption is that in the absence of external torques, spherical particles will keep their initial orientation throughout the simulation. Nevertheless, our exposition accounts for the possible presence of angular velocities in an effort to be as general as possible.

elastic and a dissipative part,

$$F_{ij}^n = -\max(F_{ij}^e + F_{ij}^d, 0), \tag{2.57}$$

where the $\max(\cdot)$ function forbids the existence of a cohesion force, and the orientation-dependent elastic component F_{ij}^e is given by the normal contact force law derived in Section 2.4.1,

$$F_{ij}^e = \frac{4}{3}\tilde{E}_*^c(\{\boldsymbol{\epsilon}_i\}, \{\boldsymbol{\epsilon}_j\}, \{\mathbf{n}_{ij}\})R_{ij}^{1/2}\delta_{ij}^{3/2}, \tag{2.58}$$

where $1/R_{ij} = (2/d_i + 2/d_j)$. Likewise, one expects the dissipative component F_{ij}^d to inherit an orientation dependence from the anisotropy of the material structure. However, deriving such a relation falls outside the scope of this chapter, and we restrict ourselves to the standard isotropic expression

$$F_{ij}^d = \gamma_n \dot{\delta}_{ij}, \tag{2.59}$$

where γ_n is a constant damping coefficient. This simplification is reasonable for flowing granular materials, where damping is known to play a negligible role within a particular range of strain rates [43]. For quasi-static problems, the form of the dissipation is inconsequential so long as one is interested in static quantities after particles have come to a rest, which is the case of our only upcoming example using a nonzero damping coefficient γ_n.

In equation (2.58), the dependence of $\tilde{E}_*^c$ on the relative orientations of bodies i and j with respect to the contact normal direction has been indicated through the Euler parameters $\{\boldsymbol{\epsilon}_i\}$, $\{\boldsymbol{\epsilon}_j\}$ and the components $\{\mathbf{n}_{ij}\}$ of the contact normal, which are readily available in the simulation. Given these inputs, we present in Appendix A.5 an algorithm to retrieve the value of $\tilde{E}_*^c$ from two (or one, if $\mathbb{C}_i = \mathbb{C}_j$) precomputed tables of values of the plane strain modulus, $[\tilde{E}_*](\cdot,\cdot;\mathbb{C}_i)$ and $[\tilde{E}_*](\cdot,\cdot;\mathbb{C}_j)$, the computation of which is described in Appendix A.4. These algorithms for the calculation of the look-up tables and contact force have been provided as a Python code in an online repository at `https://github.com/smowlavi/AnisotropicGrains.git`. Further details regarding the numerical implementation of the DEM code and parameter values are listed in Appendix A.8.

2.6.2 Static force distribution in a pyramid

As a first example, we consider a static square-based pyramid of close-packed single-crystal zirconia spheres, with ten particles along each side of the base. This system is statically indeterminate due to each interior particle possessing twelve neighbors [140]. Therefore, the equilibrium contact forces will depend on the contact stiffnesses (that is, on the composite plain strain moduli $\tilde{E}_*^c$), which in the case of anisotropic particles are a function of the contact directions and particle orientations. As we will

Torques may arise in other works as a result of the geometry or surface roughness of the particles, and it is critical to treat their orientations correctly since the anisotropic contact law is orientation-dependent.

show next, our anisotropic DEM framework enables us to investigate the relationship between the floor pressure at the base of the pyramid and the orientation of the particles.

Notice from Figure 2-4 that a sphere made of single-crystal zirconia may be thought of as having a band of high contact stiffness along its equator. We will consider four separate arrangements in which every particle is either oriented such that the strong band is roughly horizontal (orientation 1), resulting in all contacts witnessing approximately the same stiffness from that particle; or the strong band is roughly aligned with a vertical plane parallel to the $y = x$ diagonal of the square base (orientation 2), causing stiffer contacts oriented in those directions compared to those oriented in the other direction.

Figure 2-10(a) shows a three-dimensional visualization of the pyramid, which is initialized by placing the particles in a position where they barely touch their neighbors. The pyramid is then allowed to settle under the acceleration of gravity, with the contact forces oscillating during a transient phase before reaching their equilibrium values. For the case of all particles following orientation 1, Figure 2-10(b) shows this phenomenon through the time evolution of $\sum_{j \in c_i} |\mathbf{F}_{ij}|/mg$, the normalized sum of the force magnitudes that each particle experiences at all its contact points. The same quantity is shown at final time in Figure 2-10(a) as the semi-transparent color applied to each grain.

In Figure 2-11, we display the distribution of normalized reaction forces F_z/mg on the base of the pyramid, once equilibrium is reached. We consider four separate arrangements of particle orientation, which we visualize by displaying particles in orientation 2 with a diagonal line aligned along their strong band. Particles in upper layers are oriented identically to the base layer particles belonging to the same $y = x + c$ vertical plane. As expected, the reaction forces are symmetrical in the case shown in Figure 2-11(a) where all particles have orientation 1, since all the contacts see approximately the same stiffness. That symmetry is broken and a clear effect of anisotropy emerges in Figure 2-11(b), where all particles have orientation 2. Due to the stronger contacts along directions parallel to the $y = x$ plane, the two corner particles aligned along the 'strong' $y = x$ diagonal inherit a larger reaction force than the other two corner particles. The picture gets even more interesting in Figures 2-11(c) and (d), which demonstrate that it is possible to tune the reaction force beneath the pyramid by mere rotation of the constituent particles. To conclude, this simple example highlights the importance of accounting for anisotropic effects in the discrete element modeling of elastically anisotropic particles, even in situations that involve no dynamics at all.

2.6.3 Sound transmission in a granular chain

As a second example, we investigate the transmission of sound in a compressed chain of adjacent spherical particles between two fixed walls. A large body of work has researched the behavior and frequency response of such 'granular crystals' to small-amplitude dynamic displacements of the particles, where small is in comparison with the static overlap imposed between adjacent particles by the compression force. In

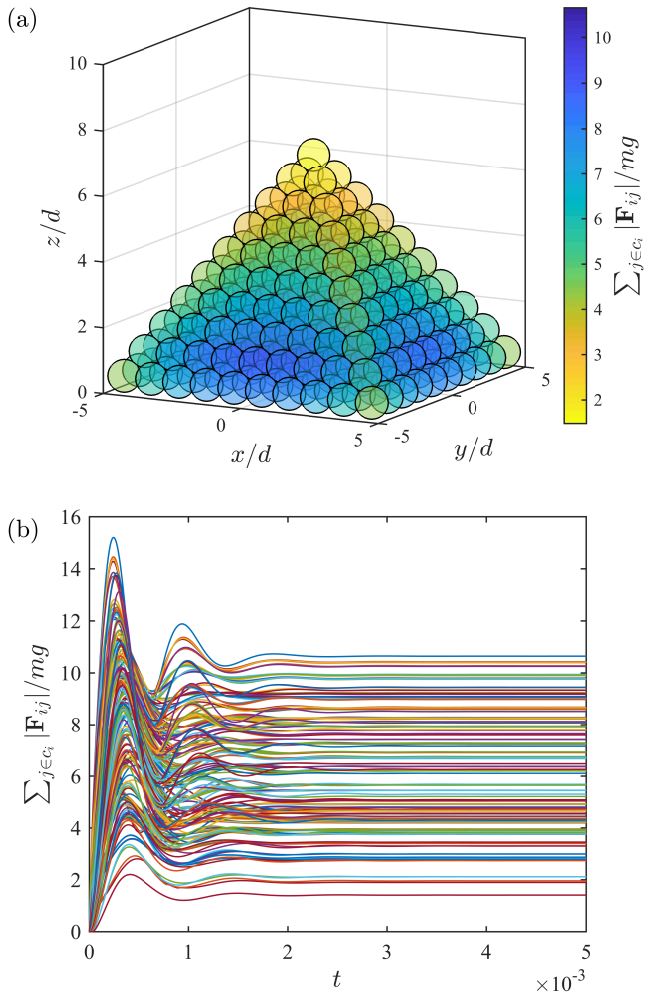

Figure 2-10: (a) Geometry of the static square-based pyramid. Each particle i is colored according to $\sum_{j\in c_i} |\mathbf{F}_{ij}|/mg$, the normalized sum of the force magnitudes that it withstands at all its contact points. (b) The same quantity is plotted over time during the settling of the pyramid.

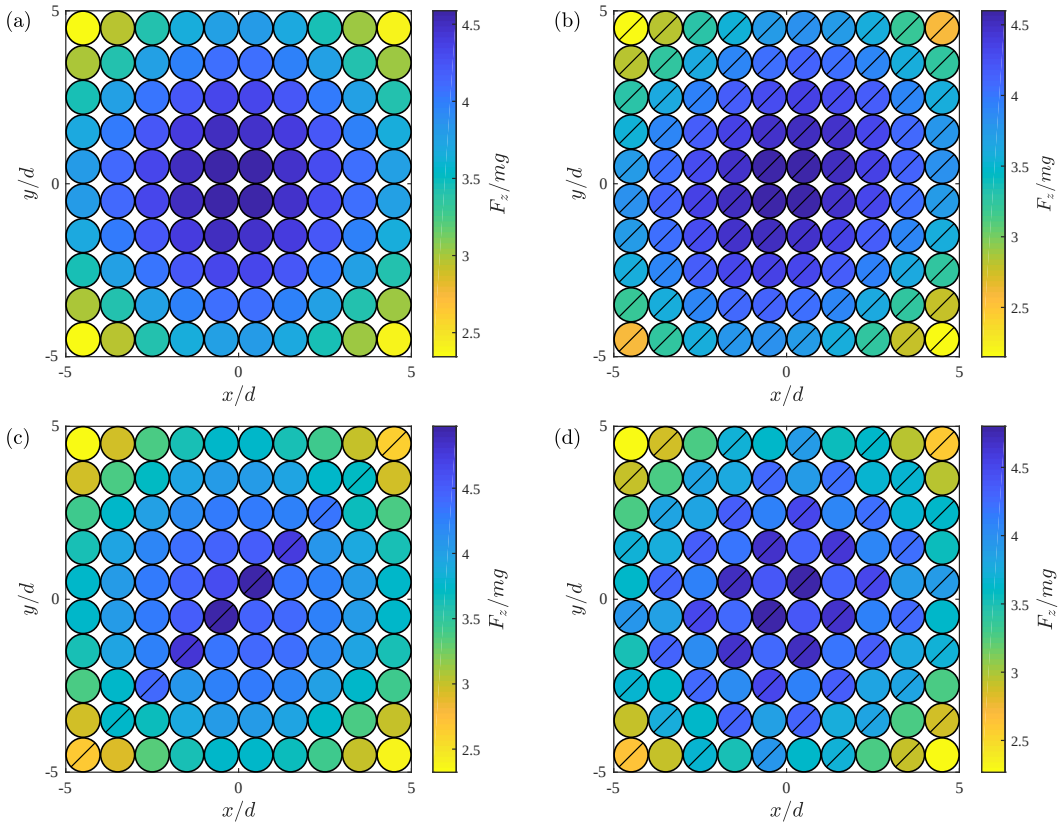

Figure 2-11: Distribution of normalized reaction forces F_z/mg on the base of the pyramid for particles oriented such that either all contacts see uniform stiffness from the particle, or contacts along directions parallel to the $y = x$ plane are stronger. Particles belonging to the second group are displayed with a diagonal line, and particles in upper layers are oriented identically to the base layer particles belonging to the same $y = x + c$ vertical plane.

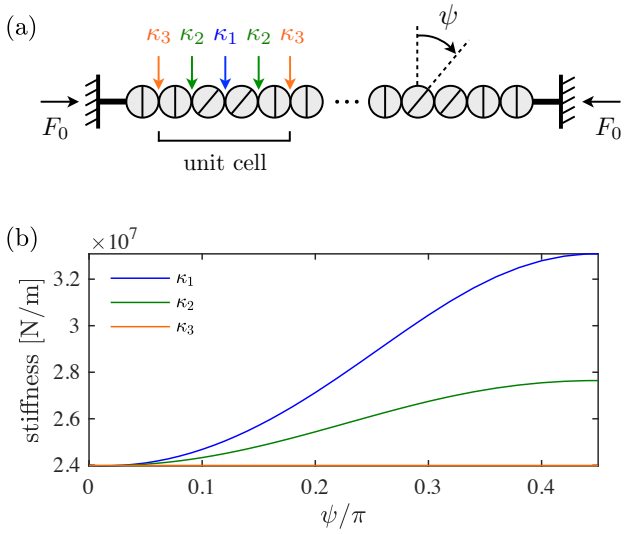

Figure 2-12: (a) Schematic of the granular chain, composed of periodically repeated four-particle unit cells and compressed between two fixed walls. The orientation ψ of the middle two particles in each unit cell is varied systematically, while that of the two edge particles is fixed. (b) Dependence of the linearized contact stiffnesses κ_1, κ_2, κ_3 on the orientation ψ of the middle particles.

particular, different authors have shown that by combining particles with different geometrical or material properties, it is possible to obtain a frequency response characterized by acoustic band gaps inside of which no frequencies are allowed [85, 82, 21], thus filtering out input frequencies. Such filters are desirable for a range of purposes ranging from acoustic filters to vibrational isolation, and the tunability of these band gaps is key to delivering optimal performance.

The existence of acoustic band gaps requires the contact stiffnesses between the grains to be non-uniform [89]. This is most simply achieved in diatomic chains consisting of particles with alternating properties, for which a single band gap appears [26, 103, 85]. A second band gap was shown in [21] to emerge in diatomic chains composed of three-particle unit cells. The tunability of these band gaps requires a change in the properties of the particles, which is typically done by altering their size, geometry or constituent material. Clearly, this is not feasible in practice when one desires to control the band gap frequencies in real-time.

As we have seen throughout this chapter, elastically anisotropic bodies exhibit an orientation-dependent contact stiffness. Here, we utilize this property to construct a monoatomic granular crystal that possesses band gaps that may be tuned by mere rotation of its constituent particles. Specifically, consider the chain of anisotropic zirconia particles pictured in Figure 2-12(a) and compressed between two fixed walls. The chain consists of periodically repeated four-particle unit cells in which the orientation ψ of the middle two particles is varied systematically while that of the two edge particles is kept fixed. The angle ψ is defined as the orientation of the strong band of the zirconia spheres (schematized in Figure 2-12(a) by the straight line within each

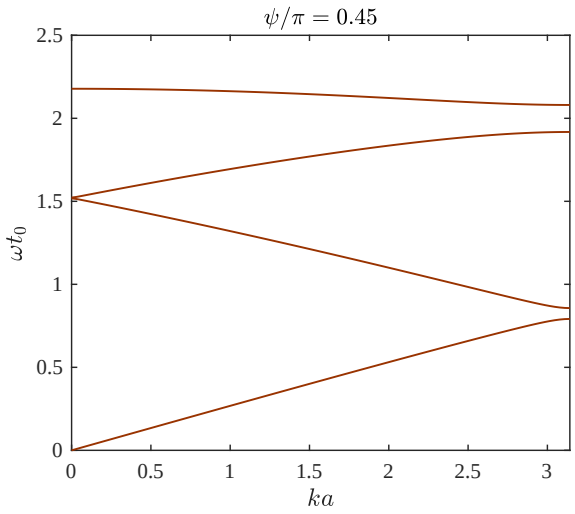

Figure 2-13: Analytical dispersion relation of the compressed granular chain for an infinite number of particles, in the case $\psi/\pi = 0.45$.

sphere) with respect to the plane orthogonal to the chain axis. The edge particles within each unit cell are oriented such that the strong band is orthogonal to the chain axis.

In the linear regime that we investigate, the relative displacement between any two adjacent particles is small with respect to their static overlap δ_{ij}^0 caused by the compression force F_0. Thus, the overlap term $\delta_{ij}^{3/2}$ occurring in the contact force law (2.58) can be linearized about δ_{ij}^0, producing a force-displacement relation that is linear with a proportionality constant termed the linearized contact stiffness. The latter is clearly a function of the composite plain strain modulus $\tilde{E}_*^c$ and therefore depends on the orientation of the particles. (For more details, the reader is invited to refer to Appendix A.9.) As pictured in Figure 2-12(a), the structure of the unit cell in our granular chain gives rise to three different linearized contact stiffnesses κ_1, κ_2, and κ_3, which depend on the orientation ψ of the middle particles according to Figure 2-12(b). Note that κ_3 is constant since it measures the stiffness between the edge particles of two adjacent unit cells, the orientations of which are fixed. Finally, we neglect dissipation effects, which in practice result in a small uniform shift of the band-gap frequencies but do not change their overall topological features [89, 20].

In order to obtain analytical insight into the frequency response of our granular crystal with four-particle unit cells, we derive in Appendix A.9 the dispersion relation of the system for an infinite number of particles, which relates the wavenumber k of propagating sound waves to their frequency ω. The dispersion relation is displayed in nondimensional form for the case $\psi/\pi = 0.45$ in Figure 2-13, where k is normalized by the equilibrium length a of each unit cell, and the corresponding ω is normalized by the ψ-independent time scale $t_0 = \sqrt{m/\kappa_3}$, with m the mass of each sphere. Compared with the three-particle unit cell studied in [21], we report the emergence of an additional fourth band of propagating frequencies above the usual acoustic

and optical bands. As a consequence, our chain of four-particle unit cells inherits three bands of forbidden frequencies, or band gaps, in which sound waves decay exponentially and cannot propagate along the chain. The second band-gap, however, has negligible width for the material properties that we consider here.

We now demonstrate the tunability of these vibrational band gaps by rotation of the middle particles in each unit cell. First, we verify the agreement between the analytical band frequencies and the behavior of a finite-length chain composed of 102 particles, which we simulate in our anisotropic DEM framework. A small initial velocity is assigned to the first sphere in the chain, reproducing the effect of an impact excitation, and the force felt by the last sphere is measured as a function of time. Figure 2-14(a) shows the resulting power spectral density for the case $\psi/\pi = 0.45$, with the shaded regions corresponding to the four bands of propagating frequencies predicted by the dispersion relation pictured in Figure 2-13. We observe excellent agreement between the theoretical predictions and numerical results, with vanishing energy of the force spectrum in the band gap regions. Next, we investigate the tunability of these band gaps by repeating the same numerical experiment for a range of orientations ψ of the middle particles in each unit cell. The resulting spectrum is displayed in Figure 2-14(b) as a filled contour plot where each column corresponds to a particular value of ψ, and demonstrates the adjustability of the band gaps by simple rotation of some of the particles. The white lines are the cut-off frequencies predicted by the dispersion relation and agree very well with the numerics. We note that the band gaps disappear as ψ goes to zero, corresponding to the limiting case of a uniform chain. Finally, it is worth keeping in mind that the dimensional band gap frequencies are a function of the precompression force through the power $-1/6$ dependence of the time scale t_0 on F_0, inherited from the dependence of κ_3 on F_0 (see equation (A.8) in Appendix A.9). As a consequence, the precompression force provides an additional control parameter to tune the band gaps, besides the angle ψ.

2.7 Conclusions and perspectives

In this chapter, we have introduced a method to resolve the normal force arising between two elastically anisotropic contacting bodies of arbitrary geometry with smooth and frictionless surfaces, with the aim of obtaining a contact law that can be easily implemented into a DEM code. We first presented a numerical procedure for the exact solution of the full linear elasticity equations, resulting in an exact anisotropic contact force law. The computational cost of this exact contact law precluded its direct implementation into a DEM code, and its dependence on four parameters at a time prevented the use of a look-up table of precomputed values.

By shortening the form of the full Green's function used in the exact solution, we then derived two simplifications to the exact contact law. *Both simplifications take the same form as the Hertzian contact law for isotropic bodies*, save for the dependence of the contact (or plane strain) modulus associated with each body on the relative orientation of the contact normal direction and on the full set of elastic constants of the body. The precise form of the contact modulus differs between the two simplifications.

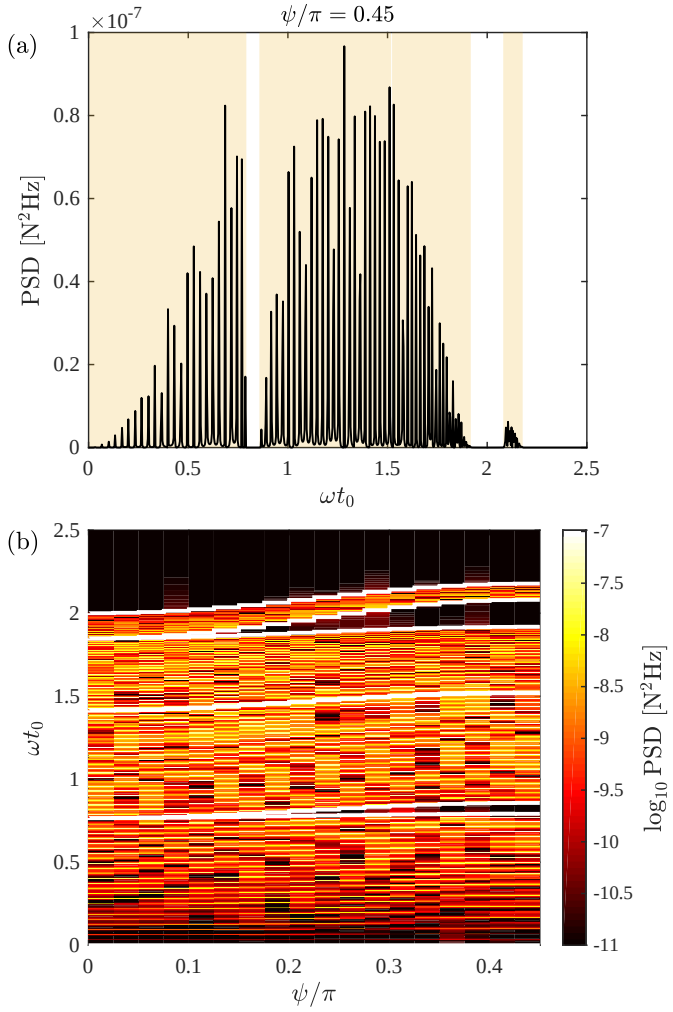

Figure 2-14: Frequency response of the compressed granular chain. (a) Power spectral density of the force observed at the last grain for $\psi/\pi = 0.45$. The shaded regions correspond to the four bands of propagating frequencies predicted by the dispersion relation pictured in Figure 2-13. (b) Power spectral density of the force observed at the last grain, plotted in log scale for a range of values of ψ/π. The white lines are the cut-off frequencies predicted by the analytical dispersion relation for an infinitely long chain.

In both cases, the parameter dependence of the computationally expensive part was reduced from four in the exact contact law down to two, a significant reduction that enabled the implementation of these simplified laws into a DEM code through the use of two-dimensional look-up tables of precomputed values of the contact modulus over all possible contact directions. Remarkably, the first of the two simplifications, which we called the truncated contact law, exhibited excellent accuracy compared to its exact counterpart, with the relative error on the predicted force remaining near or below 1% for a wide range of materials and surface geometries.

Next, we presented the implementation of the truncated contact law into a DEM code, which we leveraged to showcase two application examples in which elastic anisotropy of the particles induced changes in the macroscopic behavior of the system. The first example we considered was that of a static square-based pyramid of contacting single-crystal zirconia spheres. By changing the orientation of the particles, we demonstrated that the pressure at the base of the pyramid is affected by the anisotropy of the contact forces. We then studied the transmission of sound waves in a compressed chain of adjacent single-crystal zirconia spheres, known as a 'monoatomic granular crystal'. We leveraged the orientation-dependence of the contact stiffnesses between adjacent spheres to achieve frequency filtering characteristics that normally belong to the realm of diatomic granular crystals (assembled from two different constituent particles). More precisely, we revealed through theory and numerical computations the emergence of band gaps in which sound frequencies are unable to propagate down the chain. These band gaps are tunable by mere rotation of the particles, which offers an attractive prospect for adoption of such anisotropic granular crystals in scenarios that demand real-time control.

The present work opens the door to two distinct avenues of research. The first concerns the extension of our anisotropic contact law to frictional bodies, which can support tangential surface tractions in contrast to the frictionless bodies that we have treated. In general, the tangential force F_{ij}^t is related to the normal force F_{ij}^n through Coulomb's law, $F_{ij}^t \leq \mu F_{ij}^n$, with μ a friction coefficient [123]. In order to determine the magnitude of F_{ij}^t in the static friction case $F_{ij}^t < \mu F_{ij}^n$ as well as the onset of the dynamic friction case $F_{ij}^t = \mu F_{ij}^n$, the tangential contact law is typically regularized through a virtual tangential spring in a fashion that was pioneered by [41]. While such an approach can be readily combined with our anisotropic contact law, we mention that several authors [182, 187, 50] have developed more rigorous extensions of the tangential contact law for isotropic bodies, based on the early work of [135]. In these studies, the tangential spring becomes nonlinear and its stiffness is related to the elastic constants of the material, much in the same way that Hertzian contact theory provides a normal contact force law that is connected to the material parameters. Unlike the Hertzian normal force law, these tangential force relations depend on normal force history of the contact, which may have complex extensions in the anisotropic case. Although highly non-trivial, a generalization of the aforementioned studies to the elastically anisotropic case would be very valuable.

The second avenue of research enabled by the anisotropic contact law concerns the effect of elastic anisotropy on the behavior of granular systems, both at the microscopic and macroscopic levels. Granular materials sustain external loads through

force chains, which are, in turn, responsible for the mechanical response of the sample [198]. Considerable efforts have therefore been devoted to their characterization from both experimental [125] and theoretical [170] perspectives. Recently, Hurley *et al.* [87] measured the distribution of contact forces in an assembly of elastically anisotropic quartz grains undergoing a compression cycle, and discovered a surprising inverse relationship between macroscopic load and heterogeneity of the contact forces, despite the clear formation of force chains. Reproducing their experiment in a DEM simulation using our anisotropic contact law could possibly shed light on the potential role of anisotropy in explaining their observation. Another potential area of application outside the realm of granular materials is the mechanical behavior of rock, which can be modeled in the DEM by a heterogeneous material comprised of cemented grains whose contact force law includes both grain-based and cement-based contributions [147, 91, 33]. Although the elastic component of the grain-based portion of the normal contact law is usually considered isotropic, crystalline rocks such as granite possess a microstructure consisting of individual crystals, and would therefore benefit from the incorporation of our elastically anisotropic normal force law.

Chapter 3

Interplay between hysteresis and nonlocality during onset and arrest of flow in granular materials

3.1 Introduction

Granular materials are well-known for displaying both solid-like and fluid-like behavior depending on their internal stress state [64, 3, 175]. Flow can be induced or arrested through external loading variations, which has direct implications for a wide range of catastrophic geophysical phenomena such as landslides, avalanches and earthquakes [45, 122, 162]. The transition between solid-like and liquid-like behavior in frictional granular media is characterized by several unique macroscopic features, which have been uncovered through simple experiments in model systems [133]. Figure 3-1 showcases typical results from such experiments, where flow is triggered then arrested by ramping up and down the applied stress in (a) an annular shear cell [42], (b) a layer of grains on an inclined plane [149], and (c) a partially-filled rotating drum [38]. The features revealed in these experiments are universal to most geometries and can be outlined as follows:

(F1) the level of stress required to trigger flow is larger than that below which flow stops, leading to a hysteresis of the flow velocity as the applied stress is ramped up and down;

(F2) the onset of flow is accompanied by a finite jump in the velocity of the system;

(F3) the critical stresses for onset and arrest of flow depend on the size[1] of the system, with smaller system sizes displaying increased strengthening.

Each feature is directly relevant to geophysical events such as landslides and avalanches, since (F1) controls the mobilized mass that flows down, (F2) explains why they are so spontaneous and catastrophic, and (F3) determines the circumstances under which they might occur. The objective of the present work is to formulate a continuum

[1] By system size, we refer to the relevant length scale controlling the width of shear regions in the flow field. Depending on the geometry, this length scale can either be geometric or stress-induced.

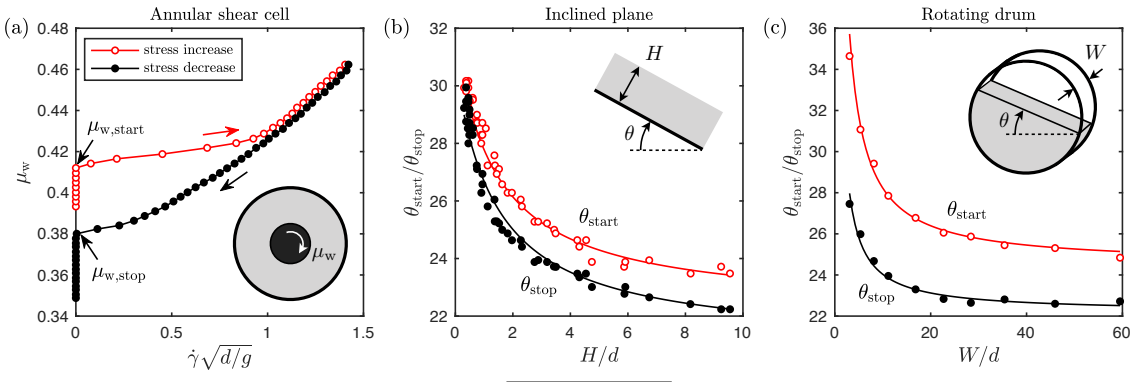

Figure 3-1: Previous experimental investigations of the flow threshold in various geometries. (a) Annular shear cell [42]: ratio of shear stress to pressure at the inner wall, μ_w, versus dimensionless mean strain rate, $\dot{\gamma}\sqrt{d/g}$, for increasing and decreasing torque applied to the inner cylinder. (b) Inclined plane [149]: angles of inclination at flow onset and arrest, θ_{start} and θ_{stop}, versus dimensionless layer thickness, H/d. (c) Rotating drum [38]: angles of inclination at flow onset and arrest, θ_{start} and θ_{stop}, versus dimensionless drum width, W/d.

model that is able to describe quantitatively the onset and arrest of flow in frictional granular materials in various two-dimensional geometries, and analyze how its constituent ingredients play a role in reproducing each of these three features, with particular focus on geometries displaying inhomogeneous flow fields.

It is now well accepted that dense and homogeneous flows of grains follow the $\mu(I)$ constitutive relationship, which states that the stress ratio μ and the inertial number I are related through a one-to-one function $\mu = \mu_{loc}(I)$ [133, 44]. In two dimensions, $\mu = \tau/P$ is the ratio of shear stress τ to pressure P, and $I = \dot{\gamma}\sqrt{m/P}$ is the strain rate $\dot{\gamma}$ nondimensionalized with a particle-wise rearrangement time scale formed by the mean grain mass m and confining pressure P. While $\mu_{loc}(I)$ has long been believed to be a monotonic function of I, several recent experiments [51, 108, 144, 160] and simulations [194, 47] have revealed the existence at very low I of a strain-rate weakening regime, wherein the stress ratio μ decreases with increasing I. One possible non-monotonic functional form is

$$\mu_{loc}(I) = \mu_s + \frac{\mu_2 - \mu_s}{(\mu_2 - \mu_s)/(bI + \chi(I;\kappa)) + 1}, \tag{3.1}$$

where μ_s, μ_2, and b are dimensionless rheological constants, $\kappa = k_n/P$ a dimensionless stiffness with k_n the grain stiffness, and χ is a decreasing function of I that accounts for the strain-rate weakening regime. The microscopic origin of the strain-rate weakening regime has recently come under debate, with some studies arguing that it is caused by inertia of the grains [151, 39, 47], while others observing velocity-weakening behavior in over-damped, inertia-less particulate media [144]. The present work is not concerned with this particular issue, and we simply leave open the possibility

for the amount of strain-rate weakening to depend on the grain stiffness through the dimensionless parameter κ entering χ. In any case, the non-monotonicity of (3.1) necessarily implies that features (F1) and (F2) above are realized in homogeneous flows: the level of stress required to trigger flow is higher than that at which flow stops, and flow onset is characterized by a velocity jump[88, 134].

However, the $\mu(I)$ constitutive relationship breaks down in inhomogeneous flows, in the sense that μ is no longer a one-to-one function of I [106, 179]. Due to the finite size of the grains, velocity fluctuations generated at an arbitrary location will spread over some grain-size-dependent correlation length and change the rheology of the neighboring material [132, 69, 155, 67], resulting in wider shear regions than are predicted by the $\mu(I)$ rheology, especially in the quasi-static limit [133]. Such spatial cooperativity at the scale of individual grains also explains why thinner layers on an inclined plane start flowing at higher inclination angles than thicker layers, despite the stress ratio μ being independent of the layer height [133]. Nonlocal rheological models, which incorporate an intrinsic length scale, have been shown to capture several of these phenomena [96]. Here, we focus on the nonlocal granular fluidity (NGF) model [98, 79], which relates the stress ratio and strain rate through a granular fluidity field $g = \dot{\gamma}/\mu$ that is governed by a reaction-diffusion partial differential equation (PDE). Our choice of the NGF model stems from its ability to reproduce the system-size dependence of the flow threshold in various geometries [97, 118], which partially explains feature (F3) above. But the current formulation of the NGF model reduces to the monotonic form of the $\mu_{\text{loc}}(I)$ relationship in homogeneous flow conditions, meaning that the model does not have a built-in mechanism to account for the remaining features (F1) and (F2).

In this chapter, we modify the NGF model so that it instead reduces to the non-monotonic form of the $\mu_{\text{loc}}(I)$ relationship, equation (3.1), in homogeneous flows. By computing time-dependent model predictions in a stress-driven planar shear configuration under gravity, we evaluate the specific ways in which nonlocality and non-monotonicity contribute to each of the three features (F1–F3) of the flow-arrest transition in inhomogeneous flows. We show that inclusion of, and interplay between both ingredients is necessary to reproduce all three features, in ways sometimes surprising: the planar shear with gravity configuration displays a finite velocity jump during onset of flow only when both non-monotonicity and nonlocality are present. In a second part, we assess the capability of the modified NGF model to predict quantitatively the behavior of dense granular materials both around the flow-arrest transition as well as in the flowing regime. To this effect, we calibrate the model using discrete element method (DEM) simulations in the simple shear geometry shown in Figure 3-2(a), and we compare predictions of the calibrated model against stress-driven DEM simulations in the other geometries displayed in Figures 3-2(b) and 3-2(c), namely plane shear under gravity and inclined plane. These two configurations are both subject to nonlocal effects as a result of the spatial inhomogeneity of their flow fields, but they critically differ in an important way – in plane shear with gravity, flow inhomogeneity is mostly a consequence of the spatial dependence of the stress ratio μ, while it is mainly caused by the rough base in inclined plane flow [118]. We observe that the accuracy of the NGF model depends on which of these two mechanisms is at play,

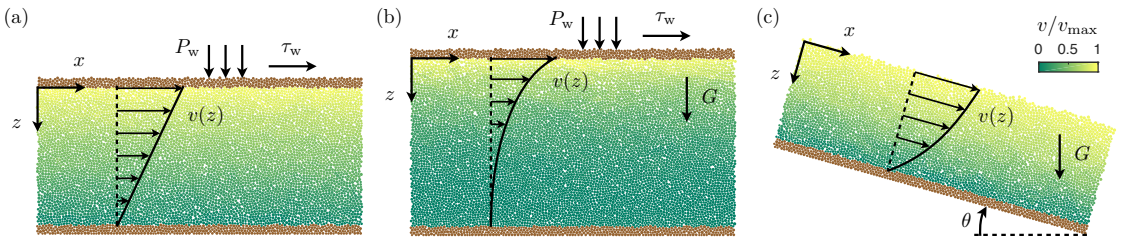

Figure 3-2: Geometries considered in this study: (a) plane simple shear, (b) plane shear under gravity, and (c) inclined plane flows. Particles that are free to flow are colored according to their relative velocity magnitude, and fixed wall particles are shown in brown.

with predictions being accurate in the case of plane shear with gravity but less so for inclined plane, which we ultimately attribute to the role played by the boundary conditions.

The remainder of this chapter is organized as follows. In Section 3.2, we present a modified NGF model that incorporates a non-monotonic local rheology, and we evaluate the combined effects of nonlocality and non-monotonicity on the features of the flow-arrest transition. We then compare in Section 3 predictions from the NGF model with DEM simulation results in stress-driven planar shear with gravity and inclined plane configurations. We close the chapter with concluding remarks in Section 4.

3.2 Nonlocal granular rheology

In this section, we discuss our nonlocal continuum modelling approach based on the nonlocal granular fluidity (NGF) model. We begin by presenting the NGF model in its current form, which does not capture the hysteresis of the flow-arrest transition. We then describe the incorporation of bistable behavior into the NGF model. Finally, we evaluate the combined effects of bistability and nonlocality on the qualitative behavior of the flow-arrest transition.

3.2.1 Nonlocal model without hysteresis

Extending earlier fluidity-based nonlocal models for concentrated emulsions [145, 69, 19] to granular materials, the NGF model introduces a positive granular fluidity field g that relates the strain rate $\dot{\gamma}$ with the stress ratio μ through the following two constitutive equations:

$$\dot{\gamma} = g\mu, \tag{3.2a}$$

$$t_0\dot{g} = A^2 d^2 \nabla^2 g - \frac{(\mu_2 - \mu_s)(\mu_s - \mu)}{\mu_2 - \mu} g - b\sqrt{\frac{m}{P}}\mu g^2, \tag{3.2b}$$

where t_0 is a constant timescale associated with the dynamics of g, and the nonlocal amplitude $A > 0$ is a dimensionless scalar parameter quantifying the strength of spatial cooperativity in the flow. As a side note, we mention that recent studies [198, 17, 101] have endowed the granular fluidity field with a clear physical meaning – g is a purely kinematic quantity related to the velocity fluctuations δv, grain size d and solid fraction ϕ through $g = (\delta v/d)F(\phi)$, where F depends solely on ϕ.

The flow rule (3.2a) states that the strain rate $\dot{\gamma}$ is directly proportional to the fluidity g. Therefore, there can only be flow provided g is nonzero, and the nonlocal granular rheology is driven by the dynamics of equation (3.2b) for the fluidity. The latter takes the form of a reaction-diffusion equation, and its behavior can be understood as follows.

In the absence of boundary effects or nonuniformities in the stress ratio μ, the g field becomes spatially uniform and (3.2b) reduces to the simple dynamical system

$$t_0\dot{g} = -\frac{(\mu_2 - \mu_s)(\mu_s - \mu)}{\mu_2 - \mu} g - b\sqrt{\frac{m}{P}}\mu g^2 \equiv F(g;\mu,P)g, \tag{3.3}$$

where $F(g;\mu,P)$ is a simple linear function of g. The steady behavior of the system is then governed by the steady-state solutions g_{loc} of (3.3), which we illustrate by the thick lines in Figure 3-3 using arbitrary values[2] for $\mu_s = 0.25$, μ_2, b, and P. When $\mu \leq \mu_s$, the function $F(g;\mu,P) < 0$ for all $g \geq 0$ as shown in Figure 3-3(a), hence (3.3) only admits the stable, arrested steady state $g_{\text{loc}} = 0$, represented by the zero green branch in Figure 3-3(b). When $\mu > \mu_s$, the linear function $F(g;\mu,P)$ acquires one positive root and $F(g = 0;\mu,P) > 0$, which has two consequences. First, the arrested steady state becomes unstable, shown by the red branch in Figure 3-3(b). Second, a stable, flowing steady-state solution $g_{\text{loc}}(\mu) > 0$ emerges, represented by the positive green branch in Figure 3-3(b). Using the flow rule (3.2a) together with the definition of the inertial number I, the flowing and arrested stable solutions can be inverted and expressed as

$$\mu_{\text{loc}}(I) = \mu_s + \frac{\mu_2 - \mu_s}{(\mu_2 - \mu_s)/bI + 1}, \tag{3.4}$$

for $I > 0$, and $\mu_{\text{loc}} \leq \mu_s$ otherwise. This relationship is pictured in green in Figure 3-3(c), together with the unstable arrested solution above μ_s in red. Therefore, the NGF model reduces to the local $\mu(I)$ rheology in steady and homogeneous flows such as plane shear without gravity.

In the presence of boundary effects or nonuniformities in the stress ratio μ, however, the diffusion term in (3.2b) spreads the granular fluidity over a cooperativity length scale proportional to the grain size d, resulting in a nonlocal flow rule (3.2a). Regions where $\mu > \mu_s$ act as stress-driven sources of granular fluidity, which is then diffused towards lower-stress regions or boundaries. Such nonlocal, cooperative effects

[2]We note that the qualitative behavior of the model is independent of its specific parameter values. The latter are therefore chosen to be reasonably close to the calibrated values obtained later in Section 3.3.1, while displaying the hysteretic behavior of the model with enough clarity in Figures 3-3 and 3-4.

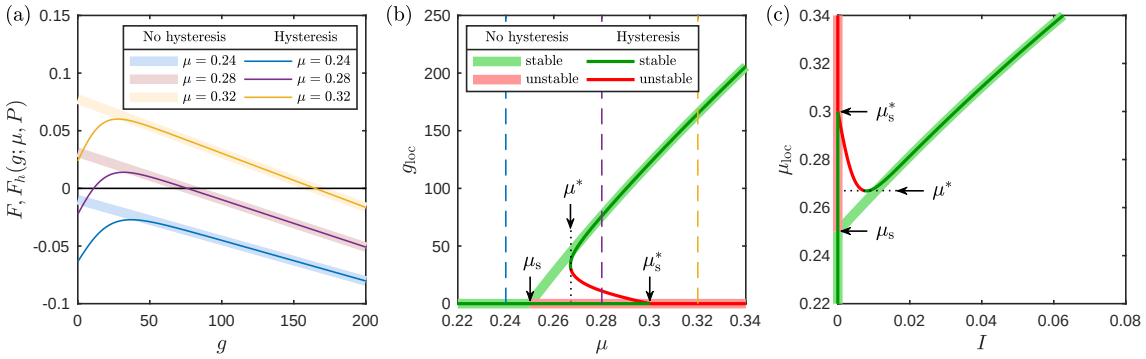

Figure 3-3: Steady-state solutions of the local limit of the NGF model for homogeneous flows, without hysteresis (thick lines) and with hysteresis (thin lines). (a) Behavior of $F(g;\mu,P)$ and $F_h(g;\mu,P)$ in (3.3) and (3.7), showing the existence of zero, one or two roots for different stress ratios μ. (b) Unstable (red) and stable (green) steady-state local solutions g_{loc} as a function of μ. (c) Resultant local rheology $\mu_{\mathrm{loc}}(I)$, with both the unstable (red) and stable (green) branches shown. The model with hysteresis displays a strain-rate weakening regime absent in the model without hysteresis.

have manifold consequences, and the NGF model explains many phenomena evading local rheological models. For instance, the model recovers the decaying motion of grains in regions where $\mu < \mu_{\mathrm{s}}$ [79, 117, 179] as well as the so-called secondary rheology, wherein flow anywhere in a granular media removes the yield stress elsewhere [80, 115]. Conversely, the model is able to explain the strengthening of the flow threshold with decreasing system size [97, 118], which is caused by boundaries or low-stress regions preventing flow in other regions where $\mu > \mu_{\mathrm{s}}$ unless μ is large enough. This last property relates to feature (F3) mentioned in the introduction. Yet, the current form of the NGF model is unable to reproduce features (F1) and (F2) due to the monotonicity of its limiting local rheology (3.4).

3.2.2 Nonlocal model with hysteresis

We now discuss the inclusion of non-monotonicity of the local rheological response into the NGF model. Taking inspiration from previous hysteretic nonlocal models [5, 114], we add a new term to the right-hand side of the fluidity equation (3.2b). The constitutive equations (3.2) become

$$\dot{\gamma} = g\mu, \tag{3.5a}$$

$$t_0\dot{g} = A^2 d^2 \nabla^2 g - \frac{(\mu_2 - \mu_{\mathrm{s}})(\mu_{\mathrm{s}} - \mu)}{\mu_2 - \mu} g - b\sqrt{\frac{m}{P}}\mu g^2 - \chi(g;\mu,P)g, \tag{3.5b}$$

where the new term $\chi(g;\mu,P)$ takes the form

$$\chi(g;\mu,P) = a\left[1-\tanh\left(c\sqrt{\frac{m}{P}}\mu g\kappa^n\right)\right], \tag{3.6}$$

with a, c, n constant scalar parameters, and $\kappa = k_n/P$ the nondimensional particle stiffness. Here, we choose to express the new term χ with a tanh function so that it vanishes for large g, which restricts its contribution to the behavior of the system near the jamming transition.

As before, we begin by evaluating the dynamics of the g field for the case of homogeneous flows, in which g is spatially uniform and (3.5b) reduces to the simple dynamical system

$$\begin{aligned} t_0\dot{g} &= -\frac{(\mu_2-\mu_s)(\mu_s-\mu)}{\mu_2-\mu}g - b\sqrt{\frac{m}{P}}\mu g^2 - \chi(g;\mu,P)g \\ &\equiv F_h(g;\mu,P)g, \end{aligned} \tag{3.7}$$

where $F_h(g;\mu,P)$ is a function of g. Contrary to the previous case without hysteresis, the presence of the χ term induces a decrease in $F_h(g;\mu,P)$ when g approaches zero, as illustrated by the thin lines in Figure 3-3(a) using the same parameter values as before and new arbitrary values for the parameters of χ. As a result, there exists a range of stress ratios $\mu^* < \mu < \mu_s^*$ in which the function $F_h(g;\mu,P)$ inherits a second positive root, so that (3.7) admits two flowing steady-state solutions $g_{\text{loc}}(\mu) > 0$, one stable and one unstable, shown by the thin lines in Figure 3-3(b). These two flowing solution branches merge at $\mu = \mu^*$. The stable branch reverts for $\mu > \mu_s^*$ to the same flowing solution as the NGF model without hysteresis, while the unstable branch merges at $\mu = \mu_s^*$ with the arrested steady-state solution $g_{\text{loc}} = 0$, which remains stable until μ exceeds μ_s^*. Collecting the pieces, the steady-state solutions of (3.7) can be expressed in terms of the inertial number I as

$$\mu_{\text{loc}}(I) = \mu_s + \frac{\mu_2-\mu_s}{(\mu_2-\mu_s)/(bI+\chi(I;\kappa))+1}, \tag{3.8}$$

for $I > 0$, and $\mu_{\text{loc}} \leq \mu_s^*$ otherwise. The function χ is now formulated in terms of I and κ as

$$\chi(I;\kappa) = a\left[1-\tanh\left(cI\kappa^n\right)\right], \tag{3.9}$$

and the static yield stress ratio μ_s^* is obtained as

$$\mu_s^* = \mu_{\text{loc}}(I\to 0) = \frac{\mu_s(\mu_2-\mu_s)+a\mu_2}{\mu_2-\mu_s+a}. \tag{3.10}$$

Thus, the modified NGF model reduces to the non-monotonic $\mu(I)$ rheology (3.1) in steady homogeneous flows, and the corresponding stable and unstable branches are displayed by the thin green and red lines in Figure 3-3(c). In the quasi-static, low-I regime, the presence of χ induces a weakening relationship between μ and I. For higher values of I, the vanishing of χ leads to a strain-rate strengthening

regime in which the non-monotonic local rheology (3.8) converges to its monotonic counterpart (3.4). The crossover between the two regimes occurs at $d\mu_{\text{loc}}/dI = 0$, which corresponds to

$$I^* = \frac{1}{c\kappa^n}\text{sech}^{-1}\sqrt{\frac{b}{ac\kappa^n}} \quad \text{and} \quad \mu^* = \mu_{\text{loc}}(I^*). \tag{3.11}$$

In agreement with force balance arguments [88], the strain-rate weakening regime is unstable while the strain-rate strengthening regime is stable. Since the latter exists for μ above μ^* and the arrested solution is stable below μ_{s}^*, there exist two stable steady-state solutions – one flowing and one arrested – in the range $\mu^* < \mu < \mu_{\text{s}}^*$. In the absence of flow gradients, this bistable behavior generates hysteresis when the stress ratio μ is ramped up and down: flow is triggered at μ_{s}^* but stops at a lower μ^*. In addition, the onset of flow is accompanied by a finite jump in the velocity of the system, as the inertial number jumps from the arrested solution to the stable flowing solution. Hence features (F1) and (F2) are accounted for, but it is unclear whether this would hold for inhomogeneous flow, in the presence of nonlocal diffusion imparted by boundaries or nonuniformities in the stress ratio.

3.2.3 Interplay between hysteresis and nonlocality

We now investigate qualitatively the combined effects of non-monotonicity and nonlocal diffusion on the characteristics of the flow-arrest transition in the presence of a spatially-varying stress ratio. To do so, we calculate quasi-steady, stress-driven predictions of the NGF model with hysteresis in the plane shear under gravity configuration pictured in Figure 3-2(b), where flow occurs along the x-direction and gravity acts orthogonally along the z-direction. A shear stress τ_{w} and pressure P_{w} are applied at the top wall, imparting under quasi-steady conditions a constant shear stress $\tau(z) = \tau_{\text{w}}$ and a nonuniform pressure $P(z) = P_{\text{w}} + \phi\rho_{\text{s}}Gz$, where ρ_{s} is the grain density, ϕ the mean area packing fraction and G the acceleration of gravity. The ratio of shear stress to pressure is thus given by

$$\mu(z) = \frac{\tau(z)}{P(z)} = \frac{\mu_{\text{w}}}{1 + z/\ell}, \tag{3.12}$$

where $\mu_{\text{w}} = \tau_{\text{w}}/P_{\text{w}}$ is the applied stress ratio at the top wall, and $\ell = P_{\text{w}}/\phi\rho_{\text{s}}G$ is a loading length scale measuring the relative importance of the pressure imparted by the top wall versus that due to the weight of the grains. Critically, ℓ is inversely proportional with the degree of nonuniformity of the stress ratio (3.12) and, thus, the strength of nonlocal effects in this geometry [150]. Results from previous DEM simulations [118] as well as our own (see Section 3.3.2) have shown that these nonlocal effects induce the same flow-arrest transition features (F1–F3) that are observed in other geometries.

We compute quasi-steady, time-dependent solutions of the NGF model with hysteresis using the same arbitrary parameters as in the previous section. Because the dynamics are uniform in the streamwise x-direction, the fluidity equation (3.5b) re-

duces to a one-dimensional PDE for $g(z,t)$, which is discretized following the procedure presented in Appendix B.2. From there, the strain rate $\dot{\gamma}(z,t)$ and therefore the velocity profile $v(z,t)$ can be computed using the flow rule (3.5a). The fluidity equation is driven by the stress ratio (3.12), for which we choose an arbitrary value $\ell = 100d$ small enough that the results are independent of the height of the domain[3]. Following previous work [118], the influence of boundaries is minimized by prescribing homogeneous Neumann boundary condition for g at both walls. Simulations begin in a flowing state at $\mu_w = 0.35$, then μ_w is progressively decreased to 0.25 before being ramped back up to 0.35 in order to induce flow arrest and restart. We ensure that the ramp rate is slow enough that it does not affect the results. Most importantly, we perform these simulations for various values of the scalar parameter A prescribing the strength of nonlocal effects, so that we can pinpoint the specific contributions of nonlocal diffusion and non-monotonicity of the limiting local rheology (3.8) to each of the three features (F1–F3).

Figures 3-4(a–c) display the time-dependent dimensionless velocity at the top wall, $\tilde{v}_w(t)$, versus the applied stress ratio, $\mu_w(t)$, for (a) $A = 0$, (b) $A = 0.03$, and (c) $A = 0.9$. Here, the dimensionless velocity is defined as $\tilde{v}_w(t) = v(z = 0, t)/\ell\sqrt{m/P_w}$. The down stress ramp is shown in blue while the up stress ramp is shown in black, as depicted by the arrows. Further, Figures 3-4(d–f) display both g fields corresponding to the two states indicated by the lone circle and cross on the down and up ramps at $\mu_w = 0.33$. Correspondingly, these fields are shown with circle or cross markers depending on the stress ramp that they belong to, and they are superimposed to the stable (green) and unstable (red) steady-state solutions g_{loc} of the local fluidity equation (3.7) under the same stress ratio and pressure fields. A movie version of Figure 3-4, which follows the state of the system as the applied stress ratio $\mu_w(t)$ is progressively decreased and increased, is also included in the electronic supplementary information.

We begin with the case $A = 0$, for which nonlocal effects are turned off and hence the fluidity equation (3.5b) is identical with its local limit (3.7). As shown in Figure 3-4(a), the non-monotonicity of the limiting local rheology (3.8) leads to different μ_w versus v_w branches in Figure 3-4(a) when the stress is ramped down or up. Indeed, the bistable behavior of (3.7) for $\mu^* < \mu < \mu_s^*$ implies that there are two stable steady-state solutions g_{loc} – one flowing and one arrested – within a range of heights, as shown by the green branches in Figure 3-4(d). When the applied stress is ramped down, the bistable region moves towards smaller (shallower) values of z, that were previously flowing; thus, the time-dependent solution for g will remain on the flowing branch. Conversely, when the applied stress is ramped up, the bistable region moves towards larger (deeper) values of z, that were previously arrested; thus, the time-dependent solution will remain on the arrested branch. This explains why the down ramp flows at a higher wall velocity v_w than the up ramp in Figure 3-4(a), which also causes flow to arrest at a lower wall stress ratio, μ^*, than that at which it restarts, μ_s^*.

[3]Previous DEM simulations[118] have shown that the vertical extent of the shear region below the top wall scales with ℓ. For small enough ℓ, the bottom region is thus quasi-static, and the results are independent of domain height.

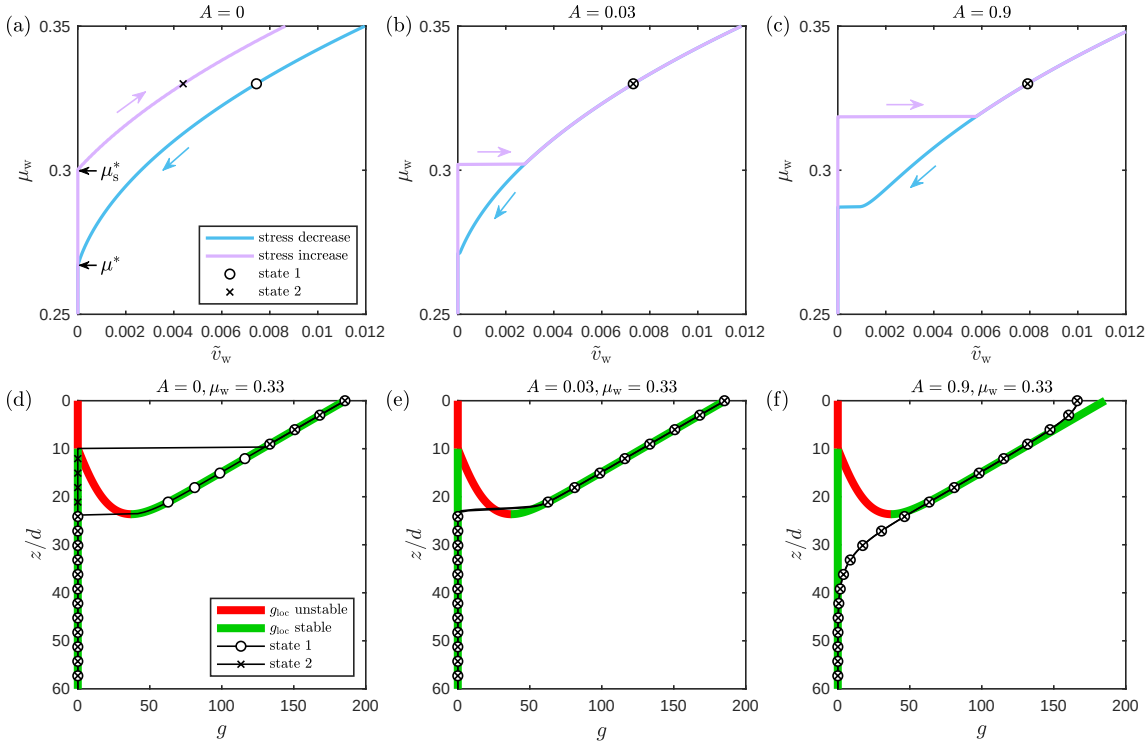

Figure 3-4: Qualitative behavior of the NGF model with hysteresis in stress-driven plane shear with gravity. (a,b,c) Time-dependent dimensionless velocity at the top wall, $\tilde{v}_w(t)$, versus the applied stress ratio at the wall, $\mu_w(t)$. (d,e,f) Fluidity field $g(z,t)$ corresponding to the state indicated in (a,b,c) by the lone circle (cross) on the down (up) branch at $\mu_w = 0.33$. The field belonging to the down (up) branch is shown with circle (cross) markers and is superimposed to the stable (green) and unstable (red) steady-state solutions g_{loc} of the local fluidity equation (3.7) under the stress ratio field (3.12). A movie version of this figure, which follows the state of the system as $\mu_w(t)$ is progressively decreased and increased, is also included in the electronic supplementary information.

Further, the onset of flow during the up ramp is characterized by a smooth increase in the top wall velocity since the thickness of the flowing layer beneath the top wall smoothly increases from zero (see movie). In conclusion, the non-monotonicity of the local rheology suffices to endow the flow-arrest transition with feature (F1), but features (F2) and (F3) are absent – there is no finite velocity jump at flow onset, and the starting and stopping stresses are identical with the local rheology predictions in Figure 3-3(c).

We then turn to the case $A = 0.03$, corresponding to a tiny amount of nonlocal effects. Figure 3-4(b) suggests that in this situation, the main role of the nonlocal diffusion term in (3.5b) is to merge the flowing section of the up stress ramp with that of the down stress ramp, which is almost unchanged from the case $A = 0$. In other words, there is only one possible flowing solution for every value of μ_w, and Figure 3-4(e) shows that in the bistable range of heights, this unique solution follows the flowing local solution g_{loc} and not the arrested one. This is a direct consequence of the regularizing effect of the diffusion term, which acts to minimize discontinuities in the fluidity profile. A crucial side effect is that a finite velocity jump emerges at flow onset, since the entire bistable region beneath the top wall suddenly jumps from the arrested to the flowing local solution (see movie). The interplay between nonlocality and non-monotonicity of the local rheology is therefore critical in achieving feature (F2), with (F3) the only one that remains unaccounted for.

Finally, we investigate the case $A = 0.9$, corresponding to the real calibrated value that we use later. Figure 3-4(b) shows that similar to the case $A = 0.03$, there is only one possible flowing solution for every value of μ_w. However, the increased strength of nonlocal diffusion leads to a different μ_w versus v_w relationship than before, with much higher wall stress ratios at flow arrest and onset. This is caused by the diffusion term spreading fluidity towards the $\mu < \mu^*$ region where the local solution is arrested, as revealed in Figure 3-4(f). Thus, when μ_w is hardly higher than the stopping and starting stress ratios observed in the case $A = 0.03$, the $\mu < \mu^*$ region acts as a fluidity sink that prevents the overall nonlocal solution from flowing. The resulting strengthening of the wall stress ratio at flow onset and arrest is dependent on the degree of nonuniformity of the stress ratio (3.12) controlled by the loading length scale ℓ, and increases with decreasing ℓ. As a result, nonlocal diffusion induces strengthening of the threshold for flow onset and arrest with reducing system size, which generalizes a similar conclusion from previous studies [97, 117] that only looked at the stopping stress.

To summarize, non-monotonicity and nonlocality are seen to contribute in different ways to the features (F1–F3) of the flow arrest transition in the case of plane shear with gravity:

(F1) hysteresis of the critical stresses for flow onset and arrest is achieved through non-monotonicity of the local rheology;

(F2) the finite velocity jump at flow onset requires an interplay between non-monotonicity and nonlocal diffusion;

(F3) increased strengthening at smaller system sizes is caused by nonlocal diffusion.

Clearly, all three features are simultaneously achievable only when both non-monotonicity and nonlocality are included in the model. Even though these results have been obtained in a plane shear under gravity configuration, they should hold in other geometries that display a similar spatially-varying stress ratio profile such as annular shear between concentric cylinders, since the mechanisms at play are similar [150, 118]. Finally, the conclusions that we have reached should apply to other nonlocal rheological models, including in particular those that treat the inertial number I as an order parameter in place of g [24, 114].

3.3 Comparisons with DEM simulations

Now that we have established that the NGF model with hysteresis is able to reproduce qualitatively the various features of the flow-arrest transition, the next step is to compare quantitatively model predictions with discrete element method (DEM) simulations in a variety of geometries. To do so, we first calibrate the rheological parameters of the model using DEM simulations of homogeneous, simple plane shear. The calibrated model is then compared with DEM simulations in plane shear under gravity and inclined plane geometries.

We first describe the general setup of our DEM simulations, which we perform in the open-source software LAMMPS [146]. We simulate 2D disks with mean diameter $d = 0.0008\,\mathrm{m}$ and aerial density $\rho_\mathrm{s} = 1.3\,\mathrm{kg/m^2}$, corresponding to a characteristic grain mass $m = \rho_\mathrm{s}\pi d^2/4$. The disk diameters are uniformly distributed within $\pm 20\%$ of d in order to prevent crystallisation. Following seminal previous work [44], we use the standard spring-dashpot model with Coulomb friction for the contact force between overlapping particles [41]. More specifically, the normal force is given by $F_n = k_n\delta_n + g_n\dot{\delta}_n$ where $\delta_n \geq 0$ is the normal contact overlap, k_n the normal stiffness and g_n the damping coefficient, which can be expressed in terms of the coefficient of restitution e for a binary collision as $g_n = -(mk_n)^{1/2}(2\ln e)/(2(\pi^2 + \ln^2 e))^{1/2}$. The tangential force is given by $F_t = k_t\delta_t$ where δ_t is the accumulated tangential contact displacement and k_t is the tangential stiffness, and its magnitude is limited by the surface friction coefficient μ_surf so that $|F_t| \leq \mu_\mathrm{surf}|F_n|$. Thus, the contact force model is fully described by the parameters k_n, k_t, e, and μ_surf, to which we assign the same values as in past studies [106, 99, 118]. Namely, we use $\mu_\mathrm{surf} = 0.4$ and we choose k_n so that $\kappa = k_n/P > 10^4$, with P the characteristic confining pressure, corresponding to the stiff grain limit [159, 44]. Further, we set $k_t/k_n = 1/2$ and $e = 0.1$, with both having little influence on the phenomenology of dense flows of stiff disks [165, 29]. Finally, we choose a time step equal to 0.1 times the binary collision time $\tau_c = (m(\pi^2 + \ln^2 e)/4k_n)^{1/2}$. At the end of each simulation, the particle-wise quantities in each saved system snapshot are coarse-grained into continuum fields according to the procedure described in Appendix B.1. Because the geometries that we investigate are homogeneous along the x-direction, this spatial averaging procedure returns an instantaneous streamwise velocity field $v(z,t)$, as well as instantaneous stress field components $\sigma_{xx}(z,t)$, $\sigma_{zz}(z,t)$, and $\sigma_{xz}(z,t)$.

3.3.1 Calibration with simple plane shear

We first simulate a simple plane shear geometry in DEM, since the homogeneity of the flow in this configuration enables us to calibrate the local part of the NGF model, given by the limiting local rheology (3.8). The configuration is pictured in Figure 3-2(a), and consists of two parallel walls of length $L = 120d$ aligned along the horizontal x-direction, and separated by a distance H along the vertical z-direction. The walls consist of glued disks, colored in brown in Figure 3-2(a), and they shear a dense collection of enclosed disks, colored according to their relative velocity magnitude in a particular snapshot of a flowing state. Periodic boundary conditions are applied along the x-direction, and the absence of gravity leads to a uniform stress ratio throughout, which is imparted by the walls. The top wall is assigned a horizontal velocity v_w that is either directly prescribed, corresponding to a velocity boundary condition, or calculated following a control scheme that simulates an applied tangential stress τ_w to the top wall through the feedback law $\dot{v}_\mathrm{w} = (\tau_\mathrm{w} - \sigma_{xz}(z = 0, t))L/M_\mathrm{w}$, corresponding to a stress boundary condition. Here, the instantaneous tangential stress σ_{xz} exerted by the flowing grains is directly evaluated at the wall, and the effective wall mass M_w acts as a damping parameter, which we take as $M_\mathrm{w} = 2000m$. Although velocity-driven DEM simulations of plane shear flow are the norm [133], such simulations miss important features of the flow-arrest transition [173]. To our knowledge, stress-driven plane shear simulations have only been implemented in very few studies either through a solid wall [186, 34], which corresponds to our setup, or through shearing of the periodic boundaries [143, 169, 172, 174]. Finally, the pressure at the top wall is maintained close to a target value P_w through a widely used feedback law [44] according to which the distance H between the walls evolves as $\dot{H} = (P_\mathrm{w} + \sigma_{zz}(z = 0, t))L/g_\mathrm{w}$, where the instantaneous normal stress σ_{zz} exerted by the flowing grains is directly evaluated at the wall, and g_w is a damping parameter that we take as $g_\mathrm{w} = 100(mk_n)^{1/2}$.

We begin by performing velocity-driven DEM simulations under various prescribed values of the top-wall velocity v_w and for two nominal system heights $H = 50d$ and $25d$. After each simulation has reached a steady state, we save 4000 system snapshots evenly distributed in time over a minimum additional top-wall shear displacement of $78H$. The instantaneous continuum velocity and stress fields produced by the coarse-graining procedure are then averaged in time. As expected from numerous previous studies [133, 44], the strain rate and the stress components are all approximately constant in the central part of the sheared layer, four grain diameters away from the walls. We therefore spatially average these quantities into the strain rate $\dot{\gamma} = \langle |dv/dz| \rangle$, shear stress $\tau = \langle \sigma_{xz} \rangle$ and pressure $P = \langle |\sigma_{zz}| \rangle$, from which we calculate the stress ratio $\mu = \tau/P$ and the inertial number $I = \dot{\gamma}\sqrt{m/P}$. Different values of the prescribed top-wall velocity produce different (μ, I) pairs, which are displayed in Figure 3-5(a) for the two nominal system heights considered. We also plot corresponding results from the DEM simulations of Liu and Henann [118], obtained under identical system parameters and for a nominal height $H = 60d$. The agreement between the two sets of data validates our simulations; furthermore, the negligible difference between the $\mu(I)$ curves pertaining to different system heights demonstrates the negligible influ-

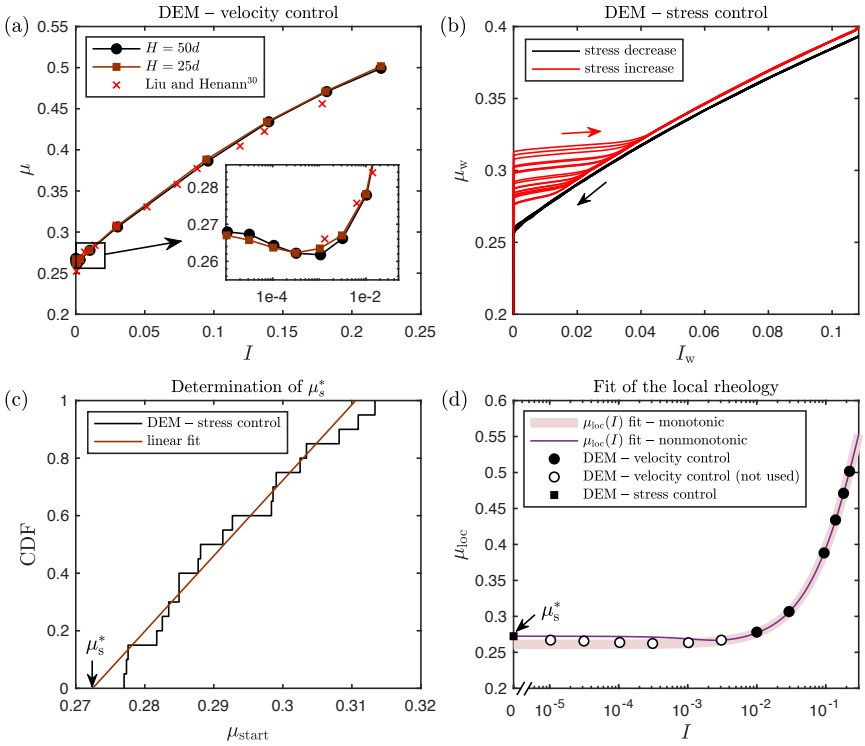

Figure 3-5: Calibration of the local part of the NGF model using simulations of plane shear without gravity. (a) Local $\mu(I)$ rheology obtained from velocity-driven DEM simulations for two system sizes, compared with data from Liu and Henann [118]. (b) Applied stress ratio at the top wall, μ_w, versus wall-based inertial number, I_w, obtained from stress-driven DEM simulations for $H = 50d$, in which the stress ratio applied to the upper wall is first slowly ramped down (black lines), then ramped back up (red lines). Different lines correspond to different initial conditions, realized by letting the system flow at $\mu = 0.4$ for varying amounts of time. (c) Determination of the local yield stress ratio μ_s^* from the stochastic μ_{start} values pertaining to the different realizations. (d) The limiting non-monotonic local rheology (3.8) is fit in two steps. First, the parameters shared with the monotonic form (3.4) are calibrated using the velocity-driven DEM data for $I \geq 10^{-2}$ (filled circles), producing the monotonic $\mu_{loc}(I)$ fit. Then, the parameters of the strain-rate weakening term $\chi(I;\kappa)$ are chosen so that μ_s^* is equal to the value extracted from the stress-driven DEM data (filled square) and the minimum of μ occurs for $10^{-3} < I^* < 10^{-2}$, finally producing the non-monotonic $\mu_{loc}(I)$ fit.

ence of the walls. Importantly, we observe the presence of a strain-rate weakening regime at low enough values of I, directly corroborating previous simulation results from DeGiuli and Wyart [47]. However, a decreasing relationship between shear stress and strain rate is mechanically unstable and typically results in the formation of shear bands that, in turn, render impossible the accurate measurement of the true strain rate in the strain-rate weakening regime [54]. Furthermore, past theoretical studies [195, 119] suggest that in this regime, systems with a nonlocal flow rule select a specific stress state independent of the nominal strain rate imparted by the walls. The NGF model, therefore, cannot be simply calibrated on velocity-driven DEM data if it is to accurately predict onset and arrest of flow under variations of the applied stress.

In order to extract the true critical stresses delineating the transition between static and flowing regimes, we run stress-driven DEM simulations of flow onset and arrest under slowly decreasing and increasing ramps of the top-wall stress ratio μ_w, for a nominal system height $H = 50d$. Specifically, we assign a time-dependent top-wall shear stress $\tau_\mathrm{w}(t) = \mu_\mathrm{w}(t)P_\mathrm{w}$ according to the control procedure described above, where $\mu_\mathrm{w}(t)$ is the target stress ratio applied to the top wall and P_w is the constant target pressure. We run 20 different simulations by letting the system flow at $\mu_\mathrm{w} = 0.4$ for a varying amount of time after it has reached steady state, effectively imparting a different initial microstructure to each simulation. The applied stress ratio μ_w is then linearly decreased from 0.4 to 0.25 over a time duration of $6.5 \cdot 10^7 \tau_c \simeq 113\,\mathrm{s}$, inducing jamming of the grains. We then let the contact forces relax by decreasing μ_w from 0.25 to 0 over a time duration of $2 \cdot 10^7 \tau_c \simeq 35\,\mathrm{s}$ and keeping μ_w at 0 for $1 \cdot 10^7 \tau_c \simeq 17\,\mathrm{s}$. Finally, μ_w is linearly ramped back up, first from 0 to 0.25 over a time duration of $2 \cdot 10^7 \tau_c \simeq 35\,\mathrm{s}$, then from 0.25 to 0.4 over $6.5 \cdot 10^7 \tau_c \simeq 113\,\mathrm{s}$, triggering onset of flow. The ramp rate is slow enough that the system can be assumed to undergo quasi-steady motion. We save 4000 system snapshots evenly distributed in time, from start to end of the stress ramp. At the end of each simulation, we calculate an instantaneous wall-based inertial number $I_\mathrm{w}(t) = v_\mathrm{w}(t)/H_\mathrm{w}\sqrt{m/P_\mathrm{w}}$, where $v_\mathrm{w}(t)$ is a moving time window average over 50 snapshots of the instantaneous wall velocity to smooth out small fluctuations, and H_w is the average true vertical distance between both walls. Thanks to the little amount of observed wall slip[4] and the quasi-steady conditions, we may consider both μ_w and I_w as smooth approximations of the highly-fluctuating instantaneous values of μ and I in the bulk. The resulting μ_w versus I_w curves are shown in Figure 3-5(b) in black and red for the decreasing and increasing stress ramps, respectively, with different curves corresponding to different simulations. Observe the similarity between these curves and the ones shown in Figure 3-2(a), with hysteresis (feature F1) and a velocity jump at flow onset (feature F2) clearly visible in both geometries. Besides, our simulations reveal that the critical stress ratio at flow onset is stochastic and noticeably higher than the threshold obtained from velocity-driven simulations in Figure 3-5(a) as I vanishes. Such variability in the transition between arrested and flowing states has been observed previously [36], and may be explained by the role played by the specific structure of the particle contact network

[4] In our DEM simulations, the relative difference between the upper wall velocity and the coarse-grained streamwise velocity of the grains near the upper wall never exceeds 3%.

[18, 37, 175]. On the other hand, the critical stress ratio at flow arrest is much more narrowly distributed and similar to the velocity-driven flow threshold.

Thus, the NGF model needs to be calibrated using data from both velocity-driven and stress-driven DEM simulations, so as to correctly predict the characteristics of both the flowing regime and the transition between arrested and flowing states. However, the critical stress ratio for flow onset observed in the stress-driven simulations is highly stochastic, while the limiting local rheology (3.8) of the model predicts a deterministic value μ_s^*. To reconcile this apparent contradiction, we note that because continuum models in general aim to reproduce the ensemble-average behavior of the discrete system across all possible realizations, we expect our NGF model to predict onset of flow so long as any measurable region of the space of ensembles initiates flow. With this in mind, we therefore reduce the stochastic critical stress from DEM to a unique deterministic value that represents the *lowest achievable* starting stress of the system as follows. For each realization, we first define μ_{start} as the observed μ_w when I_w last exceeds 10^{-3} during stress increase. We then assume that the ensemble of stochastic μ_{start} values follows a uniform probability distribution over a finite range (bounded from below by the lowest achievable starting stress), which allows us to fit a linear function to their cumulative distribution (CDF), shown in Figure 3-5(c). An estimate for the lowest achievable μ_{start} is given by the x-intercept of the fitted CDF, which we thus assign to $\mu_s^* = 0.2724$. Finally, we show in Appendix B.3 that the distribution of stochastic μ_{start} values barely changes for smaller nominal heights $H = 25d$ and $10d$, which supports our methodology of calculating the local yield stress ratio μ_s^* predicted by the model using stress-driven simulations at $H = 50d$.

The parameters of the limiting local rheology (3.8) can now be calibrated following a two-step approach pictured in Figure 3-5(d). First, the parameters shared with the monotonic form (3.4) are calibrated using the velocity-driven DEM data corresponding to $H = 50d$ and $I \geq 10^{-2}$ (filled circles), producing the monotonic $\mu_{\text{loc}}(I)$ fit. The velocity-driven DEM data for $I < 10^{-2}$ (open circles) is also shown for reference, but is not used in the calibration. Second, the parameters of the strain-rate weakening term $\chi(I;\kappa)$ are chosen so that μ_s^* is equal to the value extracted from the stress-driven DEM data (filled square) and the minimum of μ occurs for $10^{-3} < I^* < 10^{-2}$, finally producing the non-monotonic $\mu_{\text{loc}}(I)$ fit. The resulting parameter values are $\mu_s = 0.2610$, $\mu_2 = 0.9784$, $b = 1.6406$, $a = 0.0116$, $c = 50$, and $n = 1/4$. We note that once μ_s and μ_2 are known from the first step, a is obtained in the second step from μ_s^* by inverting (3.10). In the second step, we have selected the value $1/4$ from DeGiuli and Wyart [47] for the parameter n that controls the grain stiffness-dependence of the hysteresis amplitude. We have nonetheless verified that choosing instead $n = 0$, which removes the stiffness dependence, and recalibrating c accordingly produces negligible changes in our results to follow. Finally, we adopt the value $A = 0.9$ from Liu and Henann [118] for the nonlocal amplitude, which they calibrated on DEM data obtained with the same particle and contact force law properties.

3.3.2 Plane shear under gravity

We now compare the predictions of the calibrated NGF model with stress-driven DEM simulations of plane shear under gravity shown in Figure 3-2(b). We have already investigated in Section 3.2.3 the qualitative behavior of the model in this geometry, in which the gravity field imparts a nonuniform distribution of stress ratio $\mu(z)$ characterized by a loading length scale ℓ; see equation (3.12). Besides the presence of gravity, the DEM setup of the system is identical to that of the previous section, except for the nominal distance $H = 60d$ between the walls. The shear stress and pressure at the top wall are controlled according to the feedback schemes described in the previous section, and the bottom wall is fixed.

Similar to the case of simple plane shear, we perform DEM simulations of arrest and onset of flow under decreasing then increasing ramps of applied stress. Specifically, a time-varying top-wall stress ratio $\mu_{\mathrm{w}}(t)$ is prescribed through a time-dependent target shear stress $\tau_w(t) = \mu_{\mathrm{w}}(t)P_{\mathrm{w}}$ and constant target pressure P_w. The applied stress ratio $\mu_{\mathrm{w}}(t)$ follows the same protocol as in Section 3.3.1, with the only difference being that we start from, and end at, a top value of $\mu_{\mathrm{w}} = 0.45$ instead of $\mu_{\mathrm{w}} = 0.4$. Correspondingly, the duration of the decreasing and increasing sections of the stress ramp between 0.45 and 0.25 is lengthened to $9 \cdot 10^7 \tau_c \simeq 157\,\mathrm{s}$ so that the rate of change of $\mu_{\mathrm{w}}(t)$ is unaffected. As before, for each length scale ℓ we run 20 different simulations corresponding to different initial microstructures, by letting the system spend a varying amount of time in the initial shearing period at $\mu_{\mathrm{w}} = 0.45$. During that period, gravity is first turned off to ensure homogeneous mixing of the grains, before being turned back on. We save 4000 system snapshots for each simulation, from which we calculate $v_{\mathrm{w}}(t)$, a moving time window average over 50 snapshots of the instantaneous wall velocity. We then calculate the nondimensional wall velocity as $\tilde{v}_{\mathrm{w}}(t) = v_{\mathrm{w}}(t)/\ell\sqrt{m/P_{\mathrm{w}}}$. In order to compare deterministic predictions from the NGF model with the DEM results corresponding to all $N = 20$ different runs, we transform the discrete values $\tilde{v}_{\mathrm{w}}^{(i)}(t)$ for $i = 1, \ldots, N$ at each time step into a continuous probability density function (PDF) as follows:

$$f(\tilde{v}_{\mathrm{w}}) = \int_{-\infty}^{\infty} w(v - \tilde{v}_{\mathrm{w}}) \left[\frac{1}{N} \sum_{i=1}^{N} \delta(v - \tilde{v}_{\mathrm{w}}^{(i)}) \right] dv, \tag{3.13}$$

where δ is the Dirac delta function, and w is the Gaussian kernel

$$w(v) = \frac{1}{\sqrt{2\pi}L} e^{-x^2/2L^2}, \tag{3.14}$$

with L the kernel size, which we choose equal to 0.01 times the maximum observed value of $\tilde{v}_{\mathrm{w}}^{(i)}(t)$ across all runs and times. The PDF (3.13) can be conveniently expressed as

$$f(\tilde{v}_{\mathrm{w}}) = \frac{1}{N} \sum_{i=1}^{N} w(\tilde{v}_{\mathrm{w}} - \tilde{v}_{\mathrm{w}}^{(i)}), \tag{3.15}$$

and it integrates to one, as it should. Further, we also need to reduce the stochastic

transition stresses between flowing and arrested states into a unique deterministic value, which we define in a similar way to Section 3.3.1 as the *lowest achievable* value to be consistent with the methodology followed in calibrating the local yield stress ratio μ_s^* of the model. For each realization, we first calculate $\mu_{w,start}$ as the observed μ_w when $\tilde{v}_w$ last exceeds the threshold value 10^{-3} during stress increase, and $\mu_{w,stop}$ as the observed μ_w when $\tilde{v}_w$ last falls below the same threshold during stress decrease. The deterministic starting and stopping critical stress ratios are then defined as the x-intercept of a fitted linear CDF to the stochastic $\mu_{w,start}$ and $\mu_{w,stop}$ values, similar to Figure 3-5(c). In the following, we will refer to these deterministic thresholds as $\mu_{w,start}$ and $\mu_{w,stop}$.

We also compute NGF model predictions in the same geometry, using the calibrated parameter values from Section 3.3.1. As described in Section 3.2.3, the fluidity equation (3.5b) reduces to a one-dimensional PDE for $g(z,t)$ that requires the stress ratio profile $\mu(z,t)$ as input. Thanks to the homogeneous and quasi-steady conditions, the latter is given through (3.12) and set by the top-wall stress ratio $\mu_w(t)$, to which we assign the exact same temporal protocol as in the DEM simulations. Since the NGF model is not expected to be valid in the vicinity of the walls, we end the corresponding simulation domain at a distance $d_w = 2d$ away from the real walls. Furthermore, we follow previous work [118] in using homogeneous Neumann ($\partial g/\partial z = 0$) boundary conditions on g at the walls in order to minimize their influence. The details of the discretization method for the ODE governing $g(z,t)$ are presented in Appendix B.2. Once $g(z,t)$ is known, the strain rate $\dot{\gamma}(z,t)$ can be calculated using the flow rule (3.5a) and integrated into the velocity profile $v(z,t)$, taking into account a slip length equal to d_w for the velocity at the bottom wall. Finally, the top-wall velocity is extrapolated as $v_w(t) = v(z = -d_w, t) + d_w\dot{\gamma}(z = -d_w, t)$ and is nondimensionalized into $\tilde{v}_w(t)$.

Figures 3-6(a,b,d,e) display the relationship between μ_w and $\tilde{v}_w$ obtained from both DEM simulations and NGF model predictions for two different loading length scales of (a,d) $\ell = 100d$ and (b,e) $\ell = 25d$. The increasing stress ramp is shown in (a,b) while the decreasing stress ramp is shown in (d,e). The DEM results are displayed as contours of $f(\tilde{v}_w)$ corresponding to each value of μ_w, in such a way that the plots can be read as the probability of occurrence of individual realizations, with yellow color indicating high probability and blue color indicating low probability. The deterministic $\mu_{w,start}$ and $\mu_{w,stop}$ values from DEM are also displayed as filled circles, and the NGF model prediction is shown as the red line. An excellent agreement between NGF and DEM is observed in the flowing regime[5]. Similarly, the transition between arrested and flowing states occurs at similar stress levels in both cases, and displays all three features (F1–F3) identified in the introduction: hysteresis between onset and arrest, velocity jump at onset, and strengthening with smaller ℓ/d. The amount of velocity jump at flow onset exhibited by the NGF solution appears smaller than that observed in the DEM simulations; this is a consequence of the NGF model

[5]The small discrepancy observed in the case $\ell/d = 25$ is probably attributable to the choice of boundary conditions for g. An extensive discussion on the role of the latter is presented in Section 3.3.3.

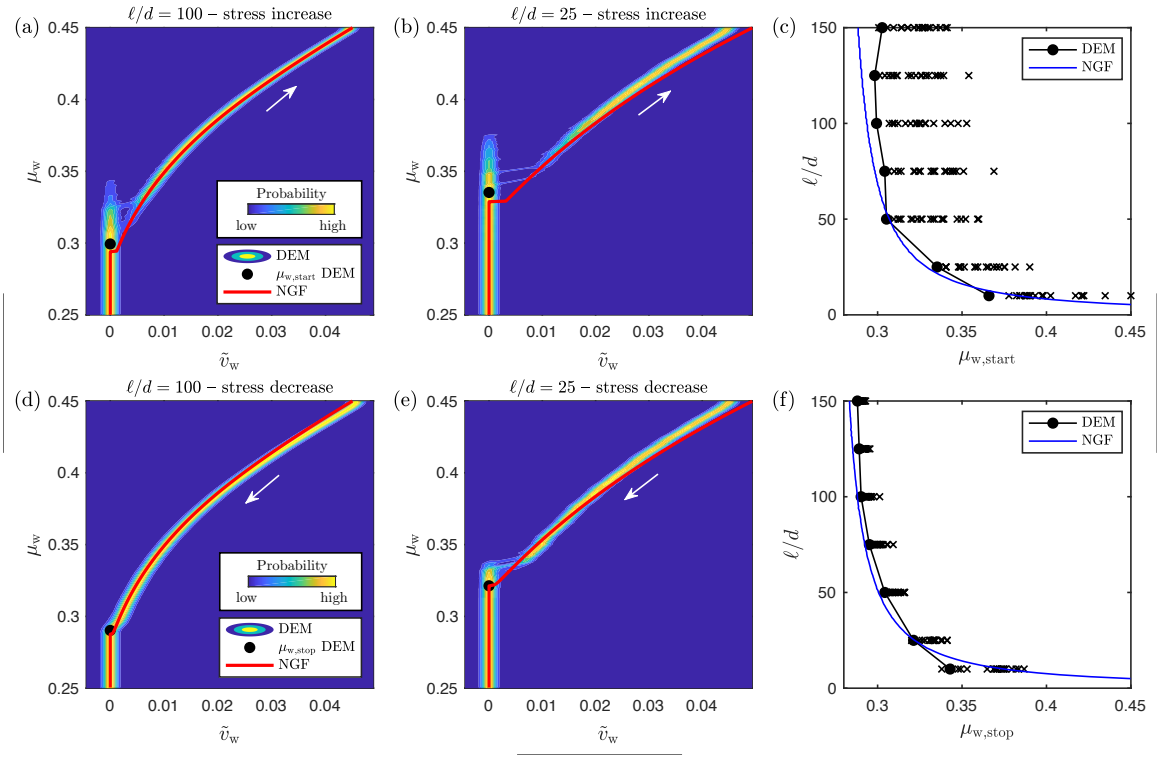

Figure 3-6: Comparison between NGF model predictions and DEM simulations for plane shear under gravity. (a,b,d,e) Relationship between stress ratio μ_w and dimensionless velocity $\tilde{v}_w$ at top wall from DEM (contour of probability values extracted from 20 different runs) and NGF (red lines) for two different loading length scales of (a,d) $\ell = 100d$ and (b,e) $\ell = 25d$, with the (a,b) increasing and (d,e) decreasing stress ramps shown separately. The filled circles display the deterministic $\mu_{w,start}$ and $\mu_{w,stop}$ critical stress ratios obtained from all DEM runs. (c,f) Critical stress ratios $\mu_{w,start}$ and $\mu_{w,stop}$ versus dimensionless loading length scale ℓ/d from DEM (filled circles) and NGF (continuous line). The crosses show the individual transition stresses pertaining to each of the 20 different runs, while the filled circles represent the deterministic value extracted from the linear fit of the CDF.

being calibrated so as to start flowing at the lowest possible critical stress based on the DEM simulations.

Finally, Figures 3-6(c,f) display the critical stresses $\mu_{\text{w,start}}$ and $\mu_{\text{w,stop}}$ versus the dimensionless loading length scale ℓ/d. Shown are the individual transition stresses from all 20 DEM runs (crosses), the deterministic values extracted from the linear fit of the CDF (filled circles), and the corresponding NGF predictions (continuous line). The NGF predictions of the critical stresses are obtained using the methodology described in Appendix B.4, which circumvents the need to run time-dependent simulations for every value of ℓ/d. Overall, the NGF model predicts a similar amount of strengthening as apparent in the DEM simulations. The slightly higher critical stresses exhibited in DEM for large loading length scales is caused by an observed change in the slope of the g field at the boundary as the pressure applied by the top wall increases. Accordingly, implementing a homogeneous Robin boundary condition for g with a finite associated length scale would lead the NGF model to predict higher values for the transition stresses, closing the gap with the DEM data. Lastly, Appendix B.5 shows that the hysteresis amplitude $\mu_{\text{w,start}} - \mu_{\text{w,stop}}$ is only weakly dependent on the dimensionless loading length scale ℓ/d.

3.3.3 Inclined plane

As a last example, we evaluate predictions from the calibrated NGF model against DEM simulations in the inclined plane configuration shown in Figure 3-2(c). A fixed basal wall of length $L = 120d$ and consisting of glued disks is inclined at an angle θ with respect to the horizontal, and is covered by a dense collection of freely moving disks forming a layer of height H. The x- and z-directions are parallel and orthogonal to the base wall, respectively, and periodic boundary conditions are applied along the x-direction. The gravity field imparts a uniform ratio of shear stress to pressure throughout the layer at steady state equal to $\mu = \tan\theta$, making this configuration inherently stress-driven. Even though μ is uniform, nonlocal effects still arise from the presence of the rough base, which acts as a sink for velocity fluctuations within the moving grains. The transition between onset and arrest of flow on an inclined plane exhibits all three features mentioned in the introduction, as documented in many past experimental and computational studies [148, 45, 165, 149, 157, 167, 160].

Following the previous cases, we run DEM simulations of flow arrest and onset by slowly decreasing then increasing the inclination angle. Specifically, we prescribe a temporal profile for $\theta(t)$ such that the stress ratio $\mu(t) = \tan\theta(t)$ follows the same protocol as in Section 3.3.1, with an initial flowing period at $\mu = 0.4$ followed by a continuous decrease to $\mu = 0$ and a continuous increase back to $\mu = 0.4$. As before, we execute 20 runs for each layer height H, each with a different time duration spent in the initial flowing regime at $\mu = 0.4$, giving a unique microstructure to every simulation before the start of the stress ramp. We save 4000 system snapshots in each simulation, from which we compute the instantaneous continuum velocity field $v(z,t)$. Anticipating that the NGF model will be run over a truncated domain ending at a distance $d_w = 2d$ away from the bottom wall and $d_s = 3d$ away from the layer's free surface, we then calculate a depth-averaged instantaneous velocity $\bar{v}(t)$ over the

corresponding truncated domain, which we smooth out using a moving time window average over 50 snapshots. We then express $\bar{v}(t)$ in terms of a dimensionless Froude number defined as $Fr(t) = \bar{v}(t)/\sqrt{GH}$. As was done in the previous section, the discrete values $Fr^{(i)}(t)$ at each time step, for $i = 1, \ldots, N$ corresponding to all $N = 20$ different runs, are transformed into a continuous PDF $f(Fr)$ through the convolution (3.13). Finally, the stochastic transition stresses between flowing and arrested states are reduced into deterministic numbers μ_{start} and μ_{stop} according to the procedure detailed in Sections 3.3.1 and 3.3.2. For each realization, we first calculate μ_{start} as the observed μ when Fr last exceeds the threshold value 10^{-2} during stress increase, and μ_{stop} as the observed μ when Fr last falls below the same threshold during stress decrease. The deterministic starting and stopping critical stress ratios are then defined as the x-intercept of a fitted linear CDF to the stochastic μ_{start} and μ_{stop} values, similar to Figure 3-5(c). In the following, we will refer to these deterministic thresholds as μ_{start} and μ_{stop}.

We compute predictions of the NGF model in the same geometry, using the calibrated parameter values from Section 3.3.1 and the same temporal protocol for $\theta(t)$ as in the DEM simulations. The fluidity equation (3.5b) reduces to a one-dimensional PDE for $g(z, t)$ that takes as input the stress ratio profile, which is still related to the inclination angle as $\mu(t) = \tan\theta(t)$ thanks to the quasi-steady conditions. The uniformity of the stress ratio implies that nonlocal effects are entirely imparted by boundaries, making the choice of boundary conditions critical. Similarly to our approach in Section 3.3.2, the domain for the NGF solution is defined to start at a distance $d_{\text{w}} = 2d$ away from the bottom wall due to the lack of validity of the NGF model near the boundary. At that location, the DEM data suggests a Robin-type homogeneous boundary condition for g, with an associated length scale δ that may be sensitive to various factors. We choose to sidestep the exact modeling of the boundary condition by considering the two edge cases of $\delta = 0$ and $\delta = \infty$, corresponding to homogeneous Dirichlet ($g = 0$) and Neumann ($\partial g/\partial z = 0$) boundary conditions, respectively. Regarding the top boundary, the DEM data shows that the strain rate vanishes about 3 grain diameters below the surface, corroborating previous studies [165, 167]. We thus end the NGF simulation domain at a distance $d_s = 3d$ away from the layer's free surface, and we prescribe a homogeneous Neumann ($\partial g/\partial z = 0$) boundary condition for g there, following Kamrin and Henann [97]. We also assign a finite pressure to the top boundary equal to the weight of the neglected layer of thickness d_s, which is approximately equal to $P(z = -d_s) = 0.8\rho_s G(2d)\cos\theta(t)$ due to the drop in packing fraction near the layer's surface. The details of the discretization method for the ODE governing $g(z, t)$ are presented in Appendix B.2. Once $g(z, t)$ is known, the strain rate $\dot{\gamma}(z, t)$ can be computed through the flow rule (3.5a). Finally, the velocity can be integrated from $\dot{\gamma}(z, t)$, taking into account a velocity slip length equal to d_{w} at the bottom boundary of the NGF domain, and depth-averaged to produce $\bar{v}(t)$ and hence $Fr(t)$.

Figures 3-7(a,b,d,e) display the relationship between μ and Fr obtained from both DEM simulations and NGF model predictions for two different layer heights at rest of (a,d) $H = 45.5d$ and (b,e) $H = 9d$. The increasing stress ramp is shown in (a,b) while the decreasing stress ramp is shown in (d,e). The DEM results are

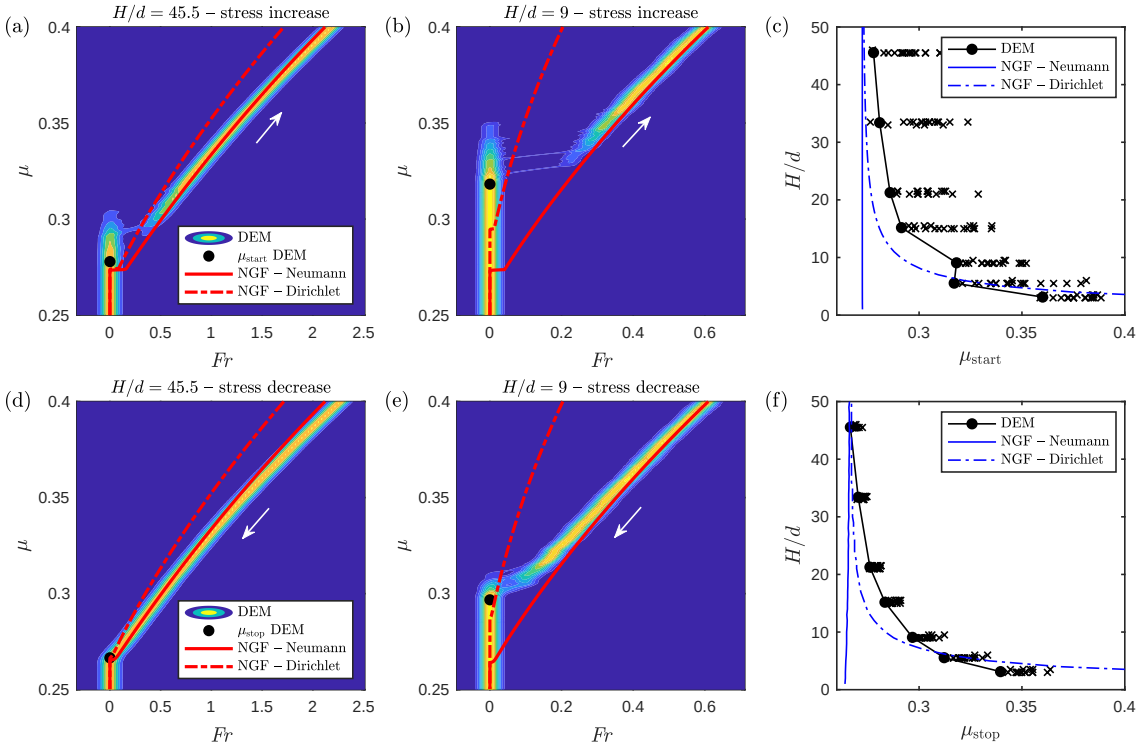

Figure 3-7: Comparison between NGF model predictions and DEM simulations for inclined plane. (a,b,d,e) Relationship between stress ratio $\mu = \tan\theta$ and Froude number Fr from DEM (contour of probability values extracted from 20 different runs) and NGF (red lines) for two different layer heights of (a,d) $H = 45.5d$ and (b,e) $H = 9d$, with the (a,b) increasing and (d,e) decreasing stress ramps shown separately. The filled circles display the deterministic μ_{start} and μ_{stop} critical stress ratios obtained from all DEM runs. (c,f) Critical stress ratios μ_{start} and μ_{stop} versus dimensionless layer height H/d from DEM (filled circles) and NGF (blue lines). The crosses show the individual transition stresses pertaining to each of the 20 different runs, while the filled circles represent the deterministic value extracted from the linear fit of the CDF.

displayed as contours of $f(Fr)$ corresponding to each value of μ, in such a way that the plots can be read as the probability of occurrence of individual realizations, with yellow color indicating high probability and blue color indicating low probability. The deterministic μ_{start} and μ_{stop} values from DEM are also displayed as filled circles, and the NGF model predictions pertaining to the two basal boundary conditions for g are shown as the continuous (homogeneous Neumann) and dash-dotted (homogeneous Dirichlet) red lines. The Neumann boundary condition in the NGF model leads to an excellent agreement with DEM in the flowing regime. However, it does not predict any strengthening of the critical transition stresses μ_{start} and μ_{stop} for smaller layer height H. The Dirichlet boundary condition, on the other hand, reproduces the strengthening of the transition stresses but fails to match the DEM results in the flowing regime.

The critical stresses μ_{start} and μ_{stop} are depicted in greater detail in Figures 3-7(c,f) versus the dimensionless layer height H/d. Shown are the individual transition stresses from all 20 DEM runs (crosses), the deterministic values extracted from the linear fit of the CDF (filled circles), and the corresponding NGF predictions using homogeneous Neumann (continuous line) or homogeneous Dirichlet (dash-dotted line) boundary conditions. As we did for planar shear with gravity, the NGF predictions of the critical stresses are obtained using the methodology described in Appendix B.4, which bypasses the need to run time-dependent simulations for every value of H/d. The Neumann boundary condition does not produce any strengthening of the critical stresses, since it kills the principal cause of flow inhomogeneity in this geometry[6]. On the other hand, the Dirichlet boundary condition generates a level of strengthening roughly comparable with the DEM data. As was the case for plane shear under gravity, Appendix B.5 shows that the hysteresis amplitude $\mu_{\text{w,start}} - \mu_{\text{w,stop}}$ is only weakly dependent on the dimensionless layer height H/d.

In summary, it appears that the critical stresses are best predicted by the NGF model endowed with a homogeneous Dirichlet boundary condition for g, while the flowing regime is most accurate when a homogeneous Neumann boundary condition is used. This dichotomy could stem either from missing ingredients in the NGF model itself or from an inaccurate description of the boundary condition. Regarding the former, a known shortcoming of the current formulation of the model is that it does not produce the widely documented collapse of the Froude number for all layer heights and angles [148, 167, 63]. We have thus modified the fluidity equation (3.5b) following the procedure given in Kamrin and Henann [97] so that the model collapses the Froude number far away from flow threshold, in the limit $\chi \to 0$. We also tried another version of that procedure replacing the quadratic term with a cubic one as in the model of Lee and Yang [114]. Yet neither of these modified models performed substantially better, with the no-slip solution still flowing significantly slower than the DEM data. This points to the boundary condition being the main culprit – incidentally, the true length scale δ at the bottom boundary is observed to increase

[6]In the presence of the Neumann boundary condition, a small amount of flow inhomogeneity is still incurred by the pressure-dependent quadratic term in the fluidity equation (3.5b), which explains the slight dependence of μ_{stop} on H/d in Figure 3-7(f).

with flow rate in the DEM data. With a velocity-dependent δ that jumps from near zero in the arrested state to a large value in the flowing state, the NGF model could potentially produce correct predictions of both the transition stresses and the flowing regime. A boundary condition of this type could be formulated as a dynamical system governing the evolution of δ in response to relevant driving quantities. As we note in the conclusion, however, formulating accurate and physically-justified fluidity boundary conditions remains a key open issue within NGF modeling, and such an endeavor is relegated to future work.

3.4 Conclusions

In this chapter, we have studied the combined role of strain-rate weakening behavior and nonlocal effects in explaining key features of the hysteretic transition between solid-like and liquid-like behavior in dense granular materials as the applied stress is ramped up and down. These features include the hysteresis of the critical stresses at flow onset and arrest, the finite jump in velocity during flow onset, and the strengthening of the critical stresses with reducing system size. In a first part, we modified the nonlocal granular fluidity (NGF) model so that it reduces to a non-monotonic local rheology in homogeneous flows. Through numerical simulations of flow onset and arrest in planar shear with gravity using the modified NGF model, we demonstrated qualitatively that the inclusion of both nonlocal effects and non-monotonicity of the local rheology is essential to account for all three features mentioned above.

In a second part, we compared quantitatively predictions of the modified NGF model with DEM simulations of flow onset and arrest in various geometries. First, we calibrated the local parameters of the model using DEM simulations of homogeneous plane shear flow. In so doing, we highlighted the importance of calibrating the local critical stress for flow onset using stress-driven simulations, since measurements based on velocity-driven simulations are unreliable in the strain-rate weakening regime. The stress-driven simulations, however, exhibited large variability in the transition stresses between arrested and flowing regimes. Thus, we developed a criterion to extract a unique deterministic value, corresponding to the lowest observable critical stress, from a large number of repeated runs. We then compared predictions of the calibrated NGF model with stress-driven DEM simulations in planar shear with gravity and inclined plane configurations. In the former case, the model gave accurate predictions of both the transition between flowing and arrested states as well as the characteristics in the flowing regime. In the latter case, the accuracy of the model predictions was strongly affected by the choice of boundary conditions, with no single choice able to reproduce both the transition stresses as well as the average velocity in the flowing regime.

These results suggest that the NGF model generally leads to more accurate predictions when nonlocal effects are generated by the spatial dependence of the μ field, as is the case for planar shear with gravity, rather than by boundaries alone, as is the case for inclined plane flow. A possible explanation stems from Liu and Henann's [118] observation that an inhomogeneous μ field leads to much stronger size-dependent strengthening than boundary effects, making the accuracy of model predictions less

reliant on the particular choice of fluidity boundary conditions when both mechanisms are present. Conversely, model predictions are much more sensitive to the specific type of boundary conditions when the latter are the sole source of nonlocal effects, which calls for better understanding of the interaction between flowing particles and the boundary. In fact, numerous experimental and computational studies have also underscored the sensitivity on wall roughness of transition stresses and velocity profiles in inclined plane flow [157, 166, 68, 190], plane shear flow without gravity [163], as well as annular shear flow [62]. Although progress has been made for flat frictional walls [7, 6], the correct modeling of boundary conditions in the general case is still an open question whose resolution would benefit all nonlocal rheological models [5, 31, 118].

Chapter 4

Solving geometry identification inverse problems via topology optimization with physics-informed neural networks

4.1 Introduction

In inverse problems, the goal is to find physical or geometric properties of a system that optimize a desired objective function under the constraint of the governing physical laws [8, 127]. Such problems concern either the identification of unobserved properties of an actual system given a set of measurement data, or the design of properties of a hypothetical system to achieve a target functionality, and they arise in various fields of engineering including solid mechanics [16, 22, 55], fluid mechanics [23, 46], optics [90, 137], and acoustics [57, 66]. Inverse problems involving identification or design of geometric properties are particularly hard to solve due to the very large number of parameters involved in describing the geometry [77], in addition to the complexity of the governing physical laws which usually take the form of partial differential equations (PDEs). An additional difficulty for topology optimization methods is to avoid restricting the search space to geometries of a given topology (number of structures or holes), as is the case in shape optimization methods, which requires a mechanism to handle merging or splitting of geometrical structures [164].

Motivated by the design of structures that exhibit minimum compliance under load, there exists various topology optimization methods for geometry design problems, although most of them are very complex as they combine traditional numerical solvers such as the finite-element or boundary-element method, adjoint techniques to evaluate the sensitivity of the objective function with respect to the shape or the topology, and gradient descent-based optimization algorithms to update the geometry at every iteration [14, 60, 1, 71]. For geometry identification problems, on the other hand, most methods are restricted to shape optimization since they require the number of shapes to be specified in advance [131, 112, 2, 188, 100, 94]. Other approaches

that can find the right number of shapes require a priori knowledge of the geometry of each shape [27, 177]. Only a few studies perform topology optimization by letting both the number and the shape of structures to be identified as part of the problem [113, 128, 129]. These studies, however, limit themselves to squares and circles and do not attempt to discover different shapes within the same geometry.

Physics-informed neural networks (PINNs) have recently emerged as a popular framework for solving inverse problems constrained by PDEs [53, 109, 152, 28]. By incorporating the residual of the PDEs together with the objective to minimize into the loss function of neural network-based approximations for the unknown state variables and properties to be found, PINNs can seamlessly blend measurement data or design objectives with physical constraints. Thus, they constitute a very user-friendly approach to solving inverse problems, and they also easily adapt to irregular domains thanks to their mesh-free nature. So far, PINNs have been applied successfully to a range of inverse problems involving identification and design of unknown physical properties or control inputs, in fields ranging from mechanics to optics and medicine [153, 161, 121, 74, 138, 32]. However, their application to geometry identification and design problems has been more limited due to the difficulty in representing shapes with well-defined boundaries. A notable exception is the recent study of Ref. [197], which applied PINNs to the problem of identifying voids and inclusions in elastic structures from boundary measurements of the response to a prescribed loading. They predefine both the number and the type (circle, ellipse, etc) of shapes to be identified, leaving the PINNs with the task of inferring the scalar parameter values defining these shapes. The practical applicability of their method is therefore restricted by the need for prior knowledge of the topology of the solution and the types of structures to expect.

Here, we introduce a topology optimization framework based on PINNs for identifying unknown geometries without any prior knowledge on the number and type of shapes. We allow for arbitrary solution topology by parametrizing the geometry using a material density field equal to 0 in one phase and 1 in the other, and defined by an underlying neural network. This neural network needs to be regularized in order to drive the material distribution towards 0 or 1 and avoid unphysical intermediate values. Thus, the key ingredient of our framework is a so-called eikonal regularization, inspired from level-set methods [141], that encourages the norm of the gradient of the underlying neural network to be unity in a narrow band along the interface. This ensures that the thickness of the interface region where the material distribution transitions between 0 and 1 is uniform everywhere, leading to well-defined boundaries throughout the domain. Following the standard PINNs procedure, the neural networks underlying the material distribution and the physical quantities are then trained to minimize a loss that enforces the governing equations as well as the measurement data to be satisfied. We apply our framework to the problem of identifying hidden voids or rigid inclusions in an elastic body using measurements of the surface deformation in response to a prescribed surface loading. Our approach successfully discovers the topology, locations and shapes of hidden structures in a variety of geometries with both linear elastic and nonlinear hyperelastic materials. This opens a pathway for physics-informed neural networks to be applied to a wide range of geom-

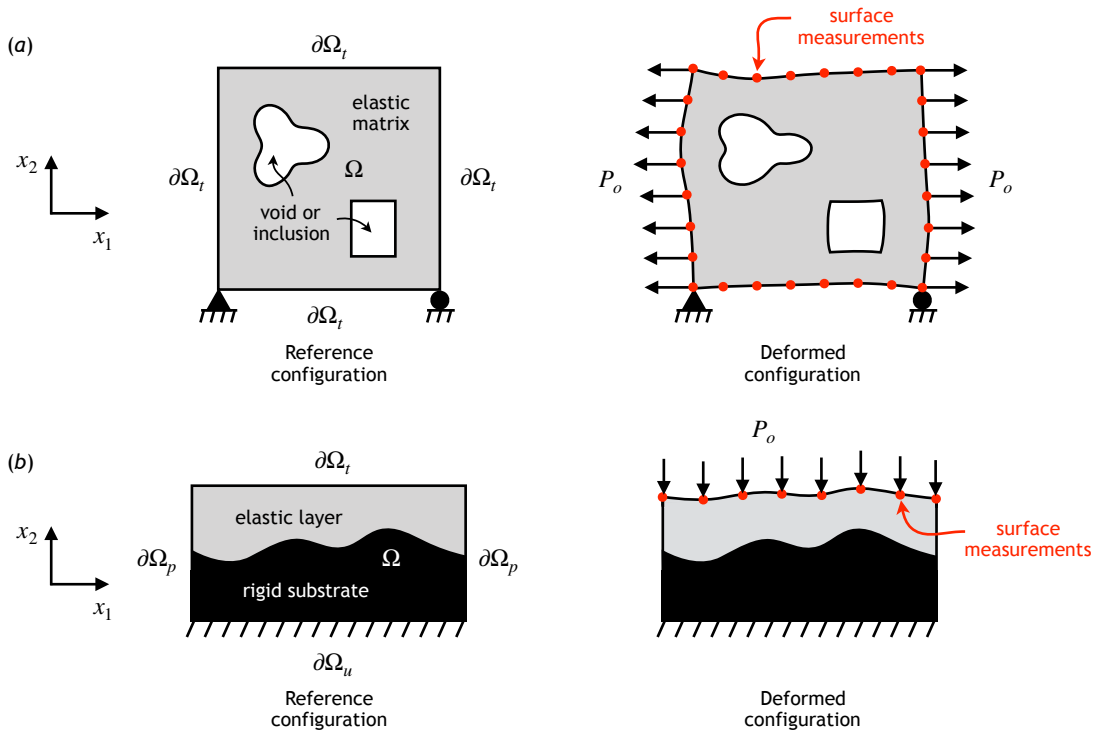

Figure 4-1: Setup of two geometry identification problems in elastic bodies under plain strain. (a) A square elastic matrix with hidden voids or rigid inclusions is pulled by a uniform traction on two opposite sides. The goal is to identify the number, locations and shapes of the voids or inclusions using measurements of the displacement along the outer boundary of the matrix. (b) An elastic layer on top of a hidden rigid substrate is compressed from the top by a uniform pressure. The goal is to identify the shape of the substrate using measurements of the displacement of the top surface.

etry identification and design problems in solid mechanics, fluid mechanics, optics, and beyond.

4.2 Methodology

4.2.1 Problem formulation

To illustrate our topology optimization framework, we consider two prototypical plane-strain elasticity problems pictured in Figure 4-1. In the first case, a square elastic matrix with hidden voids or rigid inclusions is pulled by a uniform traction P_o on two sides. The goal of the inverse problem is to identify the number, locations and shapes of the voids or inclusions using discrete measurements of the displacement of the outer boundary of the matrix. In the second case, an elastic layer on top of a hidden rigid substrate is compressed from the top by a uniform pressure P_o, with pe-

riodic lateral boundary conditions. The goal is to identify the shape of the substrate using discrete measurements of the displacement of the top surface. For both cases, we assume that we know the constitutive properties of the elastic material that is being pulled or compressed. We will consider two different types of constitutive laws: compressible linear elasticity, which characterizes the small deformation of any compressible elastic material, and incompressible nonlinear hyperelasticity, which models the large deformation of rubber-like materials. These two elasticity inverse problems are well suited for evaluating the accuracy of the topology optimization framework that we introduce in this paper, because of the existence of guarantees on the uniqueness of the solution. Indeed, Ref. [4] proves that in the case of a linear elastic material and a single void, there exists at most one cavity which yields the same surface displacements and stresses on an finite portion of the external boundary. Although this does not mean that the numerical problem is necessarily well-posed since small noise in the measurement data could cause large differences in the solution, this uniqueness result still makes these elasticity problems unique amongst inverse problems which in general do not have unique solutions [59].

In order to avoid any restriction or prior knowledge on the number or the shape of the hidden structures to be discovered, we parametrize the geometry through a discrete-valued material distribution function $\rho(\mathbf{x}) = \{0, 1\}$ that is defined at any point $\mathbf{x}$ in a global domain Ω comprising both the elastic material and the hidden voids or rigid inclusions. The material distribution is equal to 1 in the elastic solid phase and 0 in the void or rigid inclusion phase. The inverse problem can then be formulated as finding $\rho(\mathbf{x})$ for $\mathbf{x} \in \Omega$ so that the corresponding solution for the displacement fields throughout Ω, which is given by the physical governing equations of elasticity and the applied boundary conditions, matches the discrete displacement measurements at the surface. Denoting by $\mathbf{u}(\mathbf{x})$ and $\boldsymbol{\sigma}(\mathbf{x})$ the displacement and stress fields throughout Ω, these governing equations comprise the equilibrium relation

$$\nabla \cdot \boldsymbol{\sigma} = 0, \quad \mathbf{x} \in \Omega, \tag{4.1}$$

and the constitutive relation, which we write in general form as

$$F(\boldsymbol{\sigma}, \nabla \mathbf{u}, \rho) = 0, \quad \mathbf{x} \in \Omega. \tag{4.2}$$

Splitting the outer boundary $\partial\Omega$ of the domain into a portion $\partial\Omega_t$ with prescribed traction $\bar{\mathbf{t}}(\mathbf{x})$ and a portion $\partial\Omega_u$ with prescribed displacement $\bar{\mathbf{u}}(\mathbf{x})$, the applied boundary conditions will take the form

$$\mathbf{u}(\mathbf{x}) = \bar{\mathbf{u}}(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_u, \tag{4.3a}$$

$$\boldsymbol{\sigma}(\mathbf{x})\mathbf{n}(\mathbf{x}) = \bar{\mathbf{t}}(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_t, \tag{4.3b}$$

where $\mathbf{n}$ denotes the outward unit normal vector. In the case of the elastic layer, the outer boundary also comprises a portion $\partial\Omega_p$ with periodic boundary conditions on the displacement and traction. Finally, assuming that we measure surface displacements $\mathbf{u}_i^m$ at discrete locations $\mathbf{x}_i$ forming a subset $\partial\Omega^m$ of $\partial\Omega_t$, the requirement that

the predictions for $\mathbf{u}$ match the measurement data is expressed as

$$\mathbf{u}(\mathbf{x}_i) = \mathbf{u}_i^m, \quad \mathbf{x}_i \in \partial\Omega^m. \tag{4.4}$$

Appendix C.1 provides a detailed formulation of the two inverse problems that we solve, including the material-specific constitutive relations as well as the detailed boundary conditions applied in both problems.

Before proceeding further, we nondimensionalize all physical quantities and hereafter express them in dimensionless form. Specifically, we nondimensionalize tractions and stresses with the magnitude P_o of the external applied traction, and we nondimensionalize displacements with the characteristic length scale $P_o L/E$, where L is the width of the elastic matrix or elastic layer, and E is the Young's modulus of the elastic material (in the hyperelastic case, we use the equivalent Young's modulus $E = 3\nu$, where ν is the shear modulus of the hyperelastic material). Although rarely employed in the PINN community, such nondimensionalization is critical to enabling neural networks to consistently handle elasticity problems across a wide range of material moduli and applied loads. These neural networks will be defined in the next section, along with our general framework for solving the inverse problem defined above.

Since the discrete nature of the optimization problem makes the solution very challenging, we relax the binary constraint on the material distribution and allow for intermediate values of ρ between 0 and 1, which is a standard approach in topology optimization [164]. Such relaxation transforms the problem into a continuous one amenable to gradient-based optimization, but requires special treatment to ensure that the optimized material distribution converges to 0 and 1 values instead of settling on unphysical intermediate values. A common strategy in topology optimization is to penalize implicitly these intermediate values through a suitably chosen interpolation function between material properties and material density [14, 15, 23]; however, this method only works in the presence of a volume constraint (for instance, maximum allowable fraction of solid). The volume constraint is not present in the geometry identification problem, which calls for another type of regularization. Ref. [128], who also parametrized the geometry through a material distribution field, proposed to employ a total variation diminishing regularization that penalizes the gradient norm $|\nabla\rho|$ throughout the domain. When combined with PINNs, we have found this approach to produce solutions with sharp material density transitions but containing large regions of intermediate ρ values between 0 and 1. Instead, we will introduce a novel eikonal regularization inspired from level-set methods and signed distance functions to regularize the material distribution. We first present the general PINN-based topology optimization framework in Section 4.2.2, before describing the novel regularization approach in Section 4.2.3.

4.2.2 General framework

The topology optimization framework based on PINNs that we propose for solving geometry identification problems in mechanics is illustrated in Figure 1. We present the

methodology in a general setting in the next two sections, and describe in Appendix C.2 its particular application to the two problems formulated in the introduction. At the root of the framework are several deep neural networks that approximate all n physical quantities describing the problem, which we lump into a vector field $\boldsymbol{\psi}(\mathbf{x})$, and the material distribution $\rho(\mathbf{x})$. For the physical quantities, each neural network maps the spatial location $\mathbf{x} = (x_1, x_2)$ to one of the variables in $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_n)$, where the ψ_i's consist of the different components of displacement and stress. These maps can be expressed as $\psi_i(\mathbf{x}) = \bar{\psi}_i(\mathbf{x}; \boldsymbol{\theta}_i)$, where $\bar{\psi}_i$ is the map defined by the ith neural network and its associated set of trainable parameters $\boldsymbol{\theta}_i$ (see Appendix C.3.1 for details). For the material distribution, we first define a neural network with trainable parameters $\boldsymbol{\theta}_\phi$ that maps $\mathbf{x}$ to a scalar variable $\phi(\mathbf{x}) = \bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi)$. A sigmoid function is then applied to $\phi(\mathbf{x})$ to yield $\rho(\mathbf{x}) = \text{sigmoid}(\phi(\mathbf{x})/\delta) = \text{sigmoid}(\bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi)/\delta)$, which we simply write as $\rho(\mathbf{x}) = \bar{\rho}(\mathbf{x}, \boldsymbol{\theta}_\phi)$. This construction ensures that the material distribution $\rho(\mathbf{x})$ remains between 0 and 1, and δ is a transition length scale that we will comment on later.

The goal is then to find the parameters $\boldsymbol{\theta}_{\boldsymbol{\psi}} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n\}$ and $\boldsymbol{\theta}_\phi$ so that the neural network approximations for $\boldsymbol{\psi}(\mathbf{x})$ and $\rho(\mathbf{x})$ satisfy the governing equations of elasticity (4.1)-(4.2) and the applied boundary conditions (4.3), while matching the discrete measurements of displacement at the surface (4.4). This is achieved by constructing a loss function of the form

$$\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_\phi) = \lambda_{\text{meas}} \mathcal{L}_{\text{meas}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}) + \lambda_{\text{gov}} \mathcal{L}_{\text{gov}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_\phi) + \lambda_{\text{eik}} \mathcal{L}_{\text{eik}}(\boldsymbol{\theta}_\phi), \tag{4.5}$$

where $\mathcal{L}_{\text{meas}}$ and $\mathcal{L}_{\text{gov}}$ measure the degree to which the neural network approximations do not satisfy the discrete measurements and governing equations, respectively, $\mathcal{L}_{\text{eik}}$ is a crucial regularization term that drives ρ towards 0 or 1 values and that we will explain below, and the λ's are scalar weights. Assume that we have a set of measurements $(\psi_j)_i^m$ of the physical quantity ψ_j at discrete locations $\mathbf{x}_i \in \partial\Omega_j^m$ for $j = 1, \ldots, n$, where $\partial\Omega_j^m$ refers to the set of measurement locations of ψ_j. The measurement loss then takes the form

$$\mathcal{L}_{\text{meas}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}) = \sum_{j=1}^{n} \frac{1}{|\partial\Omega_j^m|} \sum_{\mathbf{x}_i \in \partial\Omega_j^m} |\bar{\psi}_j(\mathbf{x}_i; \boldsymbol{\theta}_j) - (\psi_j)_i^m|^2, \tag{4.6}$$

where $|\partial\Omega_j^m|$ denotes the size of the set $\partial\Omega_j^m$. Expressing the governing equations of elasticity in the residual form $\mathbf{r}_j(\boldsymbol{\psi}(\mathbf{x}), \rho(\mathbf{x})) = 0$ for $j = 1, \ldots, n_{\text{gov}}$, where n_{gov} is the number of governing equations, the corresponding loss takes the form

$$\mathcal{L}_{\text{gov}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_\phi) = \sum_{j=1}^{n_{\text{gov}}} \frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\mathbf{r}_j(\bar{\boldsymbol{\psi}}(\mathbf{x}_i; \boldsymbol{\theta}_{\boldsymbol{\psi}}), \bar{\rho}(\mathbf{x}_i; \boldsymbol{\theta}_\phi))|^2, \tag{4.7}$$

where Ω^d is a set of collocation points in Ω. The two sets Ω^d and $\partial\Omega^m = \partial\Omega_1^m \cup \cdots \cup \partial\Omega_n^m$ are illustrated in Figures 4-3(a) and 4-3(b) for the case of a square matrix with voids. We use automatic differentiation to calculate in a mesh-free fashion the spatial derivatives contained in the residuals $\mathbf{r}_j$ of the governing equations. Further, we design

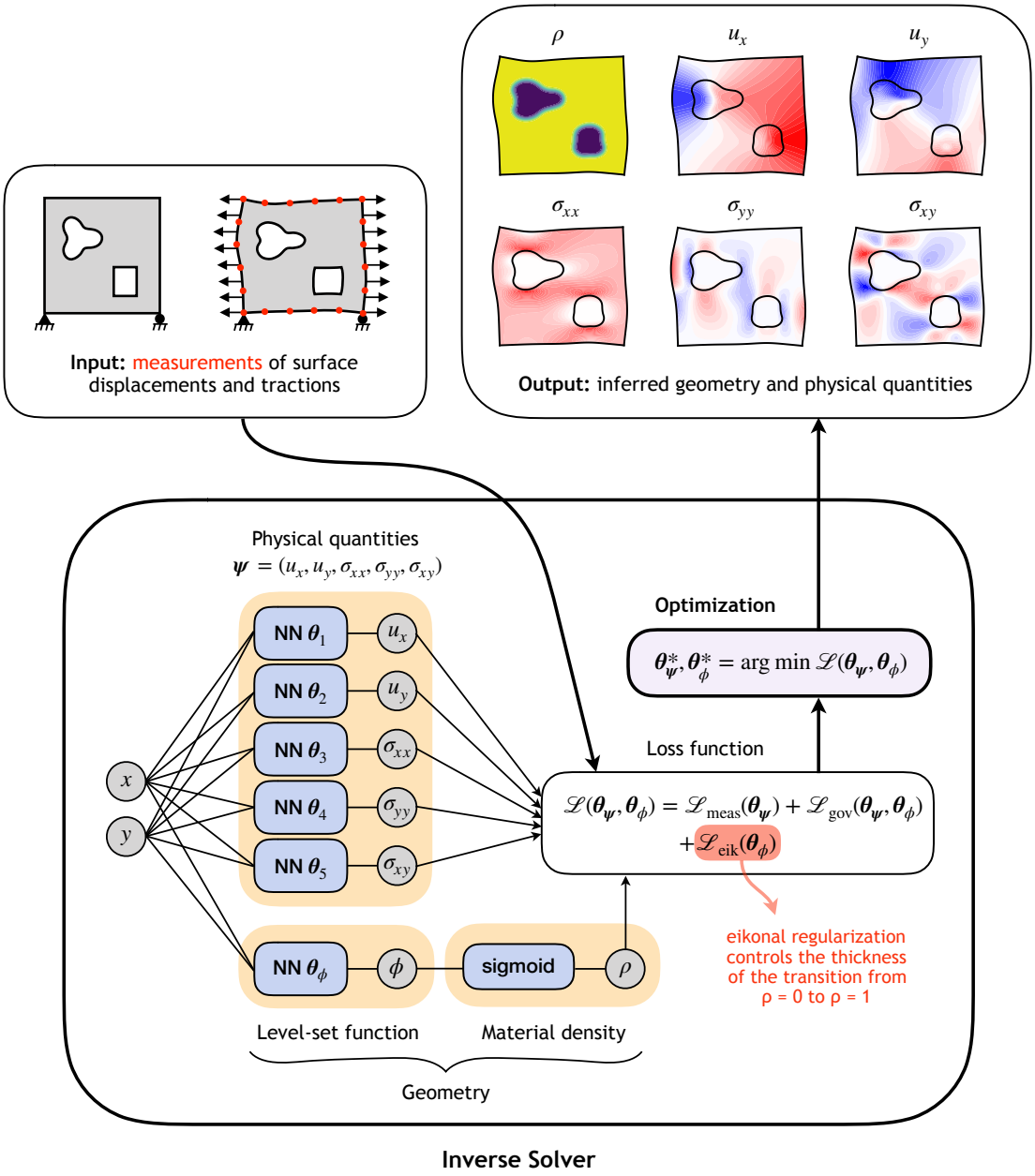

Figure 4-2: (Caption on the next page.)

Figure 4-2: Topology optimization framework for solving geometry identification inverse problems in mechanics. The geometry of the system, which is initially unknown, is parametrized by a material distribution field given through a level-set function and equal to 1 in the elastic solid and 0 in the voids or rigid inclusions. The level-set function and the physical quantities describing the problem are approximated with deep neural networks designed to inherently satisfy the applied boundary conditions. These neural networks are then trained to minimize a loss function that drives the material distribution and physical quantities towards satisfying the governing equation of elasticity while matching the discrete measurements of surface displacements and tractions. Crucially, an eikonal regularization term in the loss function ensures that the material distribution transitions between 0 and 1 over a prescribed length scale and avoids settling on intermediate values. At the end of the optimization process, the converged material distribution reveals the location and shapes of the hidden structures.

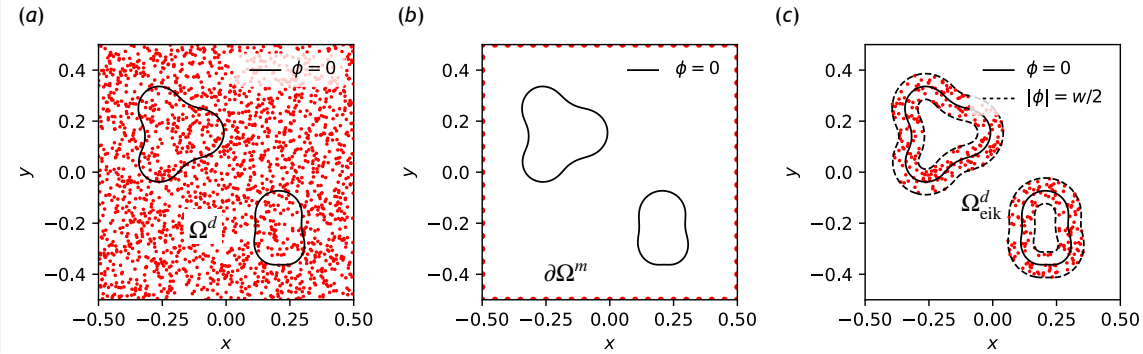

Figure 4-3: Example of the distribution of (a) collocation points in Ω^d, (b) measurement points in $\partial\Omega^m$, and (c) collocation points in Ω^d_{eik}, for the specific case of a square matrix containing two voids delimited by the zero contour level of ϕ.

the architecture of our neural networks in such a way that they inherently satisfy the boundary conditions. This approach, referred to in the literature as hard constraint as opposed to the soft constraint approach of enforcing the boundary conditions through the loss function [120, 121], is described in Appendix C.2.

Finally, the optimal parameters $\boldsymbol{\theta}^*_{\boldsymbol{\psi}}$ and $\boldsymbol{\theta}^*_{\phi}$ that solve the problem can be obtained by training the neural networks to minimize the loss (4.5) using gradient-based optimization:

$$\boldsymbol{\theta}^*_{\boldsymbol{\psi}}, \boldsymbol{\theta}^*_{\phi} = \arg\min_{\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_{\phi}} \mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_{\phi}). \tag{4.8}$$

The corresponding physical quantities $\boldsymbol{\psi}^*(\mathbf{x}) = \bar{\boldsymbol{\psi}}(\mathbf{x}; \boldsymbol{\theta}^*_{\boldsymbol{\psi}})$ and material distribution $\rho^*(\mathbf{x}) = \bar{\rho}(\mathbf{x}; \boldsymbol{\theta}^*_{\phi})$ will match the discrete measurements of displacement and traction at the surface while satisfying the governing equations of elasticity and applied boundary conditions. The number, locations and shapes of the hidden voids or rigid inclusions

is then directly obtained from the profile of the material distribution $\rho^*(\mathbf{x})$ or from the zero contour levels of the level-set function $\phi^*(\mathbf{x})$.

4.2.3 Material density regularization

We now describe the key ingredient that ensures the success of our framework. As mentioned above, the main challenge is to promote the material distribution $\rho(\mathbf{x})$ to converge towards 0 or 1 rather than values in between, which do not have physical meaning. Further, we desire the length scale of the transition region from 0 to 1 to be uniform everywhere, so that the inner boundaries defined by the distribution $\rho(\mathbf{x})$ lead to the same local physical behavior at the interface (think for instance of the vanishing fluid velocity at a solid interface).

To illustrate the problem, we show in the top row of Figure 4-4 a random initialization of the neural network $\phi(\mathbf{x}) = \bar{\phi}(\mathbf{x}, \boldsymbol{\theta}_\phi)$, the magnitude of its gradient $|\nabla\phi(\mathbf{x})|$ as well as the corresponding material distribution $\rho(\mathbf{x}) = \text{sigmoid}(\phi(\mathbf{x})/\delta)$ with $\delta = 0.01$. Thanks to the sigmoid transformation, ρ is bounded between 0 and 1, with dark blue regions ($\rho = 0$) corresponding to one phase and yellow regions ($\rho = 1$) corresponding to another. We define the transition between the two phases to be at $\rho = 0.5$, or equivalently $\phi = 0$. The zero level-set of ϕ, shown in black lines in Figure 4-4, therefore implicitly defines the interface between the two material phases. Since this is identical to the way shapes are parametrized in level-set methods [142, 141], we hereafter refer to ϕ as a level-set function. However, one observes in Figure 4-4(c) that the thickness of the transition region where ρ goes from 0 to 1 is not uniform everywhere, resulting in large zones where ρ assumes unphysical values between 0 and 1. Figure 4-4(b) reveals that this issue is caused by the nonuniformity of the gradient norm $|\nabla\phi|$ along the interface, with small and large values of $|\nabla\phi|$ leading to wide and narrow transition regions, respectively.

We propose to solve this issue by forcing the gradient norm $|\nabla\phi|$ to be unity in a narrow band of width w along the interface defined by $\phi = 0$. In this way, ϕ becomes a signed distance function to the interface in the narrow band, thereby constraining the gradient of ρ to be constant along the interface. Further, we choose $w = 10\delta$ so that $\rho = \text{sigmoid}(\pm w/2\delta) = \text{sigmoid}(\pm 5) \simeq 0$ or 1 along the edge of the narrow band. This ensures that the transition region where ρ goes from 0 to 1 will be entirely contained within the narrow band and thus have uniform thickness everywhere. We illustrate this idea in the second row of Figure 4-4, where we have forced the previous neural network $\phi(\mathbf{x}) = \bar{\phi}(\mathbf{x}, \boldsymbol{\theta}_\phi)$ to behave such that $|\nabla\phi| = 1$ in a narrow band of width $w = 10\delta = 0.1$, while keeping the same interface, defined by $\phi = 0$, as in the top row of Figure 4-4. The edges of the narrow band are displayed in dotted lines in Figure 4-4(e). Clearly, ϕ now behaves like a signed-distance function in the narrow band, which leads to a uniform interface transition thickness everywhere as observed in Figure 4-4(f).

In practice, we implement this regularization into our framework by adding the

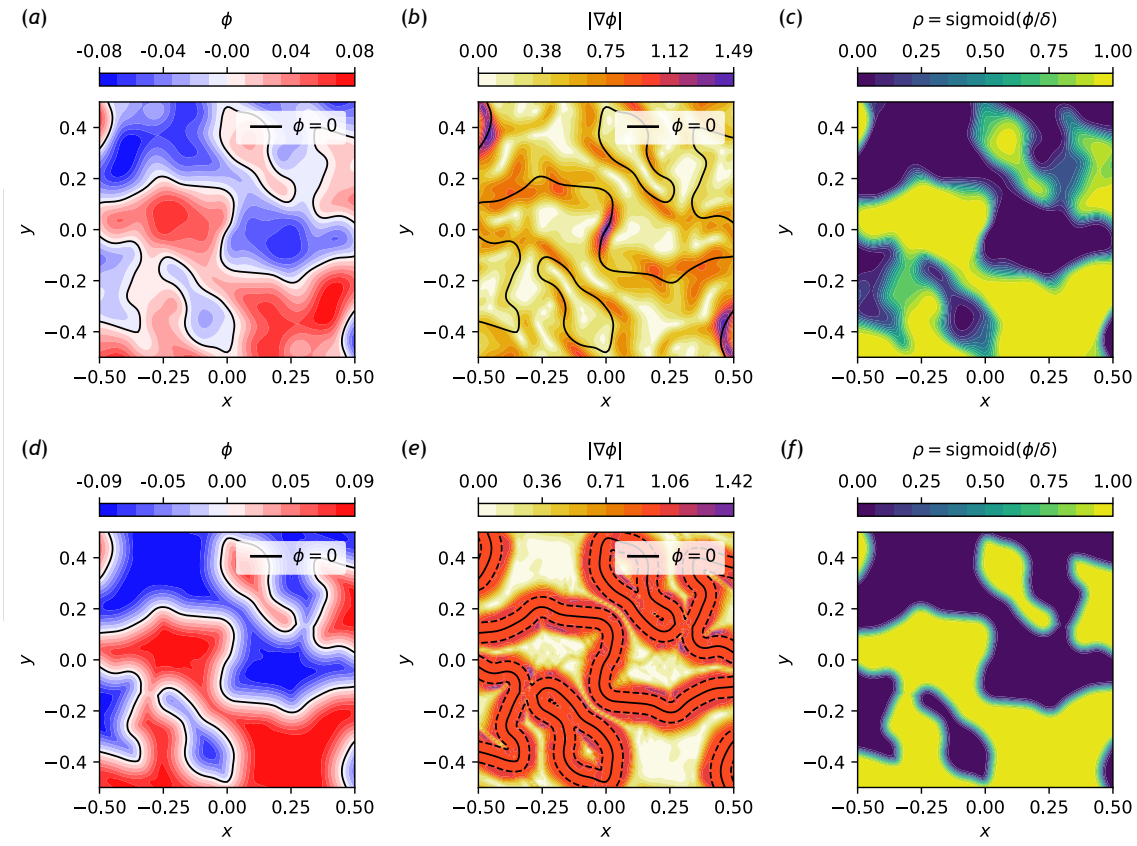

Figure 4-4: Eikonal regularization of the material distribution. (a, d) Level-set function ϕ, (b, e) magnitude of its gradient $|\nabla\phi|$, and (c, f) material distribution $\rho = \text{sigmoid}(\phi/\delta)$ defined by $(a\text{-}c)$ a random initialization of the neural network $\phi = N_\phi(\mathbf{x}, \boldsymbol{\theta}_\phi)$, and $(d\text{-}f)$ the same network trained to minimize the eikonal loss $\mathcal{L}_{\text{eik}}(\boldsymbol{\theta}_\phi)$.

loss term $\mathcal{L}_{\text{eik}}$ in (4.5), which takes the form

$$\mathcal{L}_{\text{eik}}(\boldsymbol{\theta}_\phi) = \frac{1}{|\Omega^d_{\text{eik}}|} \sum_{\mathbf{x}_i \in \Omega^d_{\text{eik}}} \left(|\nabla\phi(\mathbf{x}_i)| - 1\right)^2, \tag{4.9}$$

where $\Omega^d_{\text{eik}} = \{\mathbf{x}_i \in \Omega^d : |\phi(\mathbf{x}_i)| < w/2\}$. The aim of this term is to penalize deviations away from the constraint $|\nabla\phi| = 1$ in a narrow band of width w along the interface defined by the zero level-set $\phi(\mathbf{x}) = 0$. Because finding the true narrow band of width w at every step of the training process would be too expensive, we relax the domain over which the constraint $|\nabla\phi| = 1$ is active by evaluating the eikonal loss $\mathcal{L}_{\text{eik}}$ on the subset of points $\mathbf{x}_i$ in Ω^d that satisfy $|\phi(\mathbf{x}_i)| < w/2$. As the constraint $|\nabla\phi| = 1$ is better and better satisfied, this set of collocation points will by construction overlap with the true narrow band of width w along the zero level-set of ϕ, which is illustrated

in 4-3(c) for the case of a square matrix with two voids.

Since the constraint $|\nabla\phi| = 1$ takes the form of an eikonal equation in the narrow band, we call this approach eikonal regularization, with the subscript in $\mathcal{L}_{\text{eik}}$ referring to eikonal. In contrast to standard eikonal equations that are posed as boundary value problems, however, our regularization does not force ϕ to vanish on a specified boundary. Rather, ϕ is allowed to evolve freely during the optimization in such a way that the corresponding material distribution $\rho = \text{sigmoid}(\phi/\delta)$ and physical quantities $\boldsymbol{\psi}$ minimize the total loss (4.5).

4.3 Results

We construct a number of test cases by varying the number and shapes of the hidden voids or rigid inclusions to be discovered in the matrix problem shown in Figure 4-1(a), and by varying the shape of the rigid substrate to be discovered in the layer problem shown in Figure 4-1(b). We consider both linear elastic and nonlinear hyperelastic materials for the matrix problem, while we only consider linear elastic materials for the layer problem. As a substitute for real experiments, we use the finite element software Abaqus to compute the deformed shape of the elastic structure and generate the measurement data for each case. Using this measurement data, we run our topology optimization framework to discover the number, locations and shapes of the hidden voids or rigid inclusions, which we then compare with the ground truth to assess the efficacy of our framework. The architecture of the neural networks, the training procedure and parameter values are described in detail in Appendix C.3, while the setup of the Abaqus simulations is presented in Appendix C.4.

4.3.1 Elastic matrix

We first treat the matrix problem shown in Figure 4-1(a) in the small-deformation regime. Accordingly, we consider that the matrix consists of a linear elastic material with nondimensional Young's modulus $E = 1$ and Poisson's ratio $\nu = 0.3$. The measurement data is acquired in Abaqus by applying a nondimensional traction $P_o = 0.05$ on the left edge of the matrix.

We first consider a matrix containing two circle-shaped voids and show in Figure 4-5 the corresponding results obtained with our topology optimization framework. In (a-c), we display the evolution of the material distribution ρ during the optimization and its final state ρ^*. The evolution of the various loss components is shown in (d), while the final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$ are shown in (e,f). Finally, (g-i) show the final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2. Although the solver starts from a single void, it recognizes that the solution consists of two separate structures and splits the initial single void into two around epoch 9k. The two final holes obtained at epoch 150k match closely the true locations and shapes of the voids that were used to generate the measurement data in Abaqus. This demonstrates that our framework discovers not merely the shape but also the topology

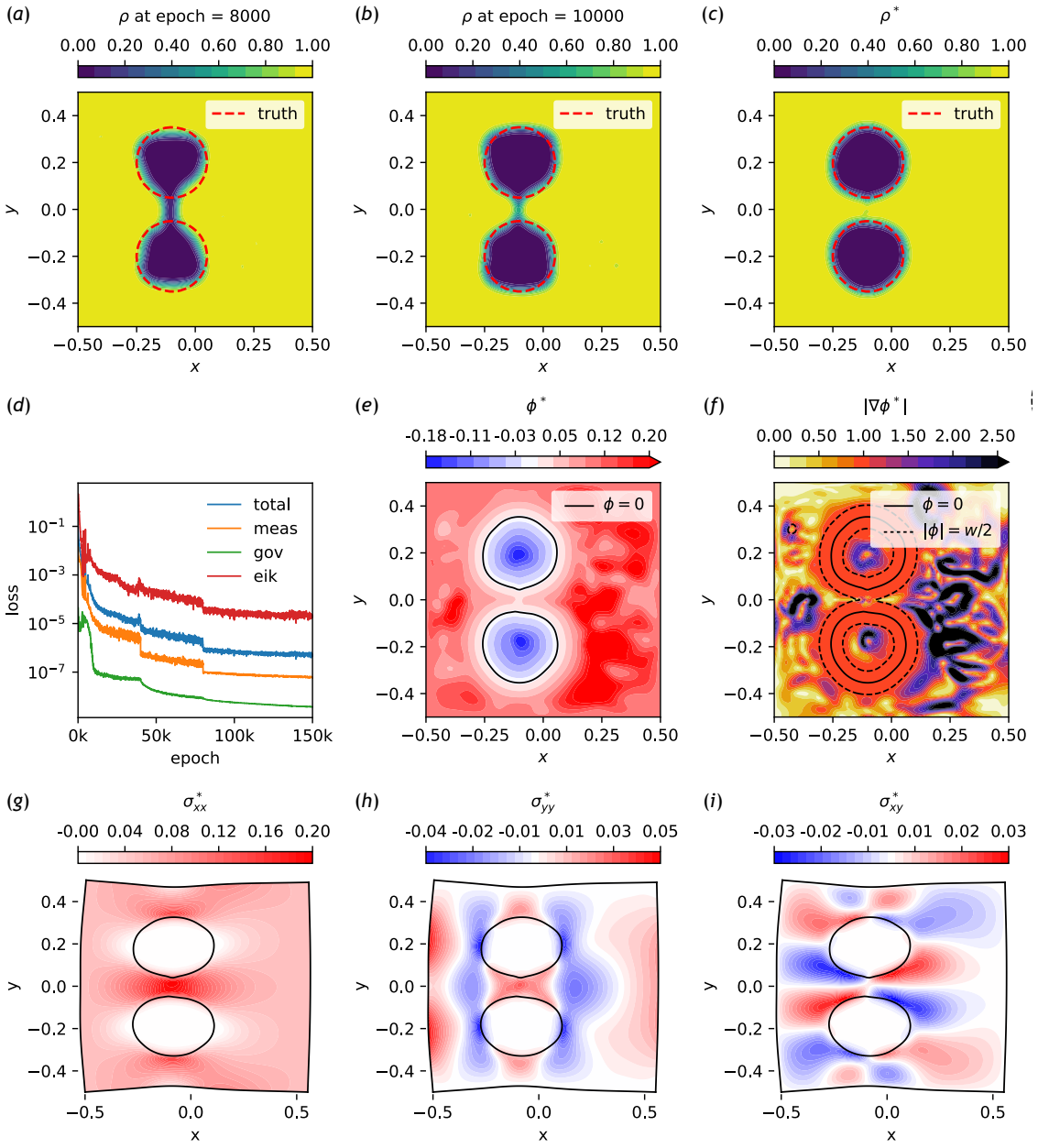

Figure 4-5: Linear elastic matrix with two circle-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

of the hidden geometry, and is therefore a true topology optimization framework. Further, we notice that by epoch 7k, the thickness of the transition region from $\rho = 0$ to $\rho = 1$ is uniform all along the interface, thanks to the Eikonal regularization that constrains the level-set function ϕ to have a unit slope in a band of width w along the interface, as can be seen in (f). We also consider a matrix containing four circle-shaped voids, with the corresponding results shown in Figure 4-6. Here too, the solver recognizes that the solution consists of four separate structures, and the final holes obtained at epoch 150k match closely the true locations and shapes of the voids.

To investigate the ability of our framework to find more complicated shapes than circles, we then consider a matrix with one rectangle-shaped and one star-shaped void. The corresponding results obtained from our topology optimization framework are presented in Figure 4-7. As in the previous case, the framework is able to detect the presence of two structures and has already split the initial single void in two by epoch 3k. Subsequent iterations allow the solver to discover the intricacies of each shape, and the final structures obtained at epoch 150k closely match the true void locations and shapes, including the aspect ratio of the rectangle and, impressively, the concavity of the star. Next, we consider a matrix containing a slit-shaped void, with the corresponding results shown in Figure 4-8. The slit is designed to resemble the location and aspect ratio of that considered in a similar setup in Ref. [197]. In their case, the geometry of the slit was parametrized by the x and y coordinates of the circles at both ends and its width known in advance. Still, finding these four scalar parameters required measurement data from 10 points located inside the matrix. On the other hand, our framework only relies on measurement data from the surface and does not know that it should be looking for a slit or even a single structure. Even then, it is able to identify the slit with remarkable accuracy. Next, we consider a matrix with a U-shaped void, with the corresponding results shown in Figure 4-9. The shape found by our framework lacks the middle part of the U. Since the middle part undergoes very little deformation and stress, its presence negligibly affects the measurements of displacement and traction at the surface, which highlights a limitation of our framework. Finally, we consider a matrix with a T-shaped void, with the corresponding results shown in Figure 4-10. The shape of the T is globally well-identified, with the exception of the two concave corners that undergo very little deformation and stress, similar to the middle part of the U in the previous example.

We then consider a matrix containing a circle-shaped rigid inclusion. The corresponding results obtained from our topology optimization framework are presented in Figure 4-11. The rigid inclusion leads to a strengthening of the matrix and, accordingly, produces a less noticeable signature on the measurements of outer surface displacement. Even then, the framework is able to identify the location and shape of the inclusion. We also consider a matrix with a square-shaped rigid inclusion, with the corresponding results shown in Figure 4-12. In this case too, we are able to find the location and approximate shape of the inclusion.

Finally, we consider the same matrix problem but in the large-deformation regime. This time, we consider that the matrix consists of a nonlinear hyperelastic material with nondimensional shear modulus $\mu = 1/2.6$, and the measurement data is acquired in Abaqus by applying a nondimensional traction $P_o = 0.2$ on the left edge of the

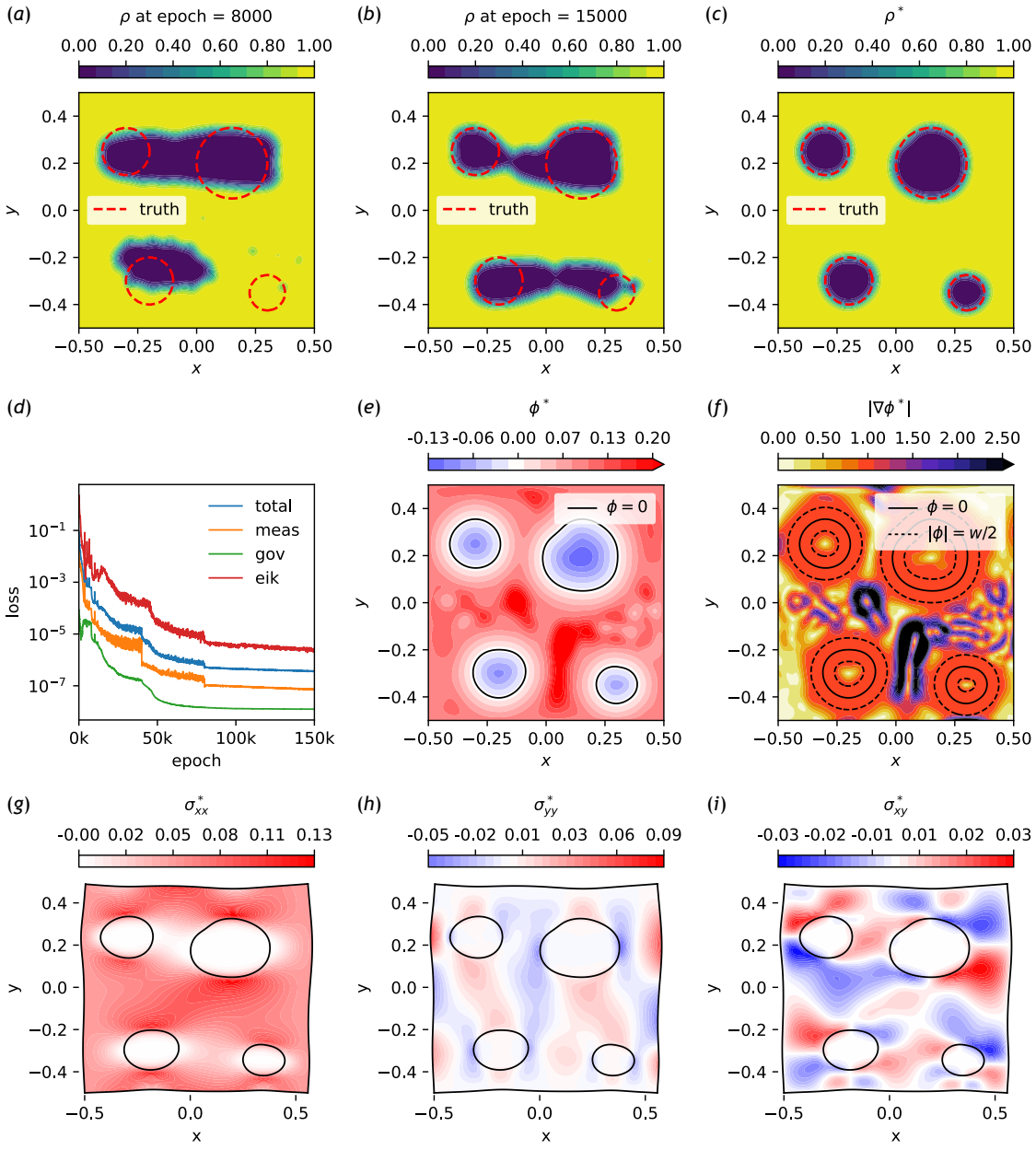

Figure 4-6: Linear elastic matrix with four circle-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

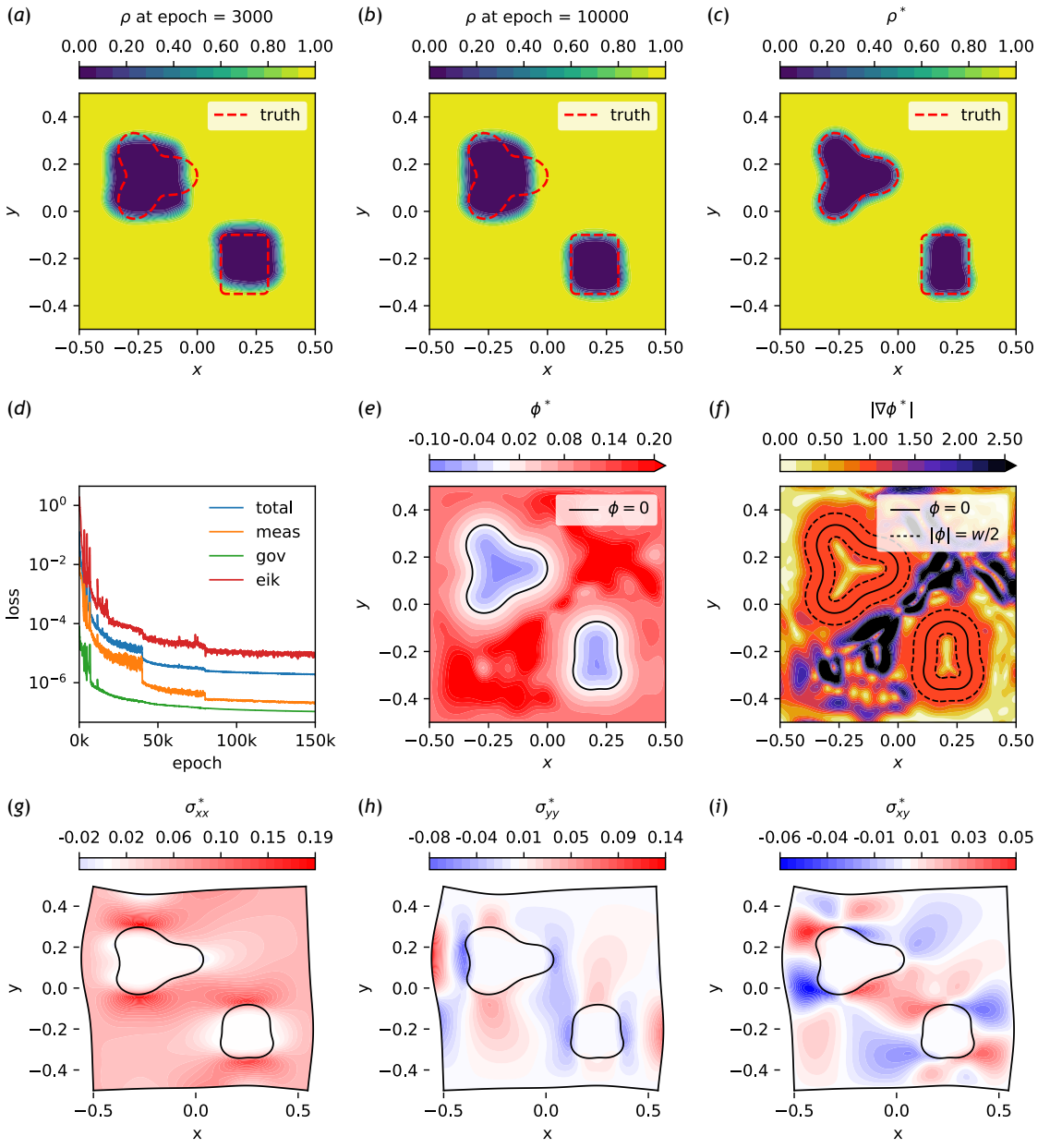

Figure 4-7: Linear elastic matrix with one rectangle-shaped and one star-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

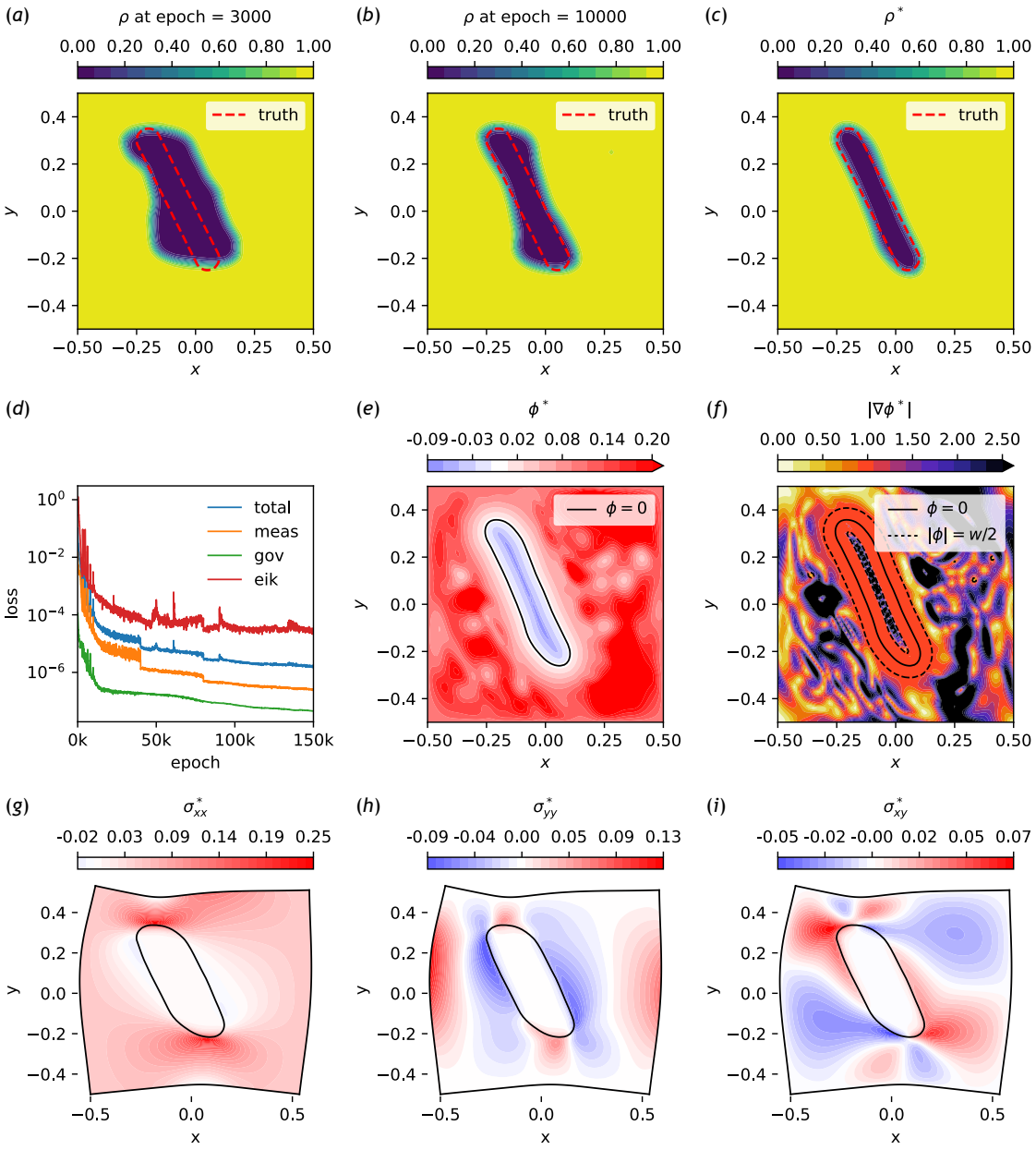

Figure 4-8: Linear elastic matrix with one slit-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

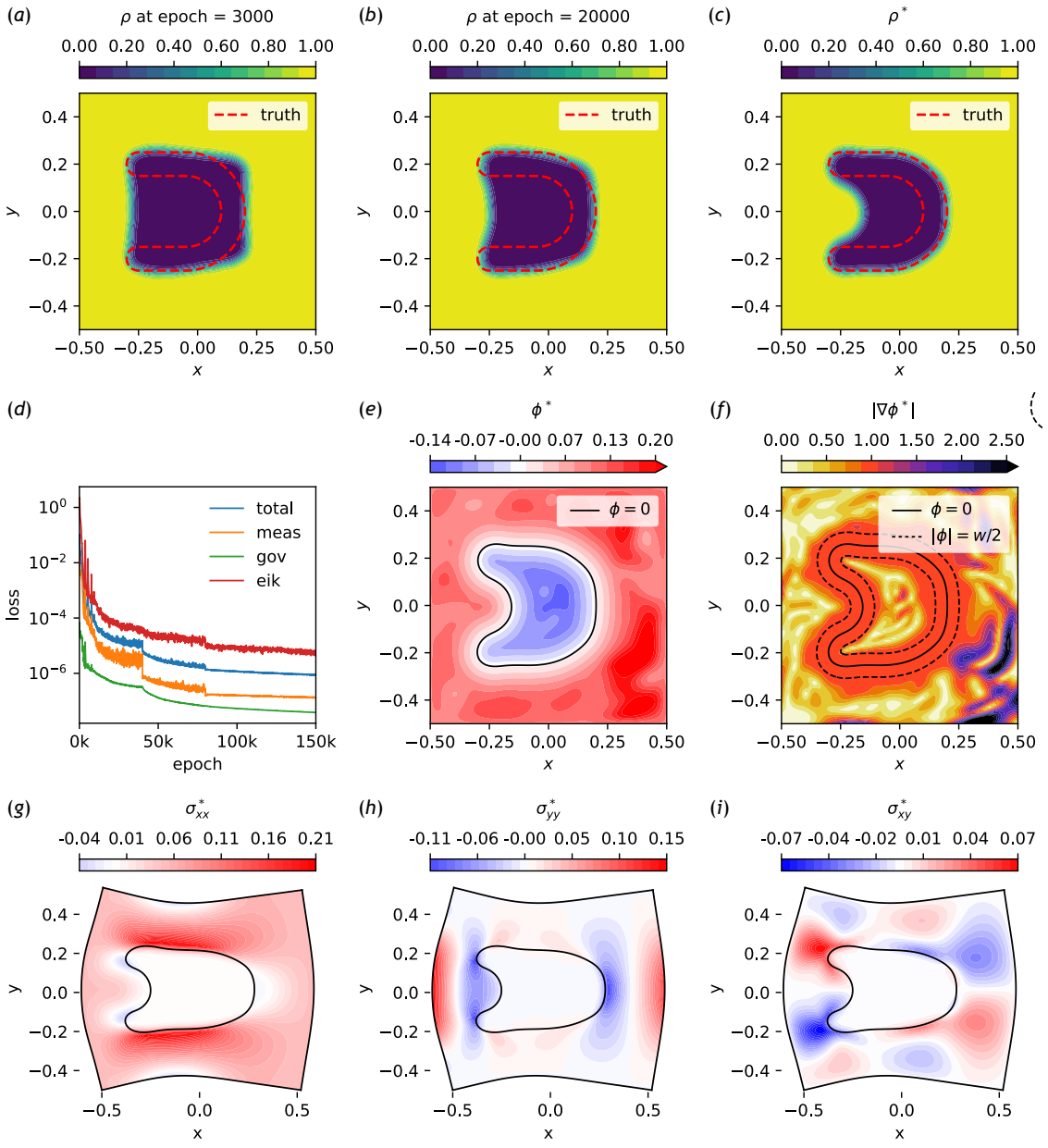

Figure 4-9: Linear elastic matrix with one U-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

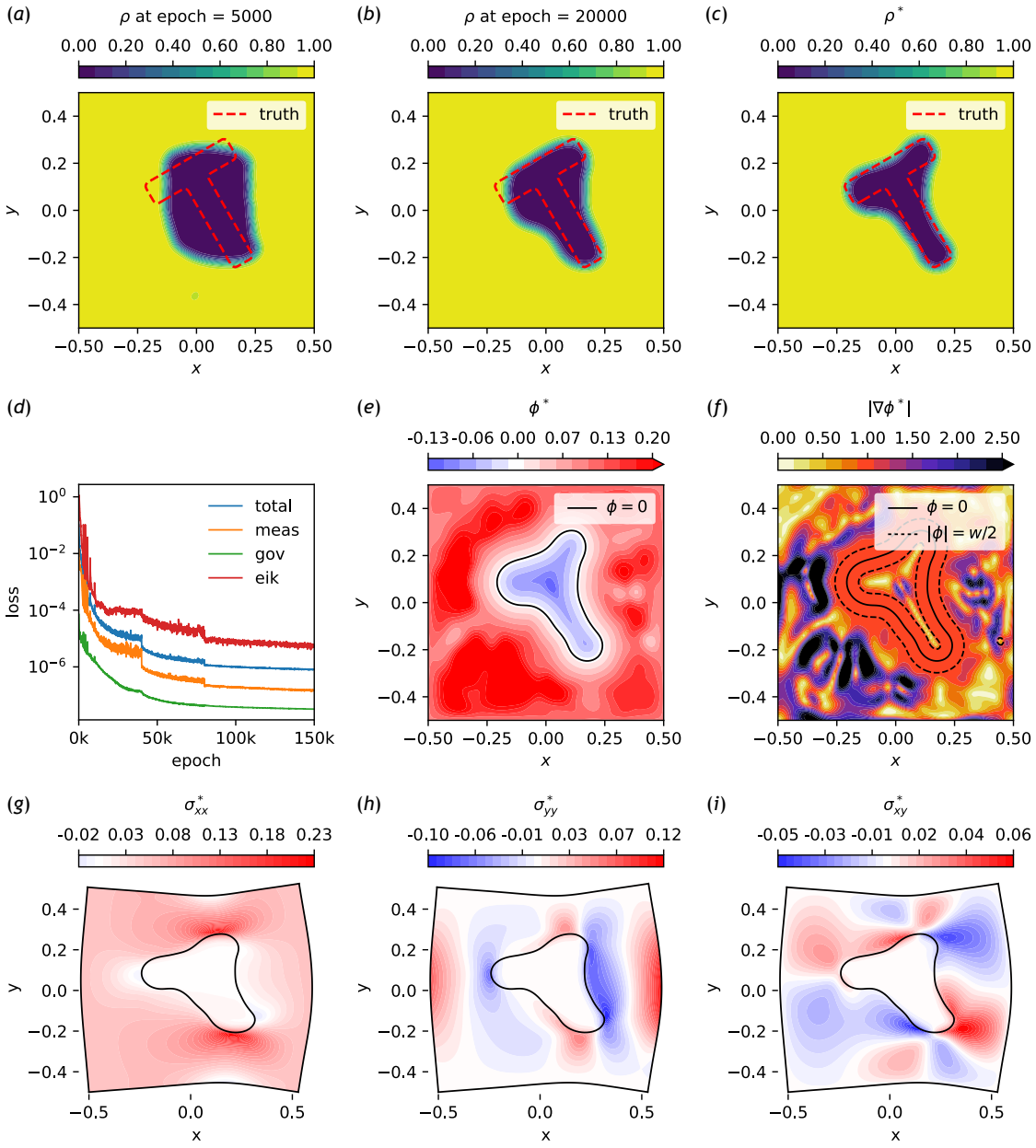

Figure 4-10: Linear elastic matrix with one T-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

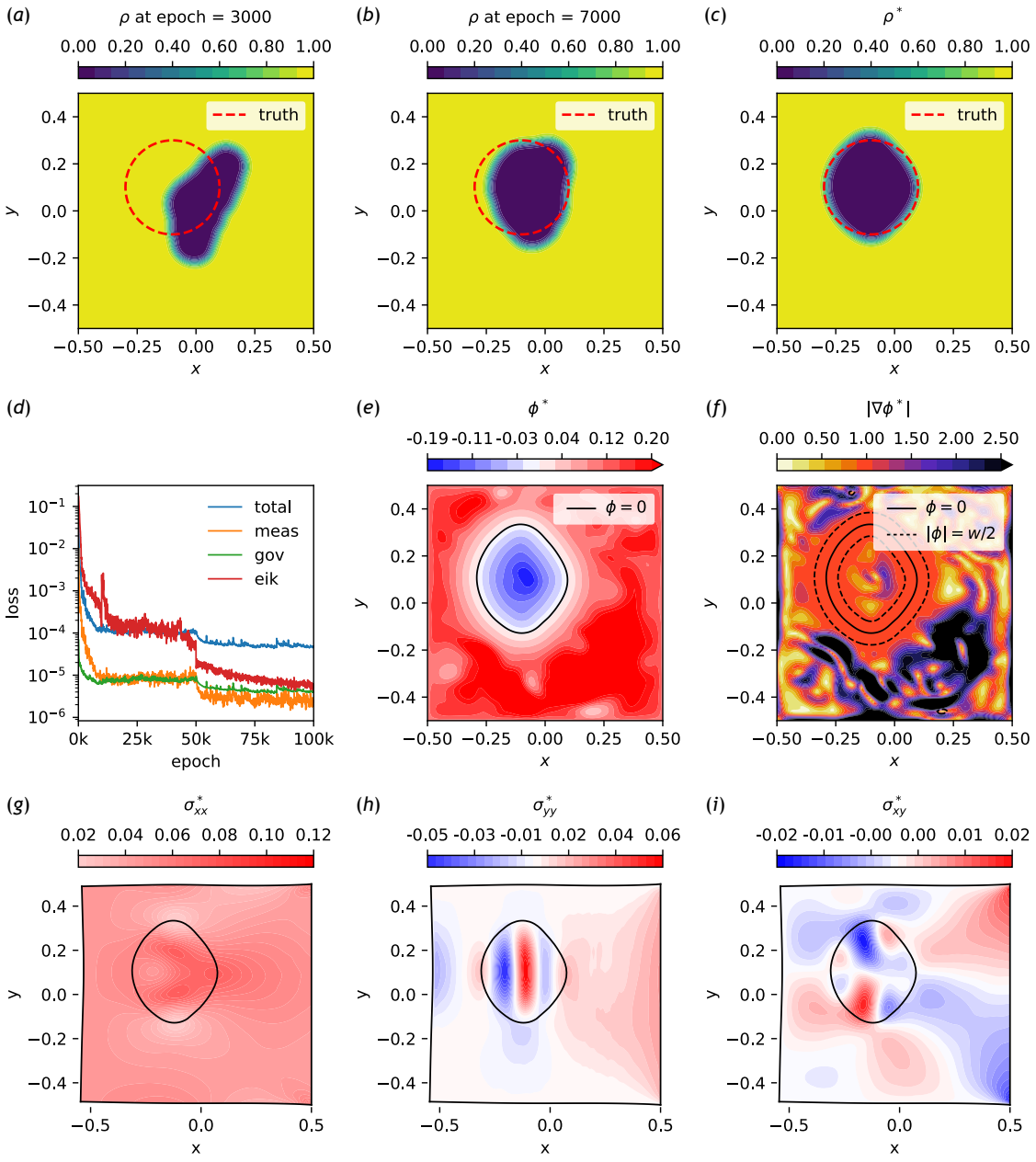

Figure 4-11: Linear elastic matrix with one circle-shaped rigid inclusion. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

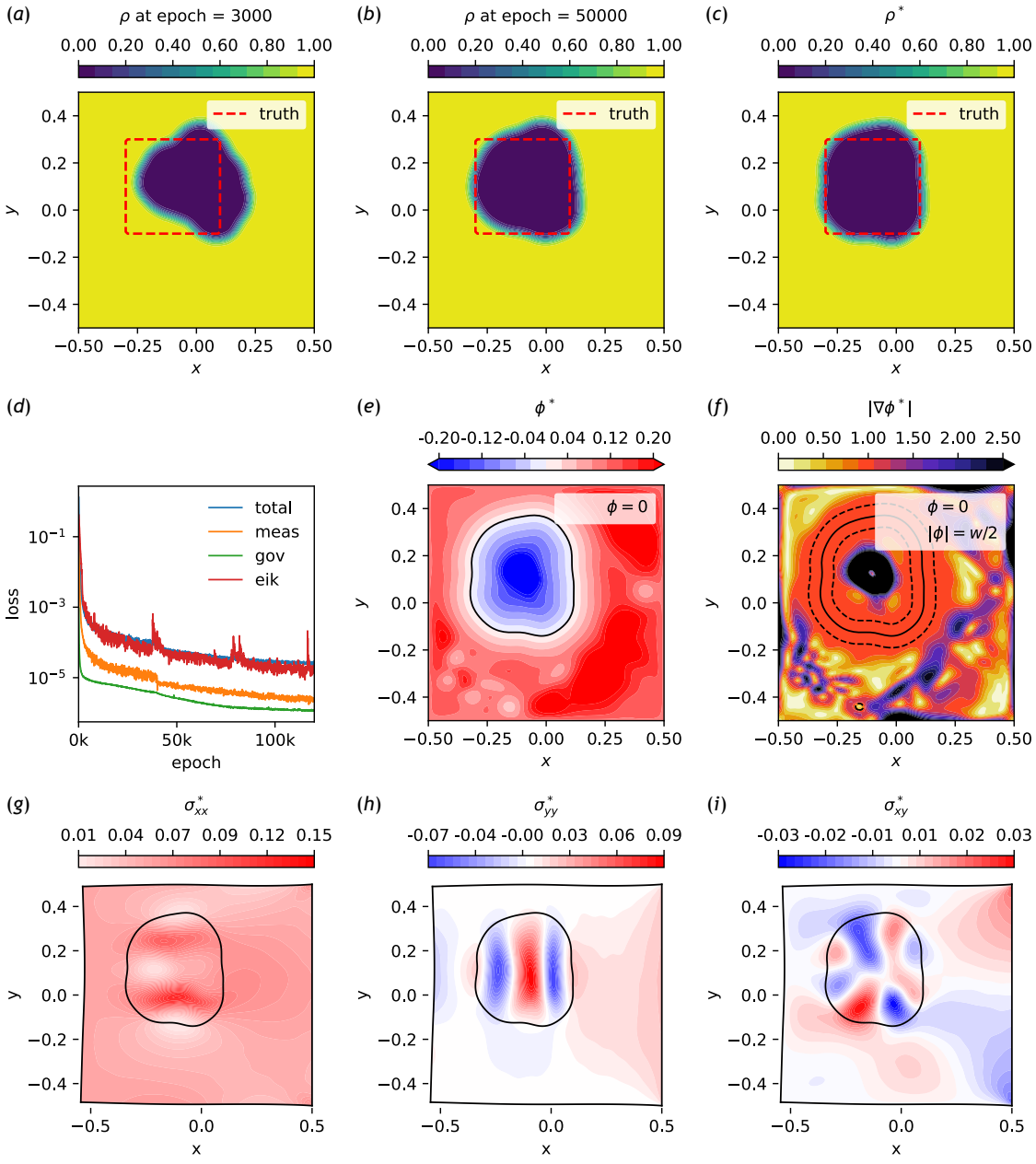

Figure 4-12: Linear elastic matrix with one square-shaped rigid inclusion. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

matrix. We investigate the same six matrix geometries containing voids as before, and we display in Figures 4-13, 4-14, 4-15, 4-16, 4-17, and 4-18 the corresponding results. All the geometries are identified equally well or almost as well as in the linear elastic case, which illustrates the ability of the framework to cope with nonlinear constitutive models.

4.3.2 Elastic layer

We now turn to the elastic layer problem shown in Figure 4-1(b), which we treat in the small-deformation regime. Accordingly, we consider that the matrix consists of a linear elastic material with nondimensional Young's modulus $E = 1$ and Poisson's ratio $\nu = 0.3$. The measurement data is acquired in Abaqus by applying a nondimensional pressure $P_o = 0.1$ on the top edge of the matrix.

We first consider a pulse-shaped rigid substrate, with the corresponding results shown in Figure 4-19. Notice that the nature of this problem is very different from the matrix one, since the hidden structure to be discovered is not entirely contained within the total domain Ω, and the measurements are only taken at the top surface instead of all four sides. Still, the topology optimization framework is able to find the correct shape of the rigid substrate by the end of the optimization process. We then consider a random-shaped rigid substrate, with the corresponding results shown in Figure 4-20. In this case too, the framework is able to identify the correct shape of the substrate, including in places that are far away from the surface.

4.4 Discussion

We have presented a topology optimization framework based on PINNs, with an eikonal regularization inspired by level-set methods and signed distance functions. Thanks to the parametrization of the geometry through a material density field that can adapt to any topology, our framework is able to discover both the number and the shape of hidden structures, without any prior knowledge required regarding the number or the type of shape to expect. In other words, it is a true topology optimization framework since it can discover the right topology of the solution beyond merely discovering the right shape(s) given a known topology. We note that this makes our approach stand out amongst other approaches in the literature that treat the same geometry identification problem, for which one needs to specify in advance either the number of structures to identify, the type of geometries to identify (e.g. circle, slit, etc), or both. Notable exceptions are the studies in Ref. [113] which parametrizes the geometry using a curve defined through a set of control nodes that can represent multiple structures at the same time, and Ref. [128] which parametrizes the geometry through a material distribution field. Ref. [113] demonstrates the detection of three circle-shaped inclusions but does not present results involving more complicated shapes, while Ref. [128] only considers examples with a single circle-shaped inclusion. By contrast, our framework is, to our knowledge, the first one to successfully detect multiple structures with each very different shapes. Moreover, thanks

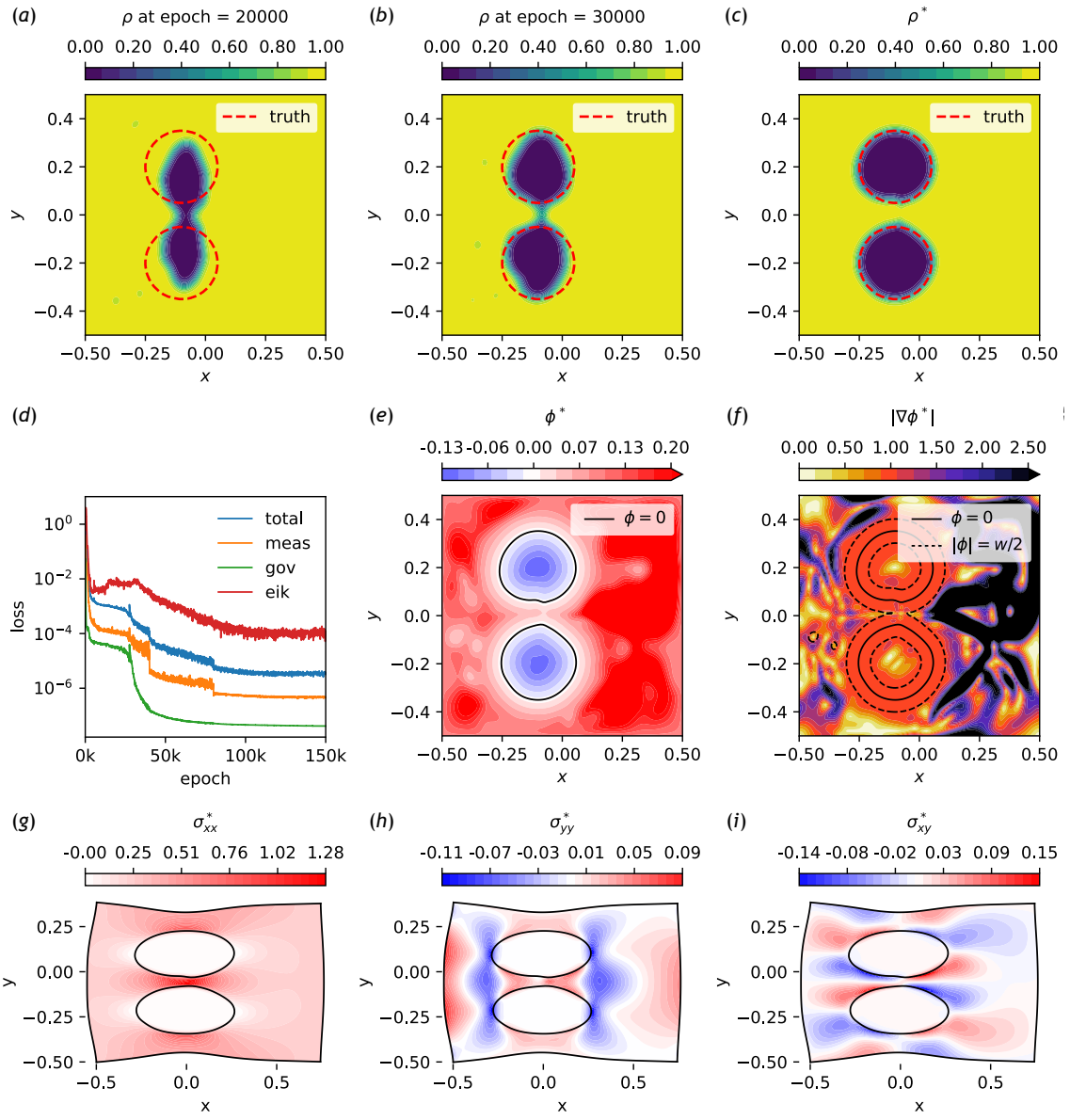

Figure 4-13: Hyperelastic matrix with two circle-shaped voids. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

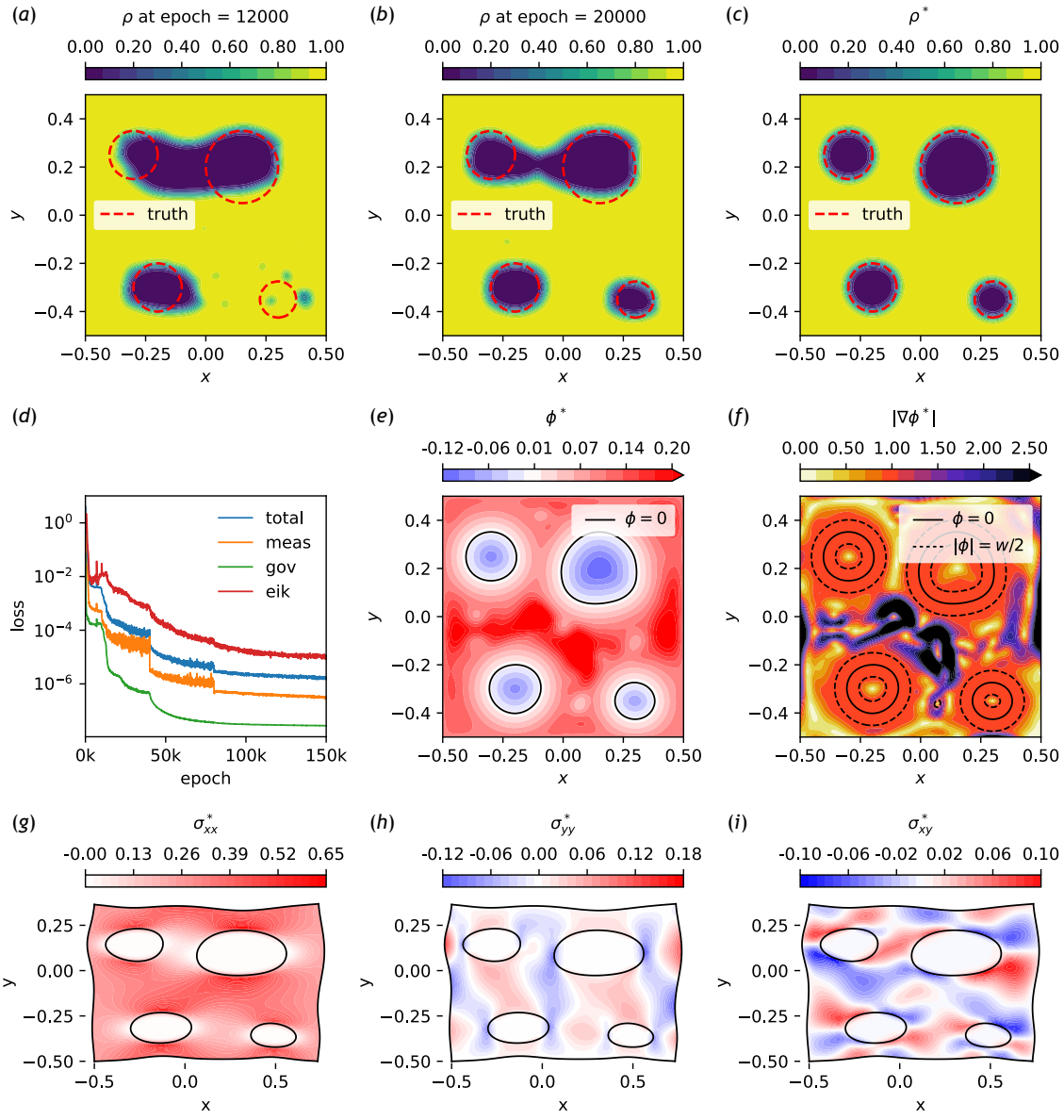

Figure 4-14: Hyperelastic matrix with four circle-shaped voids. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

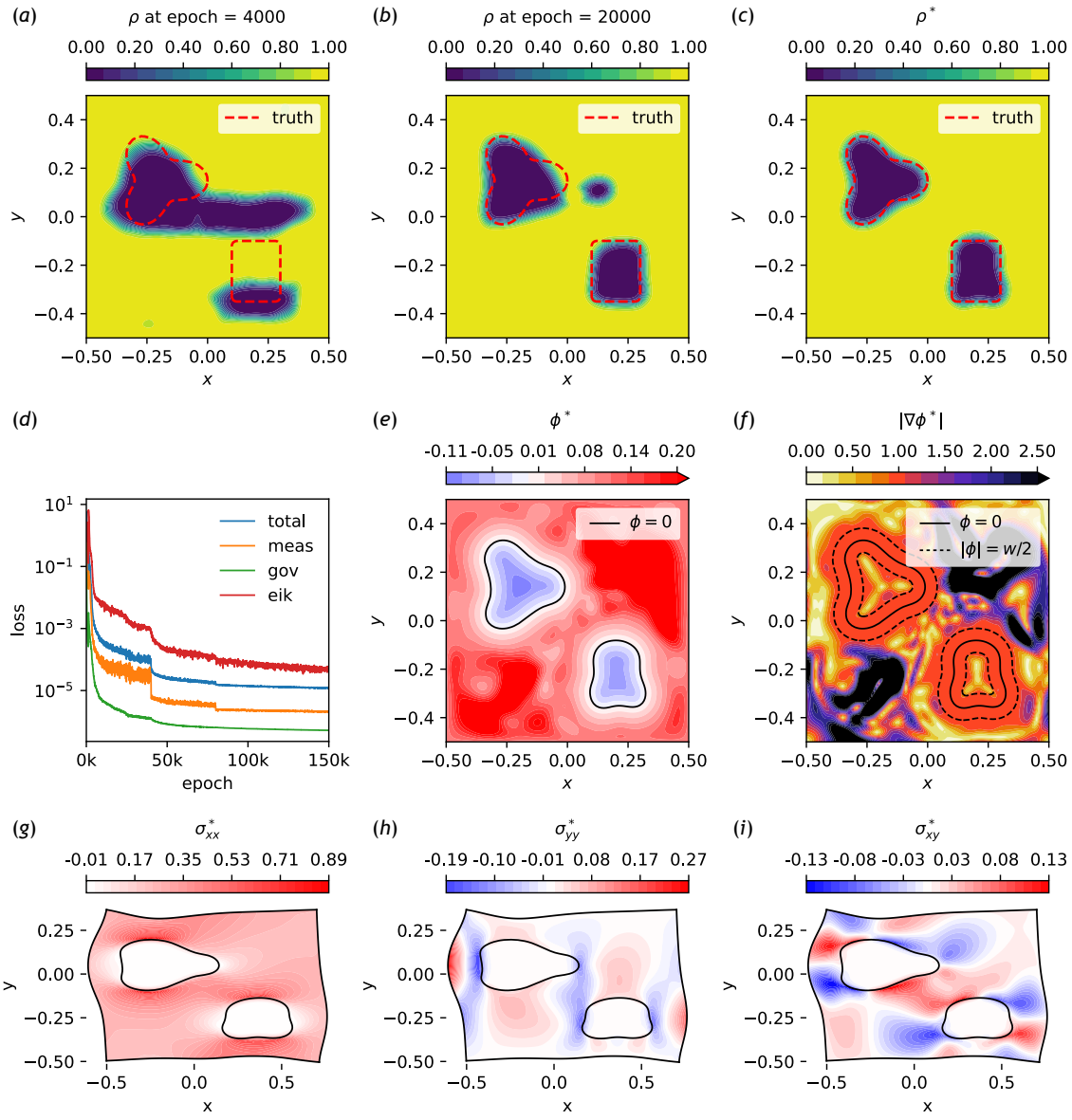

Figure 4-15: Hyperelastic matrix with one rectangle-shaped and one star-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

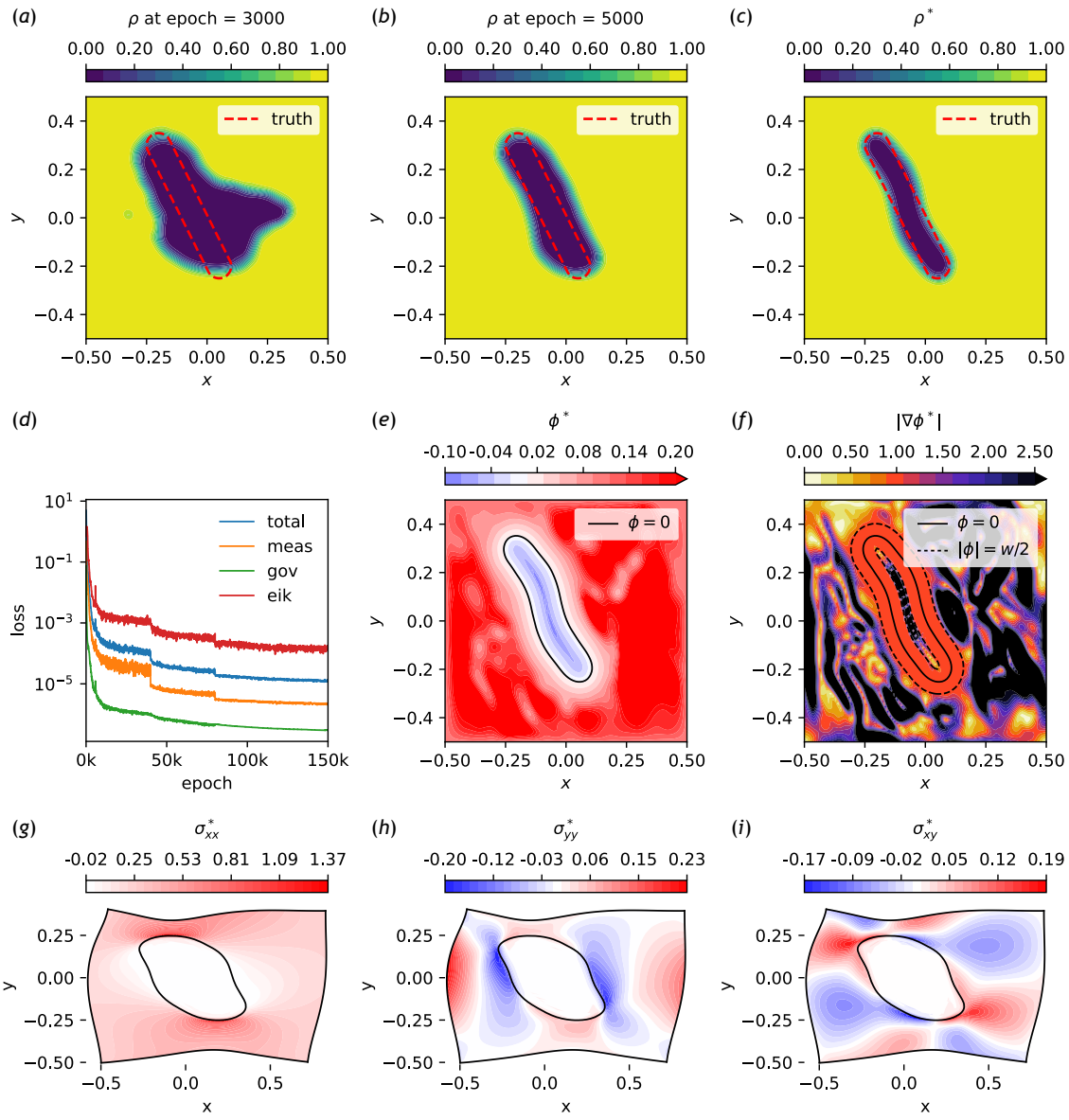

Figure 4-16: Hyperelastic matrix with one slit-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

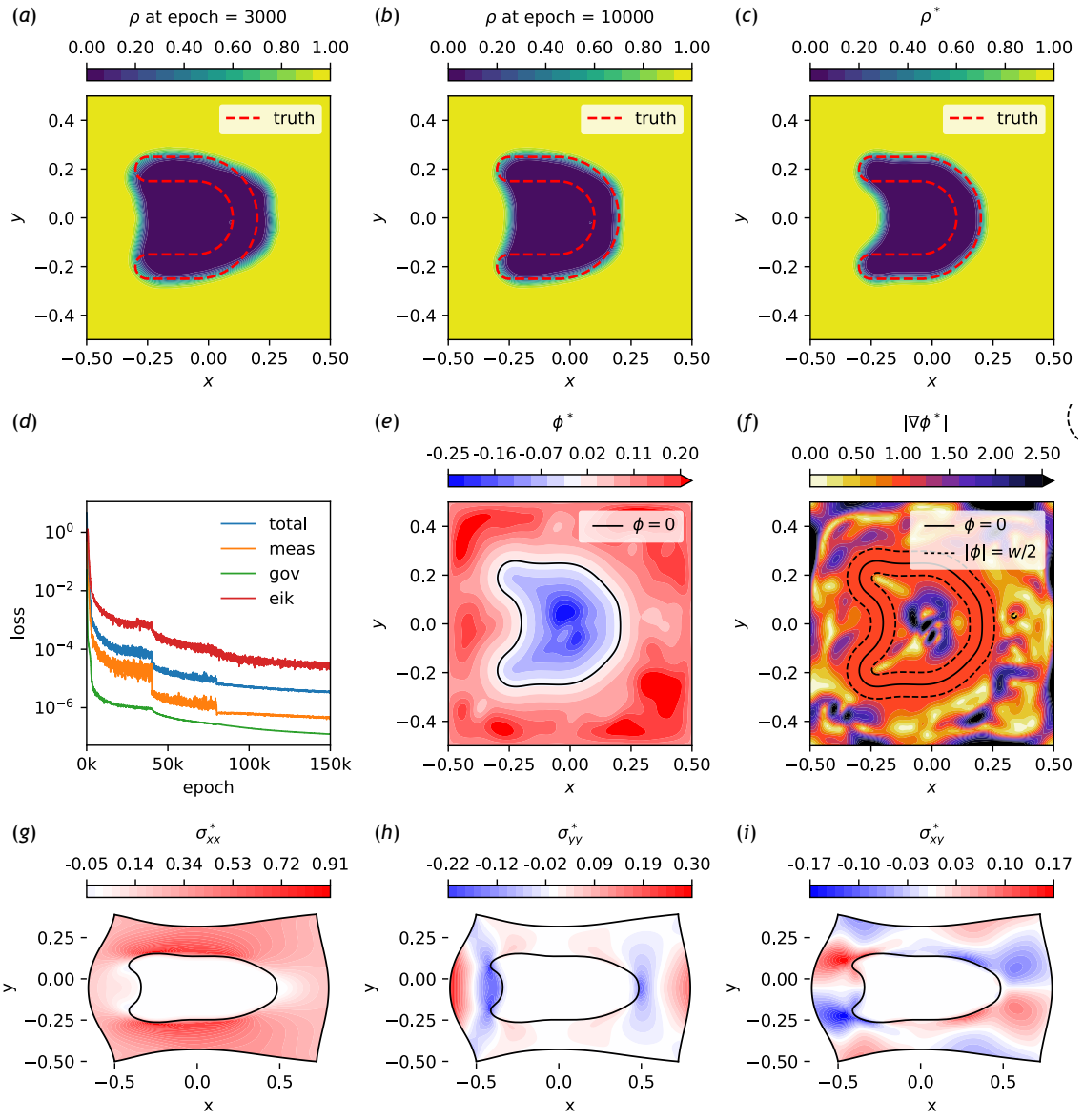

Figure 4-17: Hyperelastic matrix with one U-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

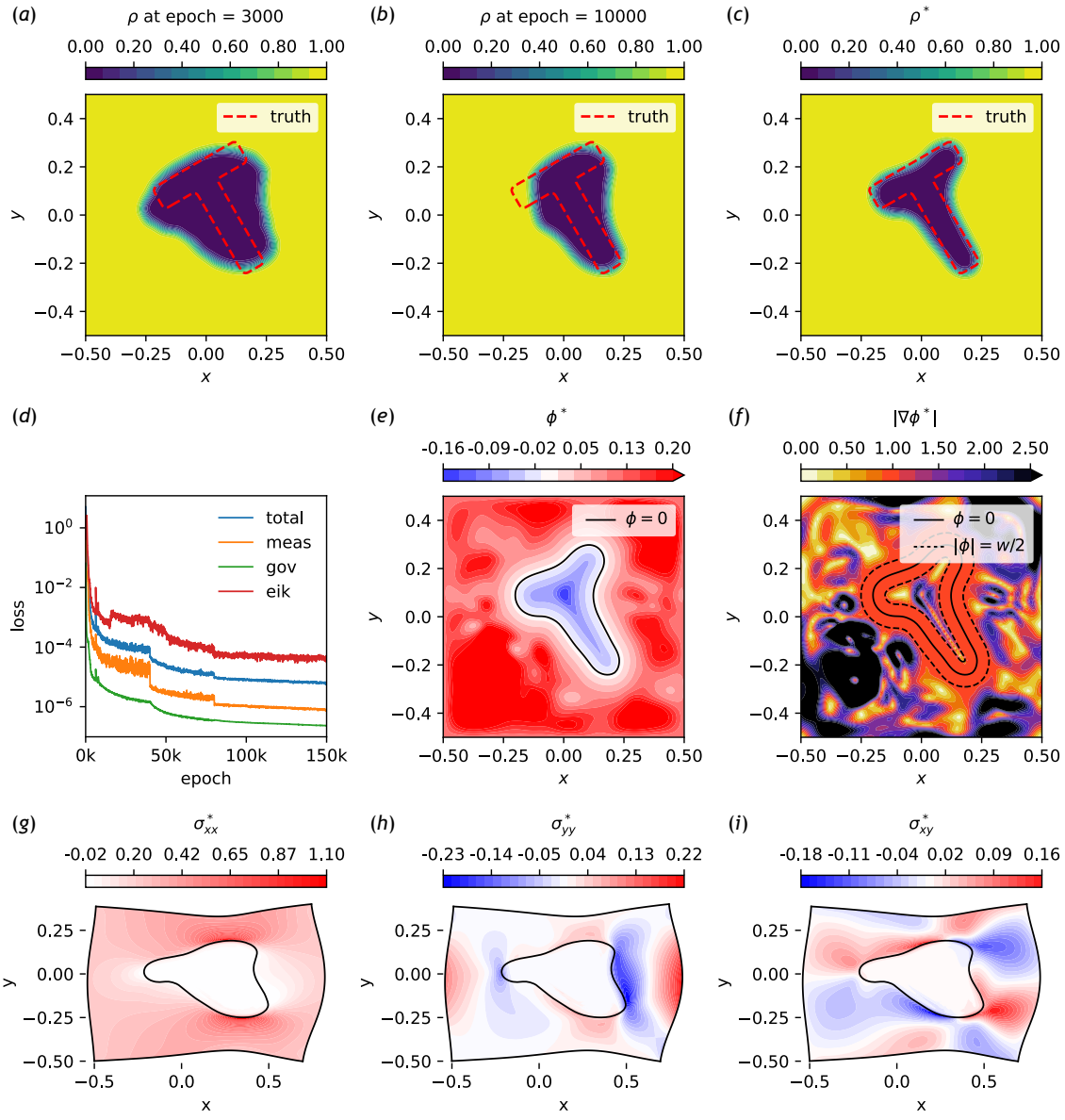

Figure 4-18: Hyperelastic matrix with one T-shaped void. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

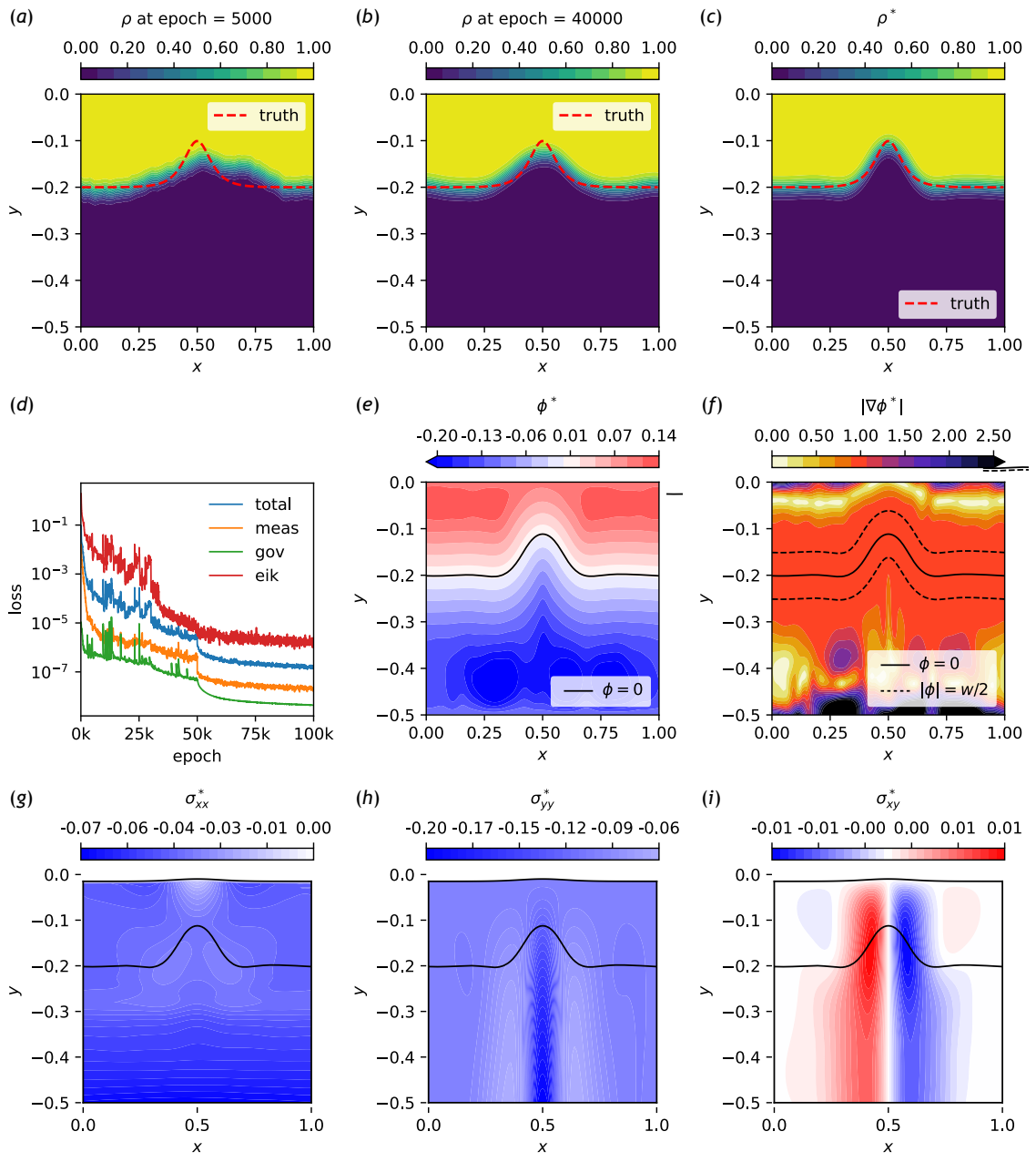

Figure 4-19: Linear elastic layer with a pulse-shaped rigid substrate. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

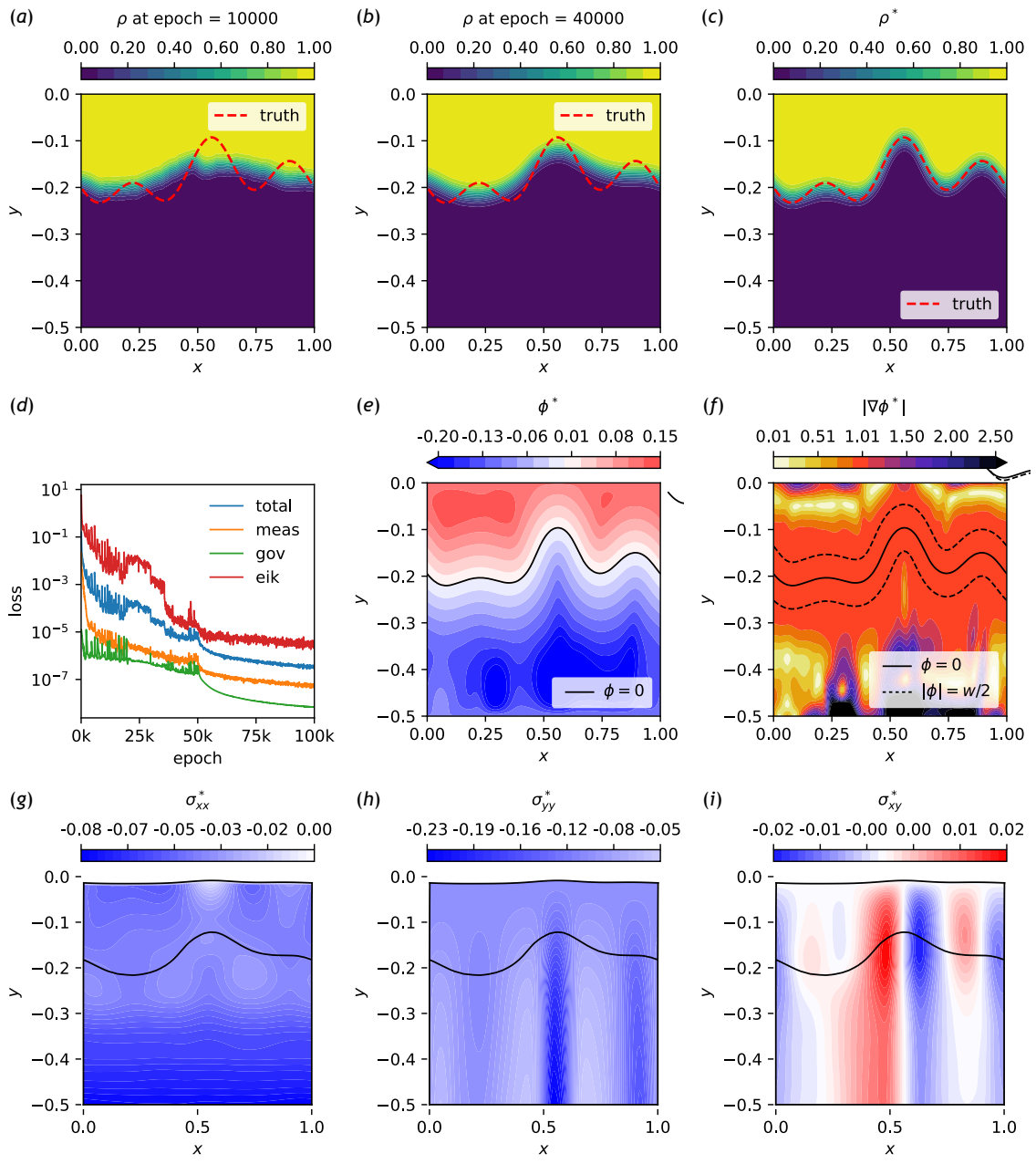

Figure 4-20: Linear elastic layer with a random-shaped rigid substrate. (*a-c*) Evolution of the material distribution ρ during the optimization, and its final state ρ^*. (*d*) Evolution of the various loss components during the optimization. (*e*, *f*) Final level-set function ϕ^* and its gradient magnitude $|\nabla\phi^*|$. (*g-i*) Final Cauchy stress components σ^*_{11}, σ^*_{22}, and σ^*_{12}, displayed in the deformed configuration obtained from the final displacement components u^*_1 and u^*_2.

to the representation of the geometry through a material distribution field and the flexibility of PINNs, extending our framework to the three-dimensional case should be straightforward.

Since we parametrize the geometry implicitly through a material density field, a major challenge is to drive the latter towards 0 or 1 and avoid intermediate values that do not have physical meaning. In this work, we have proposed to solve this problem by requiring that the level-set function underlying the material distribution satisfies the eikonal equation in a narrow band of finite width along its zero contour level. Since the zero contour level defines the location of the interface between the two phases, this ensures that the level-set function becomes a signed distance function to the interface. As a result, the thickness of the interface zone where the material density transitions between 0 and 1 is everywhere uniform, and there cannot be large zones of intermediate material density values. This eikonal constraint on the level-set function is naturally implemented in the PINN framework through the inclusion of an additional loss function, which we call eikonal regularization. This regularization bears resemblance to the work of Ref. [70], which showed that neural networks can be trained to learn signed distance functions by enforcing the Eikonal equation to be satisfied everywhere in space. However, in their case the neural network is also forced to vanish at a set of points sampled on a predefined surface of interest, while in our case the neural network approximation of the level-set function is allowed to evolve freely in such a way that the geometry defined by its zero level set solves the geometry identification problem. Another difference is that we only impose the eikonal constraint in a narrow band around the zero level set; outside the band the material density is nearly equal to 0 or 1 so the specific values of the level-set function do not matter. In fact, we have observed that imposing the eikonal constraint in the entire domain leads to worse results. Interestingly, we note that restricting the level-set function to solve the eikonal equation only in a narrow band around its zero level set is similar in spirit to narrow-band algorithms in traditional level-set methods. These algorithms nevertheless require an explicit velocity vector with which to update in a separate step the position of the zero level set, which is usually calculated using complex adjoint-based methods.

Although we have applied our framework to the problem of identifying hidden shapes and inclusions in elastic bodies, we emphasize that the methodology we propose is very general. Specifically, the idea of parametrizing geometries of arbitrary topologies with a material density field regularized with the eikonal constraint opens a pathway for PINNs to be applied to a wide range of geometry identification or design problems constrained by physical governing equations. The main requirement is that the same governing equations describe, through an explicit dependence on the material density variable, both phases of the problem under consideration. Within the realm of solid mechanics, such problems include topology optimization of structures [16] and metamaterials [95, 105], where the goal is to design structures that exhibit minimum compliance under load or other desirable mechanical, acoustic or thermal properties. In fluid mechanics, shape or topology optimization of structures is also a subject of interest, this time with the objective of minimizing the drag or maximizing the heat exchange of the fluid flowing through or around the structure [23, 136, 46].

Finally, one can also envision applications in photonic design, where the goal is to design lenses that demonstrate specified optical properties [137, 124].

Chapter 5

Conclusions

5.1 Summary

In this thesis, we have introduced analytical and numerical techniques to address several forward and inverse problems in systems ranging from a single elastic body to thousands of interacting ones. As such, a common theme between the different chapters has been the development of tools and techniques to address various engineering problems. Here, we give a high-level overview of the methodological contributions made in the different chapters. In Chapter 2, we focused on forward problems involving systems of a few hundred particles, each with elastically-anisotropic constitutive behavior. We derived analytically an orientation-dependent contact law for the force generated between elastically-anisotropic particles of any smooth shape. This contact law, which can be implemented in any DEM code, enables fast and accurate computational modeling of these systems in ways that were previously not possible. In Chapter 3, we focused on forward problems involving granular systems with millions of particles. We extended an existing continuum model for such systems in order to describe accurately the onset and arrest of flow in response to external loading variations in different geometries. Our approach rests on the integration of mechanisms for both hysteretic and nonlocal behavior, which provides a blueprint for modeling of other systems displaying similar properties, such as emulsions. In Chapter 4, we focused on inverse problems in single elastic bodies. We combined machine learning, physical governing equations and a novel regularization methodology to devise a topology optimization framework for geometry identification of hidden voids or inclusions based on surface response measurements. Our framework is very general and can be applied to a wide range of geometry identification and design problems across various fields of engineering.

5.2 Future directions

We trace out a few directions in which the different parts of the work can be extended in the future. Some of these have already been mentioned at the end of the respective chapter; some others are new.

In Chapter 2, the contact force model we have derived is limited to elastic deformation of the particles. As the force between two contacting particles increases, they eventually enter into a plastic deformation regime, which changes the relationship between force and overlap predicted by the contact force model. Thus, integrating plastic effects into our elastically anisotropic contact law is one potential avenue for future research. A second limitation of our work is that our contact law does not specify the tangential component of the contact force; addressing this shortcoming would be a valuable contribution. Finally, the custom DEM code in which we have implemented our contact law can accommodate a few particles at most; integrating the contact law into a high-performance DEM solver such as LAMMPS would be critical study the effect of elastic anisotropic on the macroscopic behavior of larger systems such as powders.

In Chapter 3, we have not been able to obtain a quantitative match between the continuum model predictions and the DEM simulations, and we have speculated that the discrepancy is due to the modeling of the boundary conditions. Addressing this question is a critical step towards practical usability of the model for engineering applications in which one needs accurate predictions both during the onset and arrest of flow as well as in the flowing regime. Furthermore, our study has revealed that the threshold for onset of flow is highly stochastic. While we have developed a procedure to adapt our deterministic continuum model to this stochastic threshold, a more accurate approach would be to make the continuum model itself stochastic, perhaps through the inclusion of stochastic term that only affects the threshold for onset of flow.

In Chapter 4, we have applied our topology optimization to linear elastic and hyperelastic materials in plane-strain configurations. An immediate next step would be to apply the framework to a wider range of constitutive models as well as to three-dimensional geometries. Other than the static loading scenario we have considered, the framework could be applied to geometry identification problems in other types of experiments such as acoustic or electromagnetic scattering. Beyond geometry identification problems, our method could also potentially be extended to inverse problems involving design of geometries to achieve targeted properties. These include structures that minimize compliance under load, airfoils that minimize drag, lenses that focus light on a target, and many other problems in a wide range of fields.

Appendix A

Supplementary information for Chapter 2

A.1 Coordinate systems and transformations

In this appendix, we introduce coordinate transformation matrices between the various reference frames that are utilized, which will come in handy when we describe the implementation of the contact force law in the following appendices. Recall Figure 2-2(a), which shows the two contacting bodies introduced in Figure 4-1, this time viewed from the global (laboratory) reference frame which is defined by the set of coordinates $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$. In Section 2.3.2, we have introduced a local set of coordinates (X_1^B, X_2^B, X_3^B) that is oriented along the material structure of a given body B, rotating with it at all times. Finally, we also need to consider the set of coordinates (x, y, z) aligned with the contact normal and tangent plane directions. Before proceeding further, we introduce three sets of orthonormal basis vectors:

- $(\mathbf{e}_1^{\mathcal{X}}, \mathbf{e}_2^{\mathcal{X}}, \mathbf{e}_3^{\mathcal{X}})$, for the global coordinate system $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$,
- $(\mathbf{e}_1^{X}, \mathbf{e}_2^{X}, \mathbf{e}_3^{X})$, for the body-centric coordinate system (X_1^B, X_2^B, X_3^B),
- $(\mathbf{e}_1^{x}, \mathbf{e}_2^{x}, \mathbf{e}_3^{x})$, for the contact coordinate system (x, y, z).

The orientation of body B – defined by coordinates (X_1^B, X_2^B, X_3^B) – with respect to the global reference frame $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$ can be parameterized by a rotation matrix $[\mathbf{R}^B]$ whose components are defined by $R_{ij}^B = \mathbf{e}_i^{\mathcal{X}} \cdot \mathbf{e}_j^{X}$. (Appendix A.7 provides a relationship between these rotation matrices and the Euler parameters introduced in Section 2.6.1 to characterize the orientation of particles in the DEM code.) Further, we also introduce a coordinate transformation matrix $[\mathbf{Q}^B]$ from the contact basis to the body-centric basis, with elements given by $Q_{ij}^B = \mathbf{e}_i^{X} \cdot \mathbf{e}_j^{x}$.

With this in hand, one can relate the components of a vector $\mathbf{v}$ in the global or contact bases to its components in the local basis of body B as

$$\{\mathbf{v}\}^X = [\mathbf{R}^B]^{\mathsf{T}} \{\mathbf{v}\}^{\mathcal{X}}, \tag{A.1a}$$

$$\{\mathbf{v}\}^X = [\mathbf{Q}^B] \{\mathbf{v}\}^{x}, \tag{A.1b}$$

where the superscripts X, $\mathcal{X}$, and x denote the components in the body-centric, global, and contact bases, respectively.

A.2 Calculation of the Green's function

Algorithm 1 presents a numerical procedure for computing the Green's function $h^B(\theta;0)$ introduced in Section 2.3.2, which is a function of the elasticity tensor $\mathbb{C}^B$ as well as the relative orientation of the contact (x,y,z) basis with respect to the body-centric (X_1^B, X_2^B, X_3^B) basis. The latter is parameterized by the coordinate transformation matrix $[\mathbf{Q}^B]$ introduced in Appendix A.1. Hereafter, we offer some complementary information on the algorithm. For the polar orientation θ and $\phi = 0$, the coordinates of the unit vectors $\mathbf{r}$ and $\mathbf{s}$ introduced in (A.2) are given in the (x,y,z) contact basis by

$$\{\mathbf{r}\}^x = (\cos\gamma\sin\theta, -\cos\gamma\cos\theta, -\sin\gamma)^\mathsf{T}, \tag{A.2a}$$

$$\{\mathbf{s}\}^x = (-\sin\gamma\sin\theta, \sin\gamma\cos\theta, -\cos\gamma)^\mathsf{T}, \tag{A.2b}$$

which is used in line 3. In line 9, we have used the fact that the unit normal $\mathbf{n}$ is related to the basis vector $\mathbf{e}_3^x$ as $\mathbf{n} = -\mathbf{e}_3^x$; therefore its coordinates in the (X_1^B, X_2^B, X_3^B) basis are given by $n_i^B = -Q_{i3}^B$. In practice, we discretize the integrals and iterate the **for** loops on lines 1 and 2 over 100 values of θ and γ, equispaced between 0 and 2π.

Algorithm 1: Calculation of the Green's function $h^B(\theta;0)$

Input: Coordinate transformation matrix $[\mathbf{Q}^B]$ from (x,y,z) basis to (X_1^B, X_2^B, X_3^B) basis, components of elasticity tensor $\mathbb{C}^B$ in (X_1^B, X_2^B, X_3^B) basis

1 **for** $\theta = 0$ **to** 2π **do**
2 **for** $\gamma = 0$ **to** 2π **do**
3 Calculate coordinates of $\mathbf{r}$, $\mathbf{s}$ in (x,y,z) basis with (A.2)
4 Transform coordinates of $\mathbf{r}$, $\mathbf{s}$ to (X_1^B, X_2^B, X_3^B) basis using (A.1b) with $[\mathbf{Q}^B]$
5 Compute integrand of (2.23) using (2.24)
6 **end**
7 Perform integral in (2.23) to get $G_{ij}(\theta;0)$
8 **end**
9 Using (2.25), compute $h^B(\theta;0) \leftarrow Q_{k3}^B G_{km}^{-1}(\theta;0) Q_{m3}^B$

Output: Green's function $h^B(\theta;0)$

A.3 Solution strategy for e and ϕ

We describe our strategy to solve numerically the coupled equations (2.33) for the eccentricity e and phase angle ϕ. First, we recast these equations as a minimization

problem for the objective function $J(e,\phi) = \log(f_1^2(e,\phi) + f_2^2(e,\phi))$, where $f_1(e,\phi)$ and $f_2(e,\phi)$ denote respectively the left-hand-sides of (2.33a) and (2.33b). We then perform a global search for the minimum of J on a coarse grid of values in the range $e \in [0, 0.8]$ and $\phi \in [-\pi/2, \pi/2]$, and feed the resulting value as an initial condition to a gradient-based constrained optimization solver. We use MATLAB's `fmincon` function, which implements an interior-point algorithm, and constrain the search over the region $e \in [0, 1]$. Since the objective function is 2π-periodic in the ϕ-direction, we have found that the optimization procedure is more robust when we leave ϕ unconstrained, and bring its value back to the interval $[-\pi/2, \pi/2]$ once the algorithm has converged.

A.4 Calculation of a look-up table for $\tilde{E}_*^B$

Given a material represented through its elasticity tensor $\mathbb{C}^B$, Algorithm 2 describes a numerical procedure for computing a look-up table of values of the equivalent plane strain modulus $\tilde{E}_*^B(\alpha^B, \beta^B)$ defined in Section 2.4.1, for all possible orientations $\alpha^B \in [0, 2\pi]$ and $\beta^B \in [0, \pi]$. We emphasize that the look-up table, $[\tilde{E}_*](\cdot,\cdot\,;\mathbb{C}^B)$, is purely a function of the elasticity tensor $\mathbb{C}^B$. As a result, a given simulation simply requires one look-up table per material present in the system. In the common case where all particles are made of the same material, only one such look-up table needs to be precomputed and stored. In our implementation, we have used 100 equispaced values for α^B and 50 for β^B. In line 7, the computation of the constant term of the Fourier series may be performed efficiently through the average $a_0^B = (2\pi)^{-1}\int_0^{2\pi} h^B(\theta;0)d\theta$. Note that the orientation of the unit vectors $\mathbf{u}$, $\mathbf{v}$ selected in line 4 of Algorithm 2 is inconsequential since only the mean component of the Green's function $h^B(\theta;0)$ is used.

Algorithm 2: Calculation of a look-up table of precomputed values of $\tilde{E}_*^B$

Input: Components of elasticity tensor $\mathbb{C}^B$ in (X_1^B, X_2^B, X_3^B) basis

1 **for** $\alpha^B = 0$ **to** 2π **do**
2 **for** $\beta^B = 0$ **to** π **do**
3 Use (2.35a) to construct $\mathbf{n}$ from Euler angles (α^B, β^B)
4 Construct $\mathbf{u}$ and $\mathbf{v}$ such that $(\mathbf{u}, \mathbf{v}, \mathbf{n})$ forms an orthonormal basis
5 Build the coordinate transformation matrix $[\mathbf{Q}^B] \leftarrow [\{\mathbf{u}\}^X, \{\mathbf{v}\}^X, \{\mathbf{n}\}^X]$
6 Call Algorithm 1 using $[\mathbf{Q}^B]$ and $\mathbb{C}^B$ to get the Green's function $h^B(\theta;0)$
7 Calculate $[\tilde{E}_*](\alpha^B, \beta^B;\mathbb{C}^B)$ from $h^B(\theta;0)$ with (2.37)
8 **end**
9 **end**

Output: Look-up table $[\tilde{E}_*](\cdot,\cdot\,;\mathbb{C}^B)$

A.5 Retrieving $\tilde{E}^c_*$ from the look-up table

We outline in Algorithm 3 a procedure for retrieving the composite plain strain modulus $\tilde{E}^c_*$ between two contacting bodies B_1 and B_2 from their orientations and the look-up table(s) precomputed by Algorithm 2. (Appendix A.7 presents formulae for obtaining the rotations matrices $[\mathbf{R}^{B_1}]$ and $[\mathbf{R}^{B_2}]$ characterizing the orientations of B_1 and B_2 from the Euler parameters utilized in Section 2.6.1.) The algorithm is outlined for the general case where B_1 and B_2 are made of different materials with elasticity tensors $\mathbb{C}^{B_1}$ and $\mathbb{C}^{B_2}$, requiring the passage of two look-up tables as an input, one corresponding to each material. Note however that if B_1 and B_2 are made of the same material, then only one look-up table is required. As pointed out in Section 2.3.2, the Green's function (2.22), and therefore the plane strain modulus $\tilde{E}^B_*$, are blind to the sign of the contact normal $\mathbf{n}$. Thus, we use in line 2 the same $\mathbf{n}$ to define the components of the contact normal direction in the reference frames of both bodies.

Algorithm 3: Retrieving the composite plain strain modulus $\tilde{E}^c_*$

Input: Rotation matrices $[\mathbf{R}^{B_1}]$ and $[\mathbf{R}^{B_2}]$ describing the orientations of bodies B_1 and B_2, components $\{\mathbf{n}\}^{\mathcal{X}}$ of contact normal direction $\mathbf{n}$ in global basis, look-up tables $[\tilde{E}_*](\cdot,\cdot;\mathbb{C}^{B_1})$ and $[\tilde{E}_*](\cdot,\cdot;\mathbb{C}^{B_2})$

1 **for** $B = B_1, B_2$ **do**
2 Transform the coordinates of $\mathbf{n}$ from the global to the body's local (X^B_1, X^B_2, X^B_3) basis using (A.1a) with $[\mathbf{R}^B]$
3 Convert these coordinates to Euler angles (α^B, β^B) using (2.35b)
4 Use (α^B, β^B) to interpolate $\tilde{E}^B_*$ from the look-up table $[\tilde{E}_*](\cdot,\cdot;\mathbb{C}^B)$
5 **end**
6 Calculate $\tilde{E}^c_*$ using (2.40)

Output: Composite plain strain modulus $\tilde{E}^c_*$

A.6 Geometric features of the exact contact law

Here, we provide further details on the geometric features of the exact contact law for the materials and indentation parameters considered in Section 2.5.2. More specifically, we show polar visualizations of the eccentricity e, orientation ϕ, and semi-major axis length a_1 of the contact area incurred by an indentation depth $\delta = 100\,\text{nm}$, for a circular gap function ($A = B = 1\,\mu\text{m}^{-1}$) in Figure A-1 and an elliptic gap function ($A = 1\,\mu\text{m}^{-1}$, $B = 2\,\mu\text{m}^{-1}$) in Figure A-2. In both figures, (a,b,c) correspond to iron, (d,e,f) to quartz, and (g,h,i) to zirconia. Note that the discontinuities of the ϕ field appearing in Figure A-1 are merely a visual artefact; indeed, the orientations $\phi = \pi/2$ and $\phi = -\pi/2$ are effectively identical as can be inferred from Figure 4-1(c). Interestingly, the contact normal direction-dependence of the fields e, ϕ, and a_1 undergoes drastic change as the gap function changes from circular to elliptic, while that of the normal force F remains relatively unaffected, as was shown in Figures 2-4 and 2-5.

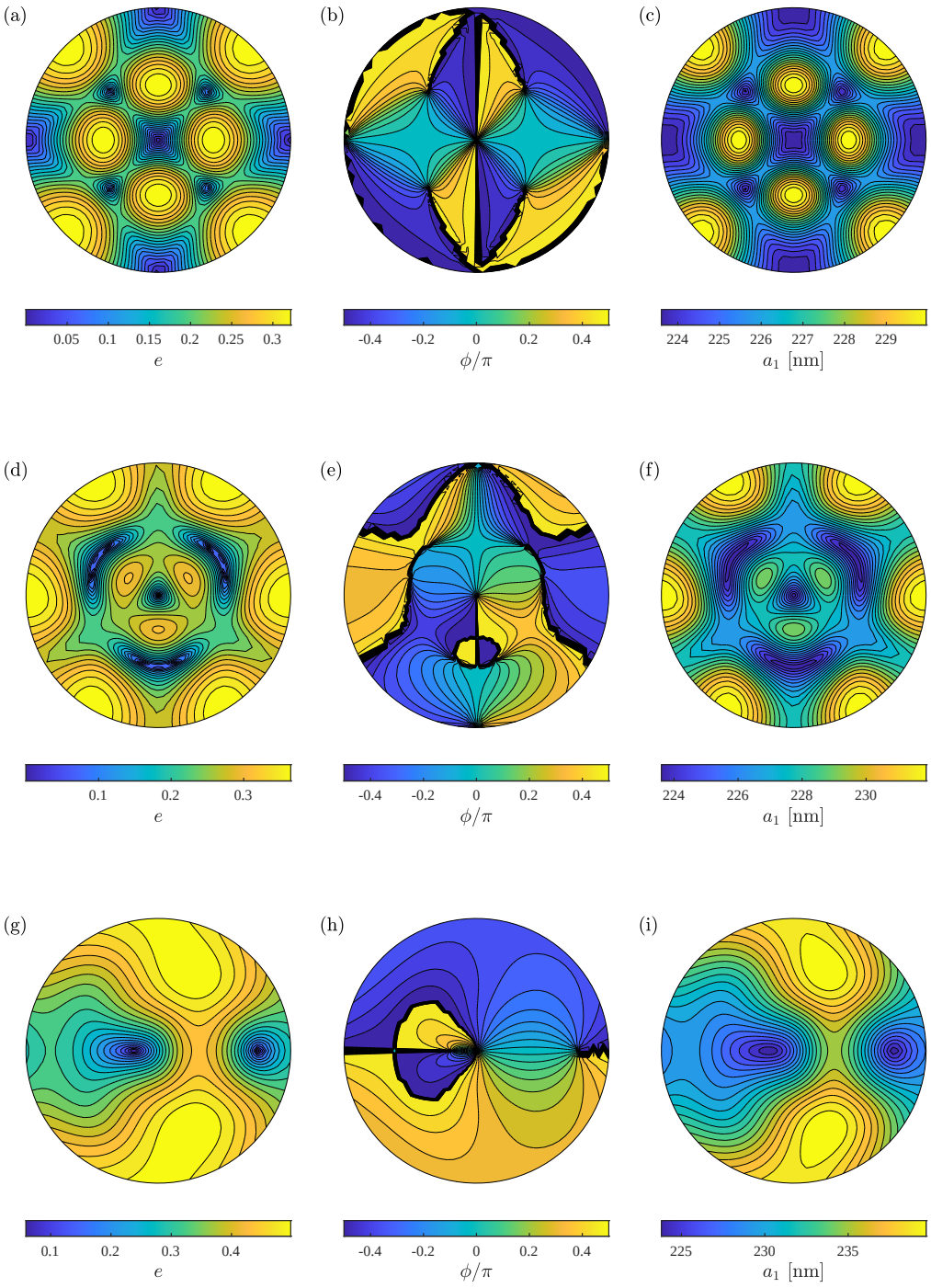

Figure A-1: Polar visualizations of the eccentricity e, orientation ϕ, and semi-major axis length a_1 of the contact area predicted by the exact solution for iron (a,b,c), quartz (d,e,f), and zirconia (g,h,i), under indentation depth $\delta = 100\,\text{nm}$ and gap function coefficients $A = B = 1\,\mu\text{m}^{-1}$.

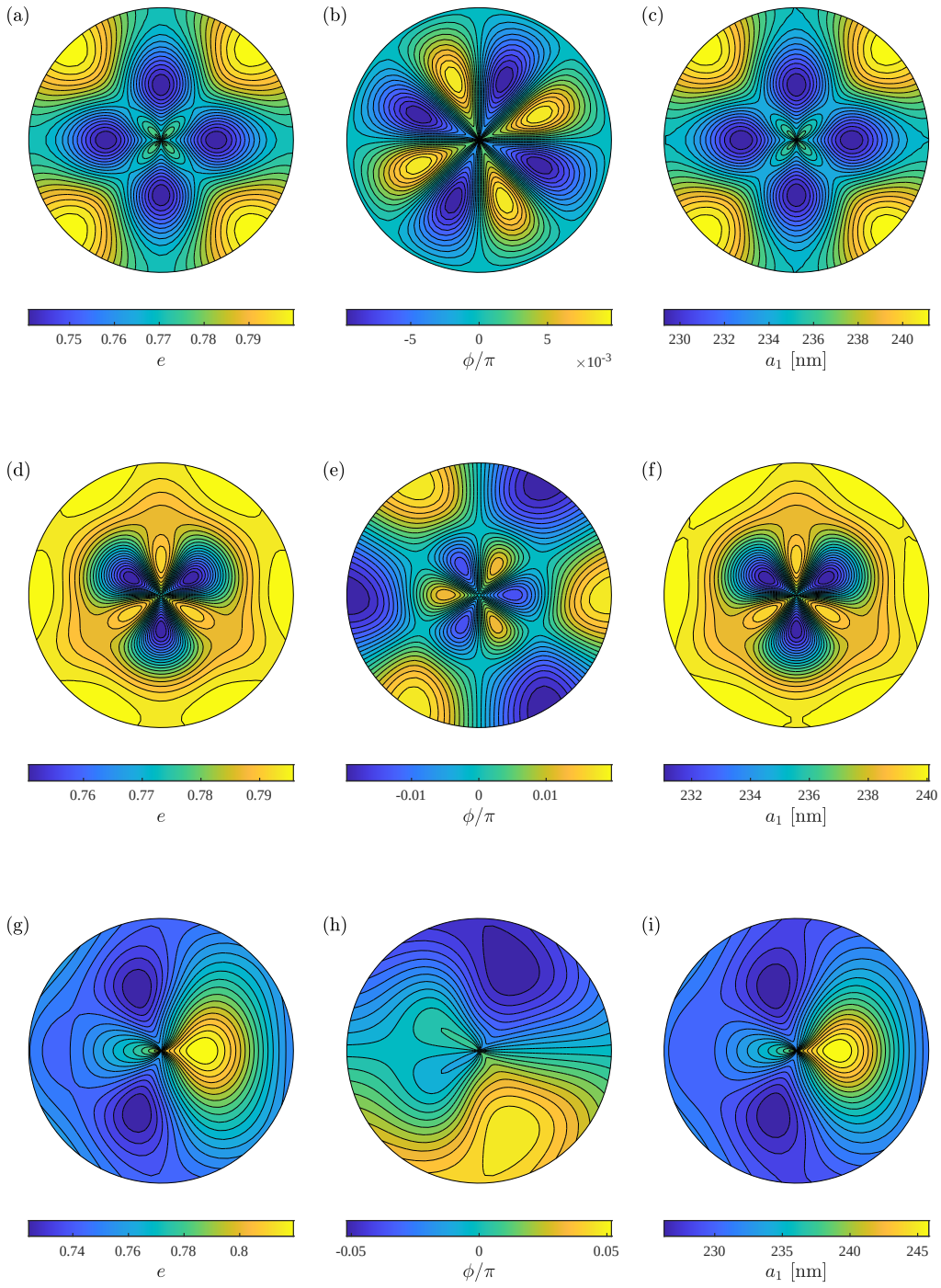

Figure A-2: Polar visualizations of the eccentricity e, orientation ϕ, and semi-major axis length a_1 of the contact area predicted by the exact solution for iron (a,b,c), quartz (d,e,f), and zirconia (g,h,i), under indentation depth $\delta = 100\,\text{nm}$ and gap function coefficients $A = 1\,\mu\text{m}^{-1}$ and $B = 2\,\mu\text{m}^{-1}$.

A.7 Relationship between Euler parameters and rotation matrices

In the DEM code presented in Section 2.6, the orientation of a given particle i is described using Euler parameters, or unit quaternions, $\{\boldsymbol{\epsilon}_i\}$ since their time integration from the angular velocity $\{\boldsymbol{\omega}_i\}$ is straightforward. In Appendices A.1 and A.5, however, the orientation is specified by a rotation matrix $[\mathbf{R}^{B_i}]$ (with B_i referring to particle i), since the latter can be utilized to transform vector components from the global to the body-centric coordinate systems. Here, we specify the simple relationship that exists between the two representations. Dropping the index i, the rotation matrix is given by the Euler parameters as [78]

$$[\mathbf{R}^B] = 2\begin{bmatrix} \epsilon_0^2+\epsilon_1^2-1/2 & \epsilon_1\epsilon_2-\epsilon_0\epsilon_3 & \epsilon_1\epsilon_3+\epsilon_0\epsilon_2 \\ \epsilon_1\epsilon_2+\epsilon_0\epsilon_3 & \epsilon_0^2+\epsilon_2^2-1/2 & \epsilon_2\epsilon_3-\epsilon_0\epsilon_1 \\ \epsilon_1\epsilon_3-\epsilon_0\epsilon_2 & \epsilon_2\epsilon_3+\epsilon_0\epsilon_1 & \epsilon_0^2+\epsilon_3^2-1/2 \end{bmatrix}, \tag{A.3}$$

while the Euler parameters are given by the rotation matrix as

$$\epsilon_0^2 = \frac{\mathrm{tr}[\mathbf{R}^B]+1}{4}, \tag{A.4a}$$

$$\epsilon_1 = \frac{R^B_{32}-R^B_{23}}{4\epsilon_0}, \tag{A.4b}$$

$$\epsilon_2 = \frac{R^B_{13}-R^B_{31}}{4\epsilon_0}, \tag{A.4c}$$

$$\epsilon_3 = \frac{R^B_{21}-R^B_{12}}{4\epsilon_0}. \tag{A.4d}$$

Note that the quadratic equation for ϵ_0 possesses two roots, and the choice of a particular root also affects the signs of ϵ_1, ϵ_2, and ϵ_3. Since the elements of $[\mathbf{R}^B]$ are quadratic in the Euler parameters, either root may be selected for ϵ_0 and still define the same physical orientation of the body.

A.8 Further details on the DEM implementation

We provide additional details regarding our DEM code. We consider spherical zirconia particles with density $m = 5680\,\mathrm{kg/m^3}$ and uniform diameter $d = 1\,\mathrm{cm}$. The elastic part of the normal force is calculated with our anisotropic contact law, using the elastic constants of zirconia given in Section 2.5. The viscous part is given a damping parameter $\gamma_n = 200\,\mathrm{Ns/m}$ in Section 2.6.2, and $\gamma_n = 0$ in Section 2.6.3. The code is implemented in MATLAB and utilizes a semi-implicit Euler method to evolve (2.52) and (2.55). The linear and angular velocities are first integrated explicitly, following which the positions and orientations are integrated using the new (end-of-time-step) linear and angular velocities. We use a time step $\Delta t = 10^{-6}\,\mathrm{s}$.

A.9 Theoretical analysis of the compressed chain

In this appendix, we derive the dispersion relation of the compressed chain of particles investigated in Section 2.6.3. Our derivation follows the exposition of [82] and [21], extending the latter to the present case of a four-particle unit cell. First, consider the force that is generated between any two particles i and j in the chain as a result of both the static force F_0 and the dynamic displacement of the particles. Following our contact force law (2.41), this force reads

$$F_{ij} = K_{ij}(\delta_{ij}^0 + \delta_{ij})^{3/2}, \tag{A.5}$$

where δ_{ij}^0, δ_{ij} are the overlaps between particles i and j due respectively to the static and dynamic force, and K_{ij} is the nonlinear contact stiffness between particles i and j, defined as

$$K_{ij} = \frac{2}{3}\tilde{E}_*^c(\alpha_i, \beta_i, \alpha_j, \beta_j)d^{1/2}, \tag{A.6}$$

with α_i, β_i, α_j, β_j the Euler angles describing the orientations of particles i and j with respect to the contact normal direction (which is parallel to the chain axis), and d the uniform diameter of the particles. Assuming that $\delta_{ij} \gg \delta_{ij}^0$, the force-overlap relationship (A.5) can be linearized about δ_{ij}^0, leading to

$$F_{ij} \simeq K_{ij}(\delta_{ij}^0)^{3/2} + \frac{3}{2}K_{ij}(\delta_{ij}^0)^{1/2}\delta_{ij} = F_0 + \kappa_{ij}\delta_{ij}, \tag{A.7}$$

where we have substituted the static force $F_0 = K_{ij}(\delta_{ij}^0)^{3/2}$ and defined the linearized stiffness κ_{ij} between particles i and j as

$$\kappa_{ij} = \frac{3}{2}K_{ij}(\delta_{ij}^0)^{1/2} = \frac{3}{2}K_{ij}^{2/3}F_0^{1/3}. \tag{A.8}$$

Going back to Figure 2-12(a), we recall that our particles are oriented in a way that gives rise to three different possible stiffnesses between any two particles. Letting these stiffnesses κ_1, κ_2, and κ_3, we can write the linearized governing equations for the infinitely long chain as

$$m\ddot{u}_{4n-3} = \kappa_3(u_{4n-4} - u_{4n-3}) - \kappa_2(u_{4n-3} - u_{4n-2}), \tag{A.9a}$$
$$m\ddot{u}_{4n-2} = \kappa_2(u_{4n-3} - u_{4n-2}) - \kappa_1(u_{4n-2} - u_{4n-1}), \tag{A.9b}$$
$$m\ddot{u}_{4n-1} = \kappa_1(u_{4n-2} - u_{4n-1}) - \kappa_2(u_{4n-1} - u_{4n}), \tag{A.9c}$$
$$m\ddot{u}_{4n} = \kappa_2(u_{4n-1} - u_{4n-0}) - \kappa_3(u_{4n} - u_{4n+1}), \tag{A.9d}$$

where m denotes the uniform mass of the particles, u_i is the dynamic displacement of particle i with respect to its static equilibrium position in the compressed chain, and n is the index of the unit cell. These are wave equations on a lattice with periodicity equal to the static unit cell length $a = 4d - \delta_1^0 - 2\delta_2^0 - \delta_3^0$, where δ_c^0 refers to the static overlap at a contact with stiffness κ_c. We therefore express the solution as a Bloch

wave expansion,

$$\{u_{4n-3}, u_{4n-2}, u_{4n-1}, u_{4n}\} = \{U, V, W, X\}e^{i(kan-\omega t)}, \tag{A.10}$$

where the wavenumber k belongs to the first Brillouin zone, $[-\pi/a, \pi/a]$. In order to find the frequency ω corresponding to each k, we substitute the expansion (A.10) into (A.9) and solve for a nontrivial solution. This results in the dispersion relation

$$m^4\omega^8 + c_6 m^3\omega^6 + c_4 m^2\omega^4 + c_2 m\omega^2 + c_0 = 0, \tag{A.11}$$

where c_6, c_4, c_2, c_0 are functions of κ_1, κ_2, κ_3 as follows:

$$c_6 = -2(\kappa_1 + 2\kappa_2 + \kappa_3), \tag{A.12a}$$

$$c_4 = -\kappa_1^2 - 2\kappa_2^2 - \kappa_3^2 + (\kappa_1 + \kappa_2)^2 + (\kappa_2 + \kappa_3)^2 + 4(\kappa_1 + \kappa_2)(\kappa_2 + \kappa_3), \tag{A.12b}$$

$$c_2 = 2(\kappa_1 + \kappa_2)(\kappa_2^2 + \kappa_3^2) + 2(\kappa_2 + \kappa_3)(\kappa_1^2 + \kappa_2^2) - 2(\kappa_1 + \kappa_2)(\kappa_2 + \kappa_3)^2 - 2(\kappa_2 + \kappa_3)(\kappa_1 + \kappa_2)^2, \tag{A.12c}$$

$$c_0 = \kappa_1^2\kappa_3^2 + \kappa_2^4 + (\kappa_1 + \kappa_2)^2(\kappa_2 + \kappa_3)^2 - \kappa_1^2(\kappa_2 + \kappa_3)^2 - \kappa_3^2(\kappa_1 + \kappa_2)^2 - 2\kappa_1\kappa_2^2\kappa_3 \cos ka = 0. \tag{A.12d}$$

The dispersion relation (A.11) possesses four ω solutions for every value of k, which are plotted in Figure 2-13 in the range $k \in [0, \pi/a]$ due to the symmetry of c_0 with respect to $k = 0$.

Appendix B

Supplementary information for Chapter 3

B.1 Coarse-graining methodology

In this appendix, we describe our coarse-graining procedure for extracting continuum velocity and stress fields from the particle-wise DEM data. The approach we follow was introduced in Zhang and Kamrin [198], building upon earlier work [106, 3, 190]. Since the geometries that we consider are homogeneous along the x-direction, the spatial averaging generates fields defined at discrete heights z_k, spaced $0.5d$ apart. For a given position z_k, we define the instantaneous velocity and stress fields as

$$\mathbf{v}(z_k, t) = \sum_{m=-M}^{M} w_m \bar{\mathbf{v}}(z_m, t), \tag{B.1}$$

$$\boldsymbol{\sigma}(z_k, t) = \sum_{m=-M}^{M} w_m \bar{\boldsymbol{\sigma}}(z_m, t), \tag{B.2}$$

where $\bar{\mathbf{v}}$ and $\bar{\boldsymbol{\sigma}}$ are sublayer-wise velocity and stress averages at the heights $z_m = z_k + (W/2)(m/M)$, each weighted by the coefficient $w_m = \cos((\pi/2)(m/n))$. Following Zhang and Kamrin [198], we choose $W = 2d$ and $M = 5$. We now let L_{im} denote the cross-sectional length between particle i and the horizontal line at height z_m. The sublayer-wise instantaneous velocity is given by

$$\bar{\mathbf{v}}(z_m, t) = \frac{\sum_i L_{im} \mathbf{v}_i(t)}{\sum_i L_{im}}, \tag{B.3}$$

where $\mathbf{v}_i$ is the velocity of grain i. Then, the sublayer-wise instantaneous stress field is defined by

$$\bar{\boldsymbol{\sigma}}(z_m, t) = \frac{\sum_i L_{im} \boldsymbol{\sigma}_i(t)}{L}, \tag{B.4}$$

where L is the domain length along the x-direction, and $\boldsymbol{\sigma}_i$ is the stress tensor associated with grain i. The latter consists of contact and kinetic contributions, and is

given by

$$\boldsymbol{\sigma}_i(t) = \frac{1}{A_i} \sum_{i \neq j} \mathbf{f}_{ij}(t) \otimes \mathbf{r}_{ij}(t) + \frac{m_i}{A_i} \delta \mathbf{v}_i(t) \otimes \delta \mathbf{v}_i(t), \tag{B.5}$$

where $A_i = \pi d_i^2/4$ and $m_i = \rho_s A_i$ are respectively the area and mass of grain i, $\mathbf{f}_{ij}$ is the contact force exerted on grain i by grain j, and $\mathbf{r}_{ij}$ is the vector pointing from the center of grain i to that of grain j. In the kinetic contribution, the velocity fluctuations are calculated as $\delta \mathbf{v}_i(t) = \mathbf{v}_i(t) - \mathbf{v}(z_i, t)$, where $\mathbf{v}(z_i, t)$ is the coarse-grained instantaneous velocity (B.1) interpolated to the vertical position z_i of grain i.

B.2 Numerical discretization

Here, we present our numerical discretization method for solving the fluidity equation (3.5b) under a time-dependent applied stress. In the problems that we consider, the equation reduces to a one-dimensional PDE for $g(z, t)$ that is driven by a prescribed stress ratio function $\mu(z, t)$. The physics governing the time scale t_0 that appears in the fluidity equation are still unknown, and we simply assign a sufficiently small value $t_0 = 10^{-4}\,\mathrm{s}$ that it does not affect the dynamics of the solution. The spatial domain is discretized into $N = 100$ nodes, and the diffusion term is evaluated using second-order finite differences. The fluidity equation is then integrated in time using an implicit Euler scheme with a time step $\Delta t = 5 \cdot 10^{-4}\,\mathrm{s}$, which we implement in MATLAB. Further, we artificially limit $g(z, t)$ to a minimum value of $10^{-2}\,\mathrm{s}^{-1}$ in order to avoid $g(z, t)$ reaching infinitesimally small values during the arrested portion of the stress ramp. Indeed, this would prevent $g(z, t)$ from growing sufficiently fast when flow onset should occur, as the applied stress is subsequently increased. We have verified that the floor value for $g(z, t)$ is small enough that it does not alter the observed transition stresses.

B.3 Size effects in simple plane shear

We report in Figure B-1 additional DEM results on the system-size dependence of the critical transition stresses μ_{start} and μ_{stop} in plane shear without gravity. Mirroring our definition of μ_{start} in Section 3.3.1, μ_{stop} is defined as the observed μ_{w} when I_{w} last falls below 10^{-3} during stress decrease. Shown are the individual transition stresses from 20 DEM runs (crosses) and the corresponding deterministic values (filled circles) extracted from the linear fit of the CDF according to the procedure outlined in Figure 3-5(c). We observe that the critical stresses are almost independent of system size, corroborating results from a previous DEM study[31].

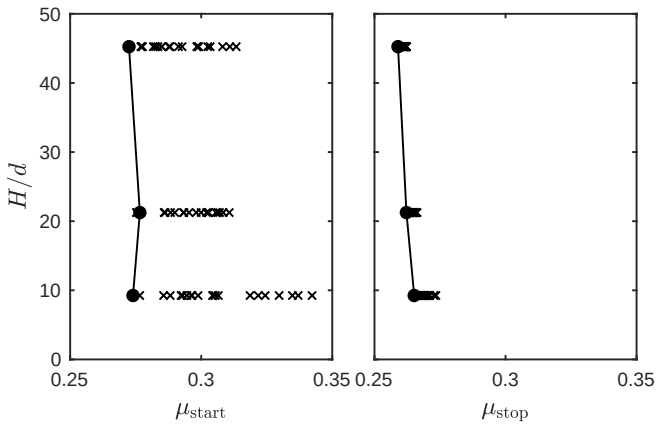

Figure B-1: Critical stresses μ_{start} and μ_{stop} in simple plane shear versus true system height H/d, obtained from 20 different DEM runs. The crosses show the individual transition stresses pertaining to every run, while the filled circles represent the deterministic value extracted from the linear fit of the CDF.

B.4 Critical stresses from NGF

In this appendix, we explain how to obtain the critical starting and stopping stresses predicted by the NGF model without computing time-dependent solutions to a slowly varying applied stress, which are computationally intensive due to the required low rate of change of the applied stress to ensure quasi-steady conditions. Let the scalar $\bar{\mu}$ denote the amplitude of the applied stress ratio $\mu(z)$ throughout the domain – for instance, $\bar{\mu}$ is μ_{w} in the case of plane shear with gravity, or $\tan\theta$ in inclined plane. For a given geometry, implying a certain distribution of the stress ratio $\mu(z)$, we then rewrite the fluidity equation (3.5b) as

$$t_0\dot{g} = \mathcal{F}(g;\bar{\mu}), \tag{B.6}$$

where the dependence on the magnitude of $\mu(z)$ has been explicitly denoted through $\bar{\mu}$. In the following, we will call g_0 any steady-state solution of (B.6).

At low $\bar{\mu}$, the arrested state $g_0 = 0$ is the only stable solution. Gradually increasing $\bar{\mu}$, flow onset occurs the moment the $g_0 = 0$ solution becomes unstable to small perturbations g', which defines the critical starting stress $\bar{\mu}_{\text{start}}$ [5]. These small perturbations are governed by the linear equation

$$t_0\dot{g}' = \mathcal{L}(g_0;\bar{\mu})g', \tag{B.7}$$

where $\mathcal{L}(g_0;\bar{\mu})$ – the linearization (also called Fréchet derivative) of $\mathcal{F}(g;\bar{\mu})$ around

g_0 – acts on the perturbation g' as

$$\mathcal{L}(g_0;\bar{\mu})g' = A^2d^2\nabla^2 g' - \frac{(\mu_2-\mu_s)(\mu_s-\mu)}{\mu_2-\mu}g' - 2b\sqrt{\frac{m}{P}}\mu g_0 g' \\ - \chi(g_0;\mu,P)g' - \frac{\partial\chi}{\partial g}(g_0;\mu,P)g_0 g'. \tag{B.8}$$

To evaluate whether perturbations grow or decay, we substitute $g' = \tilde{g}'(z)e^{\lambda t}$ into (B.7), which leads to the eigenvalue problem

$$t_0\lambda\tilde{g}'(z) = \mathcal{L}(g_0;\bar{\mu})\tilde{g}'(z) \tag{B.9}$$

for the growth rate λ. This eigenvalue problem can be solved numerically by discretizing $\mathcal{L}(g_0;\bar{\mu})$ using a finite difference approximation, giving a spectrum of eigenvalues with the one having the largest real part, λ_m, dictating the overall rate of growth or decay of the perturbation. Setting $g_0 = 0$, one can perform repeatedly this calculation for increasing values of $\bar{\mu}$ until $Re\{\lambda_m\}$ becomes positive, at which point the arrested solution loses stability and $\bar{\mu} = \bar{\mu}_{\text{start}}$.

Gradually decreasing $\bar{\mu}$ from a value above $\bar{\mu}_{\text{start}}$, flow arrest occurs the moment (B.6) ceases to admit a nonzero steady-state solution g_0, which defines the critical stopping stress $\bar{\mu}_{\text{stop}}$. To check whether that is true at a given value of $\bar{\mu}$, it suffices to perform Newton-Raphson iterations to find g_0. Starting from an initial guess g^0, the algorithm performs at each step n the update $g^{n+1} = g^n + \omega\Delta^n$, where $0 < \omega < 1$ is a relaxation parameter and the step direction Δ^n is given through the linear system

$$\mathcal{L}(g^n;\bar{\mu})\Delta^n = -\mathcal{F}(g^n;\bar{\mu}). \tag{B.10}$$

We stop the iterations when the norm of $\mathcal{F}(g^n;\bar{\mu})$ falls under a specified threshold, indicating that g^n has converged to a steady-state solution g_0 of (B.6). Thus, our strategy to find $\bar{\mu}_{\text{stop}}$ goes as follows. We start with a value of $\bar{\mu}$ above $\bar{\mu}_{\text{start}}$, for which we are guaranteed a nonzero g_0 solution. We compute the latter by letting g reach steady-state in the time-dependent solver. Then, we repeatedly compute g_0 for incrementally decreasing values of $\bar{\mu}$ through Newton-Raphson iterations, using at each step level of $\bar{\mu}$ the converged solution g_0 from the previous step as an initial guess. At some point the Newton-Raphson iterations will suddenly converge to the arrested $g_0 = 0$ solution, indicating that $\bar{\mu}$ has reached $\bar{\mu}_{\text{stop}}$.

Both the eigenvalue problem (B.9) and linear system (B.10) are implemented in MATLAB borrowing the same grid and discretized differential operators used in the time-dependent solver. In the Newton-Raphson iterations, we use $\omega = 0.5$ to balance stability and speed of convergence. Previous studies [97, 118] have shown that for some geometries and boundary conditions, there exist analytical or semi-analytical solutions for the growth rate λ and thus the threshold μ_{start}. However, such solutions are much harder to obtain for μ_{stop}, despite partial progress in that direction on a I-gradient model applied to the inclined plane scenario [114]. Therefore, we limit ourselves in this paper to the numerical methodology that we have outlined above,

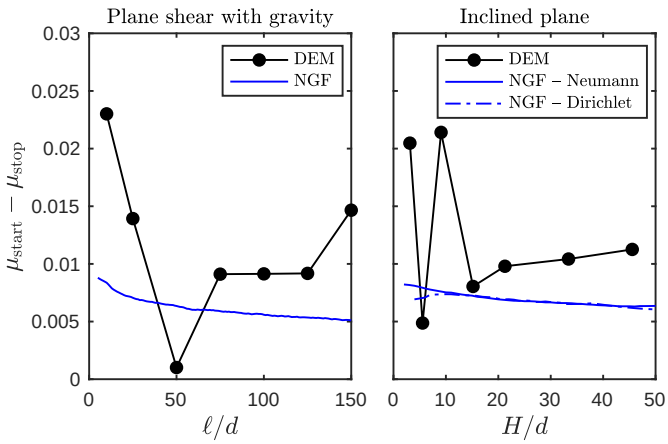

Figure B-2: Critical stress difference $\mu_{\text{start}} - \mu_{\text{stop}}$ versus dimensionless loading length scale ℓ/d for plane shear with gravity, and versus dimensionless layer height H/d for inclined plane. The DEM data (filled circles) and NGF data (blue lines) is the same as in Figures 3-6(c,f) and 3-7(c,f).

noting that it is computationally very efficient – the starting and stopping curves in Figures 3-6 and 3-7 were calculated in a few minutes on a laptop.

B.5 Hysteresis size dependence

Here, we investigate the system-size dependence of the hysteresis amplitude, measured by the difference of the starting and stopping critical stress ratios, based on the DEM simulations and NGF model predictions reported in Figures 3-6 and 3-7 in Sections 3.3.2 and 3.3.3. Figure B-2 displays $\mu_{\text{start}} - \mu_{\text{stop}}$ as a function of the dimensionless loading length scale ℓ/d for plane shear with gravity data, and as a function of the dimensionless layer height H/d for inclined plane data. The DEM and NGF data are within a comparable range and suggest a weak effect of system size on the hysteresis amplitude.

Appendix C

Supplementary information for Chapter 4

C.1 Problem formulation

Here, we provide details on the formulation of the two plane-strain elasticity inverse problems formulated in the introduction. In both problems, we define the two-dimensional domain Ω formed by the union of the elastic solid and the voids or rigid inclusions. Denoting with $\mathbf{x} = (x_1, x_2) \in \Omega$ the planar spatial coordinates, the hidden geometrical layout of voids or rigid inclusions is characterized by the material distribution $\rho(\mathbf{x})$, equal to 1 in the elastic phase and 0 in the void or rigid inclusion phase. We will now treat separately the small-deformation case where a linear elastic material is considered and the large-deformation case where an incompressible Neo-Hookean hyperelastic material is used, since the mathematical formalism and governing equations are different in the two cases.

C.1.1 Small-deformation linear elasticity

We first consider the case where the elastic material is governed by a linear elastic constitutive law with Young's modulus E and Poisson's ratio ν, which describes the small-deformation behavior of a wide range of materials. The mechanics of the two problems pictured in Figure 4-1 can then be described by a planar displacement field $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))$ and a Cauchy stress tensor $\boldsymbol{\sigma}(\mathbf{x})$ with components $\sigma_{ij}(\mathbf{x})$, $i, j = 1, 2$, for $\mathbf{x} = (x_1, x_2) \in \Omega$.

The stress tensor must satisfy the equilibrium equation

$$\nabla \cdot \boldsymbol{\sigma} = 0, \quad \mathbf{x} \in \Omega, \tag{C.1}$$

while the displacement is related to the stress through a linear stress-strain relation that depends on $\rho(\mathbf{x})$ and will be expressed differently depending on whether we are considering voids or rigid inclusions. For voids, the stress must vanish in the $\rho = 0$

regions and the stress-strain relation is therefore

$$\boldsymbol{\sigma} = \rho(\mathbf{x}) \left[\lambda \operatorname{tr}(\boldsymbol{\epsilon})\, \mathbf{I} + 2\mu\, \boldsymbol{\epsilon}\right], \quad \mathbf{x} \in \Omega, \tag{C.2}$$

where $\boldsymbol{\epsilon} = (\nabla \mathbf{u} + \nabla \mathbf{u}^T)/2$ is the infinitesimal strain tensor with components ϵ_{ij}, $i, j = 1, 2$, $\operatorname{tr}(\boldsymbol{\epsilon}) = \epsilon_{11} + \epsilon_{22}$ denotes its trace, and $\lambda = E\nu/(1+\nu)(1-2\nu)$, $\mu = E/2(1+\nu)$ are the Lamé constants. In the case of rigid inclusions, the strain must vanish in the $\rho = 0$ regions, thus we consider the stress-strain relation in the inverted form

$$\boldsymbol{\epsilon} = \rho(\mathbf{x}) \left[\frac{1+\nu}{E}\boldsymbol{\sigma} - \frac{\nu(1+\nu)}{E} \operatorname{tr}(\boldsymbol{\sigma})\, \mathbf{I}\right], \quad \mathbf{x} \in \Omega, \tag{C.3}$$

where $\operatorname{tr}(\boldsymbol{\sigma}) = \sigma_{11} + \sigma_{22}$ is the trace of the stress tensor. Note that the above relation differs from the three-dimensional one due to the plane strain assumption.

A particular loading is prescribed on the outer boundary $\partial\Omega$ of the domain, which can be decomposed into a portion $\partial\Omega_t$ with prescribed traction and a portion $\partial\Omega_u$ with prescribed displacement. The corresponding boundary conditions are

$$\mathbf{u}(\mathbf{x}) = \bar{\mathbf{u}}(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_u, \tag{C.4a}$$
$$\boldsymbol{\sigma}(\mathbf{x})\mathbf{n}(\mathbf{x}) = \bar{\mathbf{t}}(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_t, \tag{C.4b}$$

where $\bar{\mathbf{u}}(\mathbf{x})$ and $\bar{\mathbf{t}}(\mathbf{x})$ are the prescribed boundary displacement and traction, respectively, and $\mathbf{n}$ denotes the outward unit normal vector. In the case of the elastic layer, the outer boundary also comprises a portion $\partial\Omega_p$ with periodic boundary conditions on the displacement and traction. Upon application of the loading, the elastic body deforms in a way that satisfies the equilibrium equation (C.1), the constitutive relation (C.2) or (C.3), and the boundary conditions (C.4). We then measure the surface displacements $\mathbf{u}_i^m$ at discrete locations $\mathbf{x}_i$ forming a subset $\partial\Omega^m$ of $\partial\Omega_t$.

The geometry identification problem can then be stated as follows. Given the measured surface displacements $\mathbf{u}_i^m$, $i = 1, \ldots, |\partial\Omega^m|$, we want to find the hidden material distribution $\rho(\mathbf{x})$ in Ω such that the difference between the predicted and measured surface quantities vanish, that is,

$$\mathbf{u}(\mathbf{x}_i) = \mathbf{u}_i^m, \quad \mathbf{x}_i \in \partial\Omega^m. \tag{C.5}$$

For a given guess of $\rho(\mathbf{x})$, the predicted displacement and stress fields $\mathbf{u}(\mathbf{x})$ and $\boldsymbol{\sigma}(\mathbf{x})$ must satisfy the equilibrium equation (C.1), the constitutive relation (C.2) or (C.3), and the boundary conditions (C.4).

This formulation is applicable to both problems pictured in Figure 4-1, with the only difference being the domain Ω under consideration and the boundary conditions. For the elastic matrix, the domain is $\Omega = [-0.5, 0.5] \times [-0.5, 0.5]$ and the loading is defined by the boundary conditions

$$\boldsymbol{\sigma}(\mathbf{x})\mathbf{n}(\mathbf{x}) = -P_o \mathbf{e}_1, \quad \mathbf{x} \in \{-0.5, 0.5\} \times [-0.5, 0.5], \tag{C.6a}$$
$$\boldsymbol{\sigma}(\mathbf{x})\mathbf{n}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{x} \in [-0.5, 0.5] \times \{-0.5, 0.5\}. \tag{C.6b}$$

For the elastic layer, the domain is $\Omega = [0,1] \times [-0.5, 0]$ and the loading is defined by the boundary conditions

$$\boldsymbol{\sigma}(\mathbf{x})\mathbf{n}(\mathbf{x}) = -P_o \mathbf{e}_2, \quad \mathbf{x} \in [0,1] \times \{0\}, \tag{C.7a}$$

$$\mathbf{u} = \mathbf{0}, \qquad \mathbf{x} \in [0,1] \times \{-0.5\}, \tag{C.7b}$$

and periodic boundary conditions for the displacement and traction on $\mathbf{x} \in \{0,1\} \times [-0.5, 0]$.

C.1.2 Large-deformation hyperelasticity

We now consider the case where the elastic material is an incompressible Neo-Hookean hyperelastic solid with shear modulus μ. Since this constitutive law describes large deformations, we now need to distinguish between the reference (undeformed) and current (deformed) configurations. Keeping consistency with our previous notation, we will denote by $\mathbf{x} = (x_1, x_2) \in \Omega$ and $\mathbf{y} = (y_1, y_2) \in \Omega^*$ the coordinates in the reference and deformed configurations, respectively, with Ω^* the deformed image of Ω. The deformation can then be described by the displacement field $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))$ that takes an initial position $\mathbf{x} \in \Omega$ into its current location $\mathbf{y} = \mathbf{x} + \mathbf{u}(\mathbf{x}) \in \Omega^*$. Finally, in order to formulate the governing equations and boundary conditions in the reference configuration, we need to introduce the first Piola-Kirchhoff stress tensor $\mathbf{S}(\mathbf{x})$ with components $S_{ij}(\mathbf{x})$, $i, j = 1, 2$, for $\mathbf{x} = (x_1, x_2) \in \Omega$. Unlike the Cauchy stress tensor, the first Piola-Kirchhoff stress tensor is not symmetric.

The stress tensor must satisfy the equilibrium equation

$$\nabla_{\mathbf{x}} \cdot \mathbf{S} = 0, \quad \mathbf{x} \in \Omega, \tag{C.8}$$

where the derivatives in $\nabla_{\mathbf{x}}$ are taken with respect to the reference coordinates $\mathbf{x}$. As before, the stress-strain relation will depend on $\rho(\mathbf{x})$; here we only consider the presence of voids so that the stress must vanish in the $\rho = 0$ regions. The nonlinear constitutive law for incompressible Neo-Hookean materials thus takes the form

$$\mathbf{S} = \rho(\mathbf{x}) \left[-p\mathbf{F}^{-T} + \mu \mathbf{F} \right], \quad \mathbf{x} \in \Omega, \tag{C.9}$$

where $\mathbf{F}(\mathbf{x}) = \mathbf{I} + \nabla_{\mathbf{x}} \mathbf{u}(\mathbf{x})$ is the deformation gradient tensor, and $p(\mathbf{x})$ is a scalar field that serves to enforce the incompressibility constraint

$$\rho(\mathbf{x}) \left[\det(\mathbf{F}) - 1 \right] = 0, \quad \mathbf{x} \in \Omega. \tag{C.10}$$

The presence of $\rho(\mathbf{x})$ turns off the incompressibility constraint in the $\rho = 0$ regions since voids do not deform in a way that preserves volume.

Similar to the small-deformation case, the outer boundary $\partial\Omega$ of the domain can be decomposed into a portion $\partial\Omega_t$ with prescribed traction and a portion $\partial\Omega_u$ with

prescribed displacement. The corresponding boundary conditions are

$$\mathbf{u}(\mathbf{x}) = \bar{\mathbf{u}}(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_u, \tag{C.11a}$$

$$\mathbf{S}(\mathbf{x})\mathbf{n}_0(\mathbf{x}) = \bar{\mathbf{s}}(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_t, \tag{C.11b}$$

where $\bar{\mathbf{u}}(\mathbf{x})$ and $\bar{\mathbf{s}}(\mathbf{x})$ are, respectively, the prescribed boundary displacement and traction in the reference configuration, and $\mathbf{n}_0$ denotes the outward unit normal vector in the reference configuration. In the case of the elastic layer, the outer boundary also comprises a portion $\partial\Omega_p$ with periodic boundary conditions on the displacement and traction. Upon application of the loading, the elastic body deforms in a way that satisfies the equilibrium equation (C.8), the constitutive relation (C.9), and the boundary conditions (C.11). We then measure the surface displacements $\mathbf{u}_i^m$ at discrete locations $\mathbf{x}_i$ forming a subset $\partial\Omega^m$ of $\partial\Omega_t$.

The geometry identification problem can then be stated as follows. Given the measured surface displacements $\mathbf{u}_i^m$, $i = 1, \ldots, |\partial\Omega^m|$, we want to find the hidden material distribution $\rho(\mathbf{x})$ in Ω such that the difference between predicted and measured surface displacements vanish, that is,

$$\mathbf{u}(\mathbf{x}_i) = \mathbf{u}_i^m, \quad \mathbf{x}_i \in \partial\Omega^m. \tag{C.12}$$

For a given guess of $\rho(\mathbf{x})$, the predicted displacement and stress fields $\mathbf{u}(\mathbf{x})$ and $\mathbf{S}(\mathbf{x})$ and the scalar field $p(\mathbf{x})$ must satisfy the equilibrium equation (C.8), the constitutive relation (C.9), the incompressibility condition (C.10), and the boundary conditions (C.11).

We will only treat the matrix inclusion problem pictured in Figure 4-1 in this large-deformation framework. The domain and loading are identical to the definitions given in Section C.1.1, with the only difference that $\boldsymbol{\sigma}$ and $\mathbf{n}$ are replaced with $\mathbf{S}$ and $\mathbf{n}_0$, respectively.

C.2 Additional information on the solution methodology

Here, we describe in detail the application of our methodology presented in Sections 4.2.2 and 4.2.3 to the solution of the two plane-strain elasticity inverse problems formulated in the introduction and Appendix C.1. We will treat separately the small-deformation linear elasticity case and the large-deformation hyperelasticity case.

C.2.1 Small-deformation linear elasticity

The governing physical laws presented in Appendix C.1.1 show that the problem is described by the field of physical quantities $\boldsymbol{\psi} = (u_1, u_2, \sigma_{11}, \sigma_{22}, \sigma_{12})$ in addition to the

material distribution ρ. We therefore introduce the neural network approximations

$$u_1(\mathbf{x}) = \bar{u}_1(\mathbf{x}; \boldsymbol{\theta}_1), \tag{C.13a}$$
$$u_2(\mathbf{x}) = \bar{u}_2(\mathbf{x}; \boldsymbol{\theta}_2), \tag{C.13b}$$
$$\sigma_{11}(\mathbf{x}) = \bar{\sigma}_{11}(\mathbf{x}; \boldsymbol{\theta}_3), \tag{C.13c}$$
$$\sigma_{22}(\mathbf{x}) = \bar{\sigma}_{22}(\mathbf{x}; \boldsymbol{\theta}_4), \tag{C.13d}$$
$$\sigma_{12}(\mathbf{x}) = \bar{\sigma}_{12}(\mathbf{x}; \boldsymbol{\theta}_5), \tag{C.13e}$$
$$\phi(\mathbf{x}) = \bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi), \tag{C.13f}$$

and the material distribution is given by $\rho(\mathbf{x}) = \bar{\rho}(\mathbf{x}; \boldsymbol{\theta}_\phi) = \text{sigmoid}(\bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi)/\delta)$. We then formulate the loss function (4.5) by specializing the loss term expressions presented in Section 4.2.2 to the linear elasticity problem, using the governing equations given in Appendix C.1.1. Omitting the $\boldsymbol{\theta}$'s for notational simplicity, we obtain

$$\mathcal{L}_{\text{meas}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}) = \frac{1}{|\partial\Omega^m|} \sum_{\mathbf{x}_i \in \partial\Omega^m} |\bar{\mathbf{u}}(\mathbf{x}_i) - \mathbf{u}_i^m|^2, \tag{C.14a}$$

$$\mathcal{L}_{\text{gov}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_\phi) = \frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\mathbf{r}_{\text{eq}}(\bar{\boldsymbol{\sigma}}(\mathbf{x}_i))|^2 + \frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\mathbf{r}_{\text{cr}}(\bar{\mathbf{u}}(\mathbf{x}_i), \bar{\boldsymbol{\sigma}}(\mathbf{x}_i), \bar{\rho}(\mathbf{x}_i))|^2, \tag{C.14b}$$

$$\mathcal{L}_{\text{eik}}(\boldsymbol{\theta}_\phi) = \frac{1}{|\Omega^d_{\text{eik}}|} \sum_{\mathbf{x}_i \in \Omega^d_{\text{eik}}} \left(|\nabla\bar{\phi}(\mathbf{x}_i)| - 1\right)^2, \tag{C.14c}$$

where $\bar{\mathbf{u}} = (\bar{u}_1, \bar{u}_2)$ and $\bar{\boldsymbol{\sigma}}$ has components $\bar{\sigma}_{i,j}$, $i, j = 1, 2$. In (C.14b), the terms $\mathbf{r}_{\text{eq}}$ and $\mathbf{r}_{\text{cr}}$ refer to the residuals of the equilibrium equation (C.1) and the constitutive relation (C.2) or (C.3). The eikonal loss term is problem-independent and therefore identical to (4.9).

We note that instead of defining neural network approximations for the displacements and the stresses, we could define neural network approximations solely for the displacements, that is, $\boldsymbol{\psi} = (u_1, u_2)$. In this case, the loss term (C.14b) would only include the residual of the equilibrium equation (C.1), in which the stress components would be directly expressed in terms of the displacements and the material distribution using the constitutive relation (C.2). However, several recent studies [154, 74, 81, 156] have shown that the mixed formulation adopted in the present work results in superior accuracy and training performance, which could partly be explained by the fact that only first-order derivatives of the neural network outputs are involved since the displacements and stresses are only differentiated to first oder in (C.1) and (C.2). In our case, the mixed formulation holds the additional advantage that it enables us to treat rigid inclusions using the inverted constitutive relation (C.3) instead of (C.2). Finally, the mixed formulation allows us to directly integrate both displacement and traction boundary conditions into the output of the neural network approximations, as we describe in the next paragraph.

We design the architecture of the neural networks in such a way that they inherently satisfy the boundary conditions. For the elastic matrix, we do this through the

transformations

$$\bar{u}_1(\mathbf{x};\boldsymbol{\theta}_1) = \bar{u}'_1(\mathbf{x};\boldsymbol{\theta}_1), \tag{C.15a}$$
$$\bar{u}_2(\mathbf{x};\boldsymbol{\theta}_2) = \bar{u}'_2(\mathbf{x};\boldsymbol{\theta}_2), \tag{C.15b}$$
$$\bar{\sigma}_{11}(\mathbf{x};\boldsymbol{\theta}_3) = (x-0.5)(x+0.5)\,\bar{\sigma}'_{11}(\mathbf{x};\boldsymbol{\theta}_3) + P_o, \tag{C.15c}$$
$$\bar{\sigma}_{22}(\mathbf{x};\boldsymbol{\theta}_4) = (y-0.5)(y+0.5)\,\bar{\sigma}'_{22}(\mathbf{x};\boldsymbol{\theta}_4), \tag{C.15d}$$
$$\bar{\sigma}_{12}(\mathbf{x};\boldsymbol{\theta}_5) = (x-0.5)(x+0.5)(y-0.5)(y+0.5)\,\bar{\sigma}'_{12}(\mathbf{x};\boldsymbol{\theta}_5), \tag{C.15e}$$
$$\bar{\phi}(\mathbf{x};\boldsymbol{\theta}_\phi) = (x-0.5)(x+0.5)(y-0.5)(y+0.5)\,\bar{\phi}'(\mathbf{x};\boldsymbol{\theta}_\phi) + w, \tag{C.15f}$$

where the quantities with a prime denote the raw output of the neural network. In this way, the neural network approximations defined in (C.13) obey by construction the boundary conditions (C.6). Further, since we know that the elastic material is present all along the outer surface $\partial\Omega$, we defined $\bar{\phi}$ so that $\phi = w$ on $\partial\Omega$, which ensures that $\rho = \text{sigmoid}(\phi/\delta) \simeq 1$ on $\partial\Omega$ (recall that w is such that $\text{sigmoid}(w/2\delta) \simeq 1$). For the elastic layer, we introduce the transformations

$$\bar{u}_1(\mathbf{x};\boldsymbol{\theta}_1) = (y+0.5)\,\bar{u}'_1(\cos x, \sin x, y;\boldsymbol{\theta}_1), \tag{C.16a}$$
$$\bar{u}_2(\mathbf{x};\boldsymbol{\theta}_2) = (y+0.5)\,\bar{u}'_2(\cos x, \sin x, y;\boldsymbol{\theta}_2), \tag{C.16b}$$
$$\bar{\sigma}_{11}(\mathbf{x};\boldsymbol{\theta}_3) = \bar{\sigma}'_{11}(\cos x, \sin x, y;\boldsymbol{\theta}_3), \tag{C.16c}$$
$$\bar{\sigma}_{22}(\mathbf{x};\boldsymbol{\theta}_4) = y\,\bar{\sigma}'_{22}(\cos x, \sin x, y;\boldsymbol{\theta}_4) - P_o, \tag{C.16d}$$
$$\bar{\sigma}_{12}(\mathbf{x};\boldsymbol{\theta}_5) = y\,\bar{\sigma}'_{12}(\cos x, \sin x, y;\boldsymbol{\theta}_5), \tag{C.16e}$$
$$\bar{\phi}(\mathbf{x};\boldsymbol{\theta}_\phi) = y(y+0.5)\,\bar{\phi}'(\cos x, \sin x, y;\boldsymbol{\theta}_\phi) + 8w(y+0.25), \tag{C.16f}$$

so that the neural network approximations defined in (C.13) obey by construction the boundary conditions (C.7) and are periodic along the x direction. Further, since we know that the elastic material is present all along the top surface $y = 0$ and the rigid substrate is present all along the bottom surface $y = -0.5$, we define $\bar{\phi}$ so that $\phi = w$ for $y = 0$ and $\phi = -w$ for $y = -0.5$, which ensures that $\rho = \text{sigmoid}(\phi/\delta) \simeq 1$ for $y = 0$ and $\rho \simeq 0$ for $y = -0.5$.

C.2.2 Large-deformation hyperelasticity

The governing physical laws presented in Appendix C.1.2 show that the problem is described by the field of physical quantities $\boldsymbol{\psi} = (u_1, u_2, S_{11}, S_{22}, S_{12}, S_{21}, p)$ in addition to the material distribution ρ. We therefore introduce the neural network

approximations

$$u_1(\mathbf{x}) = \bar{u}_1(\mathbf{x}; \boldsymbol{\theta}_1), \tag{C.17a}$$
$$u_2(\mathbf{x}) = \bar{u}_2(\mathbf{x}; \boldsymbol{\theta}_2), \tag{C.17b}$$
$$S_{11}(\mathbf{x}) = \bar{S}_{11}(\mathbf{x}; \boldsymbol{\theta}_3), \tag{C.17c}$$
$$S_{22}(\mathbf{x}) = \bar{S}_{22}(\mathbf{x}; \boldsymbol{\theta}_4), \tag{C.17d}$$
$$S_{12}(\mathbf{x}) = \bar{S}_{12}(\mathbf{x}; \boldsymbol{\theta}_5), \tag{C.17e}$$
$$S_{21}(\mathbf{x}) = \bar{S}_{21}(\mathbf{x}; \boldsymbol{\theta}_6), \tag{C.17f}$$
$$p(\mathbf{x}) = \bar{p}(\mathbf{x}; \boldsymbol{\theta}_7), \tag{C.17g}$$
$$\phi(\mathbf{x}) = \bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi), \tag{C.17h}$$

and the material distribution is given by $\rho(\mathbf{x}) = \bar{\rho}(\mathbf{x}; \boldsymbol{\theta}_\phi) = \text{sigmoid}(\bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi)/\delta)$. We then formulate the loss function (4.5) by specializing the loss term expressions presented in Section 4.2.2 to the linear elasticity problem, using the governing equations given in Appendix C.1.1. Omitting the $\boldsymbol{\theta}$'s for notational simplicity, we obtain

$$\mathcal{L}_{\text{meas}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}) = \frac{1}{|\partial\Omega^m|} \sum_{\mathbf{x}_i \in \partial\Omega^m} |\bar{\mathbf{u}}(\mathbf{x}_i) - \mathbf{u}_i^m|^2, \tag{C.18a}$$

$$\begin{aligned}\mathcal{L}_{\text{gov}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_\phi) = &\frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\mathbf{r}_{\text{eq}}(\bar{\mathbf{S}}(\mathbf{x}_i))|^2 + \frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\mathbf{r}_{\text{inc}}(\bar{\mathbf{u}}(\mathbf{x}_i), \bar{\rho}(\mathbf{x}_i))|^2 \\ &+ \frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\mathbf{r}_{\text{cr}}(\bar{\mathbf{u}}(\mathbf{x}_i), \bar{\mathbf{S}}(\mathbf{x}_i), \bar{p}(\mathbf{x}_i), \bar{\rho}(\mathbf{x}_i))|^2,\end{aligned} \tag{C.18b}$$

$$\mathcal{L}_{\text{eik}}(\boldsymbol{\theta}_\phi) = \frac{1}{|\Omega^d_{\text{eik}}|} \sum_{\mathbf{x}_i \in \Omega^d_{\text{eik}}} \left(|\nabla\bar{\phi}(\mathbf{x}_i)| - 1\right)^2, \tag{C.18c}$$

where $\bar{\mathbf{u}} = (\bar{u}_1, \bar{u}_2)$ and $\bar{\mathbf{S}}$ has components $\bar{S}_{i,j}$, $i, j = 1, 2$. In (C.18b), the terms $\mathbf{r}_{\text{eq}}$, $\mathbf{r}_{\text{inc}}$, and $\mathbf{r}_{\text{cr}}$ refer to the residuals of the equilibrium equation (C.8), the incompressibility constraint (C.10), and the constitutive relation (C.9). The eikonal loss term is problem-independent and therefore identical to (4.9).

As in the linear elasticity case, we design the architecture of the neural networks in such a way that they inherently satisfy the boundary conditions. For the elastic

matrix problem,

$$\bar{u}_1(\mathbf{x};\boldsymbol{\theta}_1) = \bar{u}_1'(\mathbf{x};\boldsymbol{\theta}_1), \tag{C.19a}$$
$$\bar{u}_2(\mathbf{x};\boldsymbol{\theta}_2) = \bar{u}_2'(\mathbf{x};\boldsymbol{\theta}_2), \tag{C.19b}$$
$$\bar{S}_{11}(\mathbf{x};\boldsymbol{\theta}_3) = (x-0.5)(x+0.5)\,\bar{S}_{11}'(\mathbf{x};\boldsymbol{\theta}_3) + P_o, \tag{C.19c}$$
$$\bar{S}_{22}(\mathbf{x};\boldsymbol{\theta}_4) = (y-0.5)(y+0.5)\,\bar{S}_{22}'(\mathbf{x};\boldsymbol{\theta}_4), \tag{C.19d}$$
$$\bar{S}_{12}(\mathbf{x};\boldsymbol{\theta}_5) = (y-0.5)(y+0.5)\,\bar{S}_{12}'(\mathbf{x};\boldsymbol{\theta}_5), \tag{C.19e}$$
$$\bar{S}_{21}(\mathbf{x};\boldsymbol{\theta}_6) = (x-0.5)(x+0.5)\,\bar{S}_{21}'(\mathbf{x};\boldsymbol{\theta}_6), \tag{C.19f}$$
$$\bar{p}(\mathbf{x};\boldsymbol{\theta}_7) = \bar{p}'(\mathbf{x};\boldsymbol{\theta}_7), \tag{C.19g}$$
$$\bar{\phi}(\mathbf{x};\boldsymbol{\theta}_\phi) = (x-0.5)(x+0.5)(y-0.5)(y+0.5)\,\bar{\phi}'(\mathbf{x};\boldsymbol{\theta}_\phi) + w, \tag{C.19h}$$

where the quantities with a prime denote the raw output of the neural network. In this way, the neural network approximations defined in (C.17) obey by construction the boundary conditions of the problem. As before, since we know that the elastic material is present all along the outer surface $\partial\Omega$, we define $\bar{\phi}$ so that $\phi = w$ on $\partial\Omega$, which ensures that $\rho = \text{sigmoid}(\phi/\delta) \simeq 1$ on $\partial\Omega$.

C.3 Neural network architecture and training

Here, we provide implementation details regarding the architecture of the deep neural networks, the training procedure and corresponding parameter values.

C.3.1 Neural network architecture

State variable fields of the form $\psi(\mathbf{x})$ are approximated using deep fully-connected neural networks that map the location $\mathbf{x}$ to the corresponding value of ψ at that location. This map can be expressed as $\psi(\mathbf{x}) = \bar{\psi}(\mathbf{x};\boldsymbol{\theta})$, and is defined by the sequence of operations

$$\mathbf{z}^0 = \mathbf{x}, \tag{C.20a}$$
$$\mathbf{z}^k = \sigma(\mathbf{W}^k\mathbf{z}^{k-1} + \mathbf{b}^k), \quad 1 \le k \le \ell - 1, \tag{C.20b}$$
$$\psi = \mathbf{z}^\ell = \mathbf{W}^\ell\mathbf{z}^{\ell-1} + \mathbf{b}^\ell. \tag{C.20c}$$

The input $\mathbf{x}$ is propagated through ℓ layers, all of which (except the last) take the form of a linear operation composed with a nonlinear transformation. Each layer outputs a vector $\mathbf{z}^k \in \mathbb{R}^{q_k}$, where q_k is the number of 'neurons', and is defined by a weight matrix $\mathbf{W}^k \in \mathbb{R}^{q_k \times q_{k-1}}$, a bias vector $\mathbf{b}^k \in \mathbb{R}^{q_k}$, and a nonlinear activation function $\sigma(\cdot)$. Finally, the output of the last layer is assigned to ψ. The weight matrices and bias vectors, which parametrize the map from $\mathbf{x}$ to ψ, form a set of trainable parameters $\boldsymbol{\theta} = \{\mathbf{W}^k, \mathbf{b}^k\}_{k=1}^{\ell}$.

The choice of the nonlinear activation function $\sigma(\cdot)$ and the initialization procedure for the trainable parameters $\boldsymbol{\theta}$ are both important factors in determining the

performance of neural networks. While the tanh function has been a popular candidate in the context of PINNs [120], recent works by Refs. [168, 192] have shown that using sinusoidal activation functions can lead to improved training performance by promoting the emergence of small-scale features. In this work, we select the sinusoidal representation network (SIREN) architecture from Ref. [168], which combines the use of the sine as an activation function with a specific way to initialize the trainable parameters $\boldsymbol{\theta}$ that ensures that the distribution of the input to each sine activation function remains unchanged over successive layers. Specifically, each component of $\mathbf{W}^k$ is uniformly distributed between $-\sqrt{6/q_k}$ and $\sqrt{6/q_k}$ where q_k is the number of neurons in layer k, and $\mathbf{b}^k = \mathbf{0}$, for $k = 1, \ldots, \ell$. Further, the first layer of the SIREN architecture is $\mathbf{z}^1 = \sigma(\omega_0 \mathbf{W}^1 \mathbf{z}^0 + \mathbf{b}^1)$ instead of (C.20b), with the extra scalar ω_0 serving to promote higher-frequency content in the output.

C.3.2 Training procedure

We construct the total loss function (4.5) and train the neural networks in TensorFlow 2. The training is performed using ADAM, a first-order gradient-descent-based algorithm with adaptive step size [102]. Three training tricks resulted in noticeably improved training performance and consistency:

- First, we found that pretraining the level-set neural network $\phi(\mathbf{x}) = \bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi)$ in a standard supervised setting leads to much more consistent results over different initializations of the neural networks. During this pretraining step, carried out before the main optimization step in which all neural networks are trained to minimize the loss (4.5), we minimize the mean-square error

$$\mathcal{L}_{\text{sup}}(\boldsymbol{\theta}_\phi) = \frac{1}{|\Omega^d|} \sum_{\mathbf{x}_i \in \Omega^d} |\bar{\phi}(\mathbf{x}_i; \boldsymbol{\theta}_\phi) - \phi_i|, \tag{C.21}$$

  where Ω^d is the same set of collocation points as in (4.7), the supervised labels $\phi_i = |\mathbf{x}_i| - 0.25$ for the elastic matrix, and $\phi_i = y_i + 0.25$ for the elastic layer. The material density $\bar{\rho}(\mathbf{x}; \boldsymbol{\theta}_\phi) = \text{sigmoid}(\bar{\phi}(\mathbf{x}; \boldsymbol{\theta}_\phi)/\delta)$ obtained at the end of this pretraining step is one outside a circle of radius 0.25 centered at the origin for the elastic matrix, and it is one above the horizontal line $y = -0.25$ for the elastic layer. This choice for the supervised labels is justified by the fact that ρ is known to be one along the outer boundary of the domain Ω for the elastic matrix, and it is known to be one (zero) along the top (bottom) boundary of Ω for the elastic layer.

- Second, during the main optimization in which all neural networks are trained to minimize the loss (4.5), we evaluate the loss component $\mathcal{L}_{\text{gov}}$ in (4.7) using a different subset of residual points from Ω^d at every iteration. Although this practice is well-known in machine learning, where it is referred to as mini-batch learning, it is rarely employed in the PINN community except for some recent studies [193]. In our case, we choose to divide the set Ω^d into 10 different

mini-batchs (subsets) of size $|\Omega^d|/10$, which are then employed sequentially to evaluate $\mathcal{L}_{\text{gov}}$ during each subsequent gradient update

$$\boldsymbol{\theta}_{\boldsymbol{\psi}}^{k+1} = \boldsymbol{\theta}_{\boldsymbol{\psi}}^{k} - \alpha_{\boldsymbol{\psi}}(k)\nabla_{\boldsymbol{\theta}_{\boldsymbol{\psi}}}\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{\psi}}^{k}, \boldsymbol{\theta}_{\phi}^{k}), \tag{C.22a}$$

$$\boldsymbol{\theta}_{\phi}^{k+1} = \boldsymbol{\theta}_{\phi}^{k} - \alpha_{\phi}(k)\nabla_{\boldsymbol{\theta}_{\phi}}\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{\psi}}^{k}, \boldsymbol{\theta}_{\phi}^{k}). \tag{C.22b}$$

An epoch of training, which is defined as one complete pass through the whole set Ω^d, therefore consists of 10 gradient updates.

- Third, the initial nominal step size $\alpha_{\boldsymbol{\psi}}$ governing the learning rate of the physical quantities neural networks is set to be 10 times larger than its counterpart α_{ϕ} governing the learning rate of the level-set neural network. This results in a separation of time scales between the rate of change of the physical quantities neural networks and that of the level-set neural network, which is motivated by the idea that physical quantities should be given time to adapt to a given geometry before the geometry itself changes.

C.3.3 Parameter values

The parameter values described below apply to all results presented in this paper.

- **Neural network architecture.** We opted for neural networks with 4 hidden layers of 50 neurons each, which we found to be a good compromise between expressivity and training time. Further, we choose $\omega_0 = 10$ as the scalar appearing in the first layer of the SIREN architecture.

- **Residual and measurements points.** In the elastic matrix problem, we consider that the boundary displacement is measured along each of the four external boundaries at 100 equally-spaced points, which amounts to $|\partial\Omega^m| = 400$. In the elastic layer problem, we consider that the boundary displacement is measured along the top boundary at 100 equally-spaced points, which amounts to $|\partial\Omega^m| = 100$. For both geometries, the set of residual points $|\Omega^d|$ consists of 10000 points distributed in Ω with a Latin Hypercube Sampling (LHS) strategy, yielding 10 mini-batchs containing 1000 points each.

- **Training parameters.** The pretraining of the level-set neural network is carried out using the ADAM optimizer with nominal step size 10^{-3} over 800 training epochs, employing the whole set Ω^d to compute the gradient of $\mathcal{L}_{\text{sup}}$ at each update step. The main optimization, during which all neural networks are trained to minimize the total loss (4.5), is carried out using the ADAM optimizer over a total of 150k training epochs. For the first 40k epochs, we use a nominal step size 10^{-4} for the level-set neural network and 10^{-3} for the other neural networks. For the next 40k epochs, we use a nominal step size 10^{-4} for all neural networks. Finally, we use a nominal step size 10^{-5} for the remaining 70k epochs. The scalar weights in the loss (4.5) are assigned the values $\lambda_{\text{meas}} = 10$, $\lambda_{\text{gov}} = 1$, and $\lambda_{\text{eik}} = 4$. Further, we multiply the second term of $\mathcal{L}_{\text{gov}}$ in (C.14b) and (C.18b) with a scalar weight $\lambda_{\text{cr}} = 4$.

C.4 Finite-element method simulations

The finite-element method simulations that provide the boundary displacement data are performed in the software Abaqus, using its Standard (implicit) solver. For both the elastic matrix and elastic layer geometries, every considered case is meshed using a linear density of 200 elements per boundary, corresponding to between 25k to 40k total elements depending on the number and shapes of voids or inclusions. We employ bilinear quadrilateral CPE4 plain-strain elements for the small-deformation linear elastic cases, and their hybrid constant-pressure counterpart CPE4H for the large-deformation hyperelastic cases. For the elastic matrix, we apply a load $P_o/E = 0.05$ in the linear elastic case, and a load $P_o/E = 0.173$ in the hyperelastic case. For the elastic layer, we apply a load $P_o/E = 0.1$.

Bibliography

[1] G. Allaire, F. Jouve, and A.-M. Toader. Structural optimization using sensitivity analysis and a level-set method. *Journal of computational physics*, 194(1): 363–393, 2004.

[2] H. B. Ameur, M. Burger, and B. Hackl. Level set methods for geometric inverse problems in linear elasticity. *Inverse Problems*, 20(3):673, 2004.

[3] B. Andreotti, Y. Forterre, and O. Pouliquen. *Granular media: between fluid and solid*. Cambridge University Press, 2013.

[4] D. D. Ang, D. D. Trong, and M. Yamamoto. Identification of cavities inside two-dimensional heterogeneous isotropic elastic bodies. *Journal of elasticity*, 56 (3):199–212, 1999.

[5] I. S. Aranson and L. S. Tsimring. Continuum theory of partially fluidized granular flows. *Physical Review E*, 65(6):061303, 2002.

[6] R. Artoni and P. Richard. Effective wall friction in wall-bounded 3d dense granular flows. *Physical review letters*, 115(15):158001, 2015.

[7] R. Artoni, A. C. Santomaso, M. Go, and P. Canu. Scaling laws for the slip velocity in dense granular flows. *Physical review letters*, 108(23):238002, 2012.

[8] R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.

[9] J. Barber and M. Ciavarella. Contact mechanics. *International Journal of solids and structures*, 37(1-2):29–43, 2000.

[10] J. Barber and M. Ciavarella. Jkr solution for an anisotropic half space. *Journal of the Mechanics and Physics of Solids*, 64:367–376, 2014.

[11] J. Barber and F. Sturla. Application of the reciprocal theorem to some problems for the elastic half-space. *Journal of the Mechanics and Physics of Solids*, 40 (1):17–25, 1992.

[12] J. R. Barber. *Contact mechanics*, volume 250. Springer, 2018.

[13] D. M. Barnett and J. Lothe. Line force loadings on anisotropic half-spaces and wedges. *Physica Norvegica*, 8:13–22, 1975.

[14] M. P. Bendsøe. Optimal shape design as a material distribution problem. *Structural optimization*, 1(4):193–202, 1989.

[15] M. P. Bendsøe and O. Sigmund. Material interpolation schemes in topology optimization. *Archive of applied mechanics*, 69(9):635–654, 1999.

[16] M. P. Bendsoe and O. Sigmund. *Topology optimization: theory, methods, and applications*. Springer Science & Business Media, 2003.

[17] A. Bhateja and D. V. Khakhar. Rheology of dense granular flows in two dimensions: Comparison of fully two-dimensional flows to unidirectional shear flow. *Physical Review Fluids*, 3(6):062301, 2018.

[18] D. Bi, J. Zhang, B. Chakraborty, and R. P. Behringer. Jamming by shear. *Nature*, 480(7377):355–358, 2011.

[19] L. Bocquet, A. Colin, and A. Ajdari. Kinetic theory of plastic flow in soft glassy materials. *Physical review letters*, 103(3):036001, 2009.

[20] N. Boechler, G. Theocharis, S. Job, P. Kevrekidis, M. A. Porter, and C. Daraio. Discrete breathers in one-dimensional diatomic granular crystals. *Physical review letters*, 104(24):244302, 2010.

[21] N. Boechler, J. Yang, G. Theocharis, P. Kevrekidis, and C. Daraio. Tunable vibrational band gaps in one-dimensional diatomic granular crystals with three-particle unit cells. *Journal of Applied Physics*, 109(7):074906, 2011.

[22] M. Bonnet and A. Constantinescu. Inverse problems in elasticity. *Inverse problems*, 21(2):R1, 2005.

[23] T. Borrvall and J. Petersson. Topology optimization of fluids in stokes flow. *International journal for numerical methods in fluids*, 41(1):77–107, 2003.

[24] M. Bouzid, M. Trulsson, P. Claudin, E. Clément, and B. Andreotti. Nonlocal rheology of granular flows across yield conditions. *Physical review letters*, 111 (23):238301, 2013.

[25] N. V. Brilliantov, F. Spahn, J.-M. Hertzsch, and T. Pöschel. Model for collisions in granular gases. *Physical review E*, 53(5):5382, 1996.

[26] L. Brillouin. *Wave propagation in periodic structures: electric filters and crystal lattices*. Dover, 1953.

[27] T. Burczynski and W. Beluch. The identification of cracks using boundary elements and evolutionary algorithms. *Engineering Analysis with Boundary Elements*, 25(4-5):313–322, 2001.

[28] S. Cai, Z. Wang, S. Wang, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143(6), 2021.

[29] C. S. Campbell. Granular shear flows at the elastic limit. *Journal of Fluid Mechanics*, 465:261, 2002.

[30] S.-K. Chan, Y. Fang, M. Grimsditch, Z. Li, M. V. Nevitt, W. M. Robertson, and E. S. Zouboulis. Temperature dependence of the elastic moduli of monoclinic zirconia. *Journal of the American Ceramic Society*, 74(7):1742–1744, 1991.

[31] P. Chaudhuri, V. Mansard, A. Colin, and L. Bocquet. Dynamical flow arrest in confined gravity driven flows of soft jammed particles. *Physical review letters*, 109(3):036001, 2012.

[32] Y. Chen and L. Dal Negro. Physics-informed neural networks for imaging and parameter retrieval of photonic nanostructures from near-field data. *APL Photonics*, 7(1):010802, 2022.

[33] N. a. Cho, C. Martin, and D. Sego. A clumped particle model for rock. *International Journal of Rock Mechanics and Mining Sciences*, 44(7):997–1010, 2007.

[34] M. P. Ciamarra, R. Pastore, M. Nicodemi, and A. Coniglio. Jamming phase diagram for frictional particles. *Physical Review E*, 84(4):041308, 2011.

[35] A. H. Clark, L. Kondic, and R. P. Behringer. Particle scale dynamics in granular impact. *Physical review letters*, 109(23):238302, 2012.

[36] A. H. Clark, M. D. Shattuck, N. T. Ouellette, and C. S. O'Hern. Onset and cessation of motion in hydrodynamically sheared granular beds. *Physical Review E*, 92(4):042202, 2015.

[37] A. H. Clark, J. D. Thompson, M. D. Shattuck, N. T. Ouellette, and C. S. O'Hern. Critical scaling near the yielding transition in granular media. *Physical Review E*, 97(6):062901, 2018.

[38] S. Courrech du Pont. *Avalanches granulaires en milieu fluide*. PhD thesis, Université Paris XI, 2003.

[39] S. Courrech du Pont, P. Gondret, B. Perrin, and M. Rabaud. Granular avalanches in fluids. *Physical review letters*, 90(4):044301, 2003.

[40] I. R. Crystal, A. Lai, and C. A. Schuh. Cyclic martensitic transformations and damage evolution in shape memory zirconia: Single crystals vs polycrystals. *Journal of the American Ceramic Society*, 2020.

[41] P. A. Cundall and O. D. Strack. A discrete numerical model for granular assemblies. *geotechnique*, 29(1):47–65, 1979.

[42] F. Da Cruz, F. Chevoir, D. Bonn, and P. Coussot. Viscosity bifurcation in granular materials, foams, and emulsions. *Physical Review E*, 66(5):051305, 2002.

[43] F. Da Cruz, S. Emam, M. Prochnow, J.-N. Roux, and F. Chevoir. Rheophysics of dense granular materials: Discrete simulation of plane shear flows. *Physical Review E*, 72(2):021309, 2005.

[44] F. Da Cruz, S. Emam, M. Prochnow, J.-N. Roux, and F. Chevoir. Rheophysics of dense granular materials: Discrete simulation of plane shear flows. *Physical Review E*, 72(2):021309, 2005.

[45] A. Daerr and S. Douady. Two types of avalanche behaviour in granular media. *Nature*, 399(6733):241, 1999.

[46] T. Dbouk. A review about the engineering design of optimal heat transfer systems using topology optimization. *Applied Thermal Engineering*, 112:841–854, 2017.

[47] E. DeGiuli and M. Wyart. Friction law and hysteresis in granular materials. *Proceedings of the National Academy of Sciences*, 114(35):9284–9289, 2017.

[48] A. Delafargue and F.-J. Ulm. Explicit approximations of the indentation modulus of elastically orthotropic solids for conical indenters. *International journal of solids and structures*, 41(26):7351–7360, 2004.

[49] G. Dhanaraj, K. Byrappa, V. Prasad, and M. Dudley. *Springer handbook of crystal growth*. Springer Science & Business Media, 2010.

[50] A. Di Renzo and F. P. Di Maio. An improved integral non-linear model for the contact of particles in distinct element simulations. *Chemical engineering science*, 60(5):1303–1312, 2005.

[51] J. A. Dijksman, G. H. Wortel, L. T. van Dellen, O. Dauchot, and M. van Hecke. Jamming, yielding, and rheology of weakly vibrated granular media. *Physical review letters*, 107(10):108303, 2011.

[52] E. Dintwa, M. Van Zeebroeck, E. Tijskens, and H. Ramon. Torsion of viscoelastic spheres in contact. *Granular Matter*, 7(2-3):169–179, 2005.

[53] M. Dissanayake and N. Phan-Thien. Neural-network-based approximations for solving partial differential equations. *Communications in Numerical Methods in Engineering*, 10(3):195–201, 1994.

[54] T. Divoux, M. A. Fardin, S. Manneville, and S. Lerouge. Shear banding of complex fluids. *Annual Review of Fluid Mechanics*, 48:81–103, 2016.

[55] M. M. Doyley. Model-based elastography: a survey of approaches to the inverse elasticity problem. *Physics in Medicine & Biology*, 57(3):R35, 2012.

[56] Z. Du, X. M. Zeng, Q. Liu, C. A. Schuh, and C. L. Gan. Superelasticity in micro-scale shape memory ceramic particles. *Acta Materialia*, 123:255–263, 2017.

[57] M. B. Dühring, J. S. Jensen, and O. Sigmund. Acoustic design by topology optimization. *Journal of sound and vibration*, 317(3-5):557–575, 2008.

[58] A. Džiugys and B. Peters. An approach to simulate the motion of spherical and non-spherical fuel particles in combustion chambers. *Granular matter*, 3 (4):231–266, 2001.

[59] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[60] H. A. Eschenauer, V. V. Kobelev, and A. Schumacher. Bubble method for topology and shape optimization of structures. *Structural optimization*, 8(1): 42–51, 1994.

[61] D. J. Evans. On the representation of orientation space. *Molecular physics*, 34 (2):317–325, 1977.

[62] F. Fazelpour and K. E. Daniels. The effect of boundary roughness on dense granular flows. In *EPJ Web of Conferences*, volume 249, page 03014. EDP Sciences, 2021.

[63] Y. Forterre and O. Pouliquen. Long-surface-wave instability in dense granular flows. *Journal of Fluid Mechanics*, 486:21–50, 2003.

[64] Y. Forterre and O. Pouliquen. Flows of dense granular media. *Annu. Rev. Fluid Mech.*, 40:1–24, 2008.

[65] Y. Gao and G. M. Pharr. Multidimensional contact moduli of elastically anisotropic solids. *Scripta materialia*, 57(1):13–16, 2007.

[66] V. M. García-Chocano, L. Sanchis, A. Díaz-Rubio, J. Martínez-Pastor, F. Cervera, R. Llopis-Pontiveros, and J. Sánchez-Dehesa. Acoustic cloak for airborne sound by inverse design. *Applied physics letters*, 99(7):074102, 2011.

[67] J. Gaume, G. Chambon, and M. Naaim. Quasistatic to inertial transition in granular materials and the role of fluctuations. *Physical Review E*, 84(5):051304, 2011.

[68] C. Goujon, N. Thomas, and B. Dalloz-Dubrujeaud. Monodisperse dry granular flows on inclined planes: role of roughness. *The European Physical Journal E*, 11(2):147–157, 2003.

[69] J. Goyon, A. Colin, G. Ovarlez, A. Ajdari, and L. Bocquet. Spatial cooperativity in soft glassy flows. *Nature*, 454(7200):84–87, 2008.

[70] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman. Implicit geometric regularization for learning shapes. In *International Conference on Machine Learning*, pages 3789–3799. PMLR, 2020.

[71] X. Guo, W. Zhang, and W. Zhong. Doing topology optimization explicitly and geometrically?a new moving morphable components based framework. *Journal of Applied Mechanics*, 81(8), 2014.

[72] Y. Guo and J. S. Curtis. Discrete element method simulations for complex granular flows. *Annual Review of Fluid Mechanics*, 47:21–46, 2015.

[73] M. E. Gurtin, E. Fried, and L. Anand. *The mechanics and thermodynamics of continua*. Cambridge University Press, 2010.

[74] E. Haghighat, M. Raissi, A. Moure, H. Gomez, and R. Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379: 113741, 2021.

[75] S. Hall and J. Wright. Three-dimensional experimental granular mechanics. *Géotechnique Letters*, 5(4):236–242, 2015.

[76] Z. Y. Hang, M. Hassani-Gangaraj, Z. Du, C. L. Gan, and C. A. Schuh. Granular shape memory ceramic packings. *Acta Materialia*, 132:455–466, 2017.

[77] J. Haslinger and R. A. Mäkinen. *Introduction to shape optimization: theory, approximation, and computation*. SIAM, 2003.

[78] E. J. Haug. *Computer aided kinematics and dynamics of mechanical systems*, volume 1. Allyn and Bacon Boston, 1989.

[79] D. L. Henann and K. Kamrin. A predictive, size-dependent continuum model for dense granular flows. *Proceedings of the National Academy of Sciences*, 110 (17):6730–6735, 2013.

[80] D. L. Henann and K. Kamrin. Continuum modeling of secondary rheology in dense granular materials. *Physical review letters*, 113(17):178001, 2014.

[81] A. Henkes, H. Wessels, and R. Mahnken. Physics informed neural networks for continuum micromechanics. *Computer Methods in Applied Mechanics and Engineering*, 393:114790, 2022.

[82] E. Herbold, J. Kim, V. Nesterenko, S. Wang, and C. Daraio. Pulse propagation in a linear and nonlinear diatomic periodic chain: effects of acoustic frequency band-gap. *Acta Mechanica*, 205(1-4):85–103, 2009.

[83] H. Hertz. über die berührung fester elastischer körper. *Journal für die reine und angewandte Mathematik*, 92:156–171, 1882.

[84] P. Heyliger, H. Ledbetter, and S. Kim. Elastic constants of natural quartz. *The Journal of the Acoustical Society of America*, 114(2):644–650, 2003.

[85] A.-C. Hladky-Hennion, G. Allan, and M. de Billy. Localized modes in a one-dimensional diatomic chain of coupled spheres. *Journal of applied physics*, 98 (5):054909, 2005.

[86] D. Howell, R. Behringer, and C. Veje. Stress fluctuations in a 2d granular couette experiment: a continuous transition. *Physical Review Letters*, 82(26): 5241, 1999.

[87] R. Hurley, S. Hall, J. Andrade, and J. Wright. Quantifying interparticle forces and heterogeneity in 3d granular materials. *Physical review letters*, 117(9): 098005, 2016.

[88] H. Jaeger, C.-H. Liu, S. Nagel, and T. Witten. Friction in granular flows. *EPL (Europhysics Letters)*, 11(7):619, 1990.

[89] J. S. Jensen. Phononic band gaps and vibrations in one-and two-dimensional mass–spring structures. *Journal of Sound and Vibration*, 266(5):1053–1078, 2003.

[90] J. S. Jensen and O. Sigmund. Topology optimization for nano-photonics. *Laser & Photonics Reviews*, 5(2):308–321, 2011.

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

[92] K. L. Johnson. *Contact mechanics*. Cambridge university press, 1987.

[93] K. L. Johnson, K. Kendall, and a. Roberts. Surface energy and the contact of elastic solids. *Proceedings of the royal society of London. A. mathematical and physical sciences*, 324(1558):301–313, 1971.

[94] J. Jung and E. Taciroglu. Modeling and identification of an arbitrarily shaped scatterer using dynamic xfem with cubic splines. *Computer Methods in Applied Mechanics and Engineering*, 278:101–118, 2014.

[95] M. Kadic, G. W. Milton, M. van Hecke, and M. Wegener. 3d metamaterials. *Nature Reviews Physics*, 1(3):198–210, 2019.

[96] K. Kamrin. Non-locality in granular flow: Phenomenology and modeling approaches. *Frontiers in Physics*, 7:116, 2019.

[97] K. Kamrin and D. L. Henann. Nonlocal modeling of granular flows down inclines. *Soft matter*, 11(1):179–185, 2015.

[98] K. Kamrin and G. Koval. Nonlocal constitutive relation for steady granular flow. *Physical Review Letters*, 108(17):178301, 2012.

[99] K. Kamrin and G. Koval. Effect of particle surface friction on nonlocal constitutive behavior of flowing granular media. *Computational Particle Mechanics*, 1(2):169–176, 2014.

[100] A. Karageorghis, D. Lesnic, and L. Marin. The method of fundamental solutions for the detection of rigid inclusions and cavities in plane linear elastic bodies. *Computers & structures*, 106:176–188, 2012.

[101] S. Kim and K. Kamrin. Power-law scaling in granular rheology across flow geometries. *Physical Review Letters*, 125(8):088002, 2020.

[102] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[103] C. Kittel. *Introduction to solid state physics*, volume 8. Wiley New York, 1976.

[104] C. Klein and B. Dutrow. *Manual of mineral science*. Wiley New York, 2007.

[105] H. T. Kollmann, D. W. Abueidda, S. Koric, E. Guleryuz, and N. A. Sobh. Deep learning for topology optimization of 2d metamaterials. *Materials & Design*, 196:109098, 2020.

[106] G. Koval, J.-N. Roux, A. Corfdir, and F. Chevoir. Annular shear of cohesionless granular materials: From the inertial to quasistatic regime. *Physical Review E*, 79(2):021306, 2009.

[107] H. Kruggel-Emden, E. Simsek, S. Rickelt, S. Wirtz, and V. Scherer. Review and extension of normal force models for the discrete element method. *Powder Technology*, 171(3):157–173, 2007.

[108] O. Kuwano, R. Ando, and T. Hatano. Crossover from negative to positive shear rate dependence in granular friction. *Geophysical Research Letters*, 40(7): 1295–1299, 2013.

[109] I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks*, 9(5):987–1000, 1998.

[110] L. Landau and E. Lifshitz. *Theory of elasticity*. Pergamon Press, 1970.

[111] L. D. Landau, E. M. Lifšic, E. M. Lifshitz, A. M. Kosevich, and L. P. Pitaevskii. *Theory of elasticity: volume 7*, volume 7. Elsevier, 1986.

[112] H. S. Lee, Y. H. Kim, C. J. Park, and H. W. Park. A new spatial regularization scheme for the identification of the geometric shape of an inclusion in a finite body. *International Journal for Numerical Methods in Engineering*, 46(7):973–992, 1999.

[113] H. S. Lee, C. J. Park, and H. W. Park. Identification of geometric shapes and material properties of inclusions in two-dimensional finite bodies by boundary parameterization. *Computer Methods in Applied Mechanics and Engineering*, 181(1-3):1–20, 2000.

[114] K.-L. Lee and F.-L. Yang. Relaxation-type nonlocal inertial-number rheology for dry granular flows. *Physical Review E*, 96(6):062909, 2017.

[115] S. Li and D. L. Henann. Nonlocal continuum modeling of dense granular flow in a split-bottom cell with a vane-shaped intruder. *Physical Review E*, 102(2): 022908, 2020.

[116] E. W. C. Lim, Y. Zhang, and C.-H. Wang. Effects of an electrostatic field in pneumatic conveying of granular materials through inclined and vertical pipes. *Chemical Engineering Science*, 61(24):7889–7908, 2006.

[117] D. Liu and D. L. Henann. Non-local continuum modelling of steady, dense granular heap flows. *Journal of Fluid Mechanics*, 831:212–227, 2017.

[118] D. Liu and D. L. Henann. Size-dependence of the flow threshold in dense granular materials. *Soft matter*, 14(25):5294–5305, 2018.

[119] C.-Y. D. Lu, P. D. Olmsted, and R. Ball. Effects of nonlocal stress on the determination of shear banding flow. *Physical Review Letters*, 84(4):642, 2000.

[120] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021.

[121] L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, and S. G. Johnson. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6):B1105–B1132, 2021.

[122] A. Lucas, A. Mangeney, and J. P. Ampuero. Frictional velocity-weakening in landslides on earth and on other planetary bodies. *Nature communications*, 5 (1):1–9, 2014.

[123] S. Luding. Introduction to discrete element methods: basic of contact force models and how to perform the micro-macro transition to continuum theory. *European journal of environmental and civil engineering*, 12(7-8):785–826, 2008.

[124] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu. Deep learning for the design of photonic structures. *Nature Photonics*, 15(2):77–90, 2021.

[125] T. S. Majmudar and R. P. Behringer. Contact force measurements and stress-induced anisotropy in granular materials. *Nature*, 435(7045):1079–1082, 2005.

[126] J. Marshall. Discrete-element modeling of particulate aerosol flows. *Journal of Computational Physics*, 228(5):1541–1561, 2009.

[127] J. R. Martins and A. Ning. *Engineering design optimization*. Cambridge University Press, 2021.

[128] Y. Mei, R. Fulmer, V. Raja, S. Wang, and S. Goenezen. Estimating the non-homogeneous elastic modulus distribution from surface deformations. *International Journal of Solids and Structures*, 83:73–80, 2016.

[129] Y. Mei, Z. Du, D. Zhao, W. Zhang, C. Liu, and X. Guo. Moving morphable inclusion approach: an explicit framework to solve inverse problem in elasticity. *Journal of Applied Mechanics*, 88(4), 2021.

[130] C. Meier, R. Weissbach, J. Weinberg, W. A. Wall, and A. J. Hart. Modeling and characterization of cohesion in fine metal powders with a focus on additive manufacturing process simulations. *Powder technology*, 343:855–866, 2019.

[131] S. Mellings and M. Aliabadi. Flaw identification using the boundary element method. *International journal for numerical methods in engineering*, 38(3): 399–419, 1995.

[132] H. J. Melosh. Acoustic fluidization: A new geologic process? *Journal of Geophysical Research: Solid Earth*, 84(B13):7513–7520, 1979.

[133] G. MiDi. On dense granular flows. *The European Physical Journal E*, 14(4): 341–365, 2004.

[134] P. Mills, P. Rognon, and F. Chevoir. Rheology and structure of granular materials near the jamming transition. *EPL (Europhysics Letters)*, 81(6):64005, 2008.

[135] R. D. Mindlin and H. Deresiewicz. Elastic spheres in contact under varying oblique force. *J. Appl. Mech., ASME*, 20:327–344, 1953.

[136] B. Mohammadi and O. Pironneau. Shape optimization in fluid mechanics. *Annual review of fluid mechanics*, 36(1):255–279, 2004.

[137] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, 2018.

[138] S. Mowlavi and S. Nabi. Optimal control of pdes using physics-informed neural networks. *arXiv preprint arXiv:2111.09880*, 2021.

[139] D. M. Mueth, H. M. Jaeger, and S. R. Nagel. Force distribution in a granular medium. *Physical Review E*, 57(3):3164, 1998.

[140] T. Olsen and K. Kamrin. Resolving force indeterminacy in contact dynamics using compatibility conditions. *Granular Matter*, 20(4):69, 2018.

[141] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces*. Springer-Verlag, 2003.

[142] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[143] M. Otsuki and H. Hayakawa. Critical scaling near jamming transition for frictional granular particles. *Physical Review E*, 83(5):051301, 2011.

[144] H. Perrin, C. Clavaud, M. Wyart, B. Metzger, and Y. Forterre. Interparticle friction leads to nonmonotonic flow curves and hysteresis in viscous suspensions. *Physical Review X*, 9(3):031027, 2019.

[145] G. Picard, A. Ajdari, F. Lequeux, and L. Bocquet. Slow flows of yield stress fluids: Complex spatiotemporal behavior within a simple elastoplastic model. *Physical Review E*, 71(1):010501, 2005.

[146] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995.

[147] D. O. Potyondy and P. Cundall. A bonded-particle model for rock. *International journal of rock mechanics and mining sciences*, 41(8):1329–1364, 2004.

[148] O. Pouliquen. Scaling laws in granular flows down rough inclined planes. *Physics of fluids*, 11(3):542–548, 1999.

[149] O. Pouliquen and Y. Forterre. Friction law for dense granular flows: application to the motion of a mass down a rough inclined plane. *Journal of Fluid Mechanics*, 453:133–151, 2002.

[150] O. Pouliquen and Y. Forterre. A non-local rheology for dense granular flows. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1909):5091–5107, 2009.

[151] L. Quartier, B. Andreotti, S. Douady, and A. Daerr. Dynamics of a grain on a sandpile model. *Physical Review E*, 62(6):8299, 2000.

[152] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[153] M. Raissi, A. Yazdani, and G. E. Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

[154] C. Rao, H. Sun, and Y. Liu. Physics-informed deep learning for computational elastodynamics without labeled data. *Journal of Engineering Mechanics*, 147 (8):04021043, 2021.

[155] K. Reddy, Y. Forterre, and O. Pouliquen. Evidence of mechanically activated processes in slow granular flows. *Physical Review Letters*, 106(10):108301, 2011.

[156] S. Rezaei, A. Harandi, A. Moeineddin, B.-X. Xu, and S. Reese. A mixed formulation for physics-informed neural networks as a potential solver for engineering problems in heterogeneous domains: comparison with finite element method. *arXiv preprint arXiv:2206.13103*, 2022.

[157] P. Rognon. Ecoulements granulaires sur plan incliné: étude expérimentale de la transition vers l?arrêt. Master's thesis, Université de Marne la Vallée, 2002.

[158] C. A. Rotter and C. S. Smith. Ultrasonic equation of state of iron: I. low pressure, room temperature. *J. Phys. Chem. Solids*, 27(267-276):197, 1966.

[159] J.-N. Roux and G. Combe. Quasistatic rheology and the origins of strain. *Comptes Rendus Physique*, 3(2):131–140, 2002.

[160] A. Russell, C. Johnson, A. Edwards, S. Viroulet, F. Rocha, and J. Gray. Retrogressive failure of a static granular layer on an inclined plane. *Journal of Fluid Mechanics*, 869:313–340, 2019.

[161] F. Sahli Costabal, Y. Yang, P. Perdikaris, D. E. Hurtado, and E. Kuhl. Physics-informed neural networks for cardiac activation mapping. *Frontiers in Physics*, 8:42, 2020.

[162] C. H. Scholz. Earthquakes and friction laws. *Nature*, 391(6662):37–42, 1998.

[163] P. Schuhmacher, F. Radjai, and S. Roux. Wall roughness and nonlinear velocity profiles in granular shear flows. In *EPJ Web of Conferences*, volume 140, page 03090. EDP Sciences, 2017.

[164] O. Sigmund and K. Maute. Topology optimization approaches. *Structural and Multidisciplinary Optimization*, 48(6):1031–1055, 2013.

[165] L. E. Silbert, D. Ertaş, G. S. Grest, T. C. Halsey, D. Levine, and S. J. Plimpton. Granular flow down an inclined plane: Bagnold scaling and rheology. *Physical Review E*, 64(5):051302, 2001.

[166] L. E. Silbert, G. S. Grest, S. J. Plimpton, and D. Levine. Boundary effects and self-organization in dense granular flows. *Physics of fluids*, 14(8):2637–2646, 2002.

[167] L. E. Silbert, J. W. Landry, and G. S. Grest. Granular flow down a rough inclined plane: transition between thin and thick piles. *Physics of Fluids*, 15 (1):1–10, 2003.

[168] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.

[169] K. C. Smith, I. Srivastava, T. S. Fisher, and M. Alam. Variable-cell method for stress-controlled jamming of athermal, frictionless grains. *Physical Review E*, 89(4):042203, 2014.

[170] J. H. Snoeijer, T. J. Vlugt, M. van Hecke, and W. van Saarloos. Force network ensemble: a new approach to static granular matter. *Physical review letters*, 92 (5):054302, 2004.

[171] F. Soulie, F. Cherblanc, M. S. El Youssoufi, and C. Saix. Influence of liquid bridges on the mechanical behaviour of polydisperse granular materials. *International journal for numerical and analytical methods in geomechanics*, 30(3): 213–228, 2006.

[172] I. Srivastava and T. S. Fisher. Slow creep in soft granular packings. *Soft matter*, 13(18):3411–3421, 2017.

[173] I. Srivastava, L. E. Silbert, G. S. Grest, and J. B. Lechman. Flow-arrest transitions in frictional granular matter. *Physical review letters*, 122(4):048003, 2019.

[174] I. Srivastava, L. E. Silbert, G. S. Grest, and J. B. Lechman. Viscometric flow of dense granular materials under controlled pressure and shear stress. *Journal of Fluid Mechanics*, 907:A18, 2021.

[175] I. Srivastava, L. E. Silbert, J. B. Lechman, and G. S. Grest. Flow and arrest in stressed granular materials. *arXiv preprint arXiv:2104.00787*, 2021.

[176] A. Stroh. Dislocations and cracks in anisotropic elasticity. *Philosophical magazine*, 3(30):625–646, 1958.

[177] H. Sun, H. Waisman, and R. Betti. Nondestructive identification of multiple flaws using xfem and a topologically adapting artificial bee colony algorithm. *International Journal for Numerical Methods in Engineering*, 95(10):871–900, 2013.

[178] J. Swadener and G. Pharr. Indentation of elastically anisotropic half-spaces by cones and parabolae of revolution. *Philosophical Magazine A*, 81(2):447–466, 2001.

[179] Z. Tang, T. A. Brzinski, M. Shearer, and K. E. Daniels. Nonlocal rheology of dense granular flow in annular shear experiments. *Soft matter*, 14(16):3040–3048, 2018.

[180] C. Thornton and K. Yin. Impact of elastic spheres with and without adhesion. *Powder technology*, 65(1-3):153–166, 1991.

[181] J. Tomas. Adhesion of ultrafine particles—a micromechanical approach. *Chemical Engineering Science*, 62(7):1997–2010, 2007.

[182] Y. Tsuji, T. Tanaka, and T. Ishida. Lagrangian numerical simulation of plug flow of cohesionless particles in a horizontal pipe. *Powder technology*, 71(3): 239–250, 1992.

[183] J. Vlassak, M. Ciavarella, J. Barber, and X. Wang. The indentation modulus of elastically anisotropic materials for indenters of arbitrary shape. *Journal of the Mechanics and Physics of Solids*, 51(9):1701–1721, 2003.

[184] J. J. Vlassak and W. Nix. Indentation modulus of elastically anisotropic half spaces. *Philosophical Magazine A*, 67(5):1045–1056, 1993.

[185] J. J. Vlassak and W. Nix. Measuring the elastic properties of anisotropic materials by means of indentation experiments. *Journal of the Mechanics and Physics of Solids*, 42(8):1223–1245, 1994.

[186] D. Volfson, L. S. Tsimring, and I. S. Aranson. Partially fluidized shear granular flows: Continuum theory and molecular dynamics simulations. *Physical Review E*, 68(2):021301, 2003.

[187] L. Vu-Quoc and X. Zhang. An accurate and efficient tangential force–displacement model for elastic frictional contact in particle-flow simulations. *Mechanics of materials*, 31(4):235–269, 1999.

[188] H. Waisman, E. Chatzi, and A. W. Smyth. Detection and quantification of flaws in structures by the extended finite element method and genetic algorithms. *International Journal for Numerical Methods in Engineering*, 82(3):303–328, 2010.

[189] O. R. Walton and R. L. Braun. Viscosity, granular-temperature, and stress calculations for shearing assemblies of inelastic, frictional disks. *Journal of rheology*, 30(5):949–980, 1986.

[190] T. Weinhart, R. Hartkamp, A. R. Thornton, and S. Luding. Coarse-grained local and objective continuum description of three-dimensional granular flows down an inclined surface. *Physics of Fluids*, 25(7):070605, 2013.

[191] J. Willis. Hertzian contact of anisotropic bodies. *Journal of the Mechanics and Physics of Solids*, 14(3):163–176, 1966.

[192] J. C. Wong, C. Ooi, A. Gupta, and Y.-S. Ong. Learning in sinusoidal spaces with physics-informed neural networks. *arXiv preprint arXiv:2109.09338*, 2021.

[193] C. Wu, M. Zhu, Q. Tan, Y. Kartha, and L. Lu. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *arXiv preprint arXiv:2207.10289*, 2022.

[194] F.-L. Yang and Y.-T. Huang. New aspects for friction coefficients of finite granular avalanche down a flat narrow reservoir. *Granular Matter*, 18(4):1–13, 2016.

[195] X.-F. Yuan. Dynamics of a mechanical interface in shear-banded flow. *EPL (Europhysics Letters)*, 46(4):542, 1999.

[196] C. Zener. *Elasticity and anelasticity of metals*. University of Chicago press, 1948.

[197] E. Zhang, M. Dao, G. E. Karniadakis, and S. Suresh. Analyses of internal structures and defects in materials using physics-informed neural networks. *Science advances*, 8(7):eabk0644, 2022.

[198] L. Zhang, N. G. H. Nguyen, S. Lambert, F. Nicot, F. Prunier, and I. Djeran-Maigre. The role of force chains in granular materials: from statics to dynamics. *European Journal of Environmental and Civil Engineering*, 21(7-8):874–895, 2017.

[199] Q. Zheng, Z. Zhou, and A. Yu. Contact forces between viscoelastic ellipsoidal particles. *Powder technology*, 248:25–33, 2013.

[200] H. Zhu, Z. Zhou, R. Yang, and A. Yu. Discrete particle simulation of particulate systems: theoretical developments. *Chemical Engineering Science*, 62(13):3378–3396, 2007.

[201] H. Zhu, Z. Zhou, R. Yang, and A. Yu. Discrete particle simulation of particulate systems: a review of major applications and findings. *Chemical Engineering Science*, 63(23):5728–5770, 2008.