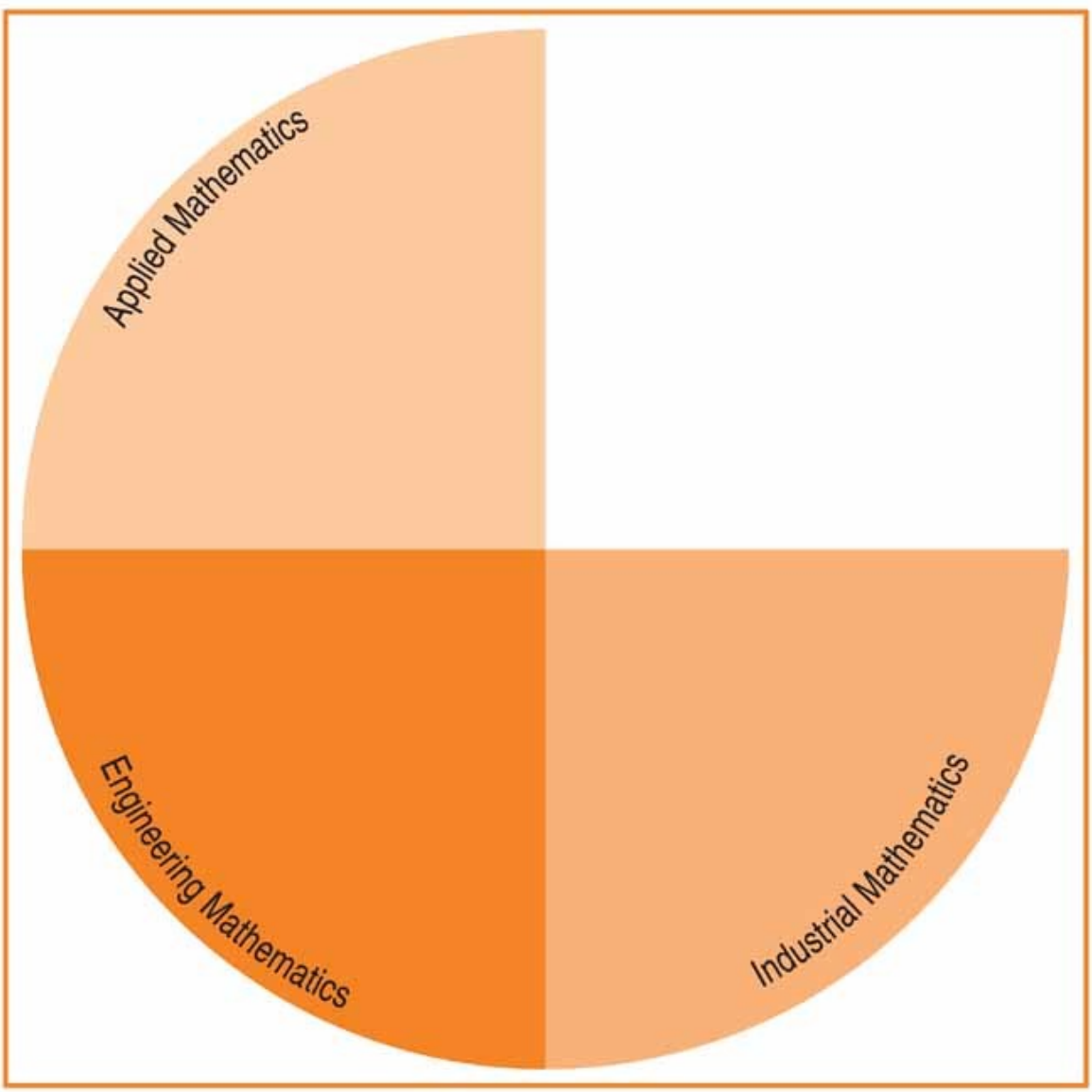
Mathematics and Mechanics of Granular Materials

Edited by
James M. Hill and A.P.S. Selvadurai



Springer

Mathematics and Mechanics of Granular Materials

Edited by

JAMES M. HILL

University of Wollongong, Australia

and

A.P.S. SELVADURAI

McGill Univeristy, Montreal, ON, Canada

Reprinted from *Journal of Engineering Mathematics,* Vol. 52, Nos. 1–3 (July 2005)

Springer

Library of Congress Cataloging-in-Publication Data

ISBN 1-4020-3781-3

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands

TABLE OF CONTENTS

Journal of Engineering Mathematics (2005) 52: 1–9
DOI 10.1007/s10665-005-2729-1

Mathematics and mechanics of granular materials

J.M. HILL and A.P.S. SELVADURAI[1]
School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW, Australia (jhill@uow.edu.au); [1]*McGill University, Montreal, QC, Canada (patrick.selvadurai@mcgill.ca)*

Received and accepted 10 February 2005

The study of granular materials has always been a topic of considerable importance in engineering. Historically, the mathematical formulation of the subject dates back to the pioneering work of C.A. Coulomb in 1776 [1]. In his now famous memoir, Coulomb postulated the conditions that should be satisfied for failure to occur in a granular material. This postulate for failure still stands as a defining point in the mathematical study of mechanics of granular materials. Coulomb largely focused on a topic of importance to that time, namely the design of earth structures to avoid collapse. As a result, the study of the deformations that lead to failure received less emphasis [2]. Recently, however, several scientific disciplines, including geomechanics, mechanical, civil and chemical engineering, physics and applied mathematics, have shown renewed interest in accurately modelling granular materials to examine, concurrently, both failure and deformations. The study of how granular materials or bulk solids flow and deform is also of practical importance for a number of industries, including mining and minerals processing, agricultural materials processing, the construction industry, foodstuff production, pharmaceutical development and nanotechnology. In these applications the granular materials involved could be as diverse as crushed ore, cereal grains, sugar, flour, tablets and nano-particulates. In each case, granular materials frequently flow through devices such as bins, hoppers and chutes and a clear knowledge of how they behave under these circumstances is invaluable for the efficient design and application of related devices.

Granular materials form an important component in modern developments in geomechanics. For the most part, geotechnical engineers are less interested in fully developed granular flows, but the deformational aspects of granular materials are highly relevant in situations that require assessment of settlements of foundations on granular media. The development of mathematically correct and physically admissible theories to describe and predict the complex behaviour of granular materials or bulk solids is therefore a topic of fundamental importance to both the engineering sciences and applied mathematics.

Modelling the flow of granular materials has been extensively studied through the use of continuum mechanics. Using this approach, one formulates governing equations for the stress and velocity fields by coupling the equations of conservation of mass and linear momentum with appropriate constitutive laws that govern the initiation of failure and the rules applicable to the flow of the granular material subsequent to its failure. For rapid granular flows that accompany a reduction in the bulk density, the behaviour of each granular particle is determined primarily by inelastic collisions with neighbouring particles, in a way analogous to colliding molecules in dense gases. In contrast, for slow dense granular flows, the dominant

mechanisms are quite different; here, the neighbouring particles continually slide and roll past each other, and friction between these particles becomes the dominant force.

The problem of modelling fully developed slow granular flows using continuum mechanics is, and continues to be, both complex and challenging. There is general agreement that stress fields within granular flows can be described by coupling the equations of linear momentum with the Coulomb–Mohr yield condition, or other forms of yield condition applicable to the myriad of granular materials that are encountered in engineering practice. However, there is little or no agreement as to how the equations for the velocity fields, that describe the deformations of fully developed flows, should be formulated, or even whether these equations should be mathematically well-posed or ill-posed. The constitutive assumption that is perhaps most widely employed by the engineering community is Saint-Venant's hypothesis, which is also referred to as the coaxiality condition. This condition states that the principal axes of the stress and strain-rate tensors should coincide. Drucker and Prager [3] were the first to formally adopt this hypothesis for the study of the mechanics of granular materials. They used the Coulomb–Mohr yield condition as a plastic potential to derive an associated flow rule. The condition of coaxiality must hold by virtue of material isotropy, and the rate-of-strain tensor depends only on the Cauchy stress tensor.

While the work of Drucker and Prager [3] marks the resurgence of the application of plasticity theories to mechanics of soils, these developments have limitations. Firstly, the theory predicts that all granular flows are accompanied by dilation or volume change, notably volume expansion, whereas in fact loose granular materials contract upon shearing, and others undergo isochoric or volume-preserving deformations. Even in situations for which dilation is appropriate, the predicted magnitude of volume increases is far in excess of those observed in most real materials. The second limitation is that for cohesionless materials; the theory predicts that the rate of specific mechanical energy dissipation is zero, which is clearly unrealistic. More sophisticated approaches attempt to overcome these difficulties by either including work-hardening/softening theories, similar to those proposed and developed by Drucker *et al.* [4], Jenike and Shield [5], Schofield and Wroth [6] or the incorporation of flow rules that are non-associated. In the former category of models, the yield condition varies with a state parameter, such as the density. For the work-hardening/softening models, the mathematical characteristics for the stress and velocity fields do not coincide, contrary to what is commonly observed experimentally; this leads to the adoption of non-associated flow rules. The subject matter in this area is extensive and no attempt will be made to provide an exhaustive review of non-associated plasticity. It is worth noting that Hill [7] proposed velocity equations for incompressible materials based on the Saint-Venant hypothesis, but, again, this theory has the undesirable property that the predicted stress and velocity characteristics do not coincide.

By abandoning the assumption of coaxiality, an alternative family of models has been derived based on a kinematic hypothesis involving the concepts of shearing motion parallel to a surface, rotation of that surface, and dilation or contraction normal to the surface. One such model is the double-shearing theory, originally proposed by Spencer [8, 9] for incompressible flows, and extended to dilatant materials by Mehrabadi and Cowin [10] and Butterfield and Harkness [11]. In this theory, the characteristic curves for the stresses and velocities coincide, and every deformation is assumed to consist of simultaneous shears along the two families of stress characteristics. These ideas build upon those of the double sliding, free rotating model, developed by de Josselin de Jong [12–14], by fixing the rotation rate as the temporal rate of change of the stress angle. To reiterate, an important advantage of the double-shearing theory over the previous coaxial theories is that it retains the assumption of

slip occurring along the stress characteristics, but does not give rise to unusually high levels of dilatancy. Spencer's [8, 9] original double-shearing theory is for incompressible materials, which in the context of fully developed granular flow is often a reasonable and a realistic assumption. Furthermore, when applied to gravity-driven flow problems [15, 16], the coaxial theory is shown to yield physically unacceptable predictions in the velocity field, whereas the double-shearing theory predicts results that are certainly reasonable. On the other hand, there are experiments, which are not consistent with predictions of Spencer's double-shearing theory, but tend to support the double-sliding, free-rotating model of de Josselin de Jong [12–14]. Research in this area must recognize the fact that there is little possibility for developing a mathematical theory of granular media for all eventualities: the materials are real and the circumstances diverse. A theory that shows promise for a given set of experimental conditions can fail for others. In any event, at this moment no single theory is clearly most applicable for describing the behaviour of fully developed flow of real granular materials. While the subject requires more reproducible non-conventional experiments to help resolve these issues, there is a serious need for in-depth mathematical and numerical analysis of the theories involved. This might include the solution of relevant boundary-value problems and initial-boundary-value problems that can allow the continuous transformation of a deformation-dominated process to a flow-dominated one, the exploration of exact and numerical solutions to the equations, and the comparison and contrasting of existing theories that will guide critical experiments of the future.

In addition to the issues raised above, a major unresolved question with Spencer's [8, 9] double-shearing theory, and most other plasticity-based theories for fully developed granular flow, is that the equations are linearly ill-posed in the sense that small perturbations to existing solutions may result in solutions that grow exponentially with time (see *e.g.* [17–19]). This characteristic places doubt on whether or not steady solutions to the governing equations actually describe real granular flows, and also leads to serious implications for numerical schemes, which do not converge in the limit as the size of a mesh discretization approaches zero. However, ill-posedness in itself is not necessarily an undesirable property for equations that describe granular deformations. In fact, it is well known that under certain circumstances granular materials exhibit unstable behaviour, in which case it is quite plausible that ill-posedness should be the norm. An example is the onset of shear-banding. Perhaps the ideal situation, as advocated by Harris [19], is a theory that contains a domain of well-posedness, in which solutions may be stable or unstable, and also a domain of ill-posedness, which corresponds to a definite physical instability. This motivation has led Harris [20, 21] to derive a single-slip model, which belongs to the class of models based upon the physical and kinematic considerations discussed above. This single-slip model is indeed well-posed under well-defined conditions and ill-posed when these conditions fail [19]. In this case the ill-posedness corresponds to the physical instability of grain separation, a process that invalidates the assumption that friction between particles is the dominant mode of momentum transfer, as opposed to inelastic collisions. There is much scope for further research in this complex and challenging field.

We note that there have been several recent attempts to model the transitional region between dense, slow granular flows and rapid, collisional flows (see, for example, [22–25]. These models combine traditional plasticity ideas with notions borrowed from the kinetic theory of gases [26]. In general, the condition of coaxiality is enforced, and again it is not entirely clear whether these theories are well-posed or ill-posed. Often in fully developed slow granular flow, there are narrow layers, referred to as shear layers, in which the material experiences intense shearing. While the models mentioned previously capture many features of fully

developed flow to varying degrees, none have the ability to accurately predict the thickness of the layers over which such intense shearing materializes. A reason for this limitation has been attributed to the fact that classical continuum models have no intrinsic length scale built into the constitutive equations. Attempts to rectify this deficiency probably date back to the work of Voigt [27] and later expanded by Cosserat and Cosserat [28] who introduced the concept of couple stresses for examining the mechanics of deformable media (see *e.g.* [29]). Here, the Cauchy stress tensor is no longer symmetric, and the conservation of angular momentum is no longer automatically satisfied but becomes a set of field equations that need to be satisfied explicitly. There are two extra field variables for Cosserat materials, namely the angular velocity and the couple-stress tensor. As a consequence of the notion of couple stresses, a length parameter or an intrinsic length scale naturally arises in the definition of constitutive relationships. The work on both micromorphic and couple-stress theories was an active area of research from the mid-1960s to the mid-1970s and the developments are summarized in [30]. A number of authors have applied these concepts to the examination of problems associated with granular media and references to recent works are given by Vardoulakis and Sulem [31]. The investigations by Mühlhaus [32], Tejchman and Wu [33], Bauer [34], Tejchman and Bauer [35], Tejchman and Gudehus [36] and others also deal with the application of higher-order formulations in elastoplasticity, in the context of the theory of hypoplasticity, which is described below, and by Mohan *et al*. [37, 38] who use more traditional ideas from plasticity. In each case, this improvement is achieved by modelling the granular material as a Cosserat (or micropolar) continuum. Mohan *et al*. [37, 38] apply an extended-associated flow rule, with the yield condition depending on the bulk density, and apply the equations to model flow through vertical channels and cylindrical Couette flow. These studies are successful in that they predict the main qualitative features of the shear layers; however, the yield condition and flow rule were chosen purely for illustrating the effectiveness of this approach.

In many civil and geotechnical engineering applications the constrained response of a granular material, such as a soil or sand, under loading is most important [29]. Examples of such situations occur with the analysis of foundations, excavations and underground structures, or simply in elemental tests. Here, the deformation of the material is contained by a surrounding material, which prevents the development of a state of plastic flow or collapse. Traditionally, a variety of elastoplastic models have been applied to problems of this nature. The history of development of theories of geomaterial behaviour that account for contained deformations of granular materials is quite extensive, and no attempt will be made here to provide an all-encompassing review. More recently, however, the constitutive theory of hypoplasticity has been developed, and has proven to be an attractive alternative to the elastoplastic models. Hypoplasticity is a natural extension of the theories of hypoelasticity developed by Truesdell [40] and the connection between the theories of hypoelasticity and theories of plasticity and of elastic-plastic flow has been discussed and investigated by Green [41, 42], Truesdell and Noll [43] and Jaunzemis [44]. Hypoplasticity in a formal sense was extensively investigated by Kolymbas [45] and many co-workers (see [46–48]). The characterizing feature of all hypoplastic theories is that the constitutive law can be written in a single nonlinear tensorial equation for the stress-rate as a function of the stress and the rate-of-deformation tensor, without reference to a yield condition or a flow rule. With hypoplasticity there is no need to decompose deformations into elastic and plastic regimes *a priori*, or to distinguish between loading and unloading; all these notions are automatically built into the theory, and arise as a consequence. Excellent reviews of hypoplasticity and its development are contained in Kolymbas [49] and Wu and Kolymbas [50].

The popularity of hypoplasticity among researchers and practitioners can be attributed to its elegance and the fact that the theory is deeply rooted in experimental observations. It is, nonetheless, a sophisticated constitutive theory, which involves complicated nonlinear constitutive relationships. When combined with the governing equations of continuum mechanics, there are little prospects for the analytical solution of real-life boundary-value problems, and progress is usually made via numerical schemes. As a result, it is often difficult to grasp the underlying mathematical structure of the equations (see [51–54]).

As mentioned previously, for each particular hypoplastic law there is a yield surface and a flow rule, but rather than being assigned in advance, they are consequences of the original constitutive relationships. Thus hypoplasticity as a theory can, in principle, be used to model fully developed granular flow. The explicit equations describing the yield condition and the flow rule can be derived from the given hypoplastic law, as illustrated by Wu and Niemunis [55] and von Wolffersdorff [56]. Von Wolffersdorff [56] has derived particular hypoplastic models that give the yield surfaces of Drucker and Prager [3] and Matsuoka and Nakai [57] as limiting cases. It is not immediately clear whether a similar derivation can be made to link hypoplasticity with other plasticity theories such as the double-shearing theory [8–10] described above. This possibility is of considerable interest, especially in light of the recent work of Spencer [58], who shows that in a strict sense, the double-shearing theory can be regarded as a special form of hypoplasticity. There is an absence of understanding of the strict connection between hypoplasticity and theories of plasticity that describe granular flow.

Shear layers often occur in the vicinity of solid boundaries, but this is not generally the case. An important property of granular materials is that shear-banding or shear layers can also occur within the bulk of the material. Shear bands are usually accompanied by localised strains, spanning several grain diameters in thickness, and as discussed above, classical continuum approaches fail to account for the dimensions of the shear bands due to the absence of an intrinsic length-scale. Furthermore, although the onset of shear-banding can be predicted [59], the ill-posedness of the governing equations prevents a complete analysis. As discussed previously, the subject of layers with intense shearing or shear-banding has received much attention by investigators who have developed approaches that incorporate Cosserat-type effects, and this is most prominent in hypoplasticity (see *e.g.* [34–36, 60]). Various hypoplastic theories have been developed and validated using finite-element techniques. The topic of mechanics and mathematics of granular materials has a rich history of involvement of researchers in the engineering sciences as well as those in the mechanics and applied-mathematics communities. These contributions are too numerous to cite as a comprehensive and complete review; readers are referred to the following Edited volumes of Symposia and Conference Proceedings for more in-depth reviews of the historical developments and the current state of advanced mathematical and mechanics approaches to the study of granular materials [61–78].

This Special Issue on the Mathematics and Mechanics of Granular Materials presents a mix of mathematical and engineering contributions to the discipline. Some of the papers, but not all, originate from four Mini-Symposia held at the 2003 ICIAM (International Congress of Industrial and Applied Mathematics) in Sydney, Australia, June 7–11, 2003. This meeting was jointly organised by the Guest Editors together with Drs. Claudio Tamagnini and Antoinette Tordesillas. The papers presented in this Special Issue cover the full range of current research activity in the area, and include general, analytical, hypoplastic, numerical and engineering contributions, but appear as follows according to the alphabetical listing of the first-named author:

1. F. Alonso-Marroquin and H.J. Herrmann, Investigation of the incremental response of soils using a discrete element model.
2. E. Bauer, Initial response of a micro-polar hypoplastic material under plane shearing.
3. R. Chambon, Some general results about second order work, uniqueness, existence and controllability.
4. G.M. Cox, S.W. McCue, N. Thamwattana and J. M. Hill, Perturbation solutions for flow through symmetrical hoppers with inserts and asymmetrical wedge hoppers.
5. B.S. Gardiner and A. Tordesillas, Micromechanical constitutive modelling of granular media: evolution and loss of contact in particle clusters.
6. D. Harris and E.F. Grekova, A hyperbolic well-posed model for the flow of granular materials.
7. S.C. Hendy, Towards a theory of granular plasticity.
8. M. Hjiaj, W. Huang, K. Krabbenhoft and S.W. Sloan, Formulation of non-standard dissipative behaviour of geomaterials.
9. W. Huang, M. Hjiaj and S.W. Sloan, Bifurcation analysis for shear localization in non-polar and micro-polar hypoplastic continua.
10. V.A. Osinov, Large-strain dynamic cavity expansion in a granular material.
11. E. Pasternak and H.-B. Muhlhaus, Generalised homogenisation procedures for granular materials.
12. J.F. Peters, Some fundamental aspects of the continuumization problem in granular media.
13. A.J.M. Spencer, Compression and shear of a layer of granular material.
14. C. Tamagnini, F. Calvetti and G. Viggiani, An assessment of plasticity theories for modelling the incrementally non–linear behavior of granular soils.
15. G.J. Weir, Incompressible granular flow from wedge-shaped hoppers.
16. H.P. Zhu and A.B. Yu, Micromechanics modeling and analysis of unsteady state granular flow in a cylindrical hopper.

Acknowledgements

The Guest Editors wish especially to acknowledge the help and assistance of a large number of people who acted as referees for this Special Issue, but particularly to Drs. Grant Cox and Scott McCue who cheerfully undertook many additional burdens.

References

1. C.A. Coulomb, Essai sur une application des règles *de maximis* & *minimis* à quelques problèmes de statique, relatifs à l'architecture, *Mémoire de Mathématique & de Physique, présentés à l'Académie Royale des Sciences par divers Savans, & lûs dans ses Assemblées*, Vol. 7, 1773, Paris (1776) pp. 343–382.
2. J. Heyman, *Coulomb's Memoir on Statics*. Cambridge: Cambridge University Press (1972) 212 pp.
3. D.C. Drucker and W. Prager, Soil mechanics and plastic analysis or limit design. *Q. Appl. Math*. 10 (1952) 157–165.
4. D.C. Drucker, R.E. Gibson and D.J. Henkel, Soil mechanics and work-hardening theories of plasticity. *Trans. ASCE* 122 (1957) 338–346.
5. A.W. Jenike and R.T. Shield, On the plastic flow of Coulomb solids beyond original failure. *J. Appl. Mech*. 26 (1959) 599–602.
6. A. Schofield and P. Wroth, *Critical State Soil Mechanics*. London: McGraw-Hill (1968) 310 pp.
7. R. Hill, *The Mathematical Theory of Plasticity*. Oxford: Clarendon Press (1950) 355 pp.
8. A.J.M. Spencer, A theory of the kinematics of ideal soils under plane strain conditions. *J. Mech. Phys. Solids* 12 (1964) 337–351.

9. A.J.M. Spencer, Deformation of ideal granular materials. In: H.G. Hopkins and M.J. Sewell (eds.), *Mechanics of Solids*. Oxford: Pergamon Press (1982) pp. 607–652.
10. M.M. Mehrabadi and S.C. Cowin, Initial planar deformation of dilatant granular materials. *J. Mech. Phys. Solids* 26 (1978) 269–284.
11. R. Butterfield and R.M. Harkness, The kinematics of Mohr-Coulomb materials. In: R.H.G. Parry (ed.), *Stress–Strain Behaviour of Soils*. Henley: Foulis (1972) pp. 220–281.
12. G. de Josselin de Jong, *Statics and Kinematics of the Failable Zone of a Granular Material*. Doctoral thesis. Delft: Uitgeverij Waltman (1959).
13. G. de Josselin de Jong, Mathematical elaboration of the double-sliding, free-rotating model. *Archs. Mech*. 29 (1977) 561–591.
14. G. de Josselin de Jong, Elasto-plastic version of the double-sliding model in undrained simple shear tests. *Geotechnique* 38 (1988) 533–555.
15. A.J.M. Spencer, Remarks on coaxiality in fully developed gravity flows of dry granular materials. In: N.A. Fleck and A.C.F. Cocks (eds.), *Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer (1997) pp. 227–238.
16. A.J.M. Spencer and N.J. Bradley, Gravity flow of granular materials in contracting cylinders and tapered tubes. *Int. J. Eng. Sci*. 40 (2002) 1529–1552.
17. A.J.M. Spencer, Instability of steady flow of granular materials. *Acta Mech*. 64 (1986) 77–87.
18. D.G. Schaeffer, Instability in the evolution equations describing incompressible granular flow. *J. Diff. Eqns.* 66 (1987) 19–50.
19. D. Harris, Ill- and well-posed models of granular flow. *Acta Mech*. 146 (2001) 199–225.
20. D. Harris, Modelling mathematically the flow of granular materials. In: N.A. Fleck and A.C.F. Cocks (eds.), *Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer (1997) pp. 239–250.
21. D. Harris, Discrete and continuum models in the mechanics of granular materials. In: R.P. Behringer, and J.T. Jenkins. (eds.), *Powders and Grains 97*. Rotterdam: A.A. Balkema (1997) pp. 247–250.
22. R. Jyotsna and K.K. Rao, A frictional-kinetic model for the flow of granular materials through a wedge-shaped hopper. *J. Fluid Mech*. 346 (1997) 239–270.
23. L.S. Mohan, P.R. Nott and K.K. Rao, Fully developed flow of coarse granular materials through a vertical channel. *Chem. Engng. Sci*. 52 (1997) 913–933.
24. S.B. Savage, Analyses of slow high-concentration flows of granular materials. *J. Fluid Mech*. 377 (1998) 1–26.
25. G.I. Tardos, S. McNamara and I. Talu, Slow and intermediate flow of a frictional bulk powder in the Couette geometry. *Powder Tech*. 131 (2003) 23–39.
26. S. Chapman and T.G. Cowling, *The Mathematical Theory of Non-Uniform Gases*. Cambridge: Cambridge University Press (1970) 423 pp.
27. W. Voigt, *Theoretische Studien über Elastizitätsverhältnisse der Kristalle*. Abh. Ges. Wiss., Gottingen. 34 (1887).
28. E. Cosserat and F. Cosserat, *Theorie des Corps Deformables*, Paris: A. Herman et Fils (1909) vi, p. 226.
29. A.P.S. Selvadurai, Bending of an infinite beam on a porous elastic medium. *Geotechnique* 23 (1973) 407–421.
30. E. Kröner (Ed.), *Mechanics of Generalized Continua, Proc. IUTAM Symposium*, Stuttgart and Freudenstadt. Berlin: Springer-Verlag (1968) 358 pp.
31. I. Vardoulakis and J. Sulem, *Bifurcation Analysis in Geomechanics*. London: Blackie Academic and Professional (1995) 462 pp.
32. H.B. Mühlhaus, Shear band analysis in granular materials by Cosserat theory. *Ing. Arch*. 56 (1986) 389–399.
33. J. Tejchman and W. Wu, Numerical study of patterning of shear bands in a Cosserat continuum. *Acta Mech*. 99 (1993) 61–74.
34. E. Bauer, Calibration of a comprehensive constitutive equation for granular materials. *Soils and Foundations* 36 (1996) 13–26.
35. J. Tejchman and E. Bauer, Numerical simulation of shear band formation with a polar hypoplastic constitutive model. *Computers and Geotechnics* 19 (1996) 221–244.
36. J. Tejchman and G. Gudehus, Shearing of a narrow granular layer with polar quantities. *Int. J. Numer. Anal. Meth. Geomech.* 25 (2001) 1–28.
37. L.S. Mohan, P.R. Nott and K.K. Rao, A frictional Cosserat model for the flow of granular materials through a vertical channel. *Acta Mech*. 138 (1999) 75–96.
38. L.S. Mohan, K.K. Rao and P.R. Nott, A frictional Cosserat model for the slow shearing of granular materials. *J. Fluid Mech*. 457 (2002) 377–409.

39. R.O. Davis and A.P.S. Selvadurai, *Plasticity and Geomechanics*. Cambridge: Cambridge University Press (2002), 287 pp.
40. C. Truesdell, Hypo-elasticity. *J. Rational Mech. Anal*. 4 (1955) 83–133, 1019–1020.
41. A.E. Green, Hypo-elasticity and Plasticity. *Proc. R. Soc. London A*. 234 (1956) 46–59.
42. A.E. Green, Hypo-elasticity and Plasticity II. *J. Rational Mech. Anal.* 5 (1956) 637–642.
43. C. Truesdell and W. Noll, *The Non-Linear Field Theories of Mechanics*. Handbuch der Physik, Vol. III/3 (S. Flugge, Ed.). Berlin: Springer-Verlag (1965) 602 pp.
44. W. Jaunzemis, *Continuum Mechanics*. New York: The Macmillan Co (1967) 604 pp.
45. D. Kolymbas, A novel constitutive law for soils, In: C.S. Desai (eds.), *Proceedings of the Second International Conference on Constitutive Laws for Engineering Materials*. Amsterdam: Elsevier (1987) pp. 319–326.
46. G. Gudehus, A comprehensive constitutive equation for granular materials. *Soils and Foundations* 36 (1996) 1–12.
47. W. Wu, *Hypoplasticity as a Mathematical Model for the Mechanical Behaviour of Granular Materials*. Publication Series of the Institute for Soil and Rock Mechanics, No. 129. Germany: University of Karlsruhe (1992).
48. W. Wu, E. Bauer and D. Kolymbas, Hypoplastic constitutive model with critical state for granular materials. *Mech. Mater.* 23 (1996) 45–69.
49. D. Kolymbas, Introduction to hypoplasticity. In: *Advances in Geotechnical Engineering and Tunnelling*, Vol. 1. Rotterdam: A.A. Balkema (2000) 94 pp.
50. W. Wu and D. Kolymbas, Hypoplasticity, then and now. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer Verlag (2000) pp. 57–105.
51. J.M. Hill and T.L. Katoanga, The velocity equations for dilatant granular flow and a new exact solution. *ZAMP* 48 (1997) 1–8.
52. J.M. Hill, Similarity 'hot-spot' solutions for a hypoplastic granular material. *Proc. R. Soc. London* A 456 (2000) 2653–2671.
53. J.M. Hill, Some symmetrical cavity problems for a hypoplastic granular material. *Q. J. Mech. Appl. Math.* 53 (2000) 111–135.
54. J.M. Hill and K.A. Williams, Dynamical uniaxial compaction of a hypoplastic granular material. *Mech. Mater.* 32 (2000) 679–691.
55. W. Wu and A. Niemunis, Failure criterion, flow rule and dissipation function derived from hypoplasticity. *Mech. Cohesive-Frict. Mater.* 1 (1996) 145–163.
56. P.-A. von Wolffersdorff, A hypoplastic relation for granular materials with a predefined limit state surface. *Mech. Cohesive-Frict. Mater.* 1 (1996) 251–271.
57. H. Matsuoka and T. Nakai, Stress-strain relationship of soil based on the SMP. In: Proceedings of Specialty Session 9, *9th International Conference on Soil Mechanics and Foundation Engineering*, Tokyo (1997) pp. 153–162.
58. A.J.M. Spencer, Double-shearing theory applied to instability and strain localization in granular materials. *J. Engng. Math*. 45 (2003) 55–74.
59. J.W. Rudnicki and J. Rice, Conditions for the localization of deformation in pressure sensitive dilatant materials. *J. Mech. Phys. Solids* 23 (1975) 371–394.
60. W. Huang and E. Bauer, Numerical investigations of shear localization in a micro-polar hypoplastic material. *Int. J. Numer. Anal. Meth. Geomech*. 27 (2003) 325–352.
61. S.C. Cowin and M. Satake (eds.), *Continuum Mechanical and Statistical Approaches in the Mechanics of Granular Materials*, Proc. U.S.-Japan Seminar, Sendai. Tokyo: Gakujutsu Bunken Fukyu-Kai (1978) 350 pp.
62. P. Vermeer and H.J. Luger (eds.), *Deformation and Failure of Granular Materials*. Proc. IUTAM Symposium, Delft. Rotterdam: A.A. Balkema (1982) 661 pp.
63. J.T. Jenkins and M. Satake (eds.), *Mechanics of Granular Materials. New Models and Constitutive Relations*. Proc. U.S./Japan Seminar, Ithaca, N.Y. Studies in Applied Mechanics, Vol.7. Amsterdam: Elsevier (1983) 364 pp.
64. M. Shahinpoor (ed.), *Advances in the Mechanics and Flow of Granular Materials*. Reston, VA: Gulf Publ. Co. (1983) 975 pp.
65. M. Satake (ed.), *Micromechanics of Granular Materials*. Proc. US-Japan Seminar, Sendai-Zao. Amsterdam: Elsevier (1988) 366 pp.
66. B.L. Keyfitz and M. Shearer (eds.), *Nonlinear Evolution Equations that Change Type*. Berlin: Springer Verlag (1991) 284 pp.

67. H. Shen, C.S. Campbell, M. Mehrabadi, C.S. Chang, and M. Satake (eds.), *Advances in Micromechanics of Granular Materials*. Proc. 2nd US/Japan Seminar, Potsdam. Amsterdam: Elsevier (1992) 462 pp.
68. M. Mehrabadi (ed.), *Mechanics of Granular Materials and Powder Systems*. New York: ASME (1992) 152 pp.
69. C.S. Chang (ed.), *Advances in Micromechanics of Granular Materials*. Amsterdam: Elsevier (1992) 462 pp.
70. D. Bideau and A. Hansen (eds.), *Disorder and Granular Media*. Random Materials and Processes Series. The Netherlands: North Holland (1993) 348 pp.
71. N.A. Fleck and A.F.C. Cocks (eds.), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer (1997) 450 pp.
72. H.J. Herrmann, J.-P. Hovi and S. Luding (eds.), *Physics of Dry Granular Media*. Dordrecht: Kluwer (1998) 710 pp.
73. D.A. Drew, D.D. Joseph and S. Passman (eds.), *Particulate Flows. Processing and Rheology*. Berlin: Springer-Verlag (1998) 142 pp.
74. K.M. Golden, G.R. Grimmett, R.D. James, G.W. Milton and P.N. Sen (eds.), *Mathematics of Multi-Scale Materials*. Berlin: Springer-Verlag (1998) 280 pp.
75. P. Vermeer, S. Diebels, W. Ehlers, H.J. Herrmann, S. Luding and E. Ramm (eds.), *Continuous and Discontinuous Modelling of Cohesive-Frictional Materials*. Berlin: Springer-Verlag (2001) 307 pp.
76. T. Pöschel and S. Luding (eds.), *Granular Gases*. Berlin: Springer Verlag (2001) 457 pp.
77. G. Capriz, V.N. Ghionna and P. Giovine (eds.), *Modeling and Mechanics of Granular and Porous Materials*. Boston: Birkhauser (2002) 369 pp.
78. H. Hinrichsen and D.E. Wolf (eds.), *The Physics of Granular Media*. New York: John Wiley (2005) 364 pp.

Journal of Engineering Mathematics (2005) 52: 11–34
DOI 10.1007/s10665-004-6675-0

The incremental response of soils. An investigation using a discrete-element model

F. ALONSO-MARROQUÍN and H.J. HERRMANN
Institute of Computer Physics, University of Stuttgart, Pffafenwaldring 27, 70569 Stuttgart, Germany (fernando@ica1.uni-stuttgart.de)

Received 21 July 2004; accepted in revised form 19 November 2004

Abstract. The incremental stress-strain relation of dense packings of polygons is investigated by using molecular-dynamics simulations. The comparison of the simulation results to the continuous theories is performed using explicit expressions for the averaged stress and strain over a representative volume element. The discussion of the incremental response raises two important questions of soil deformation: Is the incrementally nonlinear theory appropriate to describe the soil mechanical response? Does a purely elastic regime exist in the deformation of granular materials? In both cases the answer will be "no". The question of stability is also discussed in terms of the Hill condition of stability for non-associated materials. It is contended that the incremental response of soils should be revisited from micromechanical considerations. A micromechanical approach assisted by discrete element simulations is briefly outlined.

Key words: elastoplasticity, granular materials, hypoplasticity, incremental response

1. Introduction

For many years the study of the mechanical behavior of soils was developed in the framework of linear elasticity [1, Chapter 1] and the Mohr-Coulomb failure criterion [2]. However, since the start of the development of the nonlinear constitutive relations in 1968 [3], a great variety of constitutive models describing different aspects of soils have been proposed [4]. A crucial question has been brought forward: What is the most appropriate constitutive model to interpret the experimental results, or to implement a finite-element code? Or more precisely, why is the constitutive relation I am using better than that one of the fellow in the next lab?

In the last years, the discrete-element approach has been used as a tool to investigate the mechanical response of soils at the grain level [5]. Several averaging procedures have been proposed to define the stress [6–8] and the strain tensor [9,10] in terms of the contact forces and displacements at the individual grains. These methods have been used to perform a direct calculation of the incremental stress-strain relation of assemblies of disks [11] and spheres [12], without any *a priori* hypothesis about the constitutive relation. Some of the results lead to the conclusion that the nonassociated elastoplasticity theory is sufficient to describe the observed incremental behavior [11]. However, some recent investigations using three-dimensional loading paths of complex loading histories seem to contradict these results [12,13]. Since the simple spherical geometries of the grains overestimate the role of rotations in realistic soils [13], it is interesting to evaluate the incremental response using arbitrarily shaped particles.

In this paper we investigate the incremental response in the quasistatic deformation of dense assemblies of polygonal particles. The comparison of the numerical simulations with the constitutive theories is performed by introducing the concept of *Representative Volume Element* (RVE). This volume is chosen the smear out the strong fluctuations of the stress and the

deformation in the granular assembly. In the averaging, each grain is regarded as a piece of continuum. By supposing that the stress and the strain of the grain are concentrated at the small regions of the contacts, we obtain expressions for the averaged stress and strain over the RVE, in terms of the contact forces, and the individual displacements and rotations of the grains. The details of this homogenization method are presented in Section 2. A short review of incremental, rate-independent stress-strain models is presented in Section 3. We emphasize particularly the classical Drucker-Prager elastoplastic models and the recently elaborated theory of hypoplasticity. The details of the particle model are presented in Section 4. The interparticle forces include elasticity, viscous damping and friction with the possibility of slip. The system is driven by applying stress-controlled tests on a rectangular framework consisting of four walls. Some loading programs will be implemented in Section 5, in order to deal with four basic questions on the incremental response of soils: (1) The existence of tensorial zones in the incremental response, (2) the validity of the superposition principle, (3) the existence of a finite elastic regime and (4) the question of stability according to the Hill condition. A micromechanical approach for soil deformation is outlined in the concluding remarks.

2. Homogenization

The aim of this section is to calculate the macromechanical quantities, the stress and strain tensors, from micromechanical variables of the granular assembly such as contact forces, rotations and displacements of individual grains.

A particular feature of granular materials is that both the stress and the deformation gradient are very concentrated in small regions around the contacts between the grains, so that they vary strongly over short distances. The standard homogenization procedure smears out these fluctuations by averaging these quantities over a RVE. The diameter d of the RVE must be such that $\delta \ll d \ll D$, where δ is the characteristic diameter of the particles and D is the characteristic length of the continuous variables.

We use this procedure here to obtain the averages of the stress and the strain tensors over a RVE in granular materials, which will allow us to compare the molecular dynamics simulations to the constitutive theories. We regard stress and strain to be continuously distributed

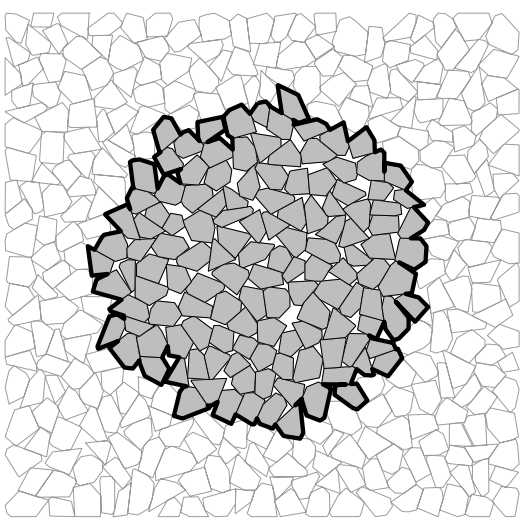


Figure 1. Representative volume element (RVE) used to calculate the incremental response.

through the grains, but concentrated at the contacts. It is important to note that this averaging procedure would not be appropriate to describe the structure of the chain forces or the shear band because typical variations of the stress correspond to few particle diameters. Different averaging procedures involving coarse-grained functions [8], or cutting the space in slide-shaped areas [10,14] allow to perform averages, and at the same time maintain these features.

We will calculate the averages around a point $\mathbf{x}_0$ of the granular sample taking a RVE calculated as follows: at the initial configuration, we select the grains whose centers of mass are less than d from $\mathbf{x}_0$. Then the RVE is taken as the volume V enclosed by the initial configuration of the grains; see Figure 1. The diameter d is taken such that the averaged quantities are not sensible to an increase of the diameter by one particle diameter.

2.1. Micromechanical stress

The Cauchy stress tensor is defined using the force acting on an area element situated on or within the grains. Let $\mathbf{f}$ be the force applied on a surface element a whose normal unit vector is $\mathbf{n}$. Then the stress is defined as the tensor satisfying [1, pp. 12–35]:

$$\sigma_{kj} n_k = \lim_{a \to 0} f_j / a, \tag{1}$$

where the Einstein summation convention is used. In absence of body forces, the equilibrium equations in every volume element lead to [1]:

$$\partial \sigma_{ij} / \partial x_i = 0. \tag{2}$$

We will calculate the average of the stress tensor $\bar{\sigma}$ over the RVE:

$$\bar{\sigma} = \frac{1}{V} \int_V \sigma \, \mathrm{d}V. \tag{3}$$

Since there is no stress in the voids of the granular media, the averaged stress can be written as the sum of integrals taken over the particles [7]

$$\bar{\sigma} = \frac{1}{V} \sum_{\alpha=1}^{N} \int_{V_\alpha} \sigma_{ij} \mathrm{d}V, \tag{4}$$

where V_α is the volume of the particle α and N is the number of particles of the RVE. By use of the identity

$$\frac{\partial (x_i \sigma_{kj})}{\partial x_k} = x_i \frac{\partial \sigma_{kj}}{\partial x_k} + \sigma_{ij}, \tag{5}$$

Equation (2), and the Gauss theorem, Equation (4) leads to [6]

$$\bar{\sigma}_{ij} = \frac{1}{V} \sum_{\alpha} \int_{V_\alpha} \frac{\partial (x_i \sigma_{kj})}{\partial x_k} \mathrm{d}V = \frac{1}{V} \sum_{\alpha} \int_{\partial V_\alpha} x_i \sigma_{kj} n_k \mathrm{d}a. \tag{6}$$

The right-hand side is the sum over the surface integrals of each grain. Further, ∂V_α represents the surface of the grain α and $\mathbf{n}$ is the unit normal vector to the surface element da.

An important feature of granular materials is that the stress acting on each grain boundary is concentrated in the small regions near to the contact points. Then we can use

the definition given in (1) to express this stress on particle α in terms of the contact forces by introducing Dirac delta functions:

$$\sigma_{kj} n_k = \sum_{\beta=1}^{N_\alpha} f_j^{\alpha\beta} \delta(\mathbf{x} - \mathbf{x}^{\alpha\beta}), \tag{7}$$

where $\mathbf{x}^{\alpha\beta}$ and $\mathbf{f}^{\alpha\beta}$ are the position and the force at the contact β, and N_α is the number of contacts of the particle α. Inserting (7) into (6), we obtain

$$\bar{\sigma}_{ij} = \frac{1}{V} \sum_{\alpha\beta} x_i^{\alpha\beta} f_j^{\alpha\beta}. \tag{8}$$

Now we decompose $\mathbf{x}^{\alpha\beta} = \mathbf{x}^{\alpha} + \boldsymbol{\ell}^{\alpha\beta}$, where $\mathbf{x}^{\alpha}$ is the position of the center of mass and $\boldsymbol{\ell}^{\alpha\beta}$ is the branch vector, connecting the center of mass of the particle to the point of application of the contact force. Imposing this decomposition in (8), and using the equilibrium equations at each particle $\sum_\beta \mathbf{f}^{\alpha\beta} = 0$, we have

$$\bar{\sigma}_{ij} = \frac{1}{V} \sum_{\alpha\beta} \ell_i^{\alpha\beta} f_j^{\alpha\beta}. \tag{9}$$

From the equilibrium equations of the torques $\sum_\beta (\ell_i^{\alpha\beta} f_j^{\alpha\beta} - \ell_j^{\alpha\beta} f_i^{\alpha\beta}) = 0$ one obtains that this tensor is symmetric, *i.e.,*

$$\bar{\sigma}_{ij} - \bar{\sigma}_{ji} = 0. \tag{10}$$

Then, the eigenvalues of this matrix are always real. This property leads to some simplifications, as we will see later.

2.2. Micromechanical strain

In elasticity theory, the strain tensor is defined as the symmetric part of the average of the displacement gradient with respect to the equilibrium configuration of the assembly. Using the law of conservation of energy, one can define the stress–strain relation in this theory [1, Section 2.2]. In granular materials, it is not possible to define the strain in this sense, because any loading involves a certain amount of plastic deformation at the contacts, so that it is not possible to define the initial configuration to calculate the strain. Nevertheless, one can define a strain tensor in the incremental sense. This is defined as the average of the displacement tensor taken from the deformation during a certain time interval.

At the micromechanical level, the deformation of the granular materials is given by a displacement field $\mathbf{u} = \mathbf{r}' - \mathbf{r}$ at each point of the assembly. Here $\mathbf{r}$ and $\mathbf{r}'$ are the positions of a material point before and after deformation. The average of the strain and rotational tensors are defined as [15]:

$$\bar{\epsilon} = \frac{1}{2}(F + F^T), \tag{11}$$

$$\bar{\omega} = \frac{1}{2}(F - F^T). \tag{12}$$

Here F^T is the transpose of the deformation gradient F, which is defined as [6]

$$F_{ij} = \frac{1}{V} \int_V \frac{\partial u_i}{\partial x_j} \mathrm{d}V. \tag{13}$$

Using the Gauss theorem, we transform it into an integral over the surface of the RVE

$$F_{ij} = \frac{1}{V} \int_{\partial V} u_i n_j \mathrm{d}a, \tag{14}$$

where ∂V is the boundary of the volume element. We express this as the sum over the boundary particles of the RVE

$$F_{ij} = \frac{1}{V} \sum_{\alpha=1}^{N_b} \int_{\partial V_\alpha} u_i n_j \mathrm{d}a, \tag{15}$$

where N_b is the number of boundary particles. It is now convenient to make some approximations. First, the displacements of the grains during deformation can be considered rigid, except for small deformations near to the contacts, which can be neglected. Then, if the displacements are small in comparison with the size of the particles, we can write the displacement of the material points inside particle α as:

$$u_i(\mathbf{x}) \approx \Delta x_i^\alpha + e_{ijk} \Delta \phi_j^\alpha (x_k - x_k^\alpha), \tag{16}$$

where $\Delta \mathbf{x}^\alpha$, $\Delta \boldsymbol{\phi}^\alpha$ and $\mathbf{x}^\alpha$ are displacement, rotation and center of mass of the particle α which contains the material point $\mathbf{x}$, and e_{ijk} is the antisymmetric unit tensor. Setting a parameterization for each surface of the boundary grains over the RVE, we can calculate the deformation gradient explicitly in terms of grain rotations and displacements by substituting (16) in (15).

In the particular case of a bidimensional assembly of polygons, the boundary of the RVE is given by a graph $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_b+1} = \mathbf{x}_1\}$ consisting of all the edges belonging to the external contour of the RVE, as shown in Figure 1. In this case, Equation (15) can be transformed into a sum of integrals over the segments of this contour.

$$F_{ij} = \frac{1}{V} \sum_{\beta=1}^{N_b} \int_{x_\beta}^{x_{\beta+1}} \left[\Delta x_i^\beta + e_{ik} \Delta \phi^\beta (x_k - x_k^\beta) \right] n_j^\beta \mathrm{d}s, \tag{17}$$

where $e_{ik} \equiv e_{i3k}$ and the unit vector $\mathbf{n}^\beta$ is perpendicular to the segment $\overrightarrow{x^\beta x^{\beta+1}}$. Here β corresponds to the index of the boundary segment; $\Delta \mathbf{x}^\beta$, $\Delta \boldsymbol{\phi}^\beta$ and $\mathbf{x}^\beta$ are displacement, rotation and center of mass of the particle which contains this segment. Finally, by using the parameterization $\mathbf{x} = \mathbf{x}^\beta + s(\mathbf{x}^{\beta+1} - \mathbf{x}^\beta)$, where $(0 < s < 1)$, we can integrate (17) to obtain

$$F_{ij} = \frac{1}{V} \sum_{\beta} \left(\Delta x_i^\beta + e_{ik} \Delta \phi^\beta \ell_k^\beta \right) N_j^\beta, \tag{18}$$

where $N_j^\beta = e_{j,k} \left(x_k^{\beta+1} - x_k^\beta \right)$ and $\boldsymbol{\ell} = (\mathbf{x}^{\beta+1} - \mathbf{x}^\beta)/2 - \mathbf{x}^\alpha$. We can calculate the stress tensor by taking the symmetric part of this tensor using Equation (11). Contrary to the strain tensor calculated for spherical particles [8,16], the individual rotations of the particles are included in the calculation of this tensor. This is borne out by the fact that for nonspherical particles the branch vector is not perpendicular to the contact normal vector, so that $e_{ik} \ell_k^\beta N_j^\beta \neq 0$.

3. Incremental theory

Since the stress and the strain are symmetric tensors, it is advantageous to simplify the notation by defining these quantities as six-dimensional vectors:

$$\tilde{\sigma} = \begin{bmatrix} \bar{\sigma}_{11} \\ \bar{\sigma}_{22} \\ \bar{\sigma}_{33} \\ \sqrt{2}\bar{\sigma}_{23} \\ \sqrt{2}\bar{\sigma}_{31} \\ \sqrt{2}\bar{\sigma}_{13} \end{bmatrix} \quad \text{and} \quad \tilde{\epsilon} = \begin{bmatrix} \bar{\epsilon}_{11} \\ \bar{\epsilon}_{22} \\ \bar{\epsilon}_{33} \\ \sqrt{2}\bar{\epsilon}_{23} \\ \sqrt{2}\bar{\epsilon}_{31} \\ \sqrt{2}\bar{\epsilon}_{13} \end{bmatrix}. \tag{19}$$

The coefficient $\sqrt{2}$ allows us to preserve the metric in this transformation: $\tilde{\sigma}_k\tilde{\sigma}_k = \bar{\sigma}_{ij}\bar{\sigma}_{ij}$. The relation between these two vectors will be established in the general context of the rate-independent incremental constitutive relations. We will focus on two particular theoretical developments: the theory of hypoplasticity and the elastoplastic models. The similarities and differences of both formulations are presented in the framework of the incremental theory that follows.

3.1. General framework

In principle, the mechanical response of granular materials can be described by a functional dependence of the stress $\tilde{\sigma}(t)$ at time t on the strain history $\{\tilde{\epsilon}(t')\}_{0<t'<t}$. However, the mathematical description of this dependence turns out to be very complicated due to the nonlinearity and irreversible behavior of these materials. An incremental relation, relating the incremental stress $\mathrm{d}\tilde{\sigma}(t) = \sigma'(t)\mathrm{d}t$ to the incremental strain $\mathrm{d}\tilde{\epsilon}(t) = \epsilon'(t)\mathrm{d}t$ and some internal variables κ that account for the deformation history, enable us to avoid these mathematical difficulties [17]. This incremental scheme is also useful for solving geotechnical problems, since the finite-element codes require that the constitutive relation be expressed incrementally.

Due to the large number of existing incremental relations, the necessity of a unified theoretical framework has been pointed out as an essential necessity to classify all the existing models [18]. In general, the incremental stress is related to the incremental strain as follows:

$$\mathcal{F}_\kappa(\mathrm{d}\tilde{\epsilon}, \mathrm{d}\tilde{\sigma}, \mathrm{d}t) = 0. \tag{20}$$

Let us look at the special case where there is no rate-dependence in the constitutive relation. This means that this relation is not influenced by the time required during any loading tests, as corresponds to the quasistatic approximation. In this case $\mathcal{F}$ is invariant with respect to $\mathrm{d}t$, and (20) can be reduced to:

$$\mathrm{d}\tilde{\epsilon} = \mathcal{G}_\kappa(\mathrm{d}\tilde{\sigma}) \tag{21}$$

In particular, the rate-independent condition implies that multiplying the loading time by a scalar λ does not affect the incremental stress-strain relation:

$$\forall\lambda, \quad \mathcal{G}_\kappa(\lambda\mathrm{d}\tilde{\sigma}) = \lambda\mathcal{G}_\kappa(\mathrm{d}\tilde{\sigma}) \tag{22}$$

The significance of this equation is that $\mathcal{G}_\kappa$ is an homogeneous function of degree one. In this case, the application of the Euler identity shows that (21) leads to

$$\mathrm{d}\tilde{\epsilon} = M_\kappa(\hat{\sigma})\mathrm{d}\tilde{\sigma}, \tag{23}$$

where $M_\kappa = \partial \mathcal{G}_\kappa / \partial(\mathrm{d}\tilde{\sigma})$ and $\hat{\sigma}$ is the unitary vector defining the direction of the incremental stress

$$\hat{\sigma} = \frac{\mathrm{d}\tilde{\sigma}}{|\mathrm{d}\tilde{\sigma}|}. \tag{24}$$

Equation (23) represents the general expression for the rate-independent constitutive relation. In order to determine the dependence of M on $\hat{\sigma}$, one can either perform experiments by taking different loading directions, or postulate explicit expressions based on a theoretical framework. The first approach will be considered in the Subsection 5.2, and the discussion about some existing theoretical models will be presented in the following.

3.2. Elastoplastic models

The classical elastoplasticity theory was established by Drucker and Prager in the context if metal plasticity [19]. Some extensions have been developed to describe sand, clays, rocks, concrete, etc. [2,20]. Here we present a short review of these developments in soil mechanics.

When a granular sample, subjected to a confining pressure, is loaded in the axial direction, it displays a typical stress-strain response as shown in the left part of Figure 2. A continuous decrease of the stiffness (*i.e.*, the slope of the stress-strain curve) is observed during loading. If the sample is unloaded, an abrupt increase in the stiffness is observed, leaving an irreversible deformation. One observes that, if the stress is changed around some region below σ_A, which is called the *yield point*, the deformation is almost linear and reversible. The first postulate of the elastoplasticity theory establishes a stress region immediately below the yield point where only elastic deformations are possible.

Postulate 1: For each stage of loading there exists a yield surface which encloses a finite region in the stress space where only reversible deformations are possible.

The simple Mohr-Coulomb model assumes that the onset of plastic deformation occurs at failure [2]. However, it has been experimentally shown that plastic deformation develops before failure [21]. In order to provide a consistent description of these experimental results with the elastoplasticity theory, it is necessary to assume that the yield function changes with the deformation process [21]. This condition is schematically shown in Figure 2. Let us assume that the sample is loaded until it reaches the stress σ_A upon which it is slightly unloaded. If the sample is reloaded, it is able to recover the same stress-strain relation of the monotonic loading once it reaches the yield point σ_A again. If one increases the load to the stress σ_B, a new elastic regime can be observed after a loading reversal. In the elastoplasticity

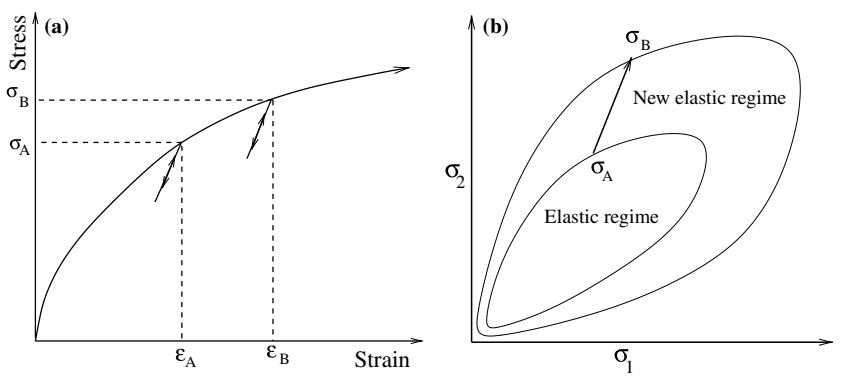


Figure 2. Evolution of the elastic regime (a) stress-strain relation (b) elastic regime in the stress space.

theory, this result is interpreted by assuming that the elastic regime is expanded to a new stress region below the yield point σ_B.

Postulate 2: The yield function remains when the deformations take place inside of the elastic regime, and it changes as the plastic deformation evolves.

The transition from the elastic to the elastoplastic response is extrapolated for more general deformations. Part (b) of Figure 2 shows the evolution of the elastic region when the yield point moves in the stress space from $\tilde{\sigma_A}$ to $\tilde{\sigma_B}$. The essential goal in the elastoplasticity theory is to find the correct description of the evolution of the elastic regime with the deformation, which is called the *hardening law*.

We will finally introduce the third basic assumption from elastoplasticity theory:

Postulate 3: The strain can be separated in an elastic (recoverable) and a plastic (unrecoverable) component:

$$\mathrm{d}\tilde{\epsilon} = \mathrm{d}\tilde{\epsilon}^e + \mathrm{d}\tilde{\epsilon}^p, \tag{25}$$

The incremental elastic strain is linked to the incremental stress by introducing an elastic tensor as

$$\mathrm{d}\tilde{\sigma} = D(\tilde{\sigma})\mathrm{d}\tilde{\epsilon}^e. \tag{26}$$

To calculate the incremental plastic strain, we introduce the *yield surface* as

$$f(\sigma, \kappa) = 0, \tag{27}$$

where κ is introduced as an internal variable, which describes the evolution of the elastic regime with the deformation. From experimental evidence, it has been shown that this variable can be taken as a function of the cumulative plastic strain ϵ^p given by [2,20]

$$\epsilon^p \equiv \int_0^t \sqrt{\mathrm{d}\epsilon_k \mathrm{d}\epsilon_k}\mathrm{d}t \tag{28}$$

When the stress state reaches the yield surface, the plastic deformation evolves. This is assumed to be derived from a scalar function of the stress as follows:

$$\mathrm{d}\epsilon_j^p = \Lambda \frac{\partial g}{\partial \sigma_j}, \tag{29}$$

where g is the so-called *plastic potential* function. Following the Drucker-Prager postulates we can show that $g = f$ [19]. However, a considerable amount of experimental evidence has shown that in soils the plastic deformation is not perpendicular to the yield surfaces [22]. It is necessary to introduce this plastic potential to extend the Drucker-Prager models to the so-called *non-associated* models.

The parameter Λ of (29) can be obtained from the so-called *consistence condition*. This condition comes from the second postulate, which establishes that, after the movement of the stress state from $\tilde{\sigma_A}$ to $\tilde{\sigma_B} = \tilde{\sigma_A} + \tilde{\mathrm{d}\sigma}$, the elastic regime must be expanded so that $\mathrm{d}f = 0$, as shown in Part (b) of Figure 2. Using the chain rule one obtains:

$$\mathrm{d}f = \frac{\partial f}{\partial \sigma_i}\mathrm{d}\sigma_i + \frac{\partial f}{\partial \kappa}\frac{\partial \kappa}{\partial \epsilon_j^p}\mathrm{d}\epsilon_j^p = 0. \tag{30}$$

Inserting (29) into (30), we obtain the parameter Λ, *viz.*

$$\Lambda = -\left(\frac{\partial f}{\partial \kappa}\frac{\partial \kappa}{\partial \epsilon_j^p}\frac{\partial g}{\partial \sigma_j}\right)^{-1}\frac{\partial f}{\partial \sigma_i}\mathrm{d}\sigma_i. \tag{31}$$

We define the vectors $N_i^y = \partial f/\partial\sigma_i$ and $N_i^f = \partial g/\partial\sigma_i$ and the unit vectors $\hat{\phi} = \mathbf{N}^y/|\mathbf{N}^y|$ and $\hat{\psi} = \mathbf{N}^f/|\mathbf{N}^f|$ as the *flow direction* and the *yield direction*. In addition, the *plastic modulus* is defined as

$$h = -\frac{1}{|\mathbf{N}^y||\mathbf{N}^f|}\frac{\partial f}{\partial\kappa}\frac{\partial\kappa}{\partial\epsilon_i^p}\frac{\partial g}{\partial\sigma_i}. \tag{32}$$

Substituting (31) in (29) and using (32), we obtain:

$$\mathrm{d}\tilde{\epsilon}^p = \frac{1}{h}\hat{\phi}\cdot\mathrm{d}\tilde{\sigma}\ \ \hat{\psi}. \tag{33}$$

Note that this equation has been calculated in the case where the stress increment takes place outside the yield surface. If the stress increment occurs inside the yield surface, the second Drucker-Prager postulate establishes that $\mathrm{d}\tilde{\epsilon}^p = 0$. Thus, the generalization of (33) is given by

$$\mathrm{d}\tilde{\epsilon}^p = \frac{1}{h}\langle\hat{\phi}\cdot\mathrm{d}\tilde{\sigma}\rangle\ \ \hat{\psi}, \tag{34}$$

where $\langle x\rangle = x$ when $x > 0$ and $\langle x\rangle = 0$ otherwise. Finally, the total strain response can be obtained from (25) and (34):

$$\mathrm{d}\tilde{\epsilon} = D^{-1}(\tilde{\sigma})\mathrm{d}\tilde{\sigma} + \frac{1}{h}\langle\hat{\phi}\cdot\mathrm{d}\tilde{\sigma}\rangle\ \ \hat{\psi}. \tag{35}$$

From this equation one can distinguish two zones in the incremental stress space where the incremental relation is linear. They are the so-called tensorial zones defined by Darve and Laouafa [17]. The existence of two tensorial zones and the continuity of the incremental response at the boundary confirm that these two zones are essential features of the elastoplastic models.

Although the elastoplasticity theory has been proved to work well for monotonically increasing loading, it shows some deficiencies in the description of complex loading histories [23, pp. 230–262]. There is also an extensive body of experimental evidence that shows that the elastic regime is extremely small and that the transition from elastic to an elastoplastic response is not as sharp as the theory predicts [24].

The *bounding surface models* have been introduced to generalize the classical elastoplastic concepts [25]. Apart from the critical-state line, these models introduce the so-called bounding surface in the stress space. For any given stress state within the surface, a proper mapping rule associates a corresponding *image* stress point on this surface. A measure of the distance between the actual and the image stress points is used to specify the plastic modulus in terms of a plastic modulus at the image stress state. Besides the versatility of this theory to describe a wide range of materials, it has the advantage that the elastic regime can be considered as vanishingly small, so that this model can be used to give a realistic description of unbounded granular soils. In the authors' opinion, the most striking aspect of the bounding-surface theory with vanishing elastic range is that it establishes a convergence point for two different approaches of constitutive modeling: the elastoplastic approaches originated from the Drucker-Prager theory, and the more recently developed hypoplastic theories.

3.3. Hypoplastic models

In recent years, an alternative approach for modeling soil behavior has been proposed, which departs from the framework of the elastoplasticity theory [26–28]. The distinctive features of this approach are:

– *The absence of the decomposition of strain in plastic and elastic components.*
– *The statement of a nonlinear dependence of the incremental response with the strain rate directions.*

The most general expression has been provided by the so-called second-order incremental nonlinear models [27]. A particular class of these models which has received special attention in recent times is provided by the theory of hypoplasticity [26,28]. A general outline of this theory was proposed by Kolymbas [26], leading to different formulations, such as the K-hypoplasticity developed in Karlsruhe [29], and the CLoE-hypoplasticity originated in Grenoble [28]. In hypoplasticity, the most general constitutive equation takes the following form:

$$\mathrm{d}\tilde{\sigma} = L(\tilde{\sigma}, v)\mathrm{d}\tilde{\epsilon} + \tilde{N}(\tilde{\sigma}, v)|\mathrm{d}\tilde{\epsilon}|, \tag{36}$$

where L is a second-order tensor and $\tilde{N}$ is a vector, both depending on the stress $\tilde{\sigma}$ and the void ratio v. Hypoplastic equations provide a simplified description of loose and dense unbounded granular materials. A reduced number of parameters are introduced, which are very easy to calibrate [30].

In the theory of hypoplasticity, the stress-strain relation is established by means of an incremental nonlinear relation without any recourse to yield or boundary surfaces. This nonlinearity is reflected in the fact that the relation between the incremental stress and the incremental strain given in (36) is always nonlinear. The incremental nonlinearity of the granular materials is still under discussion. Indeed, an important feature of the incremental nonlinear constitutive models is that they break away from the superposition principle:

$$\mathrm{d}\tilde{\sigma}(\mathrm{d}\tilde{\epsilon}_1 + \mathrm{d}\tilde{\epsilon}_2) \neq \mathrm{d}\tilde{\sigma}(\mathrm{d}\tilde{\epsilon}_1) + \mathrm{d}\tilde{\sigma}(\mathrm{d}\tilde{\epsilon}_2), \tag{37}$$

which is equivalent to:

$$\mathrm{d}\tilde{\epsilon}(\mathrm{d}\tilde{\sigma}_1 + \mathrm{d}\tilde{\sigma}_2) \neq \mathrm{d}\tilde{\epsilon}(\mathrm{d}\tilde{\sigma}_1) + \mathrm{d}\tilde{\epsilon}(\mathrm{d}\tilde{\sigma}_2) \tag{38}$$

An important consequence of this feature is addressed by Kolymbas [31, pp. 213–223] and Darve et al. [27]. They consider two stress paths; the first one is smooth and the second results from the superposition of small deviations as shown in Figure 3. The superposition principle establishes that the strain response of the stair-like path converges to the response of the proportional loading in the limit of arbitrarily small deviations. More precisely, the strain deviations $\Delta\epsilon$ and the steps of the stress increments $\Delta\sigma$ satisfy $\lim_{\Delta\sigma\to 0}\Delta\epsilon = 0$. For the hypoplastic equations, and in general for the incremental nonlinear models, this condition is never satisfied. For incremental relations with tensorial zones, this principle is satisfied whenever the increments take place inside one of these tensorial zones. It should be added that there is no experimental evidence disproving or confirming this rather questionable superposition principle.

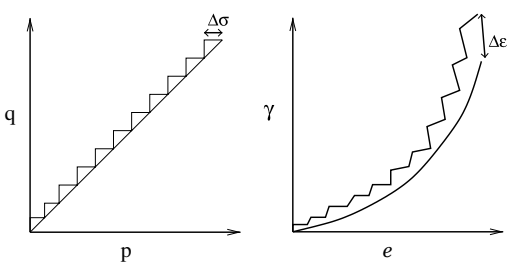


Figure 3. Smooth and stair-like stress paths and corresponding strain responses. p and q represent the pressure and the deviatoric stress. e and γ are the volumetric and deviatoric strain components.

4. Discrete model

We present here a two-dimensional discrete-element model which will be used to investigate the incremental response of granular materials. This model consists of randomly generated convex polygons, which interact via contact forces. There are some limitations in the use of such a two-dimensional code to model physical phenomena that are three-dimensional in nature. These limitations have to be kept in mind in the interpretation of the results and its comparison with the experimental data. In order to give a three-dimensional picture of this model, one can consider the polygons as a collection of bricks with randomly-shaped polygonal basis. Alternatively, one could consider the polygons as three-dimensional grains whose centers of mass all move in the same plane. In the authors' opinion, it is more sensible to consider this model as an idealized granular material that can be used to check the constitutive models.

The details of the particle generation, the contact forces, the boundary conditions and the molecular-dynamics simulations are presented in this section.

4.1. Generation of polygons

The polygons representing the particles in this model are generated by using the method of Voronoi tessellation [32]. This method is schematically shown in Figure 4. First, a regular square lattice of side ℓ is created. Then, we choose a random point in each cell of the rectangular grid. Subsequently, each polygon is constructed by assigning to each point that part of the plane that is nearer to it than to any other point. The details of the construction of the Voronoi cells can be found in the literature [33, 34].

By use of the Euler theorem, it has been shown analytically that the mean number of edges of this Voronoi construction must be 6 [34, pp. 295–296]. The number of edges of the polygons is distributed between 4 and 8 for 98·7% of the polygons. The orientational distribution of edges is isotropic, and the distribution of areas of polygons is symmetric around its mean value ℓ^2. The probabilistic distribution of areas follows approximately a Gaussian distribution with variance of $0{\cdot}36\ell^2$.

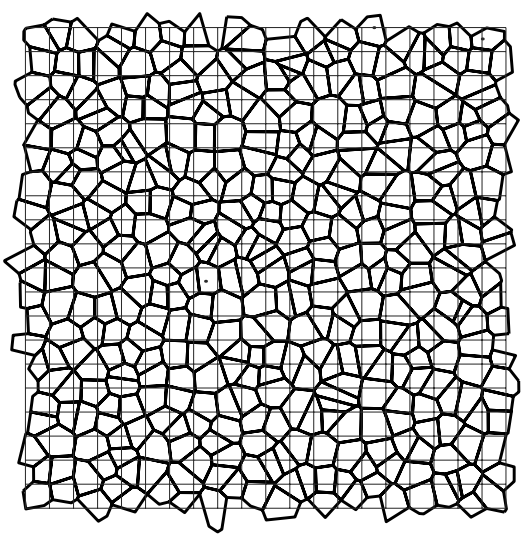


Figure 4. Voronoi construction used to generate the convex polygons. The dots indicate the point used in the tessellation. Periodic boundary conditions were used.

4.2. Contact forces

In order to calculate the forces, we assume that all the polygons have the same thickness L. The force between two polygons is written as $\mathbf{F}=\mathbf{f}L$ and the mass of the polygons is $M=mL$. In reality, when two elastic bodies come into contact, they have a slight deformation in the contact region. In the calculation of the contact force we assume that the polygons are rigid, but we allow them to overlap. Then, we calculate the force from this virtual overlap.

The first step towards the calculation of the contact force is the definition of the line representing the flattened contact surface between the two bodies in contact. This is defined from the contact points resulting from the intersection of the edges of the overlapping polygons. In most cases, we have two contact points, as shown in the left part of Figure 5. In such a case, the contact line is defined by the vector $\mathbf{C}=\overrightarrow{C_1C_2}$ connecting these two intersection points. In some pathological cases, the intersection of the polygons leads to four or six contact points. In these cases, we define the contact line by the vector $\mathbf{C}=\overrightarrow{C_1C_2}+\overrightarrow{C_3C_4}$ or $\mathbf{C}=\overrightarrow{C_1C_2}+\overrightarrow{C_3C_4}+\overrightarrow{C_5C_6}$, respectively. This choice guarantees a continuous change of the contact line, and therefore of the contact forces, during the evolution of the contact.

The contact force is separated as $\mathbf{f}^c=\mathbf{f}^e+\mathbf{f}^v$, where $\mathbf{f}^e$ and $\mathbf{f}^v$ are the elastic and viscous contribution. The elastic part of the contact force is decomposed as $\mathbf{f^e}=f_n^e\hat{n}^c+f_t^e\hat{t}^c$. The calculation of these components is explained below. The unit tangential vector is defined as $\hat{t}^c=\mathbf{C}/|\mathbf{C}|$, and the normal unit vector $\hat{n}^c$ is taken perpendicular to $\mathbf{C}$. The point of application of the contact force is taken as the center of mass of the overlapping polygons.

As opposed to the Hertz theory for round contacts, there is no exact way to calculate the normal force between interacting polygons. An alternative method has been proposed in order to model this force [35]. The normal elastic force is given by $f_n^e=-k_nA/L_c$, where k_n is the normal stiffness, A is the overlapping area and L_c is a characteristic length of the polygon pair. Our choice of L_c is $1/2(1/R_i+1/R_j)$ where R_i and R_j are the radii of the circles of the same area as the polygons. This normalization is necessary to be consistent in the units of force [32].

In order to model the quasistatic friction force, we calculate the elastic tangential force using an extension of the method proposed by Cundall-Strack [5]. An elastic force $f_t^e=-k_t\Delta x_t$ proportional to the elastic displacement is included at each contact, where k_t is the tangential stiffness. The elastic displacement Δx_t is calculated as the time integral of the tangential velocity of the contact during the time when the elastic condition $|f_t^e|<\mu f_n^e$ is satisfied. The sliding condition is imposed, keeping this force constant when $|f_t^e|=\mu f_n^e$. The

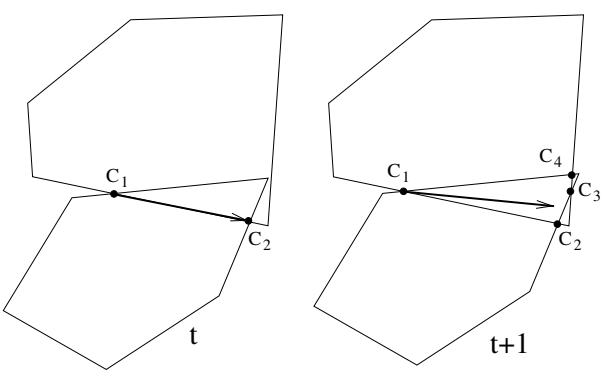


Figure 5. Contact points C_i before (left) and after the formation of a pathological contact (right). The vector denotes the contact line; t represents the time step.

straightforward calculation of this elastic displacement is given by the time integral starting at the beginning of the contact:

$$\Delta x_t^e = \int_0^t v_t^c(t')\Theta(\mu f_n^e - |f_t^e|)\mathrm{d}t', \tag{39}$$

where Θ is the Heaviside step function and $\mathbf{v}_t^c$ denotes the tangential component of the relative velocity $\mathbf{v}^c$ at the contact:

$$\mathbf{v}^c = \mathbf{v}_i - \mathbf{v}_j + \boldsymbol{\omega}_i \times \boldsymbol{\ell}_i - \boldsymbol{\omega}_j \times \boldsymbol{\ell}_j. \tag{40}$$

Here $\mathbf{v}_i$ is the linear velocity and $\boldsymbol{\omega}_i$ is the angular velocity of the particles in contact. The branch vector $\boldsymbol{\ell}_i$ connects the center of mass of particle i with the point of application of the contact force.

Damping forces are included in order to allow for rapid relaxation during the preparation of the sample, and to reduce the acoustic waves produced during the loading. These forces are calculated as

$$\mathbf{f}_v^c = -m(\gamma_n v_n^c \hat{n}^c + \gamma_t v_t^c \hat{t}^c), \tag{41}$$

$m = (1/m_i + 1/m_j)^{-1}$ being the effective mass of the polygons in contact; $\hat{n}^c$ and $\hat{t}^c$ are the normal and tangential unit vectors defined before, and γ_n and γ_t are the coefficients of viscosity. These forces introduce time-dependent effects during the cyclic loading. However, we will show that these effects can be arbitrarily reduced by increasing the loading time, which corresponds to the quasistatic approximation.

4.3. Molecular-dynamics simulation

The evolution of the position $\mathbf{x}_i$ and the orientation φ_i of the ith polygon is governed by the equations of motion:

$$\begin{aligned} m_i \ddot{\mathbf{x}}_i &= \sum_c \mathbf{f}_i^c + \sum_b \mathbf{f}_i^b, \\ I_i \ddot{\varphi}_i &= \sum_c \boldsymbol{\ell}_i^c \times \mathbf{f}_\mathbf{i}^\mathbf{c} + \sum_b \boldsymbol{\ell}_i^b \times \mathbf{f}_i^b. \end{aligned} \tag{42}$$

Here m_i and I_i are the mass and moment of inertia of the polygon i. The first summation goes over all particles in contact with this polygon. According to the previous section, these forces $\mathbf{f}^\mathbf{c}$ are given by

$$\mathbf{f}^c = -(k_n A/L_c + \gamma_n m v_n^c)\mathbf{n}^c - (\Delta x_t^c + \gamma_t m v_t^c)\mathbf{t}^c, \tag{43}$$

The second summation on the right hand of (42) goes over all the vertices of the polygons in contact with the walls. This interaction is modeled by using a simple visco-elastic force. First, we allow the polygons to penetrate the walls. Then, for each vertex of the polygon α inside of the walls we include a force

$$\mathbf{f}^b = -k_n \delta \mathbf{n} - \gamma_b m_\alpha \mathbf{v}^b, \tag{44}$$

where δ is the penetration length of the vertex, $\mathbf{n}$ is the unit normal vector to the wall, and $\mathbf{v}^b$ is the relative velocity of the vertex with respect to the wall.

We use a fifth-order Gear predictor-corrector method for solving the equation of motion [36, pp. 340–342]. This algorithm consists of three steps. The first step predicts position and

velocity of the particles by means of a Taylor expansion. The second step calculates the forces as a function of the predicted positions and velocities. The third step corrects the positions and velocities in order to optimize the stability of the algorithm. This method is much more efficient than the simple Euler approach or the Runge-Kutta method, especially for problems where very high accuracy is a requirement.

The parameters of the molecular-dynamics simulations were adjusted according to the following criteria: (1) guarantee the stability of the numerical solution, (2) optimize the time of the calculation, and (3) provide a reasonable agreement with the experimental data.

There are many parameters in the molecular-dynamics algorithm. Before choosing them, it is convenient to make a dimensional analysis. In this way, we can maintain the scale invariance of the model and reduce the parameters to a minimum of dimensionless constants. First, we introduce the following characteristic times of the simulations: the loading time t_0, the relaxation times $t_n = 1/\gamma_n$, $t_t = 1/\gamma_t$, $t_b = 1/\gamma_b$ and the characteristic period of oscillation $t_s = \sqrt{\rho \ell^2 / k_n}$ of the normal contact.

Using the Buckingham Pi theorem [37], one can show that the strain response, or any other dimensionless variable measuring the response of the assembly during loading, depends only on the following dimensionless parameters: $\alpha_1 = t_n/t_s$, $\alpha_2 = t_t/t_s$, $\alpha_3 = t_b/t_s$, $\alpha_4 = t_0/t_s$, the ratio $\zeta = k_t/k_n$ between the stiffnesses, the friction coefficient μ and the ratio σ_i/k_n between the stresses applied on the walls and the normal stiffness.

The variables α_i will be called *control parameters*. They are chosen in order to satisfy the quasistatic approximation, *i.e.*, the response of the system does not depend on the loading time, but a change of these parameters within this limit does not affect the strain response. Parameter values $\alpha_2 = 0{\cdot}1$ and $\alpha_2 = 0{\cdot}5$ were taken large enough to have a high dissipation, but not too large to keep the numerical stability of the method. The value $\alpha_3 = 0{\cdot}001$ is chosen by optimizing the time of consolidation of the sample in the Subsection 4.4. The ratio $\alpha_4 = t_0/t_s = 10{,}000$ was chosen large enough in order to avoid rate-dependence in the strain response, corresponding to the quasistatic approximation. Technically, this is performed by looking for the value of α_4 such that a reduction of it by half results in a change of the stress–strain relation less than 5%.

The parameters ζ and μ can be considered as *material parameters*. They determine the constitutive response of the system, so they must be adjusted to the experimental data. In this study, we have adjusted them by comparing the simulation of biaxial tests of dense polygonal packings with the corresponding biaxial tests with dense Hostun sand [38]. First, the initial Young modulus of the material is linearly related to the value of normal stiffness of the contact. Thus, $k_n = 160$ MPa is chosen by fitting the initial slope of the stress–strain relation in the biaxial test. Then, the Poisson ratio depends on the ratio $\zeta = k_t/k_n$. Our choice $k_t = 52{\cdot}8$MPa gives an initial Poisson ratio of 0·2. This is obtained from the initial slope of the curve of volumetric strains versus axial strain. The angles of friction and the dilatancy are increasing functions of the friction coefficient μ. Taking $\mu = 0{\cdot}25$ yields angles of friction of 30–40 degrees and dilatancy angles of 10–20 degrees, which are similar to the experimental data of river sand [39].

4.4. Sample preparation

The Voronoi construction presented above leads to samples with no porosity. This ideal case contrasts with realistic soils, where only porosities up to a certain value can be achieved. In this section, we present a method to create stable, irregular packings of polygons with a given porosity.

The porosity can be defined using the concept of void ratio. This is defined as the ratio of the volume of the void space to the volume of the solid material. It can be calculated as:

$$\nu = \frac{V_t}{V_f - V_0} - 1. \tag{45}$$

This is given in terms of the area enclosed by the walls V_t, the sum of the areas of the polygons V_f and the sum of the overlapping areas between them V_0.

Of course, there is a maximal void ratio that can be achieved, because it is impossible to pack particles with an arbitrarily high porosity. The maximal void ratio ν_m can be detected by moving the walls until a certain void ratio is reached. We find a critical value, above which the particles can be arranged without touching. Since there are no contacts, the assembly cannot support a load, and even applying gravity will cause it to compactify. For a void ratio below this critical value, there will be particle overlap, and the assembly is able to sustain a certain load. This maximal value is around 0·28.

The void ratio can be decreased by reducing the volume between the walls. The drawback of this method is that it leads to significant differences between the inner and outer parts of the boundary assembly and hence unrealistic overlaps between the particles, giving rise to enormous pressures. Alternatively, one can confine the polygons by applying gravity to the particles and on the walls. Particularly, homogeneous, isotropic assemblies, as shown in Figure 6 can be generated by a gravitational field $\mathbf{g} = -g\mathbf{r}$, where $\mathbf{r}$ is the vector connecting the center of mass of the assembly to the center of the polygon.

When the sample is consolidated, repeated shear-stress cycles are applied from the walls, superimposed to a confining pressure. The external load is imposed by applying a force $[p_c + q_c \sin(2\pi t/t_0)]W$ and $[p_c + q_c \cos(2\pi t/t_0)]H$ on the horizontal and vertical walls, respectively. Here W and H are the width and the height of the sample, respectively. If we take $p_c = 16\,\text{kPa}$ and $q_c < 0{\cdot}4p_c$, we observe that the void ratio decreases as the number of cycles increases. Void ratios of around 0·15 can be obtained. It is worth mentioning that after some thousands of cycles the void ratio is still slowly decreasing, making it difficult to identify this minimal value.

5. Simulation results

In order to investigate different aspects of the incremental response, some numerical simulations were performed. Different polygonal assemblies of 400 particles were used in the calculations. The loading between two stress states was controlled by applying forces $[\sigma_1^i + (\sigma_1^f - \sigma_1^i)r(t)]W$ and $[\sigma_2^i + (\sigma_2^f - \sigma_2^i)r(t)]H$. A smooth modulation $r(t) = (1 - \cos(2\pi t/t_0))/2$ is chosen in order to minimize the acoustic waves produced during loading. The initial void ratio is around $\nu = 0{\cdot}22$.

The components of the stress are represented by $p = (\sigma_1 + \sigma_2)/2$ and $q = (\sigma_1 - \sigma_2)/2$, where σ_1 and σ_2 are the eigenvalues of the averaged stress tensor on the RVE. Equivalently, the coordinates of the strain are given by the sum $\gamma = \epsilon_2 - \epsilon_1$ and the difference $e = -\epsilon_1 - \epsilon_2$ of the eigenvalues of the strain tensor. We use the convention that $e > 0$ means compression of the sample. The diameter of the RVE is taken as $d = 16\ell$. All the calculations were taken in the quasistatic regime.

5.1. Superposition principle

We start by exploring the validity of the superposition principle presented in Subsection 3.3. Part (a) of Figure 7 shows the loading path during the proportional loading under constant

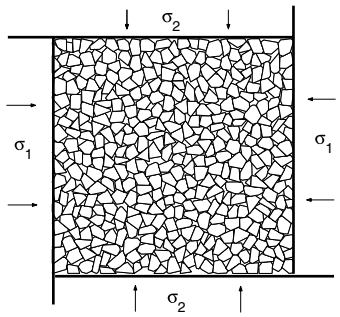


Figure 6. Polygonal assembly confined by a rectangular frame of walls.

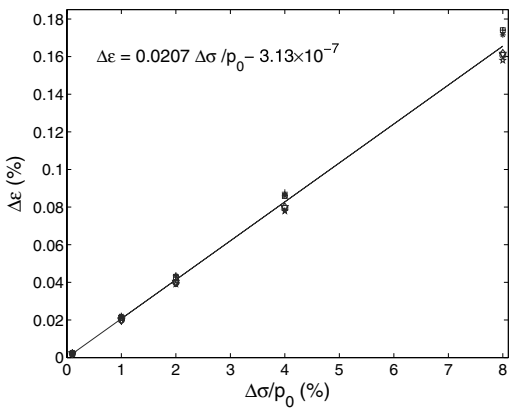


Figure 8. Distance between the response of the stair-like path and the proportional path.

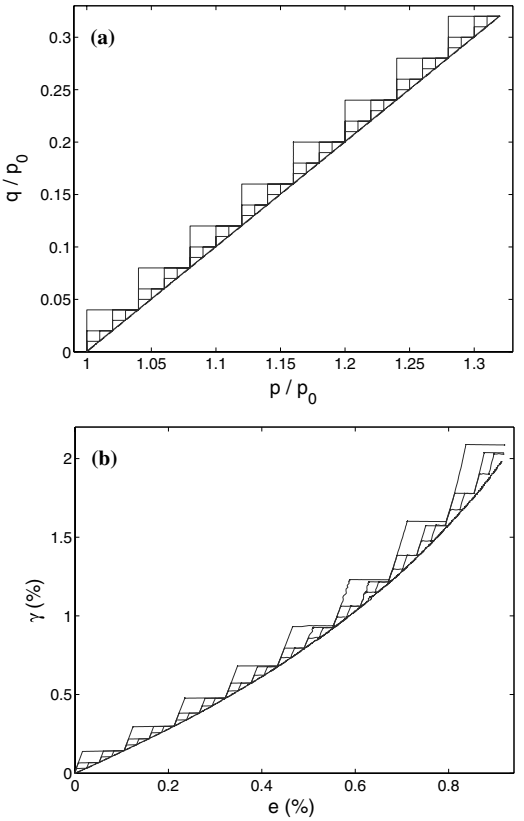


Figure 7. Numerical responses obtained from MD simulations of a rectilinear proportional loading (with constant lateral pressure) and stair-like paths. (a) Loading stress paths. (b) corresponding strain responses.

lateral pressure. This path is also decomposed into pieces divided into two parts: one is an incremental isotropic loading ($\Delta p = \Delta\sigma$ and $\Delta q = 0$), the other an incremental pure-shear loading ($\Delta q = \Delta\sigma$ and $\Delta p = 0$). The length of the steps $\Delta\sigma$ varies from to $0{\cdot}4 p_0$ to $0{\cdot}001 p_0$, where $p_0 = 640\,\text{kPa}$. Part (b) of Figure 7 shows that, as the steps decrease, the strain response converges to the response of the proportional loading. In order to verify this convergence, we calculate the difference between the strain response of the stair-like path $\gamma(e)$ and the proportional loading path $\gamma_0(e)$ as:

$$\Delta\epsilon \equiv \max_e |\gamma(e) - \gamma_0(e)|, \tag{46}$$

for different steps sizes. This is shown in Figure 8 for seven different polygonal assemblies. The linear interpolation of this data intersects the vertical axis at 3×10^{-7}. Since this value is below the error given by the quasi-static approximation, which is 3×10^{-4}, the results suggest that the responses converge to that of the proportional load. Therefore we find that, within our error bars, the superposition principle is valid.

Close inspection of the incremental response will show that the validity of the superposition principle is linked to the existence of tensorial zones in the incremental-stress space. Prior to this, a short introduction to the strain envelope responses will be given.

5.2. Incremental response

A graphical illustration of the constitutive models can be given by employing the so-called *response envelopes*. They were introduced by Gudehus [18] as a useful tool to visualize the properties of a given incremental constitutive equation. The strain-envelope response is defined as the image $\{d\tilde{\epsilon} = \mathcal{G}(d\tilde{\sigma}, \tilde{\sigma})\}$ in the strain space of the unit sphere in the stress space, which becomes a potato-like surface in the strain space.

In practice, the determination of the stress-envelope responses is difficult because it requires one to prepare many samples with identical material properties. Numerical simulations result as an alternative to the solution of this problem. They allow one to create clones of the same sample, and to perform different loading histories in each one of them.

In the case of a plane-strain test, where there is no deformation in one of the spatial directions, the strain-envelope response can be represented in a plane. According to (36), this response results in a rotated, translated ellipse in the hypoplastic theory, whereas it is given by a continuous curve consisting of two pieces of ellipses in the elastoplasticity theory, as a result of (35). It is then of obvious interest to compare these predictions with the stress-envelope response of the experimental tests.

Figure 9 shows the typical strain response resulting from different stress-controlled loadings in a dense packing of polygons. Each point is obtained from the strain response in a specific direction of the stress space, with the same stress amplitude of $10^{-4}p_0$. We take $q_0 = 0{\cdot}45p_0$ and $p_0 = 160\,\text{kPa}$ in this calculation. The best fit of these results in the envelopes response of the elastoplasticity (two pieces of ellipses). For comparison the hypoplasticity (one ellipse) is also shown in Figure 9.

From these results we conclude that the elastoplasticity theory is more accurate in describing the incremental response of our model. One can draw the same conclusion by taking different strain values with different initial stress values [40]. These results have shown that the incremental response can be accurately described using the elastoplastic relation of Equation (35). The validity of this equation is supported by the existence of a well-defined flow rule for each stress state [41].

5.3. Yield function

In Subsection 3.2, we showed that the yield surface is an essential element in the formulation of the Drucker-Prager theory. This surface encloses a hypothetical region in the stress space where only elastic deformations are possible [19]. The determination of such a yield surface is essential to determine the dependence of the strain response on the history of the deformation.

We attempt to detect the yield surface by using a standard procedure proposed in experiments with sand [24]. Figure 10 shows this procedure. Initially the sample is subjected to an isotropic pressure. Then the sample is loaded in the axial direction until it reaches the yield-stress state with pressure p and deviatoric stress q. Since plastic deformation is found at this stress value, the point (p, q) can be considered as a classical yield point. Then, the Drucker-Prager theory assumes the existence of a yield surface containing this point. In order to explore the yield surface, the sample is unloaded in the axial direction until it reaches the stress point with pressure $p - \Delta p$ and deviatoric stress $q - \Delta p$ inside the elastic regime. Then the yield surface is constructed by re-loading in different directions in the stress space. In each direction, the new yield point must be detected by a sharp change of the slope in the stress-strain curve, indicating plastic deformations.

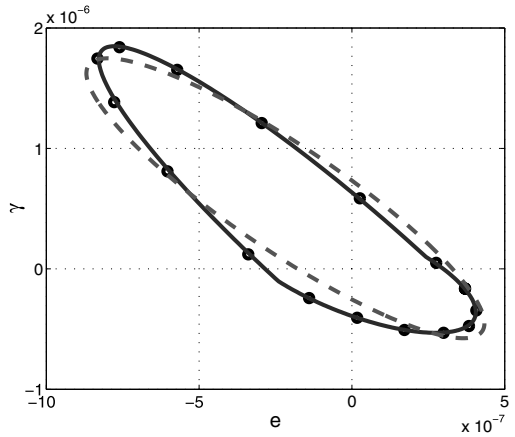


Figure 9. Numerical calculation of the incremental strain response. The dots are the numerical results. The solid curve represents the fit to the elastoplasticity theory. The dashed curve is the hypoplastic fit.

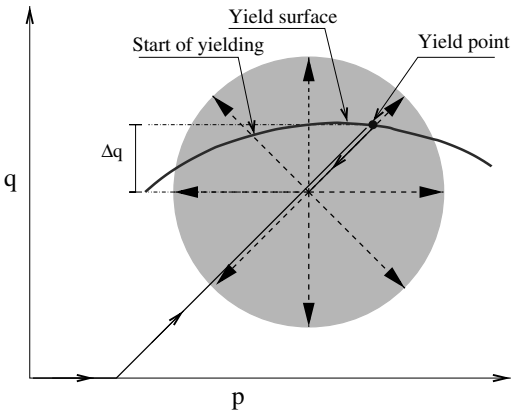


Figure 10. Method to obtain the yield surface. Load-unload-reload tests are performed taking different directions in the reload path. In each direction, the point of the reload path where the yielding begins is marked. The yield function is constructed by connecting these points.

Figure 11 shows the strain response taking different load directions in the same sample. The initial stress of the sample is given by $q_0 = 0{\cdot}5 p_0$ and $p_0 = 160\,\text{kPa}$. If the direction of the reload path is the same as that of the original load ($\theta = 45^\circ$), we observe a sharp decrease of stiffness when the load point reaches the initial yield point, which corresponds to the origin in Figure 11. However, if one takes a direction of re-loading different from 45°, the decrease of the stiffness with the loading becomes smooth. Since there is no straightforward way to identify those points where the yielding begins, the yield function, as it was introduced by Drucker and Prager [19] in order to describe a sharp transition between the elastic and plastic regions, is not consistent with our results.

Experimental studies on dry sand seem to show that the truly elastic region is probably extremely small [4]. Moreover, a micro-mechanic investigation of the mechanical response

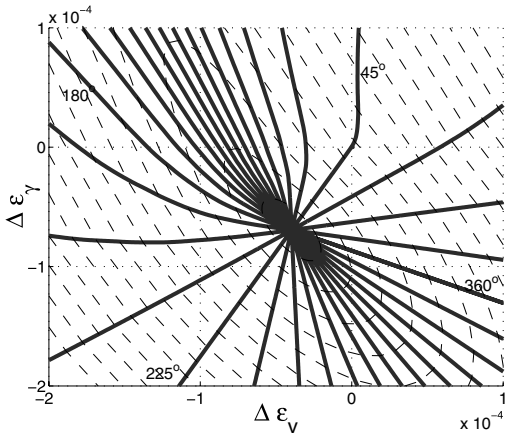


Figure 11. Strain responses according to Figure 10. The solid lines show the strain response from different reload directions. The dashed contours represent the strain envelope responses for different stress increments $|\Delta\boldsymbol{\sigma}|$.

of granular ratcheting under cyclic loading has shown that any load involves sliding contacts, and hence, plastic deformation [41]. These studies lead to the conclusion that the elastic region, used in the Drucker-Prager theory to give a dependence of the response on recent history, is not a necessary feature of granular materials.

The absence of elastic regime leads to the following question: *Is the hypoplastic theory more appropriate than the elastoplastic models to describe these simulations results?* Since hypoplastic models do not introduce any elastic regime, they seem to provide a good alternative. However, hypoplasticity departs from the superposition principle, which is not consistent with our results. An alternative approach to hypoplasticity could be made from the bounding-surface theory, by shrinking the elastic regime to the current stress point [42]. With this limit one can reproduce the observed continuous transition from the elastic to the elastoplastic behavior and at the same time keep the tensorial zones. However, it has been shown that this limit leads to a constitutive relation in terms of some internal variables, which in this theory lack physical meaning. In the authors' opinion, a micro-mechanical interpretation of these internal variables will be important to capture this essential feature of mechanics of granular materials, namely that any loading involves plastic deformation.

6. Instabilities

Instability has been one of the classical topics in soil mechanics. Localization from a previously homogeneous deformation to a narrow zone of intense shear is a common mode of failure of soils [21,39,43]. The Mohr-Coulomb criterion is typically used to understand the principal features of the localization [43]. This criterion was improved by the Drucker condition, based on the hypothesis of the normality, which results in a plastic flow perpendicular to the yield surface [19]. This theory predicts that the instability appears when the stress of the sample reaches the plastic limit surface. This surface is given by the stress states where the plastic deformation becomes infinite. In previous work, it is shown that the normality postulate is not fulfilled in the incremental response of this model, because the flow and yield directions given by (34) are different [40]. Thus, it is interesting to see if the Drucker stability criterion is still valid.

According to the flow rule of (34), the plastic-limit surface can be found by looking for the stress values where the plastic modulus vanishes. The dependence of this modulus on the stress fits the following power law relation [40]:

$$h = h_0 \left[1 - \frac{q}{q_0} \left(\frac{p_0}{p} \right)^{\vartheta} \right]^{\eta}. \tag{47}$$

This is given in terms of the following four parameters: the plastic modulus $h_0 = 14{\cdot}5 \pm 0{\cdot}05$ at $q = 0$, the constant $q_0 = 0{\cdot}85 \pm 0{\cdot}05$, and the exponents $\eta = 2{\cdot}7 \pm 0{\cdot}04$ and $\vartheta = 0{\cdot}99 \pm 0{\cdot}02$. Then, the plastic-limit surface is given by the stress states with zero plastic modulus:

$$\frac{q_p}{q_0} = \left(\frac{p}{p_0} \right)^{\vartheta}. \tag{48}$$

On the other hand, the failure surface can be defined by the limit of the stress values where the material becomes unstable. It has been shown that this is given by the following relation [40]

$$\frac{q}{q_c} = \left(\frac{p}{p_0} \right)^{\beta}. \tag{49}$$

Here $p_0 = 1{\cdot}0\,\text{MPa}$ is the reference pressure, and $q_c = 0{\cdot}78 \pm 0{\cdot}03\,\text{MPa}$. The power-law dependence on the pressure, with exponent $\beta = 0{\cdot}92 \pm 0{\cdot}02$ implies a small deviation from the Mohr-Coulomb theory where the relation is strictly linear.

By comparing (49) and (48) one observes that during loading the instabilities appear before reaching the plastic-limit surface. Theoretical studies have also shown that in the case of nonassociated materials, *i.e.*, where the flow direction does not agree with the yield direction, the instabilities can appear strictly inside of the plastic-limit surface [17]. In this context, the question of instability must be reconsidered beyond the Drucker condition.

The stability for nonassociated elastoplastic materials has been investigated by Hill, who established the following sufficient stability criterion [44].

$$\forall \mathrm{d}\tilde{\epsilon}, \quad \mathrm{d}\tilde{\sigma} \cdot \mathrm{d}\tilde{\epsilon} > 0. \tag{50}$$

The analysis of this criterion of stability will be presented here based on the constitutive relation given by (35). The stability condition of this constitutive relation can be evaluated by introducing the normalized second-order work [17]:

$$\mathrm{d}^2 W \equiv \frac{\mathrm{d}\tilde{\sigma} \cdot \mathrm{d}\tilde{\epsilon}}{|\mathrm{d}\tilde{\sigma}|^2} \tag{51}$$

The Hill condition of stability can now be obtained by substituting (35) in this expression. This results in

$$\mathrm{d}^2 W = \hat{\sigma} D^{-1} \hat{\sigma} + \frac{\langle \cos(\theta + \phi) \rangle}{h} \cos(\theta + \psi) > 0, \tag{52}$$

where $\hat{\sigma}$ is defined in (24). In the case where the Drucker normality postulate $\psi = \phi$ is valid, Equation (52) is strictly positive and, therefore, this stability condition would be valid for all the stress states inside the plastic-limit surface. Indeed, for a nonassociated flow rule as in our model, the second-order work is not strictly positive for all the load directions, and can take zero values inside the plastic-limit surface (*i.e.*, during the hardening regime where $h > 0$).

To analyze this instability, we adopt a circular representation of $\mathrm{d}^2 W$ shown in Figure 12. The dashed circles in these figures represent those regions where $\mathrm{d}^2 W < 0$. For stress ratios below $q/p = 0{\cdot}7$ we found that the second order work is strictly positive. Thus, according to the Hill stability condition, this region corresponds to stable states. For the stress ratio $q/p = 0{\cdot}8$, the second-order work becomes negative between $27^\circ < \theta < 36^\circ$ and $206^\circ < \theta < 225^\circ$. This leads to a domain of the stress space that is strictly inside the plastic-limit surface where the Hill condition of stability is not fulfilled, and therefore the material is potentially unstable.

Numerical simulations of biaxial tests show that strain localization is the most typical mode of failure [7,45,46]. The fact that it appears before the sample reaches the plastic-limit surface suggests that instability is not completely determined by the current macroscopic stress of the material, as predicted by the Drucker-Prager theory. More recent analytic [47] and experimental [38,39] works have focused on the role of the micro-structure on the localized instabilities. Frictional slips at the particles have been used to define additional degrees of freedom [47]. The introduction of the particle diameter in the constitutive relations results in a correct prediction of the shear-band thickness [15, pp. 334–381]. The new degrees of freedom of these constitutive models are still not clearly connected to the micro-mechanical variables of the grains, but with the development of numerical simulations this aspect can be better understood.

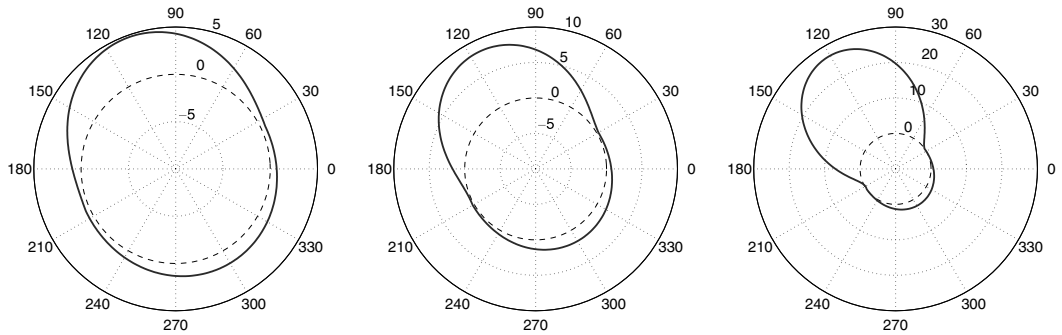


Figure 12. The solid lines show the second order work as a function of the direction of load for three different stress ratios $q/p = 0 \cdot 6$ (left), $0 \cdot 7$ (center), and $0 \cdot 65$ (right) with pressure $p = 160\,\mathrm{KPa}$. The dashed circles enclose the region where $d^2W < 0$.

7. Concluding remarks

In this paper we have obtained explicit expressions for the averaged stress and strain tensors over a RVE, in terms of the micro-mechanical variables, contact forces and the individual displacements and rotations of the grains. A short review on the incremental models has been given. The most salient aspects of both elastoplastic and hypoplastic models has been verified using molecular dynamics simulations on a polygonal packing. The results are summarized as follows:

- The elastoplasticity theory, with two tensorial zones, provides a more accurate description of the incremental response than the hypoplastic theory.
- In contradiction to the incremental nonlinear models, the simulation results show that the superposition principle is accurately satisfied.
- It is not possible to detect the finite elastic regime predicted by the elastoplasticity theory. This is due to the fact that the transition from elasticity to elastoplasticity is not as sharp as the Drucker-Prager theory predicts, but a smooth transition occurs.
- The calculation of the plastic-limit condition and the failure surface shows that the failure appears during the hardening regime $h > 0$. Using the Hill condition of stability, we may interpret the resulting instability as an effect of the non-associated flow rule of the plastic deformations.

These conclusions appear to contradict both the Drucker-Prager theory and the hypoplastic models. In future work, it would be important to revisit the question of the incremental nonlinearity of soils from micro-mechanical considerations. The recent improvements in discrete-element modeling allow one to perform this investigation. We are now ready to develop a micromechanical model giving the internal variables of the constitutive models in terms of the microstructural information, such as polydispersity of the grains, fabric coefficients, and force distributions.

To start the micromechanical investigation of those internal variables, it would be necessary to introduce an explicit relation between the incremental stress-strain relation and some statistics measuring the anisotropy of the granular assembly and the fluctuations of the contact forces. One way to do that is to introduce the statistic distribution $\Omega(\ell, \varphi, \mathbf{f})$ of the micromechanical variables. Here ℓ and φ are the magnitude and the orientation of the vector

connecting the center of mass of the grain with the point of application of the contact force $\mathbf{f}$. In the most general case, the incremental stress-strain relation can be given by

$$\mathrm{d}\sigma_{ij} = \int_{\boldsymbol{\lambda}} \mathrm{d}\boldsymbol{\lambda}\Omega(\boldsymbol{\lambda}) R_{ijkl}(\boldsymbol{\lambda})\mathrm{d}\epsilon_{kl}. \tag{53}$$

Here $\boldsymbol{\lambda} = (\ell, \varphi, f_n, f_t)$ and R_{ijkl} is a tensorial quantity, taking into account the local fluctuations of the deformation at the contacts with respect to the principal value of the averaged incremental strain tensor $\mathrm{d}\epsilon$ [48]. Note that the marginal distribution of Ω contains the basic statistics which have been intensively investigated in the microstructure of granular material: the size distribution $\Omega(\ell)$ [38,49,50], anisotropy of the contact network $\Omega(\varphi)$ [10,16,51–53] and the contact-force distribution $\Omega(\mathbf{f})$ [54–57]. A great challenge is to find explicit expressions for the incremental stress-strain response in terms of internal variables, given as a function of this distribution Ω. This investigation would be an extension of the ideas which have been proposed to relate the fabric tensor to the constitutive relation [16,51–53].

The traditional fabric tensor, measuring the distribution of the orientation of the contacts, cannot fulfill a complete micromechanical description, because it does not make a distinction between elastic and sliding contacts [16]. New structure tensors, taking into account the statistics of the subnetwork of the sliding contacts, must be introduced to give a micromechanical basis to the plastic deformation. The identification of these internal variables, the determination of their evolution equations, and their connection with the macroscopic variables would be a key step in the development of an appropriate continuous description of granular materials.

The evolution equation for these internal variables could be determined from the evolution equation of Ω during loading. This can be obtained from the conservation equations of the contacts [58,59]:

$$\frac{\partial \Omega}{\partial t} + v_i \frac{\partial \Omega}{\partial \lambda_i} = Q(\lambda). \tag{54}$$

The velocity field $\mathbf{v}(\boldsymbol{\lambda}) = \mathrm{d}\boldsymbol{\lambda}/\mathrm{d}t$ can be investigated from DEM by following the evolution of the contacts during the simulation. The source term Q takes into account the contacts arising or disappearing during the deformation of the granular assembly, as a consequence of the rearrangement of the granular skeleton and the eventual fragmentation of the grains. In future work, an important goal would be to determine the role of such micromechanical rearrangements in the overall mechanical response of granular materials.

Acknowledgments

We thank F. Darve, Y. Kishino, D. Kolymbas, F. Calvetti, Y.F. Dafalias, S. McNamara and R. Chambon for helpful discussions and acknowledge the support of the *Deutsche Forschungsgemeinschaft* within the research group *Modellierung kohäsiver Reibungsmaterialen* and the European DIGA project HPRN-CT-2002-00220.

References

1. L.D. Landau and E.M. Lifshitz, *Theory of Elasticity, Volume 7 of Course of Theoretical Physics*. Moscow: Pergamon Press (1986) 362pp.
2. P.A. Vermeer, Non-associated plasticity for soils, concrete and rock. In: H.J. Herrmann, J.-P. Hovi and S. Luding (eds), *Physics of Dry Granular Media - NATO ASI Series E350*. Dordrecht: Kluwer Academic Publishers (1998) pp. 163–193.

3. K.H. Roscoe and J.B. Burland, On the generalized stress-strain behavior of 'wet' clay. In: J. Heyman and F.A. Leckie (eds), *Engineering Plasticity*. Cambridge: Cambridge University Press (1968) pp. 535–609.
4. G. Gudehus, F. Darve and I. Vardoulakis, *Constitutive Relations of Soils*. Rotterdam: Balkema (1984) pp. 5–12.
5. P.A. Cundall and O.D.L. Strack, A discrete numerical model for granular assemblages. *Géotechnique* 29 (1979) 47–65.
6. K. Bagi, Stress and strain in granular assemblies. *Mech. Mater.* 22 (1996) 165–177.
7. P.A. Cundall, A. Drescher and O.D.L. Strack, Numerical experiments on granular assemblies; measurements and observations. In: P. Vermeer and H. Luger (eds), *IUTAM Conference on Deformation and Failure of Granular Materials*. Delf: Balkema-Rotterdam (1982) pp. 355–370.
8. C. Goldenberg and I. Goldhirsch, Force chains, microelasticity, and macroelasticity. *Phys. Rev. Lett.* 89 (2002) 084302.
9. K. Bagi, Microstructural stress tensor of granular assemblies with volume forces. *J. Appl. Mech.* 66 (1999) 934–936.
10. M. Lätzel, *From Discontinuous Models Towards A Continuum Description Of Granular Media*. PhD thesis, Universität Stuttgart (2002) 172 pp.
11. J.P. Bardet, Numerical simulations of the incremental responses of idealized granular materials. *Int. J. Plasticity* 10 (1994) 879–908.
12. Y. Kishino, On the incremental nonlinearity observed in a numerical model for granular media. *Ital. Geotech. J.* 3 (2003) 3–12.
13. F. Calvetti, G. Viggiani and C. Tamagnini, Micromechanical inspection of constitutive modeling. In: Pande and Pietruszczak (eds), *Constitutive Modeling and Analysis of Boundary Value Problems in Geotechnical Engineering*, Benevento: Hevelius Edizioni (2003) pp. 187–216.
14. M. Oda and K. Iwashita, Study on couple stress and shear band development in granular media based on numerical simulation analyses. *Int. J. Engng. Sci.* 38 (2000) 1713–1740.
15. I. Vardoulakis and J. Sulem, *Bifurcation Analysis in Geomechanics*. London: Blakie Academic & Professional (1995) 462 pp.
16. R.J. Bathurst and L. Rothenburg, Micromechanical aspects of isotropic granular assemblies with linear contact interactions. *J. Appl. Mech.* 55 (1988) 17–23.
17. F. Darve and F. Laouafa, Instabilities in granular materials and application to landslides. *Mech. of Cohes. Frict. Mater.* 5 (2000) 627–652.
18. G. Gudehus, A comparison of some constitutive laws for soils under radially symmetric loading and unloading. *Can. Geotech. J.* 20 (1979) 502–516.
19. D.C. Drucker and W. Prager, Soil mechanics and plastic analysis of limit design. *Q. Appl. Math.* 10 (1952) 157–165.
20. R. Nova and D. Wood, A constitutive model for sand in triaxial compression. *Int. J. Num. Anal. Meth. Geomech.* 3 (1979) 277–299.
21. K.H. Roscoe, The influence of the strains in soil mechanics. *Geotechnique* 20 (1970) 129–170.
22. H.B. Poorooshasb, I. Holubec and A.N. Sherbourne, Yielding and flow of sand in triaxial compression. *Can. Geotech. J.* 4 (1967) 277–398.
23. D. M. Wood, *Soil Mechanics-transient and cyclic loads*. Chichester: John Wiley and Sons Ltd. (1982) 420 pp.
24. F. Tatsouka and K. Ishihara, Yielding of sand in triaxial compression. *Soils Found.* 14 (1974) 63–76.
25. Y.F. Dafalias and E.P. Popov, A model of non-linearly hardening material for complex loading. *Acta Mech.* 21 (1975) 173–192.
26. D. Kolymbas, An outline of hypoplasticity. *Arch. Appl. Mech.* 61 (1991) 143–151.
27. F. Darve, E. Flavigny and M. Meghachou, Yield surfaces and principle of superposition: revisit through incrementally non-linear constitutive relations. *Int. J. Plast.* 11 (1995) 927–942.
28. R. Chambon, J. Desrues, W. Hammad and R. Charlier, CLoE, a new rate type constitutive model for geomaterials. Theoretical basis and implementation. *Int. J. Anal. Meth. Geomech.* 18 (1994) 253–278.
29. W. Wu, E. Bauer and D. Kolymbas, Hypoplastic constitutive model with critical state for granular materials. *Mech. Mater.* 23 (1996) 45–69.
30. I. Herle and G. Gudehus, Determination of parameters of a hypoplastic constitutive model from properties of grain assemblies. *Mech. Cohes.-Frictl. Matls*. 4 (1999) 461–486.
31. D. Kolymbas, *Modern Approaches to Plasticity*. Horton: Elsevier (1993) 489 pp.
32. F. Kun and H.J. Herrmann, Transition from damage to fragmentation in collision of solids. *Phys. Rev. E* 59 (1999) 2623–2632.
33. C. Moukarzel and H.J. Herrmann, A vectorizable random lattice. *J. Statist. Phys.* 68 (1992) 911–923.

34. A. Okabe, B. Boots and K. Sugihara, *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Wiley Series in probability and Mathematical Statistics. Chichester: John Wiley & Sons (1992) 532 pp.
35. H.J. Tillemans and H.J. Herrmann, Simulating deformations of granular solids under shear. *Physica A* 217 (1995) 261–288.
36. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*. Oxford: Oxford University Press (1987). 385 pp.
37. E. Buckingham, On physically similar systems: Illustrations of the use of dimensional equations. *Phys. Rev.* 4 (1914) 345–376.
38. T. Marcher and P.A. Vermeer, Macromodeling of softening in non-cohesive soils. In: P.A. Vermeer, S. Diebels, W. Ehlers, H.J. Herrmann, S. Luding and E. Ramm (eds), *Continuous and Discontinuous Modeling of Cohesive Frictional Materials*. Berlin: Springer (2001) pp. 89–110.
39. J. Desrues, *Localisation de la Deformation Plastique dans les Materieux Granulaires*. PhD thesis, University of Grenoble (1984).
40. F. Alonso-Marroquin and H.J. Herrmann, Calculation of the incremental stress-strain relation of a polygonal packing. *Phys. Rev. E* 66 (2002) 021301.
41. F. Alonso-Marroquin and H.J. Herrmann, Ratcheting of granular materials. *Phys. Rev. Lett.* 92 (2004) 054301.
42. Y.F. Dafalias, Bounding surface plasticity. I: Mathematical foundation and hypoplasticity. *J. Engng. Mech.* 112(9) (1986) 966–987.
43. P.A. Vermeer, A five-constant model unifying well-established concepts. In: G. Gudehus, F. Darve and I. Vardoulakis (eds), *Constitutive Relations of Soils*. Rotterdam: Balkema (1984) pp. 175–197.
44. R. Hill, A general theory of uniqueness and stability in elastic-plastic solids. *J. Geotech. Eng.* 6 (1958) 239–249.
45. J.A. Astrom, H.J. Herrmann and J. Timonen, Granular packings and fault zones. *Phys. Rev. Lett.* 84 (2000) 4638–4641.
46. F. Alonso-Marroquin, H.J. Herrmann and I. Vardoulakis, Micromechanical investigation of soil plasticity: An investigation using a discrete model of polygonal particles. In : P.A. Vermeer, W. Ehlers, H.J. Herrmann and E. Ramm (eds), *Modeling of Cohesive-Frictional Materials*. Rotterdan: Balkema (2002) pp. 45–67.
47. H.-B. Mühlhaus and I. Vardoulakis, The thickness of shear bands in granular materials. *Géotechnique* 37 (1987) 271–283.
48. F. Alonso-Marroquin, S. McNamara and H.J. Herrmann, Micromechanische Untersuchung des granulares Ratchetings. Antrag an die Deutsche Forschungsgemeinschaft, Universität Stuttgart (2003).
49. G.R. McDowell, M.D. Bolton, and D. Robertson, The fractal crushing of granular materials. *J. Mech. Phys. Solids* 44 (1996) 2079–2102.
50. M.D. Bolton, The role of micro-mechanics in soil mechanics. In: M. Hyodo and Y. Nakata (eds), *International Workshop on Soil Crushability*. Japan: Yamaguchi University (2002) pp. 166–178.
51. C. Thornton and D.J. Barnes, Computer simulated deformation of compact granular assemblies. *Acta Mech.* 64 (1986) 45–61.
52. S. Luding, Micro-macro transition for anisotropic, frictional granular packings. *Int. J. Sol. Struct.* 41 (2004) 5821–5836.
53. M. Madadi, O. Tsoungui, M. Lätzel and S. Luding, On the fabric tensor of polydisperse granular media in 2D. *Int. J. Sol. Struct.* 41 (2004) 2563–2580.
54. F. Radjai, M. Jean, J.J. Moreau and S. Roux, Force distribution in dense two-dimensional granular systems. *Phys. Rev. Lett.* 77 (1996) 274–277.
55. K. Bagi, Statistical analysis of contact force components in random granular assemblies. *Granular Matter* 5 (2003) 45–54.
56. H.M. Jaeger, S.R. Nagel and R.P. Behringer, Granular solids, liquids and gases. *Rev. Mod. Phys.* 68 (1996) 1259–1273.
57. D. Coppersmith, Model for force fluctuations in bead packs. *Phys. Rev. E* 53 (1996) 4673–4685.
58. S. Roux and F. Radjai, On the state variables of the granular materials. In: H. Aref and J. W. Philips (eds), *Mechanics of a New Millenium*. Dordrecht: Kluwer (2001) pp. 181–196.
59. S. Luding. Micro-macro models for anisotropic granular media. In: P.A. Vermeer, W. Ehlers, H.J. Herrmann and E. Ramm, *Modeling of Cohesive-Frictional Materials*. Rotterdam: Balkema (2004) pp. 195–205.

Journal of Engineering Mathematics (2005) 52: 35–51
DOI 10.1007/s10665-004-3949-5

Initial response of a micro-polar hypoplastic material under plane shearing

ERICH BAUER
Institute of General Mechanics, Graz University of Technology, A-8010 Graz, Austria (erich.bauer@tugraz.at)

Received 18 March 2004; accepted in revised form 24 September 2004

Abstract. The behavior of an infinite strip of a micro-polar hypoplastic material located between two parallel plates under plane shearing is investigated. The evolution equation of the stress tensor and the couple-stress tensor is described using tensor-valued functions, which are nonlinear and positively homogeneous of first order in the rate of deformation and the rate of curvature. For the initial response of the sheared layer an analytical solution is derived and discussed for different micro-polar boundary conditions at the bottom and top surfaces of the layer. It is shown that polar quantities appear within the shear layer from the beginning of shearing with the exception of zero couple stresses prescribed at the boundaries.

Key words: granular materials, hypoplasticity, interface behavior, micro-polar continuum, plane shearing

1. Introduction

The objective of the present paper is to investigate the initial response of an infinite granular layer located between two parallel plates under plane shearing using a continuum approach. The grains of the dry and frictional material are assumed to be incompressible as usual in the constitutive modeling of granular materials like sand or powder. Consequently, the volume change of the void space equals the volume change of the granular body and the granular body with empty voids can be modeled based on a single-component continuum. For incompressible grains inelastic deformations of the granular body are caused by sliding and rotating of particles against each other. The slide and rotation resistance is mainly determined by the shape, size, and surface roughness of the particles, the packing density, the orientation of the contact planes and the pressure level. Constitutive models based on a local or so-called classical continuum fail to describe micro-polar properties of granular materials, which are evident when shear deformation takes place. For instance, in a ring shear apparatus [1] the displacement field across the height of a granular layer under shearing is nonlinear and related to the micro-rotations of particles, which are different from the macro-rotations. Under large monotonic shearing the deformation is located within a zone of a thickness that is approximately 10–20 grain diameters [2–5]. In order to model such properties, an expanded continuum theory which reflects certain changes in the micro-structure of the material is needed. Non-local continuum models are characterized by internal length scales, additional kinematic degrees of freedom and/or higher order deformation gradients as outlined for instance in [6, 7], [8, pp. 4–24], [9–12]. In the present paper a micro-polar or so-called Cosserat continuum within the framework of hypoplasticity is used, which allows relating the characteristic length to the mean grain size in a physically natural manner.

Originally, the concept of hypoplasticity was developed based on a local or so-called classical continuum [13, 14]. The term hypoplasticity was introduced by Dafalias [15] for a certain

type of hardening plasticity and was subsequently adopted for a class of nonlinear constitutive models of the rate type, which can be understood as a generalization of the theory of hypoelasticity [16,17]. In the present paper the hypoplastic concept as proposed by Kolymbas [13] is considered, which differs fundamentally from the concept of elastoplasticity, as no decomposition of the rate of deformation into reversible and irreversible parts is needed. Inelastic material properties are modeled in hypoplasticity with inherently nonlinear isotropic tensor-valued functions depending on the rate of deformation and the rate of curvature. For instance, in the case of a non-polar hypoplastic material, the evolution of the stress tensor is represented by the sum of a tensor function which is linear in the rate of deformation, and a tensor function which is nonlinear in the rate of deformation. A flow rule and a stress limit condition are not described by separate functions in hypoplasticity but instead are included in the evolution equation for the state quantities. The advantage of the hypoplastic concept lies not only in the formulation of the constitutive equation, but also in an easy adaptation of the constitutive constants to experiments. Within the last two decades various hypoplastic material models for frictional granular materials have been proposed, predominantly within the framework of a classical continuum; see *e.g.* [13,14,18–21]. A comprehensive review of the historical development, performance and limitation of hypoplastic models is given in [19,22–24]. Extensions of the local hypoplastic model with quantities which are relevant for a micro-polar continuum, *i.e.,* rotational degrees of freedom in addition to translational degrees of freedom, non-symmetric shear stresses, couple stresses and the mean grain diameter as the internal length, are outlined in [25–31]. Both the non-polar hypoplastic model and the micro-polar model were implemented into a finite-element program and applied to different practical problems; see *e.g.* [32–35].

Analytical solutions are rare and can only be derived for certain boundary-value problems using simplified versions of hypoplastic models as shown for a non-polar hypoplastic model by Hill [36] and for a micro-polar hypoplastic model by Bauer and Huang [37], Hunag [38, pp. 71–78]. The latter was the first analytical solution given for an infinite layer of a micro-polar hypoplastic material at the onset of plane shearing. In the present paper this solution is extended to the full set of state quantities involved, and is discussed for different micro-polar boundary conditions. In contrast to earlier micro-polar hypoplastic versions by Tejchman and Gudehus [30] and Huang *et al*. [31], a simplified micro-polar hypoplastic model is employed for the present study, where the influence of the rate of deformation and rate of curvature is decoupled. In particular, the evolution equation for the non-symmetric stress tensor and the couple-stress tensor is described by tensor-valued functions which are nonlinear and homogeneous of first order only in the rate of deformation and the rate of curvature, respectively. Compared to earlier versions, an additional constitutive constant is included in the present model to refine the calibration of the volume-strain behavior.

The paper is organized as follows: In Section 2 the concept of hypoplasticity is briefly outlined for a non-polar continuum. Section 3 describes the extension to a micro-polar hypoplastic continuum which is used in Section 4 to model the plane shearing of an infinite granular layer under a constant vertical pressure. In Section 5 an analytical solution is derived for the initial response of the sheared layer and discussed for different micro-polar boundary conditions. Concluding remarks are presented in Section 6.

2. The concept of hypoplasticity

In the following the mechanical behavior of dry and cohesionless granular materials with simple grain skeletons [27] is considered under quasi-static and isothermal conditions. The

void-space between the grain skeleton is assumed to be empty and continuously connected. The ratio of the volume of the voids to the volume of solid particles is called the void ratio e, which is related to the bulk density ρ of the granular material as:

$$e = \frac{\rho_s}{\rho} - 1, \tag{2.1}$$

where ρ_s denotes the mass density of the solid grains. If the change of ρ_s under load can be neglected, the volume change of the granular material is determined by the volume change of the void space. With Equation (2.1) and $\rho_s = \text{constant}$ the balance relation of mass, *i.e.,* $\dot{\rho} + \rho \operatorname{div} \dot{\mathbf{x}} = 0$, can be rewritten as:

$$\dot{e} = (1+e) \operatorname{div} \dot{\mathbf{x}}, \tag{2.2}$$

where $\dot{\mathbf{x}}$ denotes the mean field velocity of the grain assembly. It follows from relation (2.2) that the void ratio e is not an independent process variable, and that for an empty void space with $\rho_s = \text{constant}$ a granular body can be treated as a single-component continuum.

Within the framework of a classical hypoplastic continuum, the evolution of a stress state is described by an isotropic tensor-valued function $\mathbf{H}$ depending in the simplest case on the current Cauchy stress $\mathbf{T}$ and the symmetric part of the velocity gradient or so-called rate of deformation $\mathbf{D}$, *i.e.*, $\overset{\circ}{\mathbf{T}} = \mathbf{H}(\mathbf{T}, \mathbf{D})$. In order to specify the function $\mathbf{H}$, several requirements must be fulfilled which are based on basic continuum mechanics and on the general mechanical behavior of granular materials detected in experiments. For a rate-independent material behavior the function $\mathbf{H}$ must be positively homogeneous of first order in $\mathbf{D}$, *i.e.*, $\mathbf{H}(\mathbf{T}, \lambda \mathbf{D}) = \lambda \mathbf{H}(\mathbf{T}, \mathbf{D})$ holds for any scalar $\lambda > 0$. In order to describe an inelastic behavior, the function $\mathbf{H}$ must be a nonlinear function of $\mathbf{D}$, *i.e.*, $\mathbf{H}(\mathbf{T}, \mathbf{D}) \neq -\mathbf{H}(\mathbf{T}, -\mathbf{D})$. In hypoplasticity both the homogeneity of the first order in $\mathbf{D}$ and the nonlinearity in $\mathbf{D}$ are satisfied by a decomposition of the tensor function $\mathbf{H}(\mathbf{T}, \mathbf{D})$ into the sum of the following two parts [13]:

$$\mathbf{H}(\mathbf{T}, \mathbf{D}) = \mathcal{L}(\mathbf{T}) : \mathbf{D} + \mathbf{\Lambda}(\mathbf{T}) \, \|\mathbf{D}\| \, . \tag{2.3}$$

Herein the tensor function $\mathcal{L}(\mathbf{T}) : \mathbf{D}$ is linear in $\mathbf{D}$, and the tensor function $\mathbf{\Lambda}(\mathbf{T}) \, \|\mathbf{D}\|$ is nonlinear in $\mathbf{D}$ with respect to the Euclidean norm of $\mathbf{D}$, i.e. $\|\mathbf{D}\| = \sqrt{\mathbf{D} : \mathbf{D}}$. It is easy to prove that, with the basic concept of hypoplasticity in the form of the constitutive equation (2.3), inelastic material properties are modeled. For two particular strain rates $\mathbf{D}_a$ and $\mathbf{D}_b$ with the same norm, *i.e.*, $\|\mathbf{D}_a\| = \|\mathbf{D}_b\|$, but opposite principal directions, *i.e.*, $\mathbf{D}_b = -\mathbf{D}_a$, the corresponding responses of Equation (2.3) are $\mathbf{H}(\mathbf{T}, \mathbf{D}_a) \neq -\mathbf{H}(\mathbf{T}, \mathbf{D}_b)$. Therefore an inherently inelastic material behavior is described with a single constitutive equation and there is no need to decompose the deformation into elastic and plastic parts. Limit states are also included in the constitutive equation (2.3) for particular $\mathbf{T}$ and $\mathbf{D}$ fulfilling the condition: $\mathcal{L}(\mathbf{T}) : \mathbf{D} + \mathbf{\Lambda}(\mathbf{T}) \, \|\mathbf{D}\| = \mathbf{0}$; see *e.g.* [13,39]. For a refined modeling of the mechanical properties of granular materials the tensor functions $\mathcal{L}$ and $\mathbf{\Lambda}$ in (2.3) may also depend on additional scalar- and tensor-valued state variables. For instance, the influence of pressure and on the incremental stiffness can be taken into account by scaling $\mathcal{L}$ and $\mathbf{\Lambda}$ with a pressure-dependent stiffness factor and densityfactor [19,20,40,41]; rate-dependent properties can be introduced in the nonlinear part of the constitutive relation [42,43], and with additional structure tensors initial anisotropy [44,45], cohesion [46] and a so-called inter-granular strain [47] can be modeled. In order to also take into account particle rotation and couple stresses the non-polar constitutive relation (2.3) was extended to a micro-polar continuum [28,30,31], which is discussed in more detail in the next section.

3. Micro-polar hypoplastic model

In a micro-polar continuum, *e.g.* [8], the kinematics are characterized by the velocity vector $\dot{\mathbf{x}}$ and the micro-spin vector or so-called Cosserat spin vector $\dot{\boldsymbol{\omega}}^c$. The micro-spin tensor $\mathbf{W}^c$, the rate of deformation $\mathbf{D}^c$ and the rate of curvature $\mathbf{K}$ are defined as:

$$\mathbf{W}^c = -\boldsymbol{\epsilon}\,\dot{\boldsymbol{\omega}}^c\,, \quad \mathbf{D}^c = \mathbf{L} - \mathbf{W}^c\,, \quad \mathbf{K} = \operatorname{grad}\dot{\boldsymbol{\omega}}^c\,. \tag{3.1}$$

Herein the third-order tensor $\boldsymbol{\epsilon}$ denotes the permutation tensor. The velocity gradient $\mathbf{L} = \operatorname{grad}\dot{\mathbf{x}}$ in (3.1) is related to the macro-motion and can be decomposed into the symmetric part $\mathbf{D} = (\mathbf{L} + \mathbf{L}^T)/2$ and the skew-symmetric part $\mathbf{W} = (\mathbf{L} - \mathbf{L}^T)/2$. The definition of tensor $\mathbf{D}$ is the same as for the rate of deformation in a non-polar continuum. The macro-spin tensor $\mathbf{W}$ can also be represented by the rate of the macro-spin vector $\dot{\boldsymbol{\omega}}$, *i.e.*, $\mathbf{W} = -\boldsymbol{\epsilon}\,\dot{\boldsymbol{\omega}}$. Hence definition (3.1) for the rate of deformation can alternatively be written as

$$\mathbf{D}^c = \mathbf{D} + \mathbf{W} - \mathbf{W}^c. \tag{3.2}$$

Representation (3.2) indicates that, in the case where the macro-spin is equal to the micro-spin, *i.e.*, $\dot{\boldsymbol{\omega}} = \dot{\boldsymbol{\omega}}^c$, the rate of deformation $\mathbf{D}^c$ reduces to the tensor $\mathbf{D}$ of the classical non-polar continuum. The kinematic quantities $\mathbf{D}^c$ and $\mathbf{K}$ are associated with the stress tensor $\mathbf{T}$ and the couple-stress tensor $\mathbf{M}$ defined for the current configuration. For quasi-static processes the local-equilibrium equations read:

$$\operatorname{div}\mathbf{T} + \rho\,\tilde{\mathbf{b}} = \mathbf{0}, \tag{3.3}$$

$$\operatorname{div}\mathbf{M} - \boldsymbol{\epsilon} : \mathbf{T} + \rho\,\tilde{\mathbf{m}} = \mathbf{0}. \tag{3.4}$$

Herein ρ denotes the bulk density of the material, $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{m}}$ represent the body force and body couple, respectively. Equation (3.4) indicates that the stress tensor in a micro-polar continuum is usually non-symmetric with the exception of states with $\operatorname{div}\mathbf{M} = 0$ and $\tilde{\mathbf{m}} = 0$. In order to have objective measures for the stress rate and couple-stress rate, the time derivative given by Green and Naghdi [48] is adopted, *i.e.*,

$$\mathring{\mathbf{T}} = \dot{\mathbf{T}} - \boldsymbol{\Omega}\,\mathbf{T} + \mathbf{T}\,\boldsymbol{\Omega}, \tag{3.5}$$

$$\mathring{\mathbf{M}} = \dot{\mathbf{M}} - \boldsymbol{\Omega}\,\mathbf{M} + \mathbf{M}\,\boldsymbol{\Omega}. \tag{3.6}$$

Herein the angular velocity tensor $\boldsymbol{\Omega}$ is related to the rotation tensor $\mathbf{R}$ and to the rate of rotation tensor, $\dot{\mathbf{R}}$, as $\boldsymbol{\Omega} = \dot{\mathbf{R}}\,\mathbf{R}^T$. In the present paper the following evolution equations for the stress tensor and couple-stress tensor are considered:

$$\mathring{\mathbf{T}} = f_c\,\operatorname{tr}\mathbf{T}\Big[\,\hat{a}^2\,(\mathbf{D}^c + \frac{c}{3}\,(\mathbf{I}:\mathbf{D}^c)\,\mathbf{I}) + (\hat{\mathbf{T}}:\mathbf{D}^c)\,\hat{\mathbf{T}} + \hat{a}\,(\hat{\mathbf{T}} + \hat{\mathbf{T}}^*)\,\|\mathbf{D}^c\|\,\Big], \tag{3.7}$$

$$\mathring{\mathbf{M}} = d_{50}\,f_c\,\operatorname{tr}\mathbf{T}\Big[\,a_m^2\bar{\mathbf{K}} + (\hat{\mathbf{M}}:\bar{\mathbf{K}})\,\hat{\mathbf{M}} + 2\,a_m\,\hat{\mathbf{M}}\,\|\bar{\mathbf{K}}\|\,\Big]. \tag{3.8}$$

Herein $\hat{\mathbf{T}} = \mathbf{T}/\operatorname{tr}\mathbf{T}$, $\hat{\mathbf{T}}^* = \hat{\mathbf{T}} - \mathbf{I}/3$, $\hat{\mathbf{M}} = \mathbf{M}/(d_{50}\operatorname{tr}\mathbf{T})$, $\bar{\mathbf{K}} = d_{50}\mathbf{K}$ are normalized quantities with $\operatorname{tr}\mathbf{T} = \mathbf{I}:\mathbf{T}$ and f_c, $\hat{a}$, c, a_m and d_{50} are scalar factors. The mean grain diameter d_{50} enters the constitutive equations as the internal length. In contrast to earlier micro-polar hypoplastic models a certain simplification is assumed in the present constitutive relations (3.7) and (3.8). In particular, the non-symmetric stress rate is homogeneous of first order in the rate of deformation only and the couple-stress rate is homogeneous of first order in the

rate of curvature only, which is not assumed in the more sophisticated micro-polar hypoplastic models discussed for instance by Tejchman and Gudehus [30], Huang *et al.* [31]. As outlined by Bauer [20] and Gudehus [19], factor f_c with the dimension of stress is usually related to the pressure and density-dependent stiffness under isotropic compression. Here a specific representation of f_c will not be discussed in more detail as it is not relevant to the investigations outlined in Section 5. While in earlier hypoplastic versions the transverse-strain sensitivity of the material was a pure prediction of the constitutive model, a new constant c allows a scaling of the influence of the volume strain rate $\mathbf{I}:\mathbf{D}^c$, which is included as an additional term in the constitutive equation (3.7). It can be concluded from Equation (3.7) that the constant c has no influence for isochoric deformations. In particular for critical states, which are defined for monotonic isochoric deformations without changes of the stress and couple stress state, *i.e.*, $\mathbf{I}:\mathbf{D}^c=0$, $\mathring{\mathbf{T}}=\mathbf{0}$ and $\mathring{\mathbf{M}}=\mathbf{0}$, the constitutive relations (3.7) and Equation (3.8) yield: $\|\hat{\mathbf{T}}^*\|=\hat{a}$ and $\|\hat{\mathbf{M}}\|=a_m$. Thus, the factor $\hat{a}$ is related to the stress limit condition, and factor a_m is related to the couple-stress limit condition in critical states. Factors $\hat{a}$ and a_m reflect the intergranular slide resistance and rotation resistance of particles, respectively [49]. However, for stress states which deviate from critical states, the quantity $\hat{a}$ is not only related to $\|\hat{\mathbf{T}}^*\|$ because the requirement $\hat{a}>0$ must be also fulfilled for isotropic stress states. A consistent adaptation of $\hat{a}$ to the stress limit condition by Matsuoka and Nakai [50] was, for instance, shown by Bauer [51] for a non-polar continuum, which yields for critical states under plane strain conditions:

$$\hat{a}=\sqrt{\frac{8\,\sin^2\varphi_c}{9\,(3+\sin^2\varphi_c)}},$$

and for isotropic stress states:

$$\hat{a}=\sqrt{\frac{8}{3}}\left(\frac{\sin\varphi_c}{3-\sin\varphi_c}\right),$$

Herein φ_c denotes the intergranular friction angle defined for the critical state under triaxial compression. In contrast to a non-polar continuum the stress tensor in a micro-polar continuum is usually non-symmetric, which may lead to a different representation of $\hat{a}$ [37]. For the sake of simplicity $\hat{a}$ and a_m are assumed to be constant in the present paper.

Based on the simplified micro-polar hypoplastic model, an analytical solution for the initial response of plane shearing can be derived for the full set of state quantities involved as outlined in Section 5. In this context it can be noted that the solution for the macro- and micro-spins, the rate of shear stresses and rate of the couple stresses is the same for the simplified micro-polar hypoplastic model and the models proposed by Tejchman and Gudehus [30], Huang *et al.* [31]. However, a different response is obtained for the rate of the normal stresses and the volume strain rate. The analytical solution is restricted to the initial response at the beginning of shearing. For larger shearing the underlying set of differential equations can only be solved numerically and the results are no longer the same for different models.

4. Modeling of plane shearing with dilatancy

For plane shearing of an infinite layer the field quantities are independent of the co-ordinate in the direction of shearing, *i.e.*, $\partial(.)/\partial x_1=0$ with respect to the co-ordinate system shown in Figure 1. For plane-strain conditions the field quantities are also independent of the co-ordinate x_3. Then the macro-motion can be described by $x_1=X_1+f_1(X_2,t)$, $x_2=X_2+f_2(X_2,t)$ and $x_3=X_3$, where the spatial co-ordinates x_i $(i=1,2,3)$ represent the current position of a

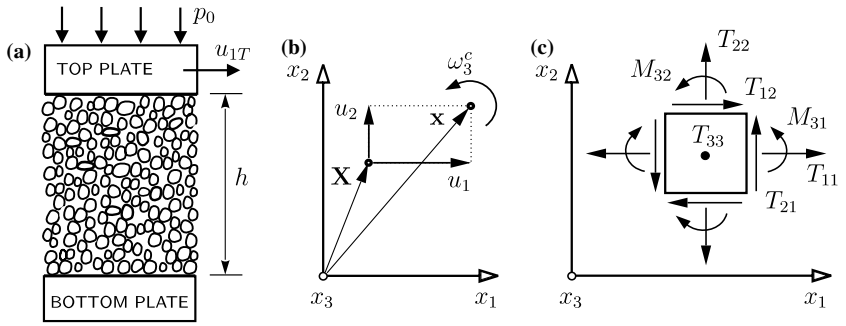


Figure 1. Modeling of plane shearing under constant vertical pressure $p_0 = -\sigma_{22}$: (a) section of the infinite granular layer between parallel plates with rough surfaces, (b) kinematics of plane shearing with dilatancy and degrees of freedom u_1, u_2 and ω_3, (c) stress components σ_{11}, σ_{22}, σ_{33}, σ_{12}, σ_{21} and couple-stress components M_{31} and M_{32} with respect to the Cartesian co-ordinate system.

particle and X_i $(i=1,2,3)$ are the corresponding co-ordinates in the initial configuration. The functions $f_i(X_2,t)$ $(i=1,2)$ are time-dependent functions with respect to $f_i(X_2,t=0)=0$ in the initial configuration at time $t=0$. The only possible micro-motion is the micro-rotation or so-called Cosserat rotation ω_3^c which is orientated perpendicular to the (x_1,x_2) plane as shown in Figure 1. The displacements $u_i = x_i - X_i$ read: $u_1 = f_1(X_2,t)$, $u_2 = f_2(X_2,t)$ and $u_3 = 0$. Since u_2 depends on X_2, the kinematics of shearing with dilatancy is taken into account. With respect to

$$g_i(X_2,t) = \frac{\partial f_i(X_2,t)}{\partial X_2} \quad \text{and} \quad \dot{g}_i(X_2,t) = \frac{d\, g_i(X_2,t)}{d\, t} \quad (i=1,2), \tag{4.1}$$

for any field variable ϕ defined in the infinite layer, the following relations are valid:

$$\frac{\partial \phi}{\partial X_1} = \frac{\partial \phi}{\partial x_1} = 0, \qquad \frac{\partial \phi}{\partial X_2} = (1+g_2)\,\frac{\partial \phi}{\partial x_2}, \qquad \frac{\partial \phi}{\partial X_3} = \frac{\partial \phi}{\partial x_3} = 0. \tag{4.2}$$

Therefore the non-vanishing components of the velocity gradient and macro-spin read:

$$L_{12} = \frac{\partial \dot{u}_1}{\partial x_2} = \frac{\dot{g}_1}{1+g_2}, \qquad L_{22} = \frac{\partial \dot{u}_2}{\partial x_2} = \frac{\dot{g}_2}{1+g_2}, \qquad W_{12} = \frac{1}{2}\,\frac{\partial \dot{u}_1}{\partial x_2} = -\dot{\omega}_3. \tag{4.3}$$

It follows from L_{22} in (4.3) that $\dot{g}_2/(1+g_2)$ is equal to the volume strain rate and therefore a measure of the dilatancy behavior of the granular layer. As body forces, body couples and inertia forces are neglected, the equilibrium equations (3.3) and (3.4) reduce to:

$$\frac{\partial T_{ij}}{\partial x_j} = 0, \qquad \frac{\partial M_{ij}}{\partial x_j} - \epsilon_{ikl}\, T_{kl} = 0. \tag{4.4}$$

With respect to the relations in (4.2) the time derivative of Equations (4.4) reads:

$$\frac{\partial \dot{T}_{ij}}{\partial x_j} - \frac{\partial T_{ij}}{\partial x_m}\,\frac{\partial \dot{u}_m}{\partial x_j} = 0, \tag{4.5}$$

and

$$\frac{\partial \dot{M}_{ij}}{\partial x_j} - \frac{\partial M_{ij}}{\partial x_m}\,\frac{\partial \dot{u}_m}{\partial x_j} - \epsilon_{ikl}\, \dot{T}_{kl} = 0. \tag{4.6}$$

For the infinite sheared layer Equations (4.5) and (4.6) yield:

$$\frac{\partial \dot{T}_{12}}{\partial x_2} = 0, \qquad \frac{\partial \dot{T}_{22}}{\partial x_2} = 0, \tag{4.7}$$

$$\frac{\partial \dot{M}_{32}}{\partial x_2} - \frac{\dot{g}_2}{1+g_2}(T_{12} - T_{21}) - (\dot{T}_{12} - \dot{T}_{21}) = 0. \tag{4.8}$$

The material derivatives $\dot{T}_{ij}$ and $\dot{M}_{ij}$ in (4.7) and (4.8) are related to the constitutive equations using the corresponding objective derivatives defined in (3.5) and (3.6). For plane shearing the components of the objective stress rate and couple stress rate read:

$$\begin{aligned}
\mathring{\hat{T}}_{11} &= f_c \left[\hat{a}^2 \, (c/3) \, D^c_{22} + \psi_1 \, \hat{T}_{11} + \hat{a} \, (2 \, \hat{T}_{11} - 1/3) \, \sqrt{\psi_2} \right], \\
\mathring{\hat{T}}_{22} &= f_c \left[\hat{a}^2 \, (D^c_{22} + (c/3) \, D^c_{22}) + \psi_1 \, \hat{T}_{22} + \hat{a} \, (2 \, \hat{T}_{22} - 1/3) \, \sqrt{\psi_2} \right], \\
\mathring{\hat{T}}_{33} &= f_c \left[\hat{a}^2 \, (c/3) \, D^c_{22} + \psi_1 \, \hat{T}_{33} + \hat{a} \, (2 \, \hat{T}_{33} - 1/3) \, \sqrt{\psi_2} \right], \\
\mathring{\hat{T}}_{12} &= f_c \left[\hat{a}^2 \, D^c_{12} + \left(\psi_1 + 2 \, \hat{a} \, \sqrt{\psi_2} \right) \hat{T}_{12} \right], \\
\mathring{\hat{T}}_{21} &= f_c \left[\hat{a}^2 \, D^c_{21} + \left(\psi_1 + 2 \, \hat{a} \, \sqrt{\psi_2} \right) \hat{T}_{21} \right], \\
\mathring{\hat{M}}_{31} &= d_{50} \, f_c \left[\left(\hat{M}_{32} \, \bar{K}_{32} + 2 \, a_m \sqrt{\bar{K}^2_{32}} \right) \hat{M}_{31} \right], \\
\mathring{\hat{M}}_{32} &= d_{50} \, f_c \left[a^2_m \, \bar{K}_{32} + \left(\hat{M}_{32} \, \bar{K}_{32} + 2 \, a_m \sqrt{\bar{K}^2_{32}} \right) \hat{M}_{32} \right],
\end{aligned} \tag{4.9}$$

with the abbreviations.

$$\psi_1 = \hat{T}_{12} \, D^c_{12} + \hat{T}_{21} \, D^c_{21} + \hat{T}_{22} \, D^c_{22}, \qquad \psi_2 = D^{c\,2}_{12} + D^{c\,2}_{21} + D^{c\,2}_{22},$$

and the kinematic quantities

$$D^c_{12} = \dot{\omega}^c_3 + \frac{\dot{g}_1}{1+g_2}, \qquad D^c_{21} = -\dot{\omega}^c_3, \qquad D^c_{22} = \frac{\dot{g}_2}{1+g_2}, \qquad \bar{K}_{32} = d_{50} \, \frac{\partial \dot{\omega}^c_3}{\partial x_2}.$$

While the symmetry condition for the infinite shear layer was already considered above, the boundary conditions at the bottom and top of the layer are still to be specified. Apart from the displacement and stress boundary conditions of the classical continuum, micro-rotation and couple-stress boundary conditions occur in a micro-polar continuum. Herein the micro-polar boundary conditions allow the modeling of the influence of the rotation resistance of particles in contact with a rough bounding structure, as has been discussed in more detail, for instance by Tejchman [28] and Bauer and Huang [34]. For the present investigation the bottom of the layer is assumed to be fixed so that neither sliding nor rotation may occur, *i.e.*,

$$x_2 = 0: \qquad \dot{u}_1 = 0, \qquad \dot{u}_2 = 0, \qquad \dot{\omega}^c_3 = 0. \tag{4.10}$$

At the top of the layer a vertical pressure is applied and kept constant, *i.e.*, $\dot{T}_{22}=0$. The horizontal velocity $\dot{u}_{1T}$ and the rate of the micro-rotation $\dot{\omega}^c_{3T}$ are prescribed, *i.e.*,

$$x_2=h: \quad \dot{u}_1=\dot{u}_{1T}, \qquad \dot{\omega}^c_3=\dot{\omega}^c_{3T}, \qquad \dot{T}_{22}=0\,. \tag{4.11}$$

5. Analytical solution for the initial response

For the investigation of the initial response of an infinite sheared layer the following homogeneous and initially isotropic state is considered:

$$T_{ij}=-p_0\,\delta_{ij}, \qquad M_{ij}=0, \tag{5.1}$$

where p_0 denotes the isotropic initial pressure. It follows from the relations in (5.1) that in the initial state the shear stresses and couple stresses are assumed to be zero; thus, the material time derivatives of the stresses and the couple stresses coincide with the objective rates at the beginning of shearing. When this property is taken into account and with respect to $g_1=g_2=0$ and the initial state assumed in (5.1), the relations for the non-zero rate of the stress and couple-stress components by (4.9) reduce to:

$$\dot{T}_{11}=\mathring{T}_{11}=f_c\left[\hat{a}^2\,\frac{c}{3}\,\dot{g}_2+\frac{\dot{g}_2}{9}+\frac{\hat{a}}{3}\sqrt{\dot{g}_2^2+(\dot{\omega}^c_3+\dot{g}_1)^2+\dot{\omega}^{c\,2}_3}\right], \tag{5.2}$$

$$\dot{T}_{22}=\mathring{T}_{22}=f_c\left[\hat{a}^2\,\left(1+\frac{c}{3}\right)\,\dot{g}_2+\frac{\dot{g}_2}{9}+\frac{\hat{a}}{3}\sqrt{\dot{g}_2^2+(\dot{\omega}^c_3+\dot{g}_1)^2+\dot{\omega}^{c\,2}_3}\right], \tag{5.3}$$

$$\dot{T}_{33}=\mathring{T}_{33}=\dot{T}_{11}, \tag{5.4}$$

$$\dot{T}_{12}=\mathring{T}_{12}=f_c\,\hat{a}^2\,(\dot{\omega}^c_3+\dot{g}_1), \tag{5.5}$$

$$\dot{T}_{21}=\mathring{T}_{21}=-f_c\,\hat{a}^2\,\dot{\omega}^c_3, \tag{5.6}$$

$$\dot{M}_{32}=\mathring{M}_{32}=f_c\,a_m^2\,d_{50}^2\,K_{32}. \tag{5.7}$$

Herein the quantities:

$$\dot{g}_1=\frac{\mathrm{d}\dot{u}_1}{\mathrm{d}x_2}=-2\,\dot{\omega}_3, \qquad \dot{g}_2=\frac{\mathrm{d}\dot{u}_2}{\mathrm{d}x_2}, \qquad \dot{\omega}^c_3 \quad \text{and} \quad K_{32}=\frac{\mathrm{d}\dot{\omega}^c_3}{\mathrm{d}x_2} \tag{5.8}$$

have to be determined for the given boundary value problem as outlined in the following section.

5.1. General solution

With respect to the requirement in (4.7) the quantity $\dot{T}_{12}$ in (5.5) is constant across the height of the layer, *i.e.*,

$$\dot{T}_{12}=f_c\,\hat{a}^2\,(\dot{\omega}^c_3+\dot{g}_1)=\chi, \tag{5.9}$$

where χ denotes a constant. By substituting the relations (5.5–5.7) in (4.8), one obtains

$$\frac{\partial}{\partial x_2}\left(f_c\,a_m^2\,d_{50}^2\,\frac{\partial\dot{\omega}^c_3}{\partial x_2}\right)-f_c\,\hat{a}^2\,(2\,\dot{\omega}^c_3+\dot{g}_1)=0. \tag{5.10}$$

For an initially homogeneous layer, f_c, $\hat{a}$ and d_{50} are independent of the co-ordinates x_i ($i = 1, 2, 3$) so that (5.10) and (5.9) lead to the following differential equation for $\dot{\omega}_3^c$

$$\eta^2 \frac{\mathrm{d}^2 \dot{\omega}_3^c}{\mathrm{d}x_2^2} - \dot{\omega}_3^c = C_1, \tag{5.11}$$

with the abbreviations

$$\eta = \frac{a_m}{\hat{a}} d_{50} \quad \text{and} \quad C_1 = \frac{\chi}{f_c \hat{a}^2} = \dot{\omega}_3^c + \dot{g}_1.$$

The general solution to the second-order differential equation (5.11) reads

$$\dot{\omega}_3^c = C_2 \cosh(x_2/\eta) + C_3 \sinh(x_2/\eta) - C_1, \tag{5.12}$$

where C_2 and C_3 are integration constants. The derivative of $\dot{\omega}_3^c$ yields the rate of curvature, *i.e.*,

$$K_{32} = \frac{\mathrm{d}\dot{\omega}_3^c}{\mathrm{d}x_2} = \frac{1}{\eta} \left[C_2 \sinh(x_2/\eta) + C_3 \cosh(x_2/\eta) \right]. \tag{5.13}$$

Inserting $\dot{g}_1 = -2\,\dot{\omega}_3$ in (5.9), one obtains for the rate of the macro-spin

$$\dot{\omega}_3 = \frac{1}{2} \left[\dot{\omega}_3^c - C_1 \right] = \frac{1}{2} \left[C_2 \cosh(x_2/\eta) + C_3 \sinh(x_2/\eta) \right] - C_1, \tag{5.14}$$

and for the quantity $\dot{g}_1$

$$\dot{g}_1 = -2\,\dot{\omega}_3 = 2\,C_1 - \left[C_2 \cosh(x_2/\eta) + C_3 \sinh(x_2/\eta) \right]. \tag{5.15}$$

With respect to (5.8) the integration of $\mathrm{d}\dot{u}_1 = -2\,\dot{\omega}_3\,\mathrm{d}x_2$ yields the initial rate of the horizontal displacement u_1 as a function of the co-ordinate x_2, *i.e.*,

$$\dot{u}_1 = 2\,C_1\,x_2 - C_2\,\eta \sinh(x_2/\eta) - C_3\,\eta \cosh(x_2/\eta) + C_4. \tag{5.16}$$

Taking into account a constant vertical pressure applied at the top of the layer, and the requirement for equilibrium across the height of the layer, one observes that $\dot{T}_{22}$ is zero within the shear layer. Inserting $\dot{T}_{22} = 0$ and $\dot{\omega}_3^c$ from (5.12) in (5.3) yields an equation for $\dot{g}_2$ with the solution

$$\dot{g}_2 = -\sqrt{\frac{C_1^2 + \left[C_2 \cosh(x_2/\eta) + C_3 \sinh(x_2/\eta) - C_1 \right]^2}{\left[\hat{a}(3+c) + 1/(3\,\hat{a}) \right]^2 - 1}}. \tag{5.17}$$

The initial rate of the vertical displacement u_2 can be obtained by integration of $\mathrm{d}\dot{u}_2 = \dot{g}_2\,\mathrm{d}x_2$, *i.e.*,

$$\dot{u}_2 = -\int \sqrt{\frac{C_1^2 + \left[C_2 \cosh(x_2/\eta) + C_3 \sinh(x_2/\eta) - C_1 \right]^2}{\left[\hat{a}(3+c) + 1/(3\,\hat{a}) \right]^2 - 1}}\,\mathrm{d}x_2 + C_5. \tag{5.18}$$

By inserting $\dot{\omega}_3^c$, K_{32}, $\dot{g}_1$ and $\dot{g}_2$ into (5.2) to (5.7), one obtains the following relations for the initial rate of the normalized quantities of the stress and couple stress components

$$\frac{\dot{T}_{11}}{f_c\,\hat{a}^2} = \left[\frac{c}{3} - \frac{1}{9\,\hat{a}^2}\right]\left[\frac{C_1^2 + [C_2\,\cosh(x_2/\eta) + C_3\,\sinh(x_2/\eta) - C_1]^2}{[\hat{a}(3+c) + 1/(3\,\hat{a})]^2 - 1}\right] + \\ + \frac{1}{3\,\hat{a}}\left[\frac{C_1^2 + [C_2\,\cosh(x_2/\eta) + C_3\,\sinh(x_2/\eta) - C_1]^2}{1 - 1/[\hat{a}(3+c) + 1/(3\,\hat{a})]^2}\right]^{1/2}, \tag{5.19}$$

$$\frac{\dot{T}_{33}}{f_c\,\hat{a}^2} = \frac{\dot{T}_{11}}{f_c\,\hat{a}^2}, \qquad \frac{\dot{T}_{12}}{f_c\,\hat{a}^2} = C_1, \tag{5.20}$$

$$\frac{\dot{T}_{21}}{f_c\,\hat{a}^2} = -\,C_2\,\cosh(x_2/\eta) - C_3\,\sinh(x_2/\eta) + C_1, \tag{5.21}$$

$$\frac{\dot{M}_{32}}{f_c\,\hat{a}\,a_m\,d_{50}} = C_2\,\sinh(x_2/\eta) + C_3\,\cosh(x_2/\eta) \tag{5.22}$$

with $\eta = (a_m/\hat{a})d_{50}$. The general solution shows that, even for the case of an initially isotropic stress state and zero couple stresses, the initial rate of the state quantities under shearing is different from the classical non-polar continuum in the case of non-vanishing constants C_2 and C_3. With the exception of $\dot{T}_{12}$ the rates of the state quantities are nonlinear functions of the co-ordinate x_2 as a result of the micro-polar quantities contained in the present micro-polar hypoplastic model. Furthermore, it follows from relation (5.17) that a transverse-strain sensitivity, which is controlled by the constitutive constant c, influences the volume strain rate $\mathbf{I}:\mathbf{D}^c = \dot{g}_2$, and consequently also the velocity $\dot{u}_2$ and the stress rates $\dot{T}_{11}$ and $\dot{T}_{33}$. It can also be noted that, for real solutions for $\dot{g}_2$, the denominator of the expression (5.17) must be greater than zero, which yields a lower bound for the calibration of the value of c.

Across the height of the sheared layer the distribution of the quantities (5.12) to (5.22) is strongly influenced by the interface behavior between the granular layer and the rough bounding plates. The interface behavior can be taken into account by specifying the constants C_i $(i = 1{\cdot}5)$ as outlined in the next section.

5.2. Influence of the micro-polar boundary conditions

The boundary conditions at the bottom and top surfaces specified in (4.10) and (4.11) yield for the constants C_i $(i = 1{\cdot}5)$:

$$C_1 = \frac{\dot{u}_{1T}/\eta + \dot{\omega}_{3T}^c\,\tanh(h/(2\,\eta))}{2\,[h/\eta - \tanh(h/(2\,\eta))]}, \qquad C_2 = C_1, \tag{5.23}$$

$$C_3 = \frac{\dot{\omega}_{3T}^c\,[(2\,h/\eta)/\sinh(h/\eta) - 1] - (\dot{u}_{1T}/\eta)\,\tanh(h/(2\,\eta))}{2\,[h/\eta - \tanh(h/(2\,\eta))]}, \tag{5.24}$$

$$C_4 = \eta\,C_3, \quad C_5 = \left[\int \dot{g}_2\,\mathrm{d}x_2\right]_{x_2=0}, \quad \text{with: } \eta = \frac{a_m}{\hat{a}}d_{50}. \tag{5.25}$$

In the following the influence of the micro-polar boundary conditions on the initial response will be discussed for a shear layer with a height of $h = 1\,\mathrm{cm}$, a prescribed horizontal shear velocity of $\dot{u}_{1T} = 1\,\mathrm{cm/s}$, a constant vertical pressure and the constitutive constants: $d_{50} = 0{\cdot}05$ cm, $\hat{a} = 0{\cdot}33$, $a_m = 1$ and $c = 0$ which are relevant for a medium quartz sand with a critical friction angle of $\varphi_c = 30^\circ$. For the present discussion the value of the constant f_c is not relevant as the rates of the stress- and couple stress components can be normalized by f_c.

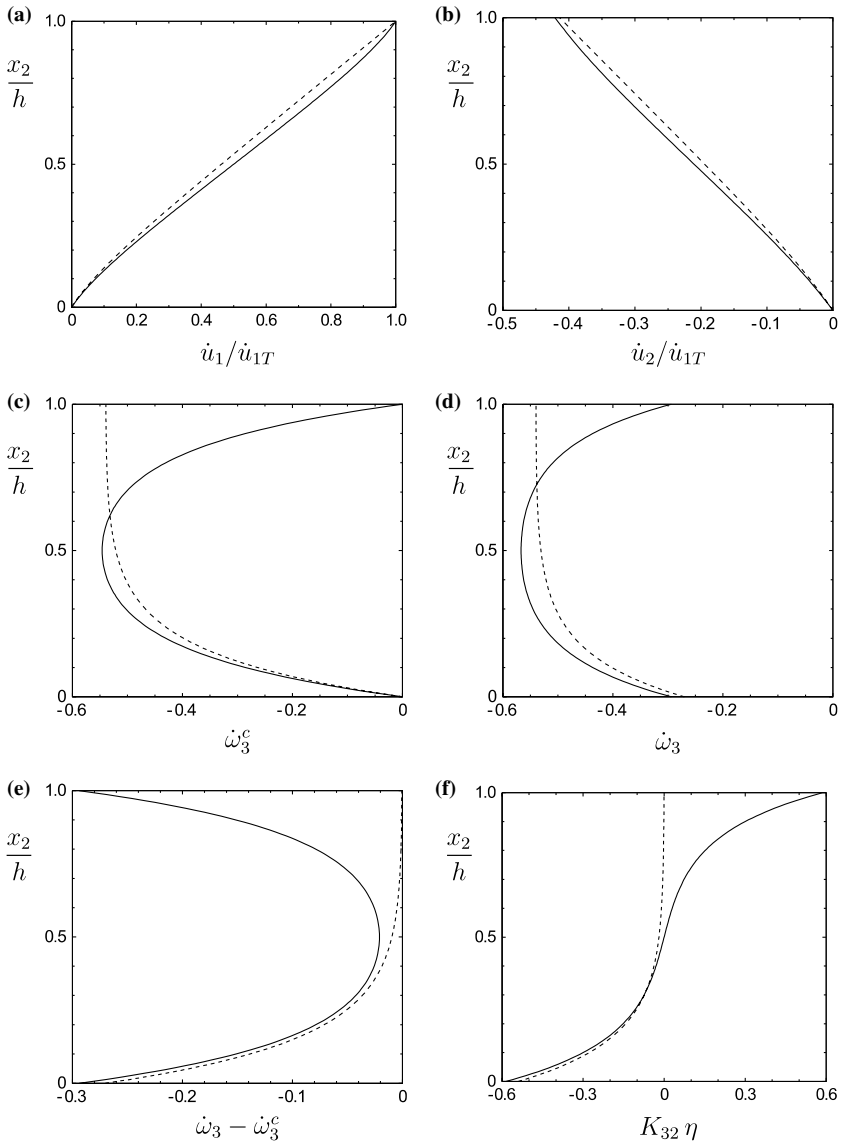


Figure 2. Initial response of a sheared layer for the boundary conditions at the bottom $x_2 = 0 : \dot{u}_1 = \dot{u}_2 = 0$, $\dot{\omega}_{3B}^c = 0$, and at the top $x_2 = h : \dot{u}_1 = \dot{u}_{1T}$, $\dot{T}_{22} = 0$, $\dot{\omega}_{3T}^c = 0$ (solid curve) and $\dot{M}_{32T} = 0$ (dotted curve). Distribution of: (a) horizontal velocity $\dot{u}_1$, (b) vertical velocity $\dot{u}_2$, (c) micro-rotation rate $\dot{\omega}_3^c$, (d) macro-rotation rate $\dot{\omega}_3$, (e) difference of macro- and micro-rotation rates $\dot{\omega}_3 - \dot{\omega}_3^c$, and (f) curvature rate K_{32}.

Figures (2) and (3) show the initial response of the shear layer for two different micro-polar boundary conditions prescribed at the top surface. The solid curves show the results obtained for $\dot{\omega}_3^c(x_2 = h) = \dot{\omega}_{3T}^c = 0$ and the dotted curves are obtained for the assumption of a zero couple stress rate $\dot{M}_{32}(x_2 = h) = \dot{M}_{32T} = 0$.

The special case $\dot{\omega}_{3T}^c = 0$ reflects the behavior of a very rough top surface without particle rotation along the interfaces. The same assumption was introduced in (4.10) for the bottom surface, so that the micro-polar boundary conditions are symmetric for this case. Figure 2a and 2b shows that the velocities $\dot{u}_1(x_2)$ and $\dot{u}_2(x_2)$ are nonlinearly distributed across the height of the shear layer. Thus, in a micro-polar continuum the deformation is inhomogeneous from the beginning of shearing. The distribution of the rate of micro-rotation

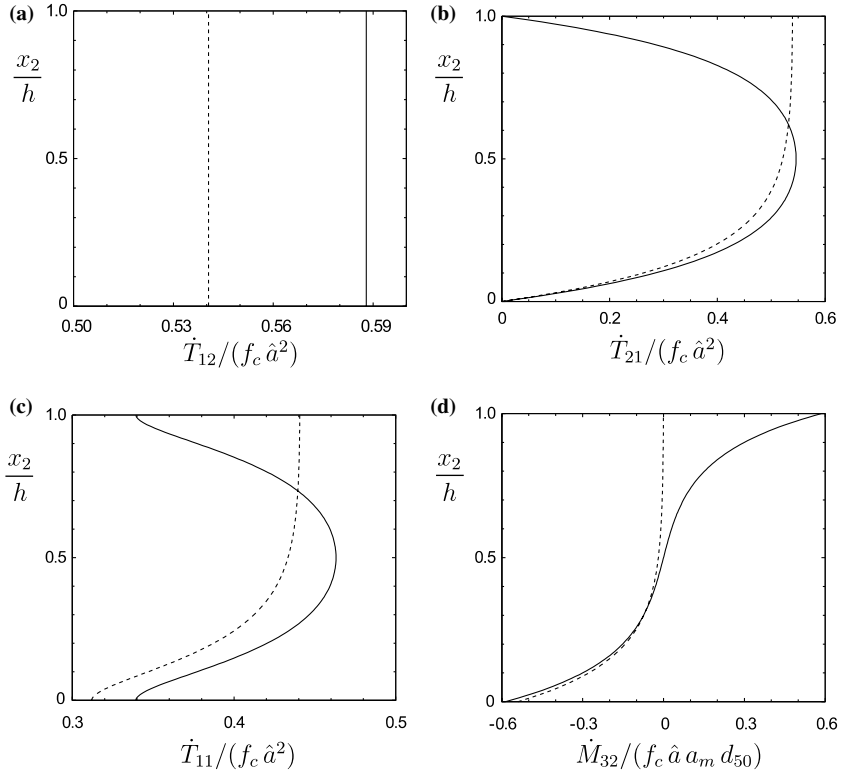


Figure 3. Initial response of a sheared layer for the boundary conditions at the bottom $x_2=0:\dot{u}_1=\dot{u}_2=0$, $\dot{\omega}^c_{3B}=0$, and at the top $x_2=h:\dot{u}_1=\dot{u}_{1T}$, $\dot{T}_{22}=0$, $\dot{\omega}^c_{3T}=0$ (solid curve) and $\dot{M}_{32T}=0$ (dotted curve). Distribution of the normalized: (a) shear stress rate $\dot{T}_{12}$, (b) shear-stress rate $\dot{T}_{21}$, (c) normal-stress rate $\dot{T}_{11}=\dot{T}_{33}$, and (d) couple-stress rate $\dot{M}_{32}$.

(Figure 2c) and the rate of macro-rotation (Figure 2d) are also nonlinear and they are different from each other (Figure 2e). With respect to the rule of signs in Figure 1a negative micro-rotation rate means clockwise rotation for a movement of the top surface to the right. The extreme values of $\dot{\omega}^c_3$ and of $\dot{\omega}_3$ occur in the middle of the layer. The gradient of the micro-rotation is termed rate of curvature, K_{32}, and shown in Figure 2f. From Figure 2e it can be concluded that, at the beginning of shearing, the influence of micro-polar properties is more pronounced close to the bottom and top boundaries of the layer. The shear-stress rate $\dot{T}_{12}$ is constant as is necessary for equilibrium (Figure 3a) and it is different from the shear stress rate $\dot{T}_{21}$ (Figure 3b). Therefore, the stress tensor also becomes non-symmetric in the case of an initially isotropic stress state. The distribution of the normal stress rate $\dot{T}_{11}$ is nonlinear (Figure 3c) and it is equal to the normal stress rate $\dot{T}_{33}$. The normalized quantity of the couple-stress rate $\dot{M}_{32}$ (Figure 3d) coincides with the normalized curvature rate K_{32} shown in Figure 2f which is due to relation (5.7). Although the couple-stress rate is zero in the middle of the layer for symmetric micro-polar boundary conditions, *i.e.*, $\dot{M}_{32}(x_2=h/2)=0$, the shear stress rate $\dot{T}_{12}(x_2=h/2)$ is different from the shear-stress rate $\dot{T}_{21}(x_2=h/2)$. Thus, $\partial\dot{M}_{32}/\partial x_2\neq 0$ holds for $x_2=h/2$ and depends on the height h of the shear layer according to

$$\left[\frac{\partial\dot{M}_{32}}{\partial x_2}\right]_{x_2=h/2}=\frac{f_c\,\hat{a}^2\,\dot{u}_1/(h/\eta)}{2[(h/\eta)\,\cosh(h/(2\eta))-\sinh(h/(2\eta))]}. \tag{5.26}$$

The results indicate that micro-polar effects appear across the entire height of the shear layer and a localization of the deformation is not manifested at the beginning of shearing.

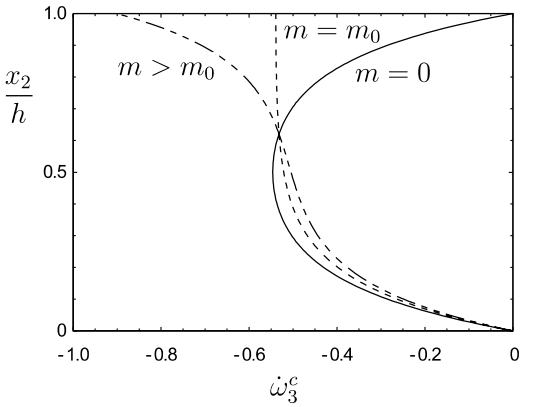


Figure 4. Distribution of initial rate of the micro-polar rotation, $\dot{\omega}_3^c$, for the boundary conditions at the bottom $x_2 = 0 : \dot{u}_1 = \dot{u}_2 = 0$, $\dot{\omega}_{3B}^c = 0$, and at the top $x_2 = h : \dot{u}_1 = \dot{u}_{1T}$, $\dot{T}_{22} = 0$, $m = 0$ (solid curve), $m = m_0$ (dotted curve) and $m > m_0$ (chain-dotted curve).

Numerical simulations of large shearing show that shear localization takes place with advanced shearing in the area of the maximum magnitude of the rate of the micro-rotations; see *e.g.* [38, pp. 79–85], [30,49,52]. Thus, the extreme value of the rate of micro-rotations obtained for the initial response (Figure 2c) is an indicator of where shear-strain localization may develop under continuous shearing.

Another case of interest arises for a zero couple-stress rate along the top surface. The boundary condition $\dot{M}_{32}(x_2 = h) = \dot{M}_{32T} = 0$ can alternatively be expressed by the corresponding micro-rotation:

$$\dot{\omega}_3^c(x_2 = h) = \dot{\omega}_{3T}^c = -m_0 \frac{\dot{u}_{1T}}{\eta}, \tag{5.27}$$

with:

$$m_0 = \frac{1 - 1/\cosh(h/\eta)}{2[h/\eta - \tanh(h/\eta)]}, \quad \eta = \frac{a_m}{\hat{a}} d_{50}.$$

Together with $\dot{\omega}_{3B}^c = 0$ prescribed at the bottom of the layer the micro-polar boundary conditions are no longer symmetric. By inserting relation (5.27) for $\dot{\omega}_{3T}^c$ in (5.23–5.25), one obtains the initial response of the shear layer for the boundary conditions $\dot{\omega}_{3B}^c = 0$ and $\dot{M}_{32T} = 0$ (dotted curves in Figure 2 and Figure 3). At the beginning of shearing the distribution of the velocities $\dot{u}_1(x_2)$ (Figure 2a) and $\dot{u}_2(x_2)$ (Figure 2b) is again nonlinear and almost the same as for the case of $\dot{\omega}_{3T}^c = 0$. The further evolution of these quantities, however, can be quite different when shear localization appears under larger shearing, as can be shown by numerical simulations; see *e.g.* [49]. In contrast with the results obtained for the boundary condition $\dot{\omega}_{3T}^c = 0$, the extreme value of the micro-rotation (Figure 2c) and macro-rotation (Figure 2d) occurs at the top of layer. Both $\mathrm{d}\dot{\omega}_3/\mathrm{d}x_2$ and $\mathrm{d}\dot{\omega}_3^c/\mathrm{d}x_2$ are zero at the top of the layer, which underlines the meaning of $\dot{M}_{32T} = 0$ as a special boundary condition, where the curvature rate $K_{32} = \mathrm{d}\dot{\omega}_3^c/\mathrm{d}x_2$ changes sign (Figure 4). The normalized quantity of the couple-stress rate $\dot{M}_{32}$ (Figure 3d) again coincides with the curvature rate K_{32} and the initial shear-stress rate $\dot{T}_{12}$ (Figure 3a) is a little lower than for the symmetric boundary conditions, $\dot{\omega}_{3B}^c = \dot{\omega}_{3T}^c = 0$. In contrast with the results obtained for symmetric boundary conditions, the maximum shear stress rate $\dot{T}_{21}$ (Figure 3b) and the maximum normal-stress rates $\dot{T}_{11} = \dot{T}_{33}$ (Figure 3c) occur at the top of the layer when a zero couple-stress rate is prescribed.

It follows from relation (5.27) that, for $\dot{M}_{32T}=0$, the micro-rotation rate $\dot{\omega}_{3T}^{c}$ along the top boundary is proportional to the prescribed horizontal shear displacement $\dot{u}_{1T}$. The evolution of the state quantities are also determined by the height of the shear layer because the factor m_0 depends on h. By replacing m_0 with an arbitrary factor m relation (5.27) agrees with the empirical formula proposed by Tejchman [28] to model the influence of a rough bounding surface on the rotation resistance of particle in contact with the bounding structure. Numerical simulations by Huang *et al.* [52] show that the location and thickness of shear-strain localization is strongly influenced by the value of the factor m. In particular for $m=0$ the case of a very rough top surface without particle rotations is modeled which shows an extreme value of micro-rotations in the middle of the shear layer. For $0<m<m_0$ the extreme value of micro-rotations and consequently the location where shear-strain localization takes place is located within the upper part of the shear layer. For $m>m_0$ shear-strain localization takes place very close to the top boundary and the thickness of the localized zone is smaller than for $m<m_0$. The influence of m on the initial response of the micro-rotations is demonstrated in Figure 4 for different values for m. It is obvious that m determines the sign of the curvature rate $K_{32}=\mathrm{d}\dot{\omega}_3^c/\mathrm{d}x_2$ and the location of maximum micro-rotation. At the top of the layer the curvature rate is positive for $m<m_0$, negative for $m>m_0$ and zero for $m=m_0$. The latter is related to $\dot{M}_{32T}=0$.

For prescribed zero couple-stress rates along the bottom and top surfaces the response of the shear layer coincides with the results obtained for the classical non-polar continuum. In particular the constants C_2 and C_3 become zero by inserting the boundary conditions $\dot{M}_{32}(x_2=0)=\dot{M}_{32}(x_2=h)=0$ in relation (5.22). Consequently relation (5.16) yields $C_4=0$ for the bottom boundary condition $\dot{u}_1(x_2=0)=0$, and with respect to the prescribed horizontal shear velocity $\dot{u}_1(x_2=h)=\dot{u}_{1T}$ at the top of the layer one obtains $C_1=\dot{u}_{1T}/(2h)$. For the bottom boundary condition $\dot{u}_2(x_2=0)=0$ relation (5.18) yields $C_5=0$. Then the couple stresses are zero, the micro-rotation rate corresponds to the macro-rotation rate, the shear stresses are symmetric and the velocity profile is linear across the height of the layer, *i.e.*,

$$\dot{\omega}_3^c=\dot{\omega}_3=-\dot{u}_{1T}/(2h), \quad K_{32}=0, \tag{5.28}$$

$$\dot{g}_1=\dot{u}_{1T}/h, \quad \dot{u}_1=(\dot{u}_{1T}/h)\,x_2, \tag{5.29}$$

$$\dot{g}_2=-\sqrt{2[\dot{u}_{1T}/(2h)]^2/[[\hat{a}(3+c)+1/(3\hat{a})]^2-1]}, \tag{5.30}$$

$$\dot{u}_2=-\sqrt{2[\dot{u}_{1T}/(2h)]^2/[[\hat{a}(3+c)+1/(3\hat{a})]^2-1]}\,x_2, \tag{5.31}$$

$$\dot{T}_{11}=\dot{T}_{33}=f_c\sqrt{2\hat{a}^2/[[\hat{a}(3+c)+1/(3\hat{a})]^2-1]}\,\dot{u}_{1T}/(2h), \tag{5.32}$$

$$\dot{T}_{12}=\dot{T}_{21}=f_c\,\hat{a}^2\,\dot{u}_{1T}/(2h), \quad \dot{M}_{32}=0. \tag{5.33}$$

It follows from (5.28) to (5.33) that for the couple-stress boundary conditions $\dot{M}_{32}(x_2=0)=\dot{M}_{32}(x_2=h)=0$ no polar properties appear within the shear layer. Numerical simulations show, however, that for continued shearing the non-polar solution is unstable. Any perturbation immediately leads to the polar solution. This phenomenon was termed "spontaneous polarisation" by Gudehus [53].

6. Conclusion

The influence of particle rotation and couple stresses in granular materials has been modeled with a micro-polar hypoplastic continuum approach. The evolution equations for the non-symmetric stress tensor and the couple-stress tensor are nonlinear isotropic tensor-valued functions, which are homogeneous of first order in the rate of deformation and rate of curvature. For an infinite strip of a micro-polar hypoplastic material located between two parallel plates under plane shearing an analytical solution has been derived. The general solution shows that polar quantities appear within the shear layer from the beginning of shearing. In contrast with the classical non-polar continuum the shear velocity and dilatancy velocity are nonlinearly distributed across the height of the shear layer, even if the material is homogeneous and non-polarized in the initial state. The macro-rotation rate is not constant and differs from the micro-rotation rate. Shear stresses are non-symmetric and the stress and couple-stress rates are distributed nonlinearly. The distribution of the state quantities within the shear layer strongly depends on the prescribed micro-polar boundary conditions which reflect the rotation resistance of the particles against the surface of the bounding structure. This is demonstrated for two different wall boundary conditions at the top of the shear layer. For the interaction with a very rough wall particle rotations are prevented, which can be modeled with locked micro-rotations along this boundary. Then the corresponding couple-stress rate shows an extreme value at this boundary. The results for the rate of the state quantities obtained for zero couple-stress rates prescribed along one of the boundaries show a non-symmetric distribution within the shear layer. The same result as for the non-polar continuum is obtained for the special case of zero couple-stress rates prescribed at both the bottom and the top boundaries. For an initially homogeneous and isotropic state, an analytical solution for the present micro-polar hypoplastic material model under plane shearing is restricted to the initial response. In general, the underlying set of differential equations for continuous shearing can only be solved numerically.

References

1. V.K. Garga and J.A. Infante Sedano, Steady state strength of sands in a constant volume ring shear apparatus. *Geotech. Testing J.* 25 (2002) 414–421.
2. K.H. Roscoe, The influence of strains in soil mechanics, 10th Rankine Lecture. *Géotechnique* 20 (1970) 129–170.
3. H.B. Mühlhaus and I. Vardoulakis, The thickness of shear bands in granular materials. *Geotechnique* 37 (1987) 271–283.
4. M. Oda, Micro-fabric and couple stress in shear bands of granular materials. In: C.C. Thornton (ed.), *Powders and Grains*, 3. Rotterdam: Balkema (1993) pp. 161–167.
5. J. Desrues, R. Chambon, M. Mokni and F. Mazerolle, Void ratio evolution inside shear bands in triaxial sand specimens studied by computed tomography. *Géotechnique* 46 (1996) 529–546
6. E. Cosserat and F. Cosserat, Théorie des Corps Deformables. *Herman et fils*, Paris (1909).
7. R.D. Mindlin, Stress functions for a Cosserat continuum. *Int. J. Solids Struct*. 1 (1965) 265–271.
8. A.C. Eringen, Polar and nonlocal field theories. *Continuum Physics* IV. New York, San Francisco, London: Academic Press (1976) 274 pp.
9. E.C. Aifantis, On the microstructural original of certain inelastic models. *J. Engng. Mat. Technol.* 106 (1984) 326–334.
10. Z.P. Bazant, T.B. Belytschko and T.P. Chang, Continuum theory for strain softening. *ASCE J. Engng. Mech.* 110 (1984) 1666–1692.
11. H.-B. Mühlhaus, Application of Cosserat theory in numerical solutions of limit load problems. *Ingenieur Archiv* 59 (1989) 124–137.

12. I. Vardoulakis and E.C. Aifantis, A gradient flow theory of plasticity for granular materials. *Acta Mech*. 87 (1991) 197–217.
13. D. Kolymbas, An outline of hypoplasticity. *Arch. Appl. Mech*. 3 (1991) 143–151.
14. F. Darve, Incrementally non-linear constitutive relationships. In: F. Darve (ed.), *Geomaterials: Constitutive Equations and Modelling*. Amsterdam: Elsevier (1991) pp. 213–237.
15. Y.F. Dafalias, Bounding surface plasticity: I. Mathematical foundation and hypoplasticity. *J. Engng. Mech., ASCE* 112 (1986) 966–987.
16. C. Truesdell and W. Noll, The non-linear field theories of mechanics. In: S. Flügge (ed.), *Encyclopedia of Physics* III/c. Heidelberg: Springer press (1965) pp. 1–602.
17. D. Kolymbas, A generalized hypoelastic constitutive law. *Proc. 11th Int. Conf. Soil Mechanics and Foundation Engineering* 5. Rotterdam: Balkema (1988) p. 2626.
18. W. Wu and E. Bauer, A simple hypoplastic constitutive model for sand. *Int. J. Num. Anal. Methods Geomech.* 18 (1994) 833–862.
19. G. Gudehus, A comprehensive constitutive equation for granular materials. *Soils and Foundations* 36 (1996) 1–12.
20. E. Bauer, Calibration of a comprehensive hypoplastic model for granular materials. *Soils and Foundations* 36 (1996) 13–26.
21. P.A. von Wolffersdorff, A hypoplastic relation for granular materials with a predefined limit state surface. *Mech. Cohesive-Frictional Materials* 1 (1996) 251–271.
22. W. Wu and D. Kolymbas, Hypoplasticity then and now. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin, Heidelberg, Newyork: Springer (2000) pp. 57–105.
23. C. Tamagnini, G. Viggiani and R. Chambon, A review of two different approaches to hypoplasticity. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin, Heidelberg, Newyork: Springer (2000) pp. 107–145.
24. E. Bauer and I. Herle, Stationary states in hypoplasticity. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin, Heidelberg, Newyork: Springer (2000) pp. 167–192.
25. J. Tejchman and E. Bauer, Numerical simulation of shear band formation with a polar hypoplastic constitutive model. *Comp. Geotech.* 19 (1996) 221–244.
26. G. Gudehus, Shear localization in simple grain skeleton with polar effect. In: T. Adachi, F. Oka and A. Yashima (eds.), *Proc. of the 4th Int. Workshop on Localization and Bifurcation Theory for Soils and Rocks*. Rotterdam: Balkema (1998) pp. 3–10.
27. G. Gudehus, Attractors, percolation thresholds and phase limits of granular soils. In: R.P. Behringer and J.T. Jenkins (eds.), *Powders and Grains*. Rotterdam: Balkema (1997) pp. 169–183.
28. J. Tejchman, Modelling of shear localisation and autogeneous dynamic effects in granular bodies. *Veröffentlichungen des Institutes für Bodenmechanik und Felsmechanik der Universität Fridericiaca in Karlsruhe*, 140 (1997) 353 pp.
29. E. Bauer and W. Huang, Numerical study of polar effects in shear zones. In: G.N. Pande, S. Pietruszczak and H.F. Schweigers (eds.), *Proc. of the 7th Int. Symp. on Num. Models in Geomechanics*. Rotterdam: Balkema (1999) pp. 133–138.
30. J. Tejchman and G. Gudehus, Shearing of a narrow granular layer with polar quantities. *Int. J. Num. Meth. Geomech.* 25 (2001) 1–28.
31. W. Huang, K. Nübel and E. Bauer, Polar extension of a hypoplastic model for granular materials with shear localization. *Mech. Materials* 34 (2002) 563–576.
32. J. Tejchman and I. Herle, A "class A" prediction of the bearing capacity of plane strain footings on sand. *Soils and Foundations* 39 (1999) 47–60.
33. K. Nübel and R. Cudmani, Examples of finite element calculations with the hypoplastic law. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin, Heidelberg, Newyork: Springer (2000) pp. 523–538.
34. E. Bauer and W. Huang, Numerical investigation of strain localization in a hypoplstic Cosserat material under shearing. In: C.S. Desai, T. Kundu, S. Harpalani, D. Contractor and J. Kemeny (eds.), *Proc. of the 10th Int. Conf. on Computer Methods and Advances in Geomechanics*. Rotterdam: Balkema (2001) pp. 525–528.
35. G. Gudehus and K. Nübel, Evolution of shear bands in sand. *Géotechnique* 54 (2004) 187–201.
36. J. Hill, Some symmetrical cavity problems for a hypoplstic granular material. *Q. J. Mech. Appl. Math.* 53 (2000) 111–135.
37. E. Bauer and W. Huang, Evolution of polar quantities in a granular Cosserat material under shearing. In: H.-B. Mühlhaus, A.V. Dyskin and E. Pasternak (eds.), *Proc. 5th Int. Workshop on Bifurcation and Localization Theory in Geomechanics*. Rotterdam: Balkema (2001) pp. 227–238.

38. W. Huang, *Hypoplastic Modelling of Shear Localisation in Granular Materials.* Dissertation. Graz University of Technology, Austria (2000) 107 pp.
39. W. Wu and A. Niemunis, Failure criterion, flow rule and dissipation function derived from hypoplasticity. *Mech. Cohesive-Frictional Materials* 1 (1996) 145–163.
40. W. Wu and E. Bauer, A hypoplastic model for barotropy and pyknotropy of granular soils. In: D. Kolymbas (ed.), *Proc. of the Int. Workshop on Modern Approaches to Plasticity*, Amsterdam Elsevier (1993) pp. 225–245.
41. W. Wu, E. Bauer and D. Kolymbas, Hypoplastic constitutive model with critical state for granular materials. *Mech. Materials* 23 (1996) 45–69.
42. W. Wu, E. Bauer, A. Niemunis and I. Herle, Visco-hypoplastic models for cohesive soils. In: D. Kolymbas (ed.), *Proc. of the Int. Workshop on Modern Approaches to Plasticity*, Amsterdam: Elsevier (1993) pp. 365–383.
43. G. Gudehus, Hypoplastic shear localisation in psammoids and peloids. *2nd Int. Symposium on Continuous and Discontinuous Modelling of Cohesive Frictional Materials*, Publication in print (2004).
44. W. Wu, Rational approach to anisotropy of sand. *Int. J. Num Anal. Methods in Geomech.* 24 (1998) 921–940.
45. E. Bauer, W. Wu and W. Huang, Influence of initially transverse isotropy on shear banding in granular materials. In: J.F. Labuz and A. Arescher (eds.), *Proc. of the Int. Workshop on Bifurcation and Instabilities in Geomechanics*. Rotterdam: Balkema (2003) pp. 161–172.
46. E. Bauer and W. Wu, Extension of Hypoplastic Constitutive Model with Respect to Cohesive Powders. In: H.J. Siriwardane and M.M. Zadan (eds.), *Proc. of the Eighth Intern. Conf. on Computer Methods and Advances in Geomechanics*. Rotterdam: Balkema (1994) pp. 531–536.
47. A. Niemunis and I. Herle, Hypoplastic model for cohesionless soils with elastic strain range. *Mech. Cohesive-Frictional Materials* 2 (1997) 279–299.
48. A.E. Green and P.M. Naghdi, A general Theory of an elastic-plastic continuum. *Arch. Rat. Mech. Anal.* 18 (1965) 251–281.
49. W. Huang and E. Bauer, Numerical investigations of shear localization in a micro-polar hypoplastic material. *Int. J. Num. Anal. Meth. Geomech.* 27 (2003) 325–352.
50. H. Matsuoka and T. Nakai, Stress-strain relationship of soil based on the 'SMP'. *Proc. of Speciality Session 9, IX Int. Conf. Soil Mech. Found. Eng.*, Tokyo (1977) pp. 153–162.
51. E. Bauer, Conditions for embedding Casagrande's critical states into hypoplasticity. *Mech. Cohesive-Frictional Materials* 5 (2000) 125–148.
52. W. Huang, E. Bauer and S. Sloan, Behaviour of interfacial layer along granular soil-structure interfaces. *Struct. Engng. Mech.* 15 (2003) 315–329.
53. G. Gudehus, Forced and spontaneous polarisation in shear zones. In: H.-B. Mühlhaus, A. V. Dyskin and E. Pasternak (eds.), *Proc. 5th Int. Workshop on bifurcation and Localization Theory in Geomechanics*, Rotterdam: Balkema (2001) pp. 45–51.

Journal of Engineering Mathematics (2005) 52: 53–61
DOI 10.1007/s10665-005-0442-8

Some theoretical results about second-order work, uniqueness, existence and controllability independent of the constitutive equation

RENÉ CHAMBON
Laboratoire 3S Université Joseph Fourier, Institut National Polytechnique, CNRS, UMR 5521, B.P. 53 38041 Grenoble Cedex France (rene.chambon@inpg.fr)

Received 6 April 2004; accepted in revised form 6 January 2005

Abstract. When they are studied as continuum media, granular materials and other soils and rocks exhibit a complex behavior. Contrary to metals, their isotropic and deviatoric behavior are coupled. This implies some mathematical difficulties concerning boundary-value problems solved with constitutive equations modelling the salient features of such geomaterials. One of the well-known consequences is that the so-called second-order work can be negative long before theoretical failure occurs. Keeping this in mind, the starting point of this work is the pioneering and illuminating work of Nova (1994), who proved that using an isotropic hardening elasto-plastic model not obeying the normality rule, it is possible to exhibit either loss of uniqueness or loss of existence of the solution of a boundary-value problem as soon as the second-order work is negative. Because the geomaterial behavior is quite difficult to model, in practice many different constitutive equations are used. It is then important to study the point raised by Nova for other constitutive equations. In this paper, his result is generalized for any inelastic rate-independent constitutive equation. Similarly the link between localization and controllability proved by Nova (1989) is extended to some extent to a general inelastic model.

Key words: existence, geomaterials, localization, non-normality, second-order work, uniqueness

1. Introduction

The behavior of geomaterials (granular materials, soils, and rocks) has some specific properties which create certain mathematical difficulties when constitutive equations modelling this behavior are used in a boundary-value problem. One of these properties can be experienced in everyday life. Walking on a beach after the last wave has filled up all the pores of the sand shows clearly that each footprint drains out the surrounding sand. This means that there is an increase in the pore volume in the surrounding sand, even though it is likely that the mean (effective) pressure increases. This phenomenon clearly shows that granular-material behavior exhibits a coupling between the isotropic volume change and the deviatoric stress. This behavior is rather different from that of metals. This complex behavior explains why so many constitutive equations are elaborated in order to model granular materials and other geomaterials. These constitutive equations are now often based on the well-known framework of classical plasticity (which here means isotropic hardening obeying a normality rule). Many of these depart strongly from classical plasticity theory such as, for instance, multi-mechanism plasticity models [1], bounding-surface plasticity models [2] hypoplasticty models – see [3] for a review – or multi-laminate [4] and microplane [5] models.

These inelastic constitutive equations are used in numerical computations assuming implicitly the well-posedness of the underlying boundary-value problem. It is our opinion that there is a need for knowledge about existence and uniqueness of solutions of boundary-value problems involving such general inelastic models. Except for classical elasto-plasticity constitutive

equations (see [6]), for non-associative plasticity (see [7]) and for hypoplasticity (see [8]), there is no simple and general (here general means independent of the geometry and the boundary conditions of the problem) results about existence and uniqueness of solutions of boundary-value problems involving inelastic constitutive equations. A common result can be deduced from the results quoted above. Considering problems solved with the so-called small-strain assumption, we observe that the positiveness of the second-order work (everywhere and for any strain rate) ensures the uniqueness of the solution for classical elastoplastic [9] and hypoplastic [8] models.

On the other hand, it has been proved by Nova for isotropic hardening elastoplastic models [10] – see also [11] – that, if the second-order work can be negative, then it is possible to construct a problem (starting from an homogeneous state) such that, for some specific boundary conditions, the uniqueness of (in this case) the homogeneous solution is lost. This result has been extended to particular cases of hypoplasticity by El Hassan [12]. More recently Niemunis gave a general proof for hypoplastic theories [13]. One of the objectives of this paper is to generalize this result to a wide class of constitutive equations. It is necessary to emphasize that, even if the constitutive equations are rather general, the problem solved is quite particular because it is related to homogeneous states only. This problem has to be related to what is often called material instability [14,9]. It is necessary to quote here the work of Petryk concerning problems similar to the one addressed here for the multi-mechanism plasticity theory. Within this specific framework his work goes beyond the scope of this paper, since he takes into account geometrical nonlinearities [15].

The second objective is to give a sufficient localization condition – *i.e.*, a condition which implies that all the equations of a Rice localization analysis [16] are fulfilled – for any rate-independent materials and to prove that this condition implies the negativeness of the second-order work.

The paper presentation is as follows. In the first part we present the problem under consideration. In particular, the controllability as defined by Nova is recalled. Next we prove that the loss of positiveness of the second-order work implies loss of controllability. The third part deals with the shear-band analysis. A conjecture is explained in a concluding-remarks section.

The following notations are used: a tensor is denoted by an underlined symbol like $\underline{\sigma}$, the component of a tensor (or vector) is denoted by the name of the tensor (or vector) accompanied by lower indices. Other indices and among them upper indices have specific meanings defined in the text. The summation convention with respect to repeated tensorial indices is used.

2. The problem under consideration

2.1. The basic assumptions

We are dealing with inelastic materials not exhibiting viscous effect in the small-strain range. The constitutive equation can thus be written in rate form as follows:

$$\dot{\underline{\sigma}} = \mathcal{F}(\dot{\underline{\varepsilon}}), \tag{1}$$

where $\dot{\underline{\sigma}}$ is the stress rate, $\dot{\underline{\varepsilon}}$ the strain rate and $\mathcal{F}$ a tensorial function depending on the state of the material. It is assumed first that $\mathcal{F}$ is invertible which means defining $\mathcal{G}$ as the inverse of $\mathcal{F}$ so that:

$$\dot{\underline{\varepsilon}} = \mathcal{G}(\dot{\underline{\sigma}}). \tag{2}$$

If this assumption does not hold, it is clear that controllability is lost. Loss of invertibility means: given $\underline{\dot{\sigma}}$, either $\mathcal{G}(\underline{\dot{\sigma}})$ does not exist or at least there are two different $\underline{\dot{\varepsilon}}$ such that Equation (1) holds. Consequently a boundary-value problem corresponding to the boundary conditions compatible with the given value of $\underline{\dot{\sigma}}$ either has no homogeneous solution or has several solutions. In both cases controllability defined hereafter is lost.

As we are studying non-viscous materials, $\mathcal{F}$ and $\mathcal{G}$ are homogeneous of the degree one with respect to their respective arguments. We will add other assumptions when this is necessary. The quantities $\underline{\dot{\sigma}}$ and $\underline{\dot{\varepsilon}}$ belong to the six-dimensional space of symmetric second-order tensors. In the following it will be useful to define an orthonormal basis for this set which means a set of six second-order tensors denoted in the following by ${}^{i}\underline{e}$ such that

$$\forall i,\ \|{}^{i}\underline{e}\| = 1, \qquad \forall i, j,\ \ i \neq j,\ \ {}^{i}\underline{e} \cdot {}^{j}\underline{e} = 0, \tag{3}$$

where $\cdot$ denotes the usual scalar product of two second-order tensors (for instance the scalar product $\underline{\dot{\sigma}} \cdot \underline{\dot{\varepsilon}} = \dot{\sigma}_{ij}\dot{\varepsilon}_{ij}$ defines the second-order work).

2.2. Controllability

In his paper Nova [10] defined controllability as follows. First he noticed that in some experiments, like the classical triaxial tests, some components of the strain and the other components of the stress are prescribed. He pointed out the practical importance of being able to perform such a test, *i.e.*, to get one (existence) and only one (uniqueness) response for such a test. He noticed then, that in some tests like the undrained ones, a linear combination of the classical strain components is prescribed. Generalizing this remark he defined a new set of strain (and stress) variables related to the strain (or stress) components via a product with a non-singular matrix the inverse of which is equal to its transpose (which means orthonormal change of basis). He proceeded by defining controllability as the ability of a material (or a model) to provide one and only one (existence and uniqueness) response to any loading path for which some strain components (in this new basis) and the other stress components are prescribed.

Finally, for a constitutive equation, controllability in other words is, existence and uniqueness of the solution of the following problem. In a given orthonormal basis some components of the strain rate and the other components of the stress rate are prescribed, and we try to solve the corresponding boundary-value problem. If there exists one and only one solution for the non-prescribed components (of the stress and the strain rates) satisfying the constitutive equation (1) or (2), then the model is said to be controllable. Otherwise it is not controllable.

3. Consequences of the non-positiveness of the second-order work

3.1. The non-positiveness of the second-order work implies non-uniqueness

Rephrasing the problem as in the previous section allows us to prove that, if the second-order work is equal to zero for some strain direction, then the constitutive equation is non-controllable. As already seen in Section 2.1, it is assumed that the constitutive Equation (1) is invertible or else the model is clearly non-controllable. If the second-order work can be equal to zero, there exists some strain rate denoted by $\underline{\dot{\varepsilon}}^{0} \neq 0$ such that:

$$\text{if} \quad \underline{\dot{\sigma}}^{0} = \mathcal{F}(\underline{\dot{\varepsilon}}^{0}), \qquad \text{then} \quad \underline{\dot{\sigma}}^{0} \cdot \underline{\dot{\varepsilon}}^{0} = 0. \tag{4}$$

This means that the second-order work is not strictly positive. Now we choose the orthonormal basis such that:

$$ {}^{1}\underline{e} = \frac{\underline{\dot{\varepsilon}}^{0}}{\|\underline{\dot{\varepsilon}}^{0}\|} \tag{5} $$

and

$$ {}^{2}\underline{e} = \frac{\underline{\dot{\sigma}}^{0}}{\|\underline{\dot{\sigma}}^{0}\|}. \tag{6} $$

Denoting in the following by ${}^{i}a$ and ${}^{i}b$ the components of $\underline{\dot{\varepsilon}}$ and $\underline{\dot{\sigma}}$, respectively, we have

$$ \underline{\dot{\varepsilon}} = {}^{1}a\,{}^{1}\underline{e} + {}^{2}a\,{}^{2}\underline{e} + {}^{3}a\,{}^{3}\underline{e} + {}^{4}a\,{}^{4}\underline{e} + {}^{5}a\,{}^{5}\underline{e} + {}^{6}a\,{}^{6}\underline{e} \tag{7} $$

and

$$ \underline{\dot{\sigma}} = {}^{1}b\,{}^{1}\underline{e} + {}^{2}b\,{}^{2}\underline{e} + {}^{3}b\,{}^{3}\underline{e} + {}^{4}b\,{}^{4}\underline{e} + {}^{5}b\,{}^{5}\underline{e} + {}^{6}b\,{}^{6}\underline{e}. \tag{8} $$

For instance, the components of $\underline{\dot{\varepsilon}}^{0}$ are

$$ \dot{\varepsilon}^{0}, 0, 0, 0, 0, 0, \tag{9} $$

defining $\dot{\varepsilon}^{0}$ and the components of $\underline{\dot{\sigma}}^{0}$ are

$$ 0, \dot{\sigma}^{0}, 0, 0, 0, 0, \tag{10} $$

defining $\dot{\sigma}^{0}$.

Controllability means that for some prescribed components ${}^{i}a$, and the prescribed complementary components ${}^{j}b$ (which then define a mixed loading path), it is possible to find one and only one set of the non-prescribed values of ${}^{i}a$ and ${}^{j}b$, such that the corresponding strain rate and stress rate meet the constitutive equation (1) or (2). Now let us consider the following problem: ${}^{1}b$, ${}^{2}a$, ${}^{3}a$, ${}^{4}a$, ${}^{5}a$ and ${}^{6}a$ are prescribed equal to zero, that is,

$$ {}^{1}b = {}^{2}a = {}^{3}a = {}^{4}a = {}^{5}a = {}^{6}a = 0. \tag{11} $$

Clearly a solution is given by

$$ {}^{1}a = 0, \quad {}^{2}b = 0, \quad {}^{3}b = 0, \quad {}^{4}b = 0, \quad {}^{5}b = 0, \quad {}^{6}b = 0. \tag{12} $$

However, another solution is:

$$ {}^{1}a = \dot{\varepsilon}^{0}, \quad {}^{2}b = \dot{\sigma}^{0}, \quad {}^{3}b = 0, \quad {}^{4}b = 0, \quad {}^{5}b = 0, \quad {}^{6}b = 0. \tag{13} $$

Moreover, due to the positive homogeneity of degree one of the constitutive equation, other solutions are

$$ {}^{1}a = \lambda\dot{\varepsilon}^{0}, \quad {}^{2}b = \lambda\dot{\sigma}^{0}, \quad {}^{3}b = 0, \quad {}^{4}b = 0, \quad {}^{5}b = 0, \quad {}^{6}b = 0, \tag{14} $$

where λ is any positive number. For the same boundary conditions, the ones corresponding to the prescribed values defined in Equation (11), there exists not only the solution given by Equation (12), but also all the solutions given by Equation (14). Clearly, first uniqueness and then controllability is lost.

3.2. The non-positiveness of the second-order work implies either non-existence or *instability* of the solutions

In the previous section, we proved that uniqueness is lost which is sufficient to prove loss of controllability. In the current section, adding some assumptions about the differentiability of the constitutive equation will allow us to prove moreover, either the loss of existence or the fact that two closed prescribed inputs give two rather different solutions. This means in the latter case discontinuity of the (mixed) response of the constitutive equation with respect to the (mixed) input loading conditions. In some sense this property is more important. In numerical computations, it is not so easy to detect non-uniqueness for a fully nonlinear problem like the one we are studying here; see, however, the algorithm proposed by Chambon [17]. On the contrary, it is easy to detect the non-existence or *instability* of the solution which is often related to non-convergence when a full Newton-Raphson method is used to solve accurately the nonlinear boundary-value problem.

3.2.1. *New assumptions*

We assume now that the functions $\mathcal{F}$ and $\mathcal{G}$ are continuous and differentiable, except in the vicinity of the null tensor. This is not so strong a restriction. However, the flow theory of plasticity does not meet this condition for a strain (or stress) rate direction corresponding to neutral loading. On the other hand, if $\dot{\underline{\varepsilon}}^0$ is not within this (very) restricted set of strain rates, the following applies also to elasto-plasticity. Towards the end of Section 3.2.3, we will extend the proof to constitutive equations involving several mechanisms (several means here two or more than two), but for the beginning we assume the following. Constitutive Equation (1) or (2) can be rewritten in the following form.

$$^{i}b = {}^{i}\varphi({}^{j}a), \tag{15}$$

where $i, j \in \{1, 2, 3, 4, 5, 6\}$ and $^{i}\varphi$ are functions that are positively homogeneous of degree one, continuous and differentiable, except for $^{j}a = 0$, $\forall j$. Let us write

$$^{ij}A = \frac{\partial\, {}^{i}\varphi}{\partial\, {}^{j}a}, \tag{16}$$

which are only defined in the vicinity of a given strain direction.

3.2.2. *Proof of the property*

We are now looking at the constitutive equation in the vicinity of $\dot{\underline{\varepsilon}}^0$.

A variation of $^{1}a = \dot{\varepsilon}^0$ means only a variation of the magnitude of $\dot{\underline{\varepsilon}}^0$. Since the constitutive equation is positively homogeneous and since $\dot{\underline{\sigma}}^0 = \mathcal{F}(\dot{\underline{\varepsilon}}^0)$, a variation of the magnitude of $\dot{\underline{\varepsilon}}^0$ implies only a variation of the magnitude of $\dot{\underline{\sigma}}^0$. This implies that:

$$^{i1}A(\dot{\underline{\varepsilon}}^0) = 0 \quad \forall i \neq 2 \tag{17}$$

and

$$^{21}A(\dot{\underline{\varepsilon}}^0) = \frac{\dot{\sigma}^0}{\dot{\varepsilon}^0}. \tag{18}$$

Let us now look for a solution of the following problem where

$$^{1}b = \alpha \quad {}^{2}a = 0 \quad {}^{3}a = \beta \quad {}^{4}a = {}^{5}a = {}^{6}a = 0 \tag{19}$$

are prescribed. The parameters α and β are assumed to be small with respect to $\dot{\varepsilon}^0$. When $\alpha=\beta=0$, we have the following solutions

$$\forall\lambda, \quad {}^1a=\lambda\dot{\varepsilon}^0 \quad {}^2b=\lambda\dot{\sigma}^0 \quad {}^3b=0 \quad {}^4b=0 \quad {}^5b=0 \quad {}^6b=0. \tag{20}$$

Thus, if $0\neq\alpha\neq\beta\neq0$, and if we are looking for a solution close to a previous one, such that $\lambda\neq0$, then we can use the derivatives of the constitutive equation defined in (16) in the vicinity of $\dot{\underline{\varepsilon}}^0$ or in the vicinity of $\lambda\dot{\underline{\varepsilon}}^0$ (Owing to positive homogeneity these derivatives are the same). This implies that necessarily

$${}^1a=\lambda\dot{\varepsilon}^0+\gamma \tag{21}$$

with γ being of the same order as α and β. So finally this implies:

$${}^1b=\alpha={}^{11}A\gamma+{}^{13}A\beta={}^{13}A\beta. \tag{22}$$

Then $\alpha={}^{13}A\beta$, which contradicts the fact that α and β are chosen independently, and so generally a solution of our problem does not exist in the vicinity of the direction of $\dot{\underline{\varepsilon}}^0$.

Let us summarize this result. There is no solution for the problem just defined in the vicinity of the direction of $\dot{\underline{\varepsilon}}^0$ $\forall\lambda\neq0$. Thus, either there is no solution at all (non existence) or, if there is a solution, it will not be in the vicinity of the direction of $\dot{\underline{\varepsilon}}^0$. This means in this case that for two close inputs given, respectively, by Equations (11) and (19) the corresponding solutions are not close to each other, which concludes the proof.

3.2.3. *Extension of the proof*

Let us now relax the assumptions of the beginning of Section 3.2.1 in order to apply our result to multi-mechanism plasticity. We assume that for some strain direction it is possible to define a finite number (say n) of zones $Z^k, k\in1,\ldots,n$ in the strain space. When $\dot{\underline{\varepsilon}}^0$ belongs to the boundary of these zones and for every zone Z^k, we can define

$${}^{ij}A^k=\frac{\partial\,{}^i\varphi}{\partial\,{}^ja}, \tag{23}$$

which are all defined in the vicinity of $\dot{\underline{\varepsilon}}^0$.

In this case, according to Section 3.2.2, if we are looking for a solution close to $\dot{\underline{\varepsilon}}^0$, we get $\alpha={}^{13}A^k\beta$ for at least one k, which still contradicts the fact that α and β are chosen independently. So it is possible to generalize the previous proof.

4. Shear-band analysis

4.1. The problem of shear banding

It is often claimed that shear banding corresponds to a zero value of the determinant of the acoustic tensor. The problem is that such a tensor can be defined using a dynamic analysis only for incrementally linear models. Generally speaking, a shear band can be generated if the following conditions hold [16]. We consider the problem of an initially homogeneous solid strained up to the current state. It is then submitted to a load rate on a straight loading path. A solution of the resulting rate-equilibrium problem corresponds to an homogeneous strain rate denoted by $\dot{\underline{\varepsilon}}^{\text{out}}$. Another solution involving the existence of a shear band is considered. It is assumed that the strain rate is equal to $\dot{\underline{\varepsilon}}^{\text{out}}$ outside a shear band and equal to

$$\dot{\underline{\varepsilon}}^{\text{in}}=\dot{\underline{\varepsilon}}^{\text{out}}+g\otimes n \tag{24}$$

inside the shear band. The vector n is normal to the shear band and g is some vector. Let $\underline{\dot{\sigma}}^{\text{in}}$ be the Cauchy stress rate with respect to a fixed frame inside the shear band and $\underline{\dot{\sigma}}^{\text{out}}$ outside. Along the boundaries of the band, equilibrium equations in a rate form can be written:

$$\underline{\dot{\sigma}}^{\text{in}} \cdot n = \underline{\dot{\sigma}}^{\text{out}} \cdot n. \tag{25}$$

Moreover the constitutive Equation (1) or (2) has to be satisfied inside and outside the band.

4.2. General criterion and consequences

For any constitutive equation shear bands are possible if there exist some n and some g such that

$$\mathcal{F}\left(\frac{1}{2}(g \otimes n + n \otimes g)\right) \cdot n = 0, \tag{26}$$

which means that

$$\underline{\dot{\sigma}} \cdot n = 0, \tag{27}$$

for the stress rate corresponding to the strain rate $\frac{1}{2}(g \otimes n + n \otimes g)$. In Equations (26) and (27), 0 like n or g is a vector.

It is easy to prove that if it is assumed that $\underline{\dot{\varepsilon}}^{\text{out}} = 0$, which means physically negligible with respect to $g \otimes n$, then Equation (26) implies Equation (24) and Equation (25). Then we have obtained a sufficient localization criterion available for any constitutive equation. Moreover, as pointed out by Nova [18] for elasto-plastic models and by Chambon [19] for hypoplastic models, shear-band localization implies that the corresponding second-order work is equal to zero, since by writing Equation (27) in component form, we have

$$\underline{\dot{\sigma}}_{ij} n_j = 0, \tag{28}$$

which implies

$$\underline{\dot{\sigma}}_{ij} n_j g_i = 0, \tag{29}$$

which is the second-order work written for the strain rate $\frac{1}{2}(g \otimes n + n \otimes g)$, since $\underline{\sigma}_{ij}$ is symmetric.

5. Concluding remarks

In Section 3, we studied homogeneous problem. This means that it is possible to put on an homogeneous sample, with boundary conditions corresponding to the prescribed components (in the basis defined above) of the strain or of the stress. Physically, this implies only the use of control devices and this can be actually encountered (for instance in undrained tests), as pointed out by Nova [18]. But we have to keep in mind that this kind of boundary conditions, as seen at the end of Section 2.2, linking different components, is not usually taken into account in classical existence and uniqueness theorem.

However, usually, existence uniqueness and controllability are proved for very particular constitutive equations. Indeed, the theorems proved here are almost independent of the constitutive equation.

Since we have proved recently that, for materials not obeying the normality rule, the second-order work can become negative strictly inside the limit surface [20], it is then possible to lose the well-posedness of the rate boundary-value problem far in advance of what is

classically seen as rupture. Let us emphasize that it is possible to define normality in a general manner without reference to a particular constitutive equation; see [20, Section 4.3.2]. So the previous proposition is true irrespective of the constitutive equation.

Al these discussions are very important for geomaterials which are well known not to obey normality rules, again without any reference to a particular law.

Clearly the loss of controllability of some homogeneous problem, when the second-order work can be negative, does not imply that uniqueness (or existence) is lost if the second-order work can be equal to zero or even negative in a subset of a studied domain. This statement is sustained by the stability study of Dascalu *et al.* [21] which proved for a linear system that the system can remain stable (which implies existence and uniqueness of the static system) if a limited fault undergoes softening provided that the corresponding (negative) softening modulus has a small absolute value.

We have studied only a rate problem here. It would be desirable to deal with the more interesting initial-boundary-value problem; however, the latter problem is more difficult to tackle.

Finally let us conclude by stating a conjecture. For a reasonable constitutive equation (continuous and smooth enough), when the second order is positive everywhere, and for any strain rate, then, independently of the boundary conditions (with the same restrictions as for classical elastic computations) and of the shape of the studied domain, the small-strain-rate boundary-value problem is well-posed. If, on the contrary, for some points of the domain and some strain directions, the second-order work is negative, it is possible that the corresponding small-strain-rate boundary-value problem is ill-posed.

References

1. W.T. Koiter, General theorems for elstic plastic solids. In: I.N. Sneddon, R. Hill (eds.), *Progress in Solid Mechnics* 1 Amsterdam: North Holland Publishing Comp. (1960) pp. 165–221.
2. Y. Dafalias and I. Popov, Plastic internal variables formalism of cyclic plasticity. *J. Appl. Mech.* 98 (1976) 645–650.
3. C. Tamagnini, G. Viggiani and R. Chambon, A review of two different approaches to hypoplasticity. In: D. Kolymbas (ed), *Constitutive Modelling of Granular Materials including Developments and Perspectives of Hypoplasticity*. Berlin: Springer (1999) pp. 107–145.
4. O.C. Zienkiewicz and G.N. Pande, Time-dependent multi-laminate model of rocks. A numerical study of deformation and failure of rock masses. *Int. Num. Analyt. Methods Geomech.* 1 (1977) 219–247.
5. M. Jirasek and Z.P. Bazant, *Inelastic Analysis of Structures*. Chichester: John Wiley and sons (2002) 758 pp.
6. R. Hill, Aspect of invariance in solids mechanics. *Adv. Appl. Mech.* 18 (1978) 1–75.
7. B. Raniecki and O.T. Bruhns, Bounds to bifurcation stresses in solids with non associated plastic flow rule at finite strain. *J. Mech. Phys. Solids* 29 (1981) 153–172.
8. R. Chambon and D. Caillerie, Existence and uniqueness theorems for boundary-value problems involving incrementally non linear models. *Int. J. Solids Struct.* 36 (1999) 5089–5099.
9. D. Bigoni, Bifurcation and instability on non associative elastoplastic solids. In: H. Petryk (ed.), *CISM Material Instabilities in Elastic and Plastic Solids.* Wien, New York: Springer (1999) pp. 1–52.
10. R. Nova, Controllability of the incremental response of soil specimens subjected to arbitrary loading programmes. *J. Mech. Behav. Materials* 5 (1994) 221–243.
11. S. Imposimato and R. Nova, An investigation on the uniqueness of the incremental response of elstoplastic models for virgin sand. *Mech. Cohesive Frictl. Material* 3 (1998) 65–87.
12. N. El Hassan, *Modélisation théorique et numérique de la localisation de la déformation dans les géomatériaux*. PhD thesis, Université Joseph Fourier Grenoble France (1997) 345 pp.
13. A. Niemunis, *Extended Hypoplasticiy for Soils*. Habilitation thesis, Bochum, Germany (2002) 191 pp.
14. H. Petryk, Theory of material instability in incrementally nonlinear plasticity. In: H. Petryk (ed.), *CISM Material Instabilities in Elastic and Plastic Solids*. Wien, New York: Springer (1999) pp. 261–331.

15. H. Petryk, General conditions for uniqueness in materials with multiple mechanisms in inelastic deformation. *J. Mech. Phys. Solids* 48 (2000) 367–396.
16. J. Rice, The localization of plastic deformation. In: W.D. Koiter (ed.), *Proceedings of the International Congress of Theoretical and Applied Mechanics*. Amsterdam: North Holland Publishing Comp. (1976) pp. 207–220.
17. R. Chambon, S. Crochepeyre and R. Charlier, An algorthm and a method to search bifurcation points in non linear problems. *Int. J. Num. Methods Engng.* 51 (2001) 315–332.
18. R. Nova, Liquefaction, stability, bifurcation, of soil via strainhardening plasticity. In: Z. Sikora (ed.), *Proceedings of the International Workshop on Bifurcation and Localization in Soil and Rocks*. Gdansk (1989) pp. 117–131.
19. R. Chambon and S. Crochepeyre, Daphnis a new model for the description of post-localization behavior: application to sands. *Mech. Cohesive Frictl. Material.* 3 (1998) 127–153.
20. R. Chambon and V. Roger, Mohr-Culomb mini CLoE model: uniqueness and localization studies, links with normality rule. *Int. J. Num. Analyt. Methods Geomech.* 27 (2003) 49–68.
21. C. Dascalu, I. R. Ionescu and M. Campillo, Fault finitness and initiation of dynamic shear instability. *Earth Planet. Sci. Lett.* 177 (2000) 163–176.

Journal of Engineering Mathematics (2005) 52: 63–91
DOI 10.1007/s10665-004-7952-7

Perturbation solutions for flow through symmetrical hoppers with inserts and asymmetrical wedge hoppers

G.M. COX*, S.W. MCCUE[1], N. THAMWATTANA and J.M. HILL
School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, 2522, N.S.W., Australia
**Author for correspondence (E-mail: gcox@uow.edu.au); [1]Currrent address: School of Science, Griffith University, Nathan 4111, QLD Australia*

Received 19 July 2004; accepted in revised form 20 December 2004

Abstract. Under certain circumstances, an industrial hopper which operates under the "funnel-flow" regime can be converted to the "mass-flow" regime with the addition of a flow-corrective insert. This paper is concerned with calculating granular flow patterns near the outlet of hoppers that incorporate a particular type of insert, the cone-in-cone insert. The flow is considered to be quasi-static, and governed by the Coulomb–Mohr yield condition together with the non-dilatant double-shearing theory. In two-dimensions, the hoppers are wedge-shaped, and as such the formulation for the wedge-in-wedge hopper also includes the case of asymmetrical hoppers. A perturbation approach, valid for high angles of internal friction, is used for both two-dimensional and axially symmetric flows, with analytic results possible for both leading order and correction terms. This perturbation scheme is compared with numerical solutions to the governing equations, and is shown to work very well for angles of internal friction in excess of 45°.

Key words: asymmetric hoppers, Coulomb-Mohr yield condition, granular flow, hopper inserts, perturbation solution

1. Introduction

Granular materials are used extensively throughout the world in most industrial processes, where these materials are stored in hoppers (or silos), which are large containers designed to provide easy retrieval. Disruptions to the flow of the material from a hopper often arise due to phenomena such as arching (see [1] for a thorough review) and rat-holing (see [2–4]). For both phenomena the underlying mechanisms of initiation and formation are not properly understood, and to gain a better understanding of these situations, improved models are necessary. The ability to accurately predict the flow distribution of material from a hopper is of great importance, especially when disruptions to flow reduce the productivity and profitability of the industrial process concerned. Ideally, hoppers should be designed to operate in a prescribed manner, giving reliable flow-rates, which in turn leads to an increase in both the profit and viability of the industrial process.

In general, the desired flow pattern in an industrial hopper is where the entire material in the hopper is simultaneously in motion, commonly referred to as "mass-flow". Conversely, an undesired flow pattern is where flow occurs only in a central region that is surrounded by a stagnant region, and is referred to as "funnel-flow". Parameters that contribute to whether the flow pattern is mass-flow or funnel-flow include the slope and smoothness of the walls and the cohesiveness and the inter-particle friction of the material. As such, studying the flow behaviour for varying parameters should lead to a deeper understanding of hopper flow of granular

materials, enabling improved predictions as to whether a particular hopper with a particular material will give rise to mass-flow or funnel-flow.

One of the methods of converting a funnel-flow hopper into a mass-flow hopper is to use an insert (for experimental studies, see: [5–8]; and for numerical simulations, see: [9,10]). An insert is an apparatus (usually metal), which is fastened at a fixed position in a silo or hopper, in order to influence the flow of granular material in such a way as to prohibit the formation of funnel-flow. The three most common types of inserts are "cone-in-cone", "inverted cone" and "double cone", as shown in Figure 1. Here we are primarily interested in investigating the effect of the cone-in-cone inserts [5,9,10]. As described by Dantoin et al. [11], funnel-flow hoppers can cause serious problems to industrial processes. In [11], a situation is described where due to the build-up of stagnant material in a coal funnel-flow bunker at a power plant, an explosion occurred that caused in excess of $US 4 million in damages and loss production. Further, after the event, another $US 1·2 millon was spent in upgrading the funnel-flow bunker to a mass-flow bunker. Accordingly, any improvement in our understanding of the flow patterns of hoppers with inserts has the potential to provide the particulate industries with a much cheaper option in regard to converting a funnel-flow hopper into a mass-flow hopper.

Gravity flow of granular material from a hopper has been studied extensively throughout the literature, with a review of some developments given in [12], and some more recent studies in [13,14], for example. Early progress for this problem was made in [15–18], where the so-called "radial stress field" solutions are studied for quasi-static flow of incompressible materials which obey the Coulomb–Mohr yield condition. These solutions are similarity solutions, and are valid in the neighbourhood of the hopper outlet. In this case the equilibrium equations and the Coulomb–Mohr yield condition reduce to two highly non-linear coupled ordinary differential equations, which in general can only be solved numerically. Recently, some exact parametric solutions to these equations have been determined by Cox and Hill [19] and Hill and Cox [20] for the limiting case of $\sin\phi = 1$, where ϕ is the angle of internal friction of the material. Furthermore, these exact solutions have been recently used by Thamwattana and Hill [21] as leading order terms in a regular perturbation valid for $1 - \sin\phi \ll 1$. There do exist granular materials that possess angles of internal friction around 60 to 65°, as supported by the experimental data given in [22, p. 23], [23–25]. Such materials give rise to values of $\sin\phi$ around 0·87 to 0·91, and we refer to materials with $1 - \sin\phi \ll 1$ as being "highly frictional". Of course, we must be cautious when studying high-friction granular materials, as the physics involved in taking the limit $\phi \to \pi/2$ is not well-understood (and could be the subject of experimental investigations). However, in the present study, as with [21], we are interested in deriving solutions for $1 - \sin\phi$ small but finite, so the question of the validity of setting $\phi = \pi/2$ is not relevant here.

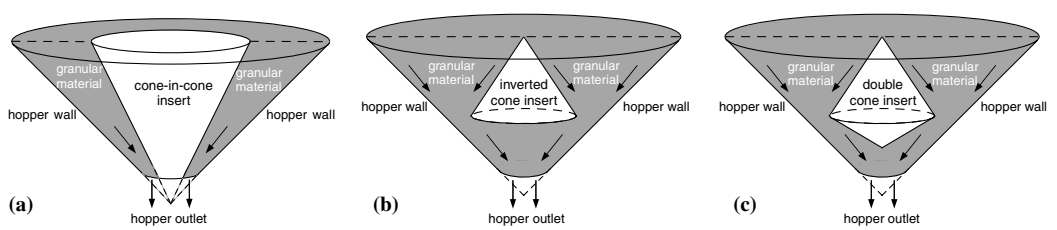


Figure 1. Schematic diagram of conical hopper with a cone-in-cone insert (a), an inverted cone insert (b), and a double-cone insert (c).

The use of the Coulomb–Mohr yield condition to predict stress fields in quasi-static granular flows is quite widely accepted, and produces results that are well in accord with experimental data. However, the correct formulation of the accompanying velocity equations is still controversial. To accompany the radial stress fields described above, [15,17] use the coaxial flow rule to describe the velocity fields. This approach assumes the principal axes of stress and strain-rate coincide. An alternate flow rule comes from the double-shearing theory, originally proposed by Spencer [26,27]. In this theory, every deformation is assumed to consist of simultaneous shears along the two families of stress lines, and (for quasi-static flow) the characteristic curves for the stresses and velocities coincide. For the flow of material near the outlet of a hopper, [28,29] show the coaxial theory used in [15,17] yields physically unacceptable predictions in the velocity field, whereas the double-shearing theory predicts results which are certainly reasonable. Further, [30] compares the double-shearing theory with experimental results for this problem, and the agreement is excellent. Exact velocity fields for the limiting case of $\phi = \pi/2$ have been calculated in [31] using the double-shearing theory. These solutions have yet to be employed as leading-order terms in a perturbation series, and this is one of the goals of the present study.

In this paper, we consider the two-dimensional problems of quasi-static granular flow through asymmetrical wedge hoppers and hoppers with wedge-in-wedge inserts, as well as the axially symmetric problem of flow through conical hoppers with cone-in-cone inserts. In the following section we formulate the governing ordinary differential equations which apply near the outlet of the hoppers, and derive the appropriate boundary conditions in each case. In Sections 3 and 4 we attack the problems by considering a regular perturbation series, following [21], by assuming that the quantity $1-\sin\phi \ll 1$, remembering that ϕ is the angle of internal friction. The first two terms of the expansion are found parametrically, and it is shown that these analytic results provide excellent approximations to the exact numerical solutions for values of ϕ in excess of 45°, particularly for steep hoppers, and the necessary physical condition of the rate of work being non-negative is examined in Section 5. Finally, in Section 6 we make some concluding remarks.

2. Mathematical formulation

In this section we briefly state the governing equations for steady quasi-static granular flow through two-dimensional and axially symmetric hoppers under gravity. In two-dimensions, boundary conditions are given for hoppers whose walls have unequal slopes, a situation which includes the case of a wedge-in-wedge insert. For axially symmetric flows, boundary conditions are given for the case in which there is a cone-in-cone insert.

2.1. Governing equations in two-dimensions

We consider here steady flow of a granular material in two dimensions, and use cylindrical polar coordinates (r,θ) as defined by Figures 2(a) and (b). By assuming the flow is quasi-static, the inertia terms in the momentum equations may be neglected, so that the components of the Cauchy stress tensor $\sigma_{rr}, \sigma_{\theta\theta}$ and $\sigma_{r\theta}$ satisfy the equilibrium equations

$$\frac{\partial\sigma_{rr}}{\partial r}+\frac{1}{r}\frac{\partial\sigma_{r\theta}}{\partial\theta}+\frac{\sigma_{rr}-\sigma_{\theta\theta}}{r}=\rho g\sin\theta, \quad \frac{\partial\sigma_{r\theta}}{\partial r}+\frac{1}{r}\frac{\partial\sigma_{\theta\theta}}{\partial\theta}+\frac{2\sigma_{r\theta}}{r}=\rho g\cos\theta, \tag{2.1}$$

where ρ is the bulk density and g is acceleration due to gravity. In this study the material is taken to be incompressible, so that the density ρ is constant. We note the stress components

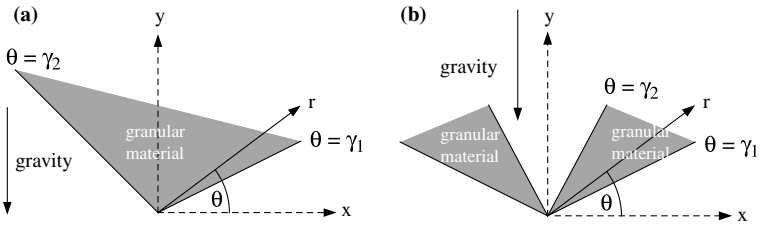


Figure 2. Coordinate system for a two-dimensional asymmetrical wedge hopper (a), and a two-dimensional hopper with a wedge-in-wedge insert (b).

are assumed to be positive in tension, so that a positive force will produce a positive extension.

To close the system of equations, we assume the material yields according to the Coulomb–Mohr yield condition. Thus, for every surface through an arbitrary point within the material, the magnitude of the tangential component of the traction vector τ_n is bounded by the critical value

$$|\tau_n| \leq c - \sigma_n \tan\phi, \tag{2.2}$$

where σ_n is the normal component of tensive traction on the surface. Here, $c \geq 0$ denotes the cohesion of the material, while $0 \leq \phi \leq \pi/2$ is the angle of internal friction, both assumed constant. We note if equality holds in (2.2), then the material yields at that point along the particular surface whose tangential and normal components of traction are given by τ_n and σ_n, respectively.

Now, to utilize the above equations, we introduce the stress angle ψ defined by

$$\tan 2\psi = \frac{2\sigma_{r\theta}}{\sigma_{rr} - \sigma_{\theta\theta}}, \tag{2.3}$$

where physically ψ corresponds to the angle between the direction corresponding to the maximum principal stress and the r-axis, in the direction of increasing θ. Next, upon introducing the generally positive stress invariants p and q defined by

$$p = -\frac{1}{2}(\sigma_I + \sigma_{III}) = -\frac{1}{2}(\sigma_{rr} + \sigma_{\theta\theta}), \quad q = \frac{1}{2}(\sigma_I - \sigma_{III}) = \frac{1}{2}\left\{(\sigma_{rr} - \sigma_{\theta\theta})^2 + 4\sigma_{r\theta}^2\right\}^{1/2}, \tag{2.4}$$

the usual stress decomposition arises

$$\sigma_{rr} = -p + q\cos 2\psi, \quad \sigma_{\theta\theta} = -p - q\cos 2\psi, \quad \sigma_{r\theta} = q\sin 2\psi, \tag{2.5}$$

where σ_I and σ_{III} denote the maximum and minimum principal stresses respectively, and physically speaking, p represents an average pressure, while q is the maximum magnitude of the shear stress. This stress decomposition enables the equilibrium equations to be expressed in terms of an equivalent formulation using p, q and ψ.

To express the Coulomb–Mohr yield condition using these three variables, we note that for an arbitrary surface whose unit normal $\mathbf{n}$ makes an angle of δ with the positive x-axis, the normal tensive component and the magnitude of the tangential component of traction are given by

$$\sigma_n = -p + q\cos 2(\delta - \psi), \quad |\tau_n| = q|\sin 2(\delta - \psi)|. \tag{2.6}$$

As the quantity $|\tau_n| + \sigma_n \tan\phi$ attains its maximum possible value when

$$\delta = \psi \pm \left(\frac{\pi}{4} - \frac{\phi}{2}\right), \tag{2.7}$$

then from (2.2), yield only occurs along surfaces whose normal **n** makes an angle of δ given by (2.7). Further, we denote the normal tensive component and the magnitude of the tangential component of traction of such surfaces by σ and τ, respectively, so that from (2.6) and (2.7) we find

$$\sigma = -p + q\sin\phi, \quad \tau = q\cos\phi. \tag{2.8}$$

As a result, for yielding material, we can express the Coulomb–Mohr yield condition in the more useful form of

$$q = p\sin\phi + c\cos\phi. \tag{2.9}$$

Now, [16, 17] show that a radial stress field solution is a reasonable approximation for the stress distribution near the outlet of a hopper. This involves a radial wedge field solution of the form

$$\psi = \psi(\theta), \quad q = \rho g r F(\theta), \tag{2.10}$$

so from (2.5) and (2.9) we find that the equilibrium Equations (2.1) give rise to two non-linear coupled ordinary differential equations, namely

$$\frac{\mathrm{d}F}{\mathrm{d}\theta} = \frac{F\sin 2\psi - \beta\cos(\theta + 2\psi)}{\beta + \cos 2\psi}, \quad 1 + \frac{\mathrm{d}\psi}{\mathrm{d}\theta} = \frac{F(\beta^{-1} - \beta) + \sin\theta + \beta\sin(\theta + 2\psi)}{2F(\beta + \cos 2\psi)}, \tag{2.11}$$

where here and throughout the paper we use the notation $\beta = \sin\phi$. We can eliminate F from the governing Equations (2.11), and deduce the single second-order ordinary differential equation for $\psi(\theta)$

$$\begin{aligned}(\beta + \cos 2\psi)[\sin\theta + \beta\sin(\theta + 2\psi)]\psi'' = 2(1 + \psi')\{\sin 2\psi[\sin\theta + \beta\sin(\theta + 2\psi)]\psi' + \\ + 2\beta\cos(\theta + 2\psi)(\beta + \cos 2\psi)\psi' + (3\beta^2 + 2\beta\cos 2\psi - 1)\cos(\theta + 2\psi)\},\end{aligned} \tag{2.12}$$

where the primes denote differentiation with respect to θ. With ψ determined, the function F can be recovered from

$$F = \frac{\beta[\sin\theta + \beta\sin(\theta + 2\psi)]}{2\beta(\beta + \cos 2\psi)(1 + \psi') + \beta^2 - 1}. \tag{2.13}$$

As mentioned in the Introduction, the above equations are generally accepted as a reasonable basis for the determination of the plane-strain stress field for gravity flow from a hopper; however, the prescription of the governing equations for the determination of the velocity field is not as readily agreed upon. In this study we assume the velocity profile is governed by the non-dilatant double-shearing theory derived by Spencer [26, 27], which is based on the idea that deformation arises as a result of shear along the surfaces defined by (2.7) on which the critical shear stress is mobilized, and which coincide with the slip-lines in quasi-static flows. In particular, if $v_r(r, \theta)$ and $v_\theta(r, \theta)$ are the components of velocity in the r and θ, directions, respectively, then the double-shearing theory provides that they satisfy the following equations

$$\begin{aligned}&\frac{\partial v_r}{\partial r} + \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r}{r} = 0, \\ &\left(\frac{\partial v_\theta}{\partial r} + \frac{1}{r}\frac{\partial v_r}{\partial \theta} - \frac{v_\theta}{r}\right)\cos 2\psi - \left(\frac{\partial v_r}{\partial r} - \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} - \frac{v_r}{r}\right)\sin 2\psi = \beta\left(\frac{\partial v_\theta}{\partial r} - \frac{1}{r}\frac{\partial v_r}{\partial \theta} - \frac{v_\theta}{r} - 2\Omega\right),\end{aligned} \tag{2.14}$$

where the quantity Ω is defined to be

$$\Omega = v_r \frac{\partial \psi}{\partial r} + \frac{v_\theta}{r} \frac{\partial \psi}{\partial \theta}. \tag{2.15}$$

Following [15], for hopper flow we assume the particular velocity profile

$$v_r(r, \theta) = \frac{u(\theta)}{r}, \quad v_\theta(r, \theta) = 0, \tag{2.16}$$

which ensures that flow can only occur in the radial direction, and $(2.14)_1$ is automatically satisfied. In this case, recalling that $\psi = \psi(\theta)$, then $(2.14)_2$ becomes simply

$$\frac{\mathrm{d}u}{\mathrm{d}\theta}(\beta + \cos 2\psi) = -2u \sin 2\psi, \tag{2.17}$$

which implies that

$$u = \bar{u} \exp\left\{ 2 \int_\theta^{\theta_c} \frac{\sin 2\psi}{\beta + \cos 2\psi} \, \mathrm{d}\theta \right\}, \tag{2.18}$$

for some constants θ_c and $\bar{u}$ such that $\bar{u} = u(\theta_c)$. We note that $\bar{u}$ is the arbitrary constant of integration arising from solving (2.17), while θ_c is a constant that has been introduced to ensure that $u/\bar{u} = 1$ at some specific angle $\theta = \theta_c$ (which is done in order to facilitate the comparison of the perturbation solution with a numerical solution). In particular, we choose $\theta_c = \pi/2$ for the asymmetrical wedge hopper, and $\theta_c = \gamma_2$ for the hopper with a wedge-in-wedge insert.

2.2. Boundary conditions for two-dimensional flow

In this subsection we specify appropriate boundary conditions for the problems of flow in an asymmetrical wedge hopper and flow in a hopper with a wedge-in-wedge insert, as shown in Figure 2(a) and (b). For the latter problem we need only consider half of the hopper, since the flow field is symmetric. The two mathematical problems are essentially the same, with the exception being the physical range of values of γ_1 and γ_2. In particular, for the asymmetrical wedge hopper we have $0 < \gamma_1 < \pi/2 < \gamma_2 < \pi$, while for the hopper with an insert we instead have $0 < \gamma_1 < \gamma_2 < \pi/2$ (with the other half of the flow given by reflection). We note that the problem of an asymmetrical wedge hopper is in some respects similar to the non-axially symmetric hopper flows considered recently in [14].

In all the problems we are considering in this study, it is assumed that the cohesion is zero ($c = 0$), and as the material flows along the sidewalls of the hopper and the sidewalls of the insert, it is reasonable that a Coulomb friction condition should apply along these walls. In this case, upon examining the geometry of the problems as depicted in Figure 2, we find that the two-dimensional Cauchy stress components must satisfy the boundary conditions

$$\sigma_{r\theta} = \sigma_{\theta\theta} \tan \mu_1 \text{ at } \theta = \gamma_1, \quad \sigma_{r\theta} = -\sigma_{\theta\theta} \tan \mu_2 \text{ at } \theta = \gamma_2, \tag{2.19}$$

where μ_1 and μ_2 denote the angles of wall friction of the sidewall of the hopper along $\theta = \gamma_1$ and $\theta = \gamma_2$, respectively. Thus, from (2.5) and (2.9) we find for a cohesionless material that the two-dimensional boundary conditions (2.19) become

$$\sin[\mu_1 + 2\psi(\gamma_1)] = -\frac{\sin \mu_1}{\sin \phi} \text{ at } \theta = \gamma_1, \quad \sin[\mu_2 - 2\psi(\gamma_2)] = -\frac{\sin \mu_2}{\sin \phi} \text{ at } \theta = \gamma_2, \tag{2.20}$$

which are valid provided $\mu_1 < \phi$ and $\mu_2 < \phi$. If the material is such that $\mu_1 \geq \phi$ or $\mu_2 \geq \phi$, then the material will slip on itself at the wall, and here the wall is referred to as being "perfectly rough". We do not consider this case in the present study. We note that if we assume

the cohesion is non-zero, then the Coulomb friction conditions cannot be satisfied (unless $\phi = \pi/2$), as detailed in [32].

In summary, the problem of computing the stress and velocity fields near the outlet of a two-dimensional asymmetrical wedge hopper reduces to solving the non-linear ordinary differential equation (2.12) subject to the boundary conditions (2.20), with $F(\theta)$ and $u(\theta)$ then given by (2.13) and (2.18), respectively. In general, this can only be achieved by solving (2.12) numerically, with a finite-element scheme, for example. However for large values of the angle of internal friction ϕ we seek analytic progress in Section 3 using perturbation analysis.

2.3. Governing equations with axial symmetry

For axially symmetric flows through conical hoppers it is appropriate to use spherical coordinates (R, Θ, Φ), as defined by Figure 3. In this case the four independent components of the stress tensor, denoted by $\sigma_{RR}, \sigma_{\Theta\Theta}, \sigma_{\Phi\Phi}$ and $\sigma_{R\Theta}$, are independent of the variable Φ. We will employ the usual stress decomposition

$$\sigma_{RR} = -P + Q\cos 2\Psi, \quad \sigma_{\Theta\Theta} = -P - Q\cos 2\Psi, \quad \sigma_{R\Theta} = Q\sin 2\Psi, \tag{2.21}$$

where P and Q are the stress invariants given by

$$P = -\frac{1}{2}(\sigma_I + \sigma_{III}) = -\frac{1}{2}(\sigma_{RR} + \sigma_{\Theta\Theta}), \quad Q = \frac{1}{2}(\sigma_I - \sigma_{III}) = \frac{1}{2}\left\{(\sigma_{RR} - \sigma_{\Theta\Theta})^2 + 4\sigma_{R\Theta}^2\right\}^{1/2}, \tag{2.22}$$

with σ_I and σ_{III} denoting the maximum and minimum principal stresses, respectively. Further, Ψ is the stress angle defined by

$$\tan 2\Psi = \frac{2\sigma_{R\Theta}}{\sigma_{RR} - \sigma_{\Theta\Theta}}, \tag{2.23}$$

where physically Ψ corresponds to the angle between the direction corresponding to the maximum principal stress and the R-axis, in the direction of increasing Θ.

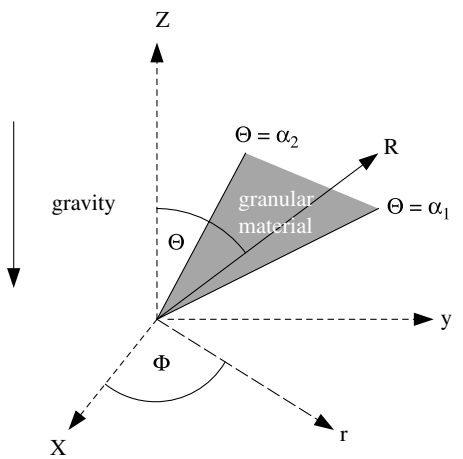


Figure 3. Coordinate system for a three-dimensional hopper with a cone-in-cone insert.

For quasi-static flow, the stress components satisfy the equilibrium equations

$$\begin{aligned}
&\frac{\partial \sigma_{RR}}{\partial R}+\frac{1}{R}\frac{\partial \sigma_{R\Theta}}{\partial \Theta}+\frac{2\sigma_{RR}-\sigma_{\Theta\Theta}-\sigma_{\Phi\Phi}}{R}+\frac{\sigma_{R\Theta}}{R}\cot\Theta=\rho g\cos\Theta,\\
&\frac{\partial \sigma_{R\Theta}}{\partial R}+\frac{1}{R}\frac{\partial \sigma_{\Theta\Theta}}{\partial \Theta}+\frac{\sigma_{\Theta\Theta}-\sigma_{\Phi\Phi}}{R}\cot\Theta+\frac{3\sigma_{R\Theta}}{R}=-\rho g\sin\Theta,
\end{aligned} \tag{2.24}$$

which can be rewritten in terms P, Q, Ψ and $\sigma_{\Phi\Phi}$, but we leave out the details here. In addition, we have the Coulomb–Mohr yield condition (2.2), which is written in terms of P and Q as

$$Q=P\sin\phi+c\cos\phi. \tag{2.25}$$

We therefore have three Equations (2.24) and (2.25) to describe the stress field, with the four unknowns P, Q, Ψ (or σ_{RR}, $\sigma_{\Theta\Theta}$, $\sigma_{R\Theta}$) and $\sigma_{\Phi\Phi}$.

To close the system of equations we need to make an assumption about the hoop stress, in order to determine an expression for $\sigma_{\Phi\Phi}$ in terms of P, Q and Ψ. It has been stated in [33] that the plastic regimes which agree with the Haar-von Kármán hypothesis will give rise to solutions that are most likely to be of the greatest significance to axially symmetric problems of interest. In particular, the heuristic Haar-von Kármán principle states that under an axially symmetric condition the hoop stress is equal to either the maximum or minimum principal stress. This condition gives rise to the idea of Haar-von Kármán regimes, and in particular, either $\sigma_I=\sigma_{\Phi\Phi}=\sigma_{II}>\sigma_{III}$ or $\sigma_I>\sigma_{II}=\sigma_{\Phi\Phi}=\sigma_{III}$, where σ_{II} denotes the intermediate principal stress, and in both cases the hoop stress $\sigma_{\Phi\Phi}$ is assumed to be a principal stress. Here, we choose the former, from which we may deduce

$$\sigma_{\Phi\Phi}=-P+Q. \tag{2.26}$$

We note that this choice differs from the traditional view (see [15,16,28]), in which more physically realistic results for the converging conical hopper problem are believed to be produced by choosing $\sigma_{\Phi\Phi}=\sigma_{III}$, leading to

$$\sigma_{\Phi\Phi}=-P-Q, \tag{2.27}$$

while (2.26) is believed as being more applicable for diverging flow. However, with (2.27) we have been unable to determine an exact parametric solution for the limiting case of $\phi=\pi/2$, and as such, analytic progress with a perturbation scheme for $1-\sin\phi\ll 1$ has not been possible. We therefore adopt (2.26) and we keep in mind that the results obtained may not be as physically applicable as those obtained when $\sigma_{\Phi\Phi}$ is given by (2.27).

Now, following [15–17], we assume a stress field solution of the form

$$\Psi=\Psi(\Theta),\quad Q=\rho g R G(\Theta), \tag{2.28}$$

so from (2.21), (2.25) and (2.26) we find that the equilibrium Equations (2.24) give rise to two non-linear coupled ordinary equations, namely

$$\begin{aligned}
&\frac{\mathrm{d}G}{\mathrm{d}\theta}=\frac{2G\cos\Psi\{\sin\Psi-\beta\,\mathrm{cosec}\,\Theta\cos(\Theta+\Psi)\}+\beta\sin(\Theta+2\Psi)}{\beta+\cos 2\Psi},\\
&1+\frac{\mathrm{d}\Psi}{\mathrm{d}\Theta}=\frac{G(1-\beta)\{\beta^{-1}(1+2\beta)-\mathrm{cosec}\,\Theta\sin(\Theta+2\Psi)\}+\cos\Theta+\beta\cos(\Theta+2\Psi)}{2G(\beta+\cos 2\Psi)},
\end{aligned} \tag{2.29}$$

where again we adopt the notation $\beta=\sin\phi$. We can eliminate G from the governing Equations (2.29), and deduce the single second-order ordinary differential equation for $\Psi(\Theta)$

$$
\begin{aligned}
&2(\beta+\cos 2\Psi)[\cos\Theta+\beta\cos(\Theta+2\Psi)]\Psi'' \\
&\quad=2(1+\Psi')\Big\{\operatorname{cosec}\Theta\Big[\beta(\beta+\cos 2\Psi)[1+\cos\{2(\Theta+\Psi)\}]- \\
&\qquad-[\cos\Theta+\beta\cos(\Theta+2\Psi)][2\sin\Theta\sin 2\Psi+(1-\beta)\cos(\Theta+2\Psi)]+ \\
&\qquad+2(1-\beta)\sin(\Theta+2\Psi)[(1+2\beta)\sin\Theta-\beta\sin(\Theta+2\Psi)]\Big]+ \\
&\qquad+2(1+\Psi')\Big[\sin 2\Psi[\cos\Theta+\beta\cos(\Theta+2\Psi)]-2\beta\sin(\Theta+2\Psi)(\beta+\cos 2\Psi)\Big]\Big\}+ \\
&\qquad+(1-\beta)\operatorname{cosec}^2\Theta\Big\{\sin\{2(\Theta+\Psi)\}[\cos\Theta+\beta\cos(\Theta+2\Psi)]- \\
&\qquad-[1+\cos\{2(\Theta+\Psi)\}][(1+2\beta)\sin\Theta-\beta\sin(\Theta+2\Psi)]\Big\},
\end{aligned}
\tag{2.30}
$$

where prime denotes differentiation with respect to Θ. With Ψ determined, the function G can be recovered from

$$
G=\frac{\beta[\cos\Theta+\beta\cos(\Theta+2\Psi)]}{2\beta(\beta+\cos 2\Psi)(1+\Psi')-(1-\beta)[1+2\beta-\beta\operatorname{cosec}\Theta\sin(\Theta+2\Psi)]}. \tag{2.31}
$$

We note the choice of the Haar-von Kármán regime (2.27) leads to a system of coupled nonlinear ordinary differential equations analogous to (2.29), where the only differences involve sign changes (see [19] for full details). For this case, [34] identify a simple exact solution that provides an envelope for solutions in (Ψ,Θ) space for the problem of converging flow through a conical hopper without an insert. This envelope can be very useful, as it provides a bound for the hopper angle. If the hopper angle is larger than that given by the envelope curve, then the material must be in funnel-flow. Unfortunately, for the choice of (2.26), the corresponding exact solution (see Equation (2.29) in [19]) does not take real values, and hence cannot be used as an envelope curve. We therefore do not have a simple criterion for mass-flow when adopting (2.26).

Now, upon assuming the velocity profile is governed by the non-dilatant double-shearing theory, and if $V_R(R,\Theta)$ and $V_\Theta(R,\Theta)$ are the components of velocity in the R and Θ directions respectively, we obtain

$$
\begin{aligned}
&\frac{\partial V_R}{\partial R}+\frac{1}{R}\frac{\partial V_\Theta}{\partial\Theta}+\frac{2V_R}{R}+\frac{V_\Theta}{R}\cot\Theta=0, \\
&\left(\frac{\partial V_\Theta}{\partial R}+\frac{1}{R}\frac{\partial V_R}{\partial\Theta}-\frac{V_\Theta}{R}\right)\cos 2\Psi-\left(\frac{\partial V_R}{\partial R}-\frac{1}{R}\frac{\partial V_\Theta}{\partial\Theta}-\frac{V_R}{R}\right)\sin 2\Psi=\beta\left(\frac{\partial V_\Theta}{\partial R}-\frac{1}{R}\frac{\partial V_R}{\partial\Theta}-\frac{V_\Theta}{R}-2\Omega\right),
\end{aligned}
\tag{2.32}
$$

where the quantity Ω is defined to be

$$
\Omega=V_R\frac{\partial\Psi}{\partial R}+\frac{V_\Theta}{R}\frac{\partial\Psi}{\partial\Theta}. \tag{2.33}
$$

Following [15], we assume the velocity profile

$$
V_R(R,\Theta)=\frac{U(\Theta)}{R^2},\ V_\Theta(R,\Theta)=0, \tag{2.34}
$$

which ensures that flow can only occur in the radial direction and $(2.32)_1$ is automatically satisfied. In this case, recalling that $\Psi=\Psi(\Theta)$, then $(2.32)_2$ becomes simply

$$
\frac{\mathrm{d}U}{\mathrm{d}\Theta}(\beta+\cos 2\Psi)=-3U\sin 2\Psi, \tag{2.35}
$$

which implies that

$$
U=\bar{U}\exp\left\{-3\int_{\Theta_c}^{\Theta}\frac{\sin 2\Psi}{\beta+\cos 2\Psi}\,\mathrm{d}\Theta\right\}, \tag{2.36}
$$

for some constants Θ_c and $\bar{U}$ such that $\bar{U} = U(\Theta_c)$. We note that $\bar{U}$ is the arbitrary constant of integration arising from solving (2.35), while Θ_c is a known constant that has been introduced to ensure that $U/\bar{U} = 1$ at some specific angle $\Theta = \Theta_c$. In particular, we choose $\Theta_c = \alpha_2$ for the hopper with a cone-in-cone insert.

2.4. Boundary conditions for axially symmetric flow

In this subsection we specify appropriate boundary conditions for the problem of flow in a hopper with a cone-in-cone insert, as shown in Figure 3. We note that the physical range of values of α_1 and α_2 must satisfy $0 < \alpha_2 < \alpha_1 < \pi/2$.

As with the two-dimensional problem, we assume that the cohesion is zero ($c = 0$), and that a Coulomb friction condition applies along the hopper walls. It follows from the problem geometry (see Figure 3) that the axially symmetric Cauchy stress components must satisfy

$$\sigma_{R\Theta} = -\sigma_{\Theta\Theta} \tan \mu_1 \text{ at } \Theta = \alpha_1, \quad \sigma_{R\Theta} = \sigma_{\Theta\Theta} \tan \mu_2 \text{ at } \Theta = \alpha_2, \tag{2.37}$$

where μ_1 and μ_2 denote the angles of wall friction of the sidewall of the hopper along $\Theta = \alpha_1$ and $\Theta = \alpha_2$, respectively. Thus, from (2.21) and (2.25) we find for a cohesionless material that the axially symmetric boundary conditions (2.37) become

$$\sin[\mu_1 - 2\Psi(\alpha_1)] = -\frac{\sin \mu_1}{\sin \phi} \text{ at } \Theta = \alpha_1, \quad \sin[\mu_2 + 2\Psi(\alpha_2)] = -\frac{\sin \mu_2}{\sin \phi} \text{ at } \Theta = \alpha_2, \tag{2.38}$$

which are valid provided $\mu_1 < \phi$ and $\mu_2 < \phi$. As with the two-dimensional case, we mention that, if either $\mu_1 \geq \phi$ or $\mu_2 \geq \phi$, then the material will slip on itself at the appropriate wall, the wall being referred to as "perfectly rough". Since we are primarily concerned with materials with high angles of internal friction, we do not consider this case here.

In summary, the problem for computing the stress and velocity fields near the outlet of an axially symmetric hopper with a cone-in-cone insert reduces to solving the non-linear ordinary differential equation (2.30) subject to the boundary conditions (2.38), with $G(\Theta)$ and $U(\Theta)$ then given by (2.31) and (2.36), respectively. In general, this can only be achieved by solving (2.30) numerically, however, in Section 4 we derive analytic solutions via a perturbation scheme valid when the quantity $1 - \sin \phi \ll 1$.

3. Highly frictional limit in two-dimensions

In this section we seek analytic solutions to the two-dimensional problems formulated in Section 2 by considering the asymptotic limit $\phi \to \pi/2$. For $\phi = \pi/2$ there has been a great deal of success recently in deriving exact solutions for gravity-driven quasi-static flow (see [35], for example), and for the particular case of the similarity solutions described by (2.10) and (2.16), these solutions have been used as leading-order terms in perturbation series for $1 - \sin \phi \ll 1$ in [21]. Here we extend this analysis to hold for asymmetrical wedge hoppers and hoppers with wedge-in-wedge inserts.

3.1. Perturbation analysis

We analyse the governing Equations (2.12), (2.13) and (2.18) by writing out the solutions in the form

$$\psi = \psi_0(\theta) + \epsilon \psi_1(\theta) + O(\epsilon^2), \quad F = F_0(\theta) + \epsilon F_1(\theta) + O(\epsilon^2), \quad u = u_0(\theta) + \epsilon u_1(\theta) + O(\epsilon^2), \tag{3.1}$$

where $\epsilon = 1-\beta = 1-\sin\phi$ and $\epsilon \ll 1$. To leading order we find that ψ_0 satisfies the non-linear ordinary differential equation

$$\psi_0'' = 2(1+\psi_0')\left\{\cot(\theta+\psi_0)(1+\psi_0') - \tan\psi_0\right\}, \tag{3.2}$$

and F_0 and u_0 are given in terms of ψ_0 by

$$F_0 = \frac{\sin(\theta+\psi_0)}{2(1+\psi_0')\cos\psi_0},\ u_0 = \bar{u}\exp\left\{2\int_\theta^{\theta_c}\tan\psi_0\,\mathrm{d}\theta\right\}. \tag{3.3}$$

The leading-order terms ψ_0, F_0 and u_0 are evidently the solutions for the ideal limit $\phi = \pi/2$. These equations were first solved by Hill and Cox [20], and the solution procedure is briefly described here. We make the transformation

$$h(\xi) = \cot(\theta+\psi_0), \quad \xi = \tan\theta, \tag{3.4}$$

and (3.2) becomes

$$(h+\xi)h'' + 2h' = 0, \tag{3.5}$$

where the primes here indicate derivatives with respect to ξ. This equation can be solved by introducing the substitutions $\nu(\xi) = h+\xi$, $\omega(\nu) = 1-\nu'(\xi)$, for some intermediate variable ν, the result being that the solution is given parametrically by

$$h = \cot(\theta+\psi_0) = \frac{I(\omega)}{C_2}, \quad \xi = \tan\theta = \frac{2\omega^{-1/2}\mathrm{e}^{\omega/2} - I(\omega)}{C_2}, \tag{3.6}$$

where ω acts as a parameter, $I(\omega)$ is the integral defined by

$$I(\omega) = \int_0^\omega t^{-1/2}\mathrm{e}^{t/2}\mathrm{d}t + C_1, \tag{3.7}$$

and C_1 and C_2 are arbitrary constants. It follows that the solutions for F_0 and u_0 are given parametrically by

$$F_0 = \frac{1}{4}\frac{\omega^{-1/2}\mathrm{e}^{-\omega/2}[C_2^2 + I^2(\omega)]}{\left\{C_2^2 + [2\omega^{-1/2}\mathrm{e}^{\omega/2} - I(\omega)]^2\right\}^{1/2}}, \quad u_0 = \frac{\bar{u}\,\omega\left\{C_2^2 + [2\omega^{-1/2}\mathrm{e}^{\omega/2} - I(\omega)]^2\right\}}{\omega_c\left\{C_2^2 + [2\omega_c^{-1/2}\mathrm{e}^{\omega_c/2} - I(\omega_c)]^2\right\}}, \tag{3.8}$$

where the parameter value $\omega = \omega_c$ corresponds to $\theta = \theta_c$.

The correction term for the stress angle ψ_1 is found to satisfy the linear equation

$$\begin{aligned}\psi_1'' = {} & 2\psi_1'\left\{2(\psi_0'+1)\cot(\theta+\psi_0) - \tan\psi_0\right\} + \psi_1\{\psi_0''[3\tan\psi_0 - \cot(\theta+\psi_0)] - \\ & -2(\psi_0'+1)^2[1+3\tan\psi_0\cot(\theta+\psi_0)] + 2(\psi_0'+1)[2\tan^2\psi_0 - \tan\psi_0\cot(\theta+\psi_0) - 1]\} + \\ & +(\psi_0'+1)^2[\tan\psi_0\mathrm{cosec}^2(\theta+\psi_0) + \sec^2\psi_0(\tan\psi_0 - \cot(\theta+\psi_0)] - (\psi_0'+1)\sec^2\psi_0\cot(\theta+\psi_0),\end{aligned} \tag{3.9}$$

where ψ_0 is given by the leading-order solution (3.6). The method for solving this ordinary differential equation is presented in Appendix A, along with the analysis for determining F_1 and u_1. The result is that the correction terms are given by

$$\psi_1 = \frac{1}{C_2^2 + I^2(\omega)} \left\{ C_4(1-\omega) + \tfrac{1}{4}[2\omega^{1/2}\mathrm{e}^{\omega/2} + (1-\omega)I(\omega)] \left[\int_0^\omega (1-t)K(t)\mathrm{d}t + C_3 \right] - \right.$$
$$\left. -\tfrac{1}{4}(1-\omega) \int_0^\omega [2t^{1/2}\mathrm{e}^{t/2} + (1-t)I(t)]K(t)\mathrm{d}t \right\},$$
$$F_1 = F_0 \frac{\omega^{1/2}\mathrm{e}^{-\omega/2}}{8C_2^2}[C_2^2 + I^2(\omega)] \left\{ 8C_2 \frac{\mathrm{d}\psi_1}{\mathrm{d}\omega} + 4C_2\psi_1 + (1+\omega)\omega^{-1/2}\mathrm{e}^{-\omega/2}[C_2^2 + I^2(\omega)] - \right. \quad (3.10)$$
$$\left. -6I(\omega) + \frac{8\omega^{-1/2}\mathrm{e}^{\omega/2}I^2(\omega)}{C_2^2 + I^2(\omega)} \right\},$$
$$u_1 = \frac{u_0}{8C_2^2} \left\{ \mathrm{e}^{-\omega_c}[C_2^2 + I^2(\omega_c)]^2 - \mathrm{e}^{-\omega}[C_2^2 + I^2(\omega)]^2 \right.$$
$$\left. +2\int_{\omega_c}^\omega t^{-1/2}\mathrm{e}^{-t/2}[C_2^2 + I^2(t)][2C_2\psi_1 + I(t)]\mathrm{d}t \right\},$$

where C_3 and C_4 are constants of integration, F_0 given by $(3.8)_1$, u_0 given by $(3.8)_2$, and the function K given by the expression

$$K(\omega) = -\frac{1}{2C_2}\omega^{-1/2}\mathrm{e}^{-\omega/2}I(\omega)[C_2^2 + I^2(\omega)] - \frac{2}{C_2}\omega^{-1/2}\mathrm{e}^{\omega/2}[2\omega^{-1/2}\mathrm{e}^{\omega/2} - I(\omega)] +$$
$$+\frac{1}{4C_2}\omega\mathrm{e}^{-\omega}[C_2^2 + (2\omega^{-1/2}\mathrm{e}^{\omega/2} - I(\omega))^2]^2. \quad (3.11)$$

3.2. Application of boundary conditions

By substituting the perturbation series $(3.1)_1$ in the boundary conditions (2.20), we find that

$$\psi_0 = -\mu_1, \quad \psi_1 = -\tfrac{1}{2}\tan\mu_1 \text{ on } \theta = \gamma_1,$$
$$\psi_0 = \mu_2, \quad \psi_1 = \tfrac{1}{2}\tan\mu_2 \text{ on } \theta = \gamma_2. \quad (3.12)$$

In this subsection we use these conditions to determine the values of the constants C_1 and C_2 in the leading-order solution ψ_0 given by (3.6)–(3.7), and the constants C_3 and C_4 in the correction term ψ_1 in $(3.10)_1$. Throughout this section we will use the integral

$$J(\omega) = \int_0^\omega t^{-1/2}\mathrm{e}^{t/2}\ \mathrm{d}t, \quad (3.13)$$

so that from (3.7) we have $I(\omega) = J(\omega) + C_1$.

3.2.1. *Wedge-in-wedge insert*

For the case of a wedge-in-wedge insert (see Figure 2) we need only consider the domain $\gamma_1 \le \theta \le \gamma_2$, since the flow is symmetric about $\theta = \pi/2$. We suppose that the parameter ω in (3.6) takes the values $\omega = \omega_1$ and $\omega = \omega_2$ at $\theta = \gamma_1$ and $\theta = \gamma_2$, respectively. By applying the boundary conditions $(3.12)_1$ and $(3.12)_3$, we find the constants C_1 and C_2 are given in terms of these parameter values by

$$C_1 = C_2\cot(\gamma_2 + \mu_2) - J(\omega_2),\ C_2 = \frac{2\omega_1^{-1/2}\mathrm{e}^{\omega_1/2} - J(\omega_1) + J(\omega_2)}{\tan\gamma_1 + \cot(\gamma_2 + \mu_2)},$$

while the values of ω_1 and ω_2 are determined by the pair of transcendental equations

$$\tfrac{1}{2}\omega_1^{1/2}\mathrm{e}^{-\omega_1/2}[J(\omega_1)-J(\omega_2)]=\frac{\cot(\gamma_1-\mu_1)-\cot(\gamma_2+\mu_2)}{\tan\gamma_1+\cot(\gamma_1-\mu_1)},$$

$$\left(\frac{\omega_2}{\omega_1}\right)^{1/2}\mathrm{e}^{(\omega_1-\omega_2)/2}=\frac{\tan\gamma_1+\cot(\gamma_1-\mu_1)}{\tan\gamma_2+\cot(\gamma_2+\mu_2)}.$$

The other two boundary conditions $(3.12)_2$ and $(3.12)_4$ provide the following equations for C_3 and C_4

$$C_3=\frac{1}{mM(\omega_2)-M(\omega_1)}\left\{2(m\tan\mu_2[C_2^2+I^2(\omega_2)]+\tan\mu_1[C_2^2+I^2(\omega_1)])+\right.$$
$$\left.+(1-\omega_1)\int_{\omega_1}^{\omega_2}M(t)K(t)\mathrm{d}t-mM(\omega_2)\int_0^{\omega_2}(1-t)K(t)\mathrm{d}t+M(\omega_1)\int_0^{\omega_1}(1-t)K(t)\mathrm{d}t\right\},$$
$$C_4=\frac{1}{mM(\omega_2)-M(\omega_1)}\left\{-\frac{1}{2(1-\omega_2)}(\tan\mu_1M(\omega_2)[C_2^2+I^2(\omega_1)]+\tan\mu_2M(\omega_1)[C_2^2+I^2(\omega_2)])-\right.$$
$$\left.-\frac{M(\omega_1)M(\omega_2)}{4(1-\omega_2)}\int_{\omega_2}^{\omega_1}(1-t)K(t)\mathrm{d}t+\frac{1}{4}\left[mM(\omega_2)\int_0^{\omega_1}M(t)K(t)\mathrm{d}t-M(\omega_1)\int_0^{\omega_2}M(t)K(t)\mathrm{d}t\right]\right\},$$

where $M(\omega)$ and m are defined by

$$M(\omega)=2\omega^{1/2}\mathrm{e}^{\omega/2}+(1-\omega)I(\omega),\qquad m=\frac{1-\omega_1}{1-\omega_2},$$

respectively.

3.2.2. *Asymmetrical wedge hopper*

For the case of an asymmetrical wedge hopper, the application of the boundary conditions is not as straightforward. Here we must split up the domain into two parts, namely $\gamma_1\le\theta\le\pi/2$ and $\pi/2\le\theta\le\gamma_2$, and treat each region separately. The solutions in each part are then matched by ensuring that the stress angle ψ and its first derivative ψ' are continuous at $\theta=\pi/2$.

For the first region, $\gamma_1\le\theta\le\pi/2$, we suppose the relevant solutions are given by (3.6)–$(3.8)_2$ and $(3.10)_1$–(3.11), with $\omega=0$ corresponding to $\theta=\pi/2$ and $\omega=\omega_1$ corresponding to $\theta=\gamma_1$. In this case, from (3.6) and $(3.12)_1$, we find

$$C_2\tan\gamma_1=2\omega_1^{-1/2}\mathrm{e}^{\omega_1/2}-J(\omega_1)-C_1,\qquad C_2\cot(\gamma_1-\mu_1)=J(\omega_1)+C_1,\tag{3.14}$$

which upon solving for C_1 and C_2, gives

$$C_1=\frac{2\omega_1^{-1/2}\mathrm{e}^{\omega_1/2}\cot(\gamma_1-\mu_1)}{\tan\gamma_1+\cot(\gamma_1-\mu_1)}-J(\omega_1),\qquad C_2=\frac{2\omega_1^{-1/2}\mathrm{e}^{\omega_1/2}}{\tan\gamma_1+\cot(\gamma_1-\mu_1)}.\tag{3.15}$$

The other condition on $\theta=\gamma_1$ $(3.12)_2$ provides the equation

$$-\tfrac{1}{2}\tan\mu_1[C_2^2+I^2(\omega_1)]=C_4(1-\omega_1)-\tfrac{1}{4}(1-\omega_1)\int_0^{\omega_1}[2t^{1/2}\mathrm{e}^{t/2}+(1-t)I(t)]K(t)\mathrm{d}t$$
$$+\tfrac{1}{4}[2\omega_1^{1/2}\mathrm{e}^{\omega_1/2}+(1-\omega_1)I(\omega_1)]\left[\int_0^{\omega_1}(1-t)K(t)\mathrm{d}t+C_3\right],\tag{3.16}$$

which we will come back to.

For the other region in the hopper $\pi/2\le\theta\le\gamma_2$, we suppose that the general solutions are also given by (3.6)–$(3.8)_2$ and $(3.10)_1$–(3.11), except now we assign different labels to the

parameter and the constants of integration. Instead of the parameter ω we shall use s, and instead of the constants C_1, C_2, C_3, C_4 we shall use D_1, D_2, D_3, D_4, respectively. So for $\pi/2 \leq \theta \leq \gamma_2$ by $I(s)$ we mean $J(s) + D_1$, where $J(s)$ is defined in (3.13). For reference, these solutions are listed by Equations (B.1)–(B.4) in Appendix B. Note that in this region the parameter s takes the value $s = 0$ at $\theta = \pi/2$. We denote the value at $\theta = \gamma_2$ to be $s = s_2$. From (B.1) and $(3.12)_3$, we find

$$D_2 \tan\gamma_2 = 2s_2^{-1/2} e^{s_2/2} - J(s_2) - D_1, \quad D_2 \cot(\gamma_2 + \mu_2) = J(s_2) + D_1, \tag{3.17}$$

which upon solving for D_1 and D_2, gives

$$D_1 = \frac{2s_2^{-1/2} e^{s_2/2} \cot(\gamma_2 + \mu_2)}{\tan\gamma_2 + \cot(\gamma_2 + \mu_2)} - J(s_2), \; D_2 = \frac{2s_2^{-1/2} e^{s_2/2}}{\tan\gamma_2 + \cot(\gamma_2 + \mu_2)}. \tag{3.18}$$

Thus, once ω_1 and s_2 are known, (3.15) and (3.18) constitute four expressions for the arbitrary constants C_1, C_2, D_1 and D_2. The condition $(3.12)_4$ on $\theta = \gamma_2$ becomes

$$\begin{aligned} \tfrac{1}{2}\tan\mu_2[D_2^2 + I^2(s_2)] = D_4(1 - s_2) - \tfrac{1}{4}(1 - s_2)\int_0^{s_2} [2t^{1/2}e^{t/2} + (1-t)I(t)]K(t)\mathrm{d}t + \\ + \tfrac{1}{4}[2s_2^{1/2}e^{s_2/2} + (1 - s_2)I(s_2)]\left[\int_0^{s_2}(1-t)K(t)\mathrm{d}t + D_3\right]. \end{aligned} \tag{3.19}$$

Next, we need to ensure that $\psi(\theta)$ and $\psi'(\theta)$ remain continuous throughout the entire domain. Clearly, these quantities are continuous within the two regions $\gamma_1 \leq \theta < \pi/2$ and $\pi/2 < \theta \leq \gamma_2$, and as such, we need only examine the boundary between the two the solutions at $\theta = \pi/2$. From (3.6) and (B.1), we find at $\theta = \pi/2$ ($\omega = s = 0$) that

$$\cot[\pi/2 + \psi_0(\pi/2)] = \frac{C_1}{C_2} = \frac{D_1}{D_2}, \tag{3.20}$$

so it must be that

$$C_1 = \frac{C_2 D_1}{D_2}. \tag{3.21}$$

Further, we want $\mathrm{d}\psi_0/\mathrm{d}\theta$ to also be continuous at $\theta = \pi/2$. To do this, we first note that

$$\frac{\mathrm{d}\psi_0}{\mathrm{d}\theta} + 1 = \frac{4e^{\omega} + C_2^2\omega - 4\omega^{1/2}e^{\omega/2}(J(\omega) + C_1) + \omega(J(\omega) + C_1)^2}{C_2^2 \mathrm{cosec}^2(\theta_0 + \psi_0)}, \quad \gamma_1 \leq \theta \leq \pi/2,$$

$$\frac{\mathrm{d}\psi_0}{\mathrm{d}\theta} + 1 = \frac{4e^{s} + D_2^2 s - 4s^{1/2}e^{s/2}(J(s) + D_1) + s(J(s) + D_1)^2}{D_2^2 \mathrm{cosec}^2(\theta + \psi_0)}, \quad \pi/2 \leq \theta \leq \gamma_2,$$

so that at $\theta = \pi/2$ ($\omega = s = 0$) we find

$$\left.\frac{\mathrm{d}\psi_0}{\mathrm{d}\theta}\right|_{\theta=\pi/2} + 1 = \frac{4}{C_2^2 \mathrm{cosec}^2[\pi/2 + \psi_0(\pi/2)]} = \frac{4}{D_2^2 \mathrm{cosec}^2[\pi/2 + \psi_0(\pi/2)]}. \tag{3.22}$$

Clearly, for $\mathrm{d}\psi_0/\mathrm{d}\theta$ to be continuous, from (3.22) we require $C_2^2 = D_2^2$, so that

$$C_2 = -D_2, \tag{3.23}$$

and consequently, from (3.21) we find

$$C_1 = -D_1. \tag{3.24}$$

Similarly, by taking the limits $\omega \to 0$ and $s \to 0$ in $(3.10)_1$ and $(B.3)_1$ we find

$$\psi_1(\pi/2) = \frac{4C_4 + C_1C_3}{4(C_1^2 + C_2^2)} = \frac{4D_4 + D_1D_3}{4(D_1^2 + D_2^2)},$$

$$\left.\frac{\mathrm{d}\psi_1}{\mathrm{d}\theta}\right|_{\theta=\pi/2} = -\frac{2(C_3 - 4C_1C_4 - C_3C_1^2)}{C_2(C_1^2 + C_2^2)} = -\frac{2(D_3 - 4D_1D_4 - D_3D_1^2)}{D_2(D_1^2 + D_2^2)},$$

so that by making use of (3.23) and (3.24), we have

$$C_3 = -D_3, \quad C_4 = D_4. \tag{3.25}$$

Hence, from $(3.15)_2$, $(3.18)_2$ and (3.23), we get

$$\omega_1^{-1/2}\mathrm{e}^{\omega_1/2} = -\left[\frac{\tan\gamma_1 + \cot(\gamma_1 - \mu_1)}{\tan\gamma_2 + \cot(\gamma_2 + \mu_2)}\right] s_2^{-1/2}\mathrm{e}^{s_2/2}, \tag{3.26}$$

and from $(3.15)_1$, $(3.18)_1$ and (3.24), we obtain

$$J(\omega_1) + J(s_2) = 2\left[\frac{\cot(\gamma_2 + \mu_2) - \cot(\gamma_1 - \mu_1)}{\tan\gamma_2 + \cot(\gamma_2 + \mu_2)}\right] s_2^{-1/2}\mathrm{e}^{s_2/2}, \tag{3.27}$$

and as such, (3.26) and (3.27) constitute as two transcendental equations for ω_1 and s_2. Furthermore, together with (3.25), the conditions (3.16) and (3.19) provide simultaneous equations for C_3, C_4, D_3, D_4.

3.3. Results

In this subsection, we compare the analytic results obtained from the perturbation scheme (3.1) with the numerical results for the problem of two-dimensional flow from a hopper with a wedge-in-wedge insert, as depicted in Figure 2(b). In what follows, we choose $\gamma_1 = \pi/3$ and $\gamma_2 = 5\pi/12$, noting from [5] that the angle $\pi/2 - \gamma_1$ is generally chosen to be twice the angle $\pi/2 - \gamma_2$. Furthermore, we assume the typical values $\mu_1 = \mu_2 = \pi/12$, so that both the insert and the hopper are assumed to be made of the same material.

Figures 4, 5 and 6 show, respectively, plots of ψ, F and $u/\bar{u}$ versus the angle $\pi/2 - \theta$ for the four values of the angle of internal friction $\phi = 7\pi/18$, $\pi/3$, $\pi/4$ and $\pi/6$. In each case the solid curves correspond to the leading-order terms (3.6) and (3.8) (which are also the solutions for the limiting case $\phi = \pi/2$), while the dotted curves correspond to the first two terms in the perturbation expansion (3.1), with the correction terms given by (3.10). Finally, the dashed curves denote the full numerical solution of the governing Equations (2.12), (2.13) and (2.17). Here, the numerical solution to the second-order ordinary differential Equation (2.12) is solved using a non-linear finite-difference scheme, as described in [36, page 601], while u is found by applying a fourth-order Runge–Kutta scheme, as given is [36, page 259]. From Figure 6 we see that the material is flowing the fastest near the middle of the material, while it slows down towards both the insert and the hopper walls, due to the material being "held up" on the walls. Further, we note that due to gravity the material is moving faster at the insert wall than the hopper wall.

From these three figures, it is clear that the perturbation scheme performs extremely well for high angles of internal friction such as $\phi = 7\pi/18$ and $\phi = \pi/3$, which is to be expected. However, somewhat surprisingly, the scheme still provides an excellent approximation for the moderately high value $\phi = \pi/4$, and even provides a reasonable estimate for $\phi = \pi/6$, which is by no means a high angle of internal friction. We comment that qualitatively similar results

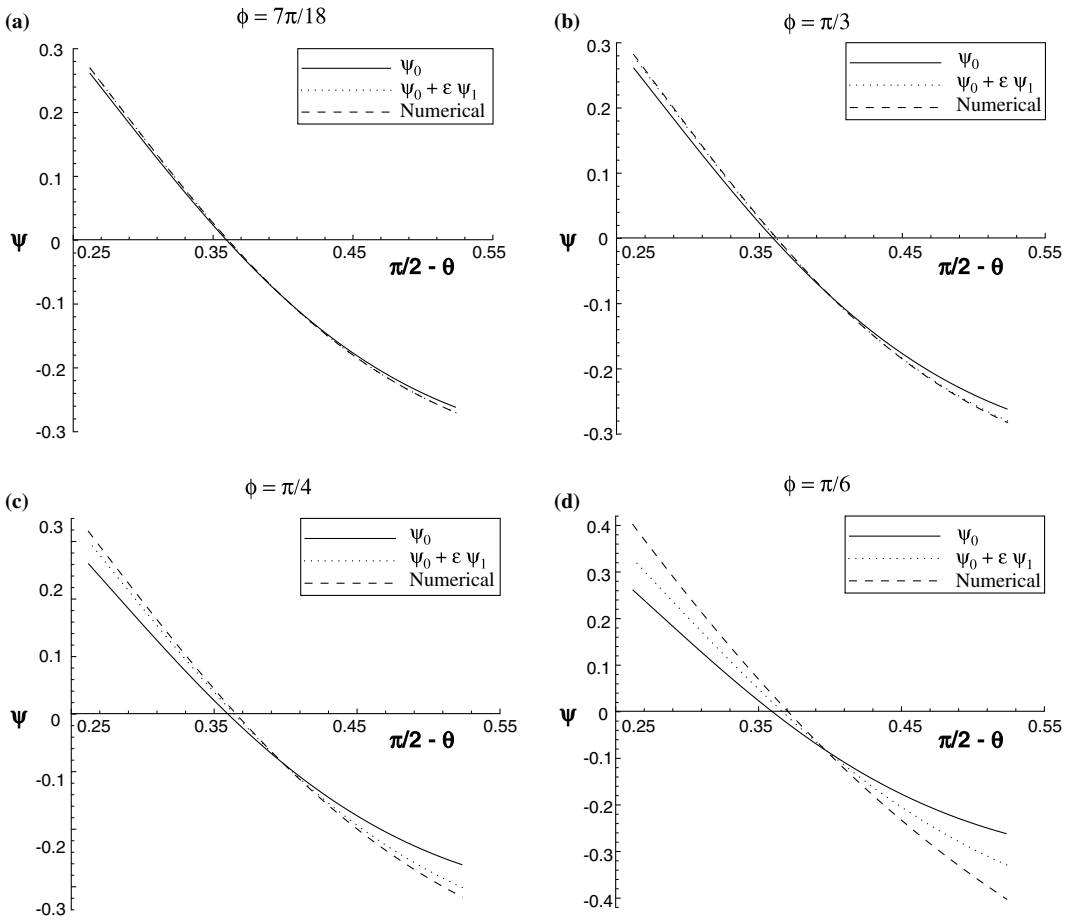


Figure 4. Comparison of the numerical, the zeroth-order and the full perturbation solutions for $\psi(\pi/2-\theta)$ for $\phi=7\pi/18,\ \pi/3,\ \pi/4$ and $\pi/6$.

are found for different sets of α_1, α_2, μ_1 and μ_2, with the general trend being that the perturbation scheme works better for steeper hoppers.

For the sake of brevity we do not provide figures for the two-dimensional asymmetrical wedge hopper. The results are, not surprisingly, qualitatively similar to those found for the symmetric hopper, and we refer the reader to [21].

4. Highly frictional limit with axial symmetry

This section is dedicated to analysing the axially symmetric problem formulated in Sections 2.3 and 2.4 under the asymptotic limit $\phi \to \pi/2$. It is the analog of Section 3. We extend the analysis of Thamwattana and Hill [21] to hold for hoppers with cone-in-cone inserts, and we also determine the associated velocity profile according to the non-dilatant double-shearing theory.

4.1. Perturbation analysis

We seek solutions to (2.30), (2.31) and (2.36) of the form

$$\Psi = \Psi_0(\Theta) + \epsilon\Psi_1(\Theta) + O(\epsilon^2),\ G = G_0(\Theta) + \epsilon G_1(\Theta) + O(\epsilon^2),\ U = U_0(\Theta) + \epsilon U_1(\Theta) + O(\epsilon^2), \tag{4.1}$$

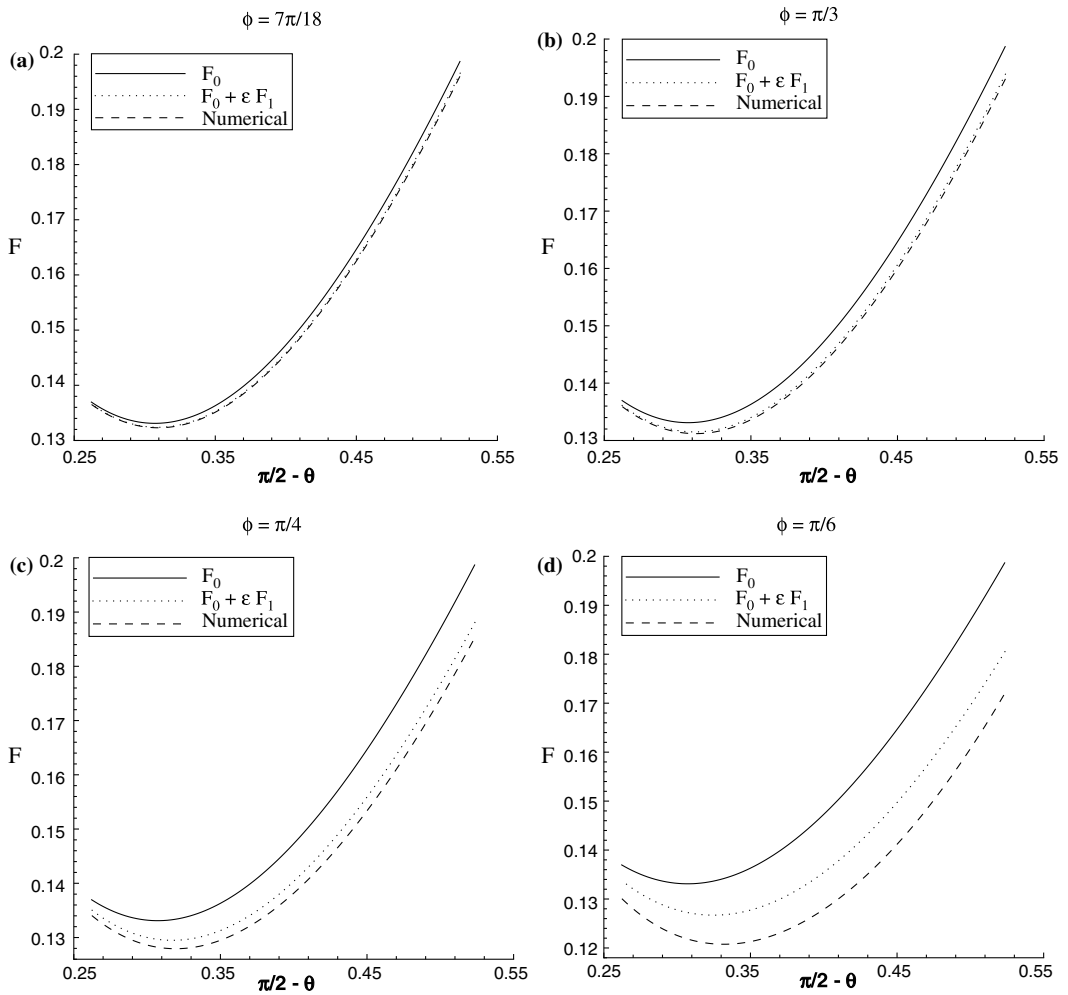


Figure 5. Comparison of the numerical, the zeroth-order and the full perturbation solutions for $F(\pi/2-\theta)$ for $\phi=7\pi/18,\ \pi/3,\ \pi/4$ and $\pi/6$.

where $\epsilon=1-\beta=1-\sin\phi$ and $\epsilon\ll1$. To leading order, Ψ_0 satisfies the equation

$$\Psi_0''=(1+\Psi_0')\{\cot\Theta-3\tan\Psi_0-2(1+\Psi_0')\tan(\Theta+\Psi_0)\}, \tag{4.2}$$

while G_0 and U_0 are given by

$$G_0=\frac{\cos(\Theta+\Psi_0)}{2(1+\Psi_0')\cos\Psi_0},\qquad U_0=\bar{U}\exp\left\{-3\int_{\Theta_c}^{\Theta}\tan\Psi_0\mathrm{d}\Theta\right\}. \tag{4.3}$$

The leading-order terms Ψ_0, F_0 and U_0 represent solutions for the ideal limit $\phi=\pi/2$, with (4.2) and (4.3) first solved by Cox and Hill [19]. The solution procedure is analogous to that described earlier for (3.2) and (3.3), and produces the parametric result

$$H=\tan(\Theta+\Psi_0)=\frac{I(\omega)}{C_2},\ \xi=\cot\Theta=\frac{3\omega^{-1/3}\mathrm{e}^{\omega/3}-I(\omega)}{C_2}, \tag{4.4}$$

where ω is the parameter, $I(\omega)$ the integral defined by

$$I(\omega)=\int_0^{\omega}t^{-1/3}\mathrm{e}^{t/3}\ \mathrm{d}t+C_1, \tag{4.5}$$

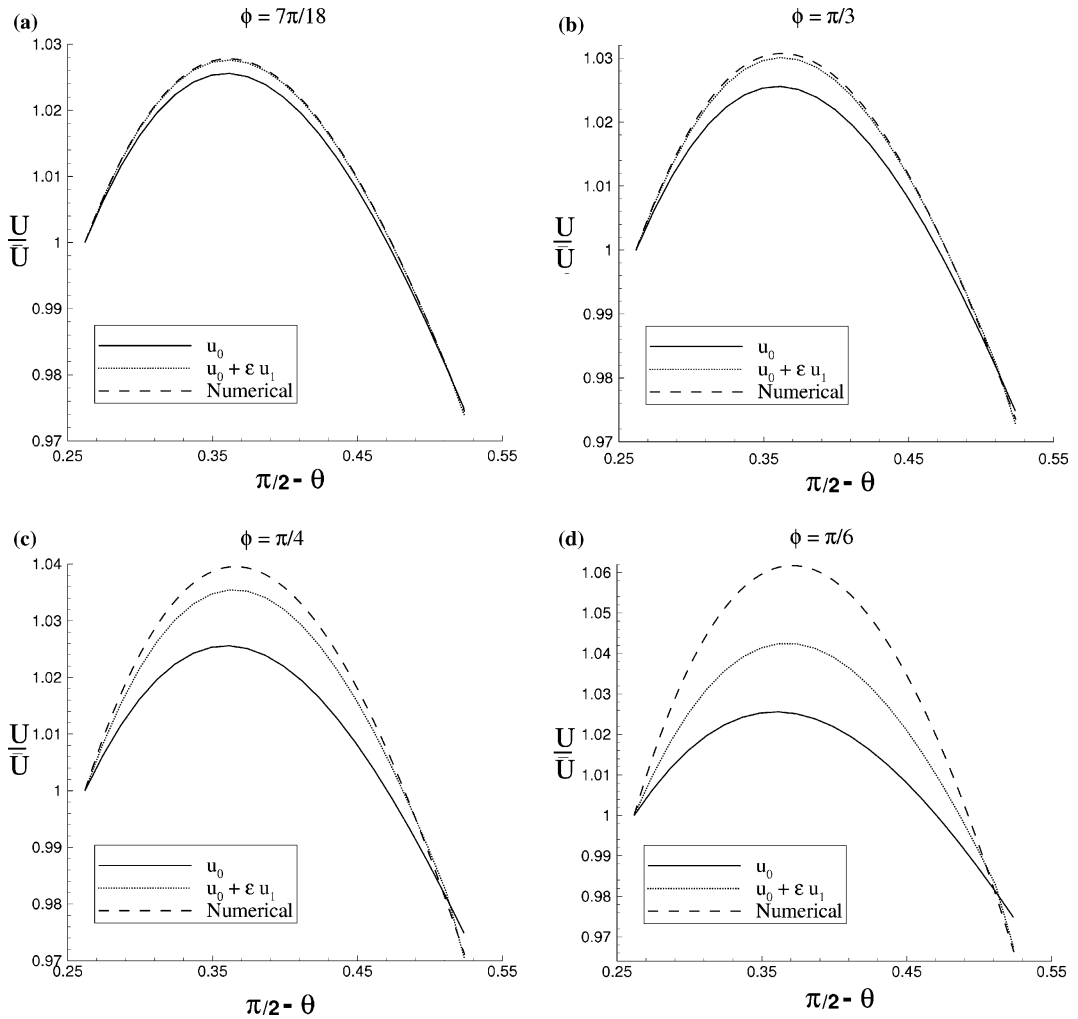


Figure 6. Comparison of the numerical, the zeroth-order and the full perturbation solutions for $u(\pi/2-\theta)/\bar{u}$ for $\phi=7\pi/18,\ \pi/3,\ \pi/4$ and $\pi/6$.

and C_1 and C_2 constants of integration. Furthermore, G_0 and U_0 are given parametrically by

$$G_0=\frac{1}{6}\frac{\omega^{-2/3}\mathrm{e}^{-\omega/3}[C_2^2+I^2(\omega)]}{\{C_2^2+[3\omega^{-1/3}\mathrm{e}^{\omega/3}-I(\omega)]^2\}^{1/2}},\qquad U_0=\frac{\bar{U}\omega\{C_2^2+[3\omega^{-1/3}\mathrm{e}^{\omega/3}-I(\omega)]^2\}^{3/2}}{\omega_c\{C_2^2+[3\omega_c^{-1/3}\mathrm{e}^{\omega_c/3}-I(\omega_c)]^2\}^{3/2}},\tag{4.6}$$

where the parameter value $\omega=\omega_c$ corresponds to $\Theta=\Theta_c$.

The correction term for the stress angle Ψ_1 is found to satisfy the linear equation

$$\begin{aligned}2\Psi_1''=&2\Psi_1'\{-4(1+\Psi_0')\tan(\Theta+\Psi_0)+\cot\theta-3\tan\Psi_0\}+2\Psi_1\{\Psi_0''[\tan(\Theta+\Psi_0)+3\tan\Psi_0]\\&2(1+\Psi_0')[(1+\Psi_0')(3\cos\Theta\sec\Psi_0\sec(\Theta+\Psi_0)-4)+\tan(\Theta+\Psi_0)(2\tan\Psi_0-\cot\Theta)+\\&+3\tan^2\Psi_0^2-\cot\Theta\tan\Psi_0-1]\}+\Psi_0''[2-\tan\Psi_0\tan(\Theta+\Psi_0)+\tan^2\Psi_0]+\\&+2(1+\Psi_0')^2[\tan(\Theta+\Psi_0)(3+2\tan^2\Psi_0)+\tan\Psi_0(2+\tan^2\Psi_0)]+\\&+(1+\Psi_0')[-2(2+\tan^2\Psi_0)(\cot\Theta-\tan\Psi_0)-(\tan\Psi_0+\tan(\Theta+\Psi_0))(1+\cot\Theta\tan\Psi_0)+\\&+\tan\Psi_0(5+3\tan^2\Psi_0)+\tan(\Theta+\Psi_0)(3+\tan^2\Psi_0)]+2\tan\Psi_0(\cot\Theta-\tan\Psi_0)^2,\end{aligned}\tag{4.7}$$

where Ψ_0 is given by the leading-order solution (4.4). The method for solving this ordinary differential equation is presented in Appendix C, along with the analysis for determining G_1 and U_1. The result is that the correction terms are given by

$$\Psi_1 = \frac{1}{C_2^2 + I^2(\omega)} \left\{ C_4(1-\omega) + \frac{1}{9}[3\omega^{2/3}\mathrm{e}^{\omega/3} + (1-\omega)I(\omega)] \left[\int_0^{\omega} (1-t)K(t)\mathrm{d}t + C_3 \right] \right.$$
$$\left. -\frac{1}{9}(1-\omega)\int_0^{\omega} [3t^{2/3}\mathrm{e}^{t/3} + (1-t)I(t)]K(t)\mathrm{d}t \right\},$$
$$G_1 = -\frac{\omega^{-1/3}\mathrm{e}^{-2\omega/3}G_0}{18C_2^2} \left\{ 18C_2\omega^{2/3}\mathrm{e}^{\omega/3}[C_2^2 + I^2(\omega)] \left(\frac{\mathrm{d}\Psi_1}{\mathrm{d}\omega} + \frac{\Psi_1}{3} \right) - [C_2^2 + I^2(\omega)]^2 \right.$$
$$\left. -[3\omega^{-1/3}\mathrm{e}^{\omega/3}I(\omega) - C_2^2 - I^2(\omega)][6\omega^{2/3}\mathrm{e}^{\omega/3}I(\omega) - (1+\omega)[C_2^2 + I^2(\omega)]] \right\},$$
$$U_1 = -\frac{U_0}{6C_2^2}\int_{\omega_c}^{\omega} t^{-2/3}\mathrm{e}^{-t/3}[C_2^2 + I^2(t)][2C_2\Psi_1 + I(t)]\mathrm{d}t$$
$$+\frac{U_0}{18C_2^2}\int_{\omega_c}^{\omega} t^{-1/3}\mathrm{e}^{-2t/3}[C_2^2 + I^2(t)]^2\mathrm{d}t, \tag{4.8}$$

where C_3 and C_4 are constants of integration, G_0 given by $(4.6)_1$, U_0 given by $(4.6)_2$, and the function K given by the expression

$$K(\omega) = \frac{C_2}{\omega} + \frac{[2I(\omega) - 3\omega^{-1/3}\mathrm{e}^{\omega/3}]^2}{C_2\omega} + \frac{3}{C_2}\omega^{-1/3}\mathrm{e}^{\omega/3}[3\omega^{-1/3}\mathrm{e}^{\omega/3} - I(\omega)] +$$
$$+\frac{1}{9C_2}\omega^{-4/3}\mathrm{e}^{-2\omega/3}[C_2^2 + I^2(\omega)][3\omega^{-1/3}\mathrm{e}^{\omega/3}I(\omega) - C_2^2 - I^2(\omega)] -$$
$$-\frac{1}{9C_2}\omega^{-1/3}\mathrm{e}^{-2\omega/3}(1+\omega)[C_2^2 + (3\omega^{-1/3}\mathrm{e}^{\omega/3} - I(\omega))^2]^2. \tag{4.9}$$

4.2. Application of boundary conditions

From $(4.1)_1$ and (2.38) we see that the appropriate boundary conditions are

$$\begin{aligned} &\Psi_0 = \mu_1, \quad \Psi_1 = \tfrac{1}{2}\tan\mu_1, \qquad \text{on} \quad \Theta = \alpha_1, \\ &\Psi_0 = -\mu_2, \ \Psi_1 = -\tfrac{1}{2}\tan\mu_2, \quad \text{on} \quad \Theta = \alpha_2, \end{aligned} \tag{4.10}$$

which are used to determine the values of the constants C_1 and C_2 in the leading-order solution Ψ_0 given by (4.4)–(4.5), and the constants C_3 and C_4 in the correction term Ψ_1 in $(4.8)_1$. The analysis for this procedure is given in Appendix D.

4.3. Results

In order to examine the accuracy and applicability of the full perturbation solution given by (4.1) of an axially symmetric hopper with a cone-in-cone insert, as depicted in Figure 3, we choose $\alpha_1 = \pi/6$ and $\alpha_2 = \pi/12$ ($\alpha_1 = 30^\circ$ and $\alpha_2 = 15^\circ$), noting from [5] that the angle that the sidewall of the hopper makes with the z-axis is generally assumed to be twice the angle that the insert makes with the z-axis. Further, we assume $\mu_1 = \mu_2 = \pi/12$, so that the surface on both sidewalls possess the same level of friction, and for four values of the angle of internal friction, namely $\phi = 7\pi/18,\ \pi/3,\ \pi/4$ and $\pi/6$, Figures 7, 8 and 9, respectively, show the comparison of $\Psi(\Theta)$, $G(\Theta)$ and $U(\Theta)/\bar{U}$ as determined from a full numerical solution of the governing Equations (2.30), (2.31) and (2.35), the zeroth-order approximation given by (4.4) and (4.6), and the full perturbation solution given by (4.1) and (4.8). We note that ψ is a

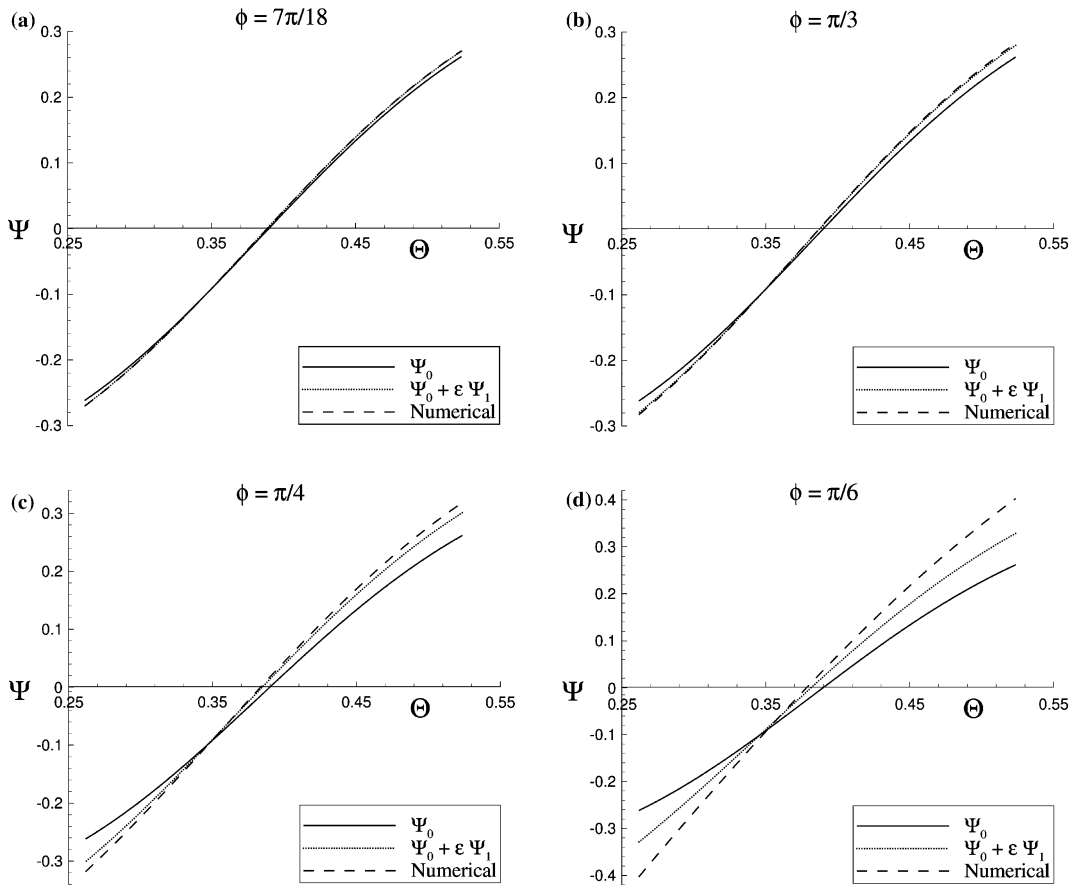


Figure 7. Comparison of the numerical, the zeroth-order and the full perturbation solutions for $\Psi(\Theta)$ for $\phi = 7\pi/18$, $\pi/3$, $\pi/4$ and $\pi/6$.

decreasing function of $\pi/2-\theta$ in Figure 4, while Ψ is an increasing function Θ in Figure 7. This behaviour arises since ψ is measured from the r-axis in the direction of increasing θ (not $\pi/2-\theta$), while Ψ is measured from the R-axis in the direction of increasing Θ. From Figure 9 we again find that near the middle of the material the magnitude of the velocity is greatest, while due to the material being "held up" on the walls, the velocity decreases towards both the insert and the hopper walls. Finally, due to gravity the material is again moving faster at the insert wall than the hopper wall.

From the figures, we see that the full perturbation solution gives an excellent estimate to the full numerical solution for values of angle of internal friction as low as $\phi = \pi/4$, while still providing a reasonable estimate for an angle of internal friction equal to $\phi = \pi/6$. For all cases considered, the full perturbation solution gives a much improved estimate than the zeroth-order solution, and we note that the steeper the sidewall of the hopper, the better the estimate from the full perturbation solution.

5. Rate of work

In this section we check the necessary physical condition of the rate of work being non-negative for both the two-dimensional and axially symmetric perturbation solutions presented in Sections 3 and 4, respectively.

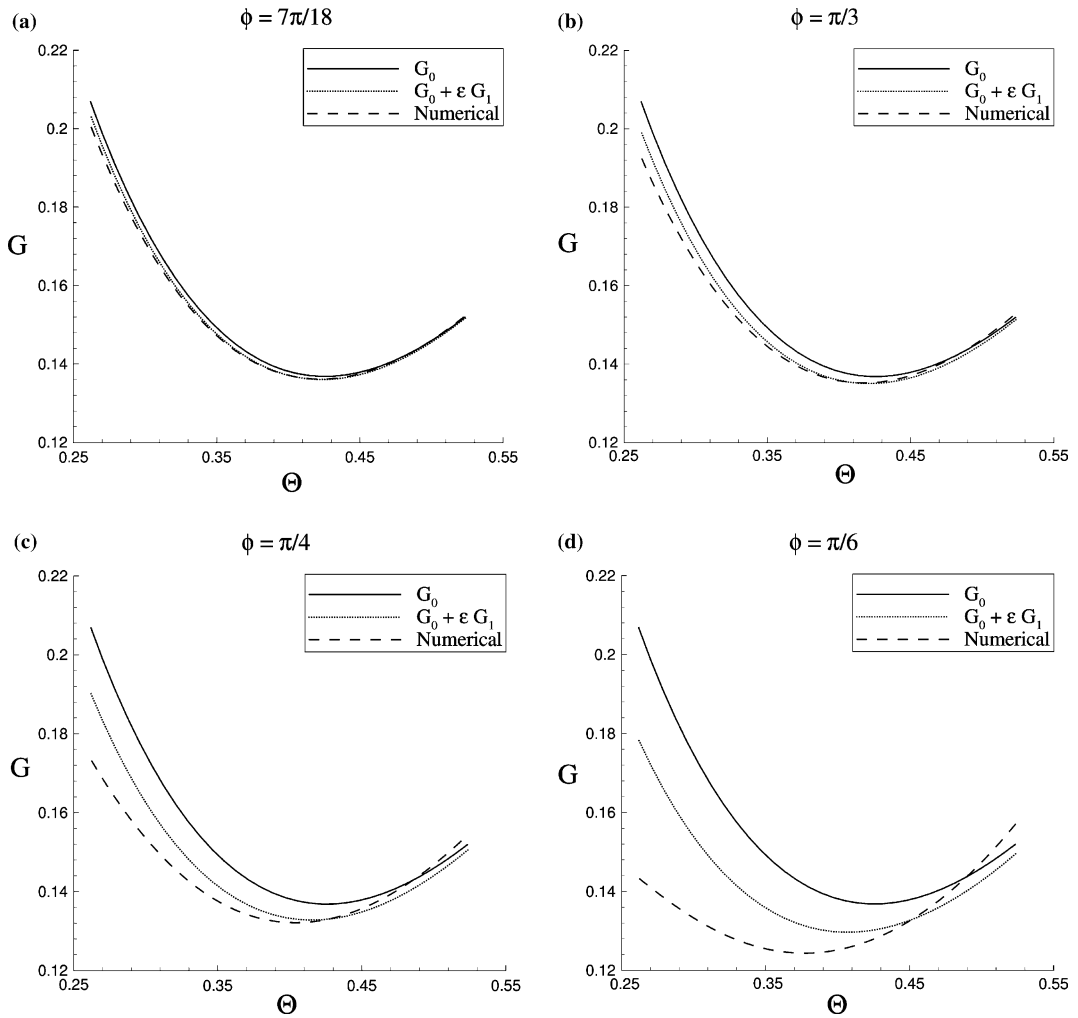


Figure 8. Comparison of the numerical, the zeroth-order and the full perturbation solutions for $G(\Theta)$ for $\phi = 7\pi/18,\ \pi/3,\ \pi/4$ and $\pi/6$.

5.1. Rate of work for the two-dimensional problem

To examine the rate of work for steady quasi-static plane strain gravity flow through a two-dimensional hopper with a wedge-in-wedge insert, we find from [37, page 482] that for the solution to be physical, the rate of plastic work is required to remain non-negative, resulting in the inequality

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \mathrm{tr}(\boldsymbol{\sigma d}) \geq 0, \tag{5.11}$$

where $\mathrm{d}W/\mathrm{d}t$ denotes the rate of plastic work, tr denotes the usual trace of a tensor, $\boldsymbol{\sigma}$ is the Cauchy stress tensor and $\boldsymbol{d}$ is the strain-rate tensor. In terms of cylindrical polar coordinates (r, θ) for two-dimensional plane strain flow, as defined by Figure 2(b), the non-zero components of $\boldsymbol{\sigma}$ are σ_{rr}, $\sigma_{r\theta}$, $\sigma_{\theta\theta}$ and σ_{zz}, while the non-zero components of $\boldsymbol{d}$ are given by

$$d_{rr} = \frac{\partial v_r}{\partial r}, \quad d_{r\theta} = \frac{1}{2}\left[\frac{1}{r}\frac{\partial v_r}{\partial \theta} + \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r}\right], \quad d_{\theta\theta} = \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r}{r}. \tag{5.12}$$

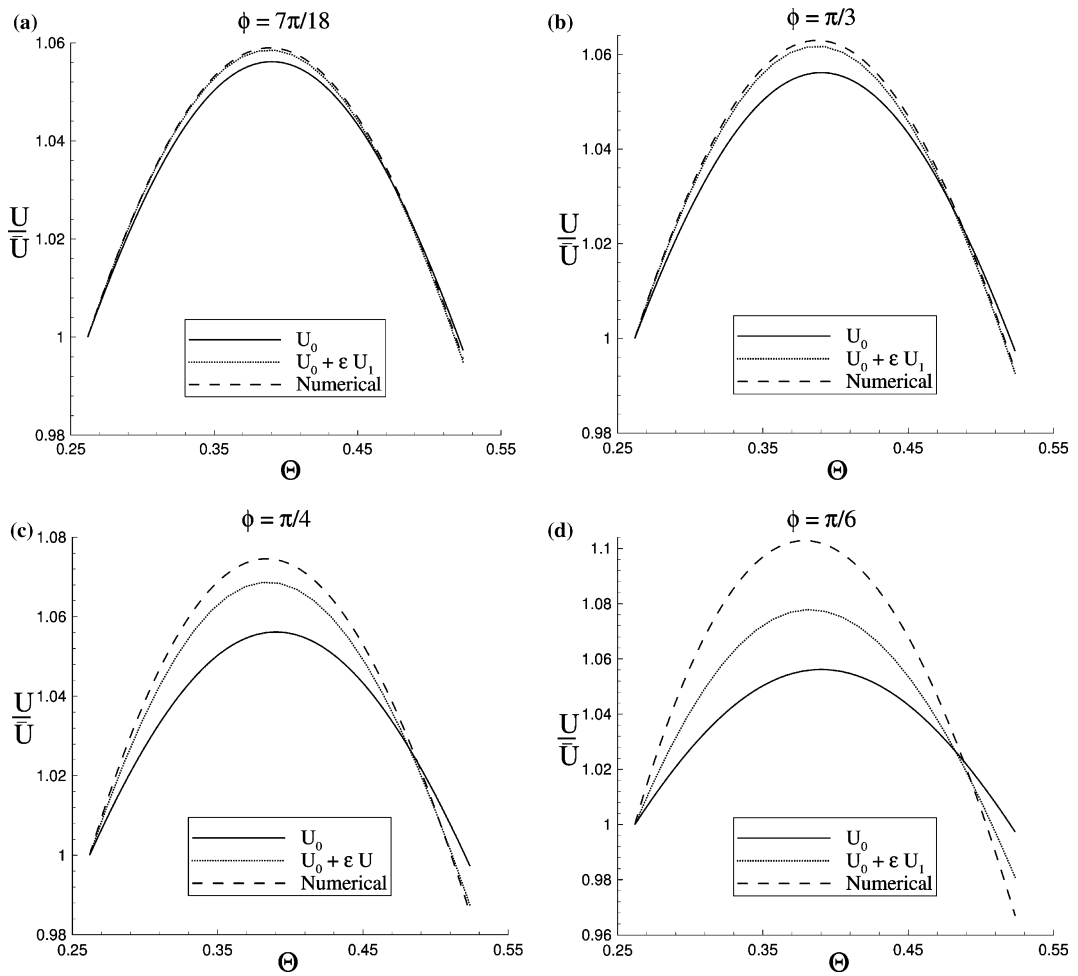


Figure 9. Comparison of the numerical, the zeroth-order and the full perturbation solutions for $U(\Theta)/\bar{U}$ for $\phi = 7\pi/18,\ \pi/3,\ \pi/4$ and $\pi/6$.

Thus, from (2.5), (5.11)–(5.12), we require

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}t} &= \sigma_{rr}\frac{\partial v_r}{\partial r} + \sigma_{r\theta}\left[\frac{1}{r}\frac{\partial v_r}{\partial \theta} + \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r}\right] + \sigma_{\theta\theta}\left[\frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r}{r}\right] \\ &= -p\left[\frac{\partial v_r}{\partial r} + \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r}{r}\right] \\ &\quad + q\left\{\left[\frac{\partial v_r}{\partial r} - \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} - \frac{v_r}{r}\right]\cos 2\psi + \left[\frac{\partial v_\theta}{\partial r} + \frac{1}{r}\frac{\partial v_r}{\partial \theta} - \frac{v_\theta}{r}\right]\sin 2\psi\right\} \geq 0, \end{aligned} \tag{5.13}$$

so that from $(2.14)_1$ and (2.16), the inequality (5.13) becomes

$$\sin 2\psi \frac{\mathrm{d}u}{\mathrm{d}\theta} - 2u\cos 2\psi \geq 0, \tag{5.14}$$

noting that $q \geq 0$. Further, from (2.17), (5.14) simplifies to give

$$\frac{1 + \sin\phi\cos 2\psi}{\sin\phi + \cos 2\psi} \geq 0,$$

since $u \leq 0$ (the flow of material is towards the apex of the hopper). We can immediately see that for the extreme case $\phi = \pi/2$ the rate of working is always positive. Given we have restricted ourselves to the range $-\pi \leq \psi \leq \pi$, for $\phi < \pi/2$ it follows that we must have

$$|\psi| < \frac{\pi}{4} + \frac{1}{2}\phi \tag{5.15}$$

for the rate of energy dissipation to remain non-negative.

For the two-dimensional problem considered in this paper, it is straightforward to test the rate of working condition (5.15). All that is required is to check the plots of ψ versus θ, like those presented in Figure 4. In fact, given the boundary conditions (2.20), we know that

$$\psi(\gamma_1) > -\left(\frac{\pi}{4} + \frac{1}{2}\phi\right), \quad \psi(\gamma_2) < \frac{\pi}{4} + \frac{1}{2}\phi,$$

so a sufficient condition for (5.15) to hold is that there are no local minima or maxima over the range $\gamma_1 < \theta < \gamma_2$. We have checked that the condition (5.15) holds true for a variety of parameter values, as required.

5.2. Rate of work for axially symmetric problem

For axially symmetric flow in terms of spherical polar coordinates (R, Θ, Φ) as defined by Figure 3, the non-zero components of $\boldsymbol{\sigma}$ are denoted by σ_{RR}, $\sigma_{R\Theta}$, $\sigma_{\Theta\Theta}$ and $\sigma_{\Phi\Phi}$ while the non-zero components of the rate-of-deformation tensor $\boldsymbol{d}$ are given by

$$d_{RR} = \frac{\partial V_R}{\partial R}, \quad d_{R\Theta} = \frac{1}{2}\left[\frac{1}{R}\frac{\partial V_R}{\partial \Theta} + \frac{\partial V_\Theta}{\partial R} - \frac{V_\Theta}{R}\right], \quad d_{\Theta\Theta} = \frac{1}{R}\frac{\partial V_\Theta}{\partial \Theta} + \frac{V_R}{R}, \quad d_{\Phi\Phi} = \frac{V_\Theta}{R}\cot\Theta + \frac{V_R}{R}.$$

Thus, from (2.21), (2.26) and (5.11), we require

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}t} &= \sigma_{RR}\frac{\partial V_R}{\partial R} + \sigma_{R\Theta}\left[\frac{1}{R}\frac{\partial V_R}{\partial \Theta} + \frac{\partial V_\Theta}{\partial R} - \frac{V_\Theta}{R}\right] + \sigma_{\Theta\Theta}\left[\frac{1}{R}\frac{\partial V_\Theta}{\partial \Theta} + \frac{V_R}{R}\right] + \sigma_{\Phi\Phi}\left[\frac{V_\Theta}{R}\cot\Theta + \frac{V_R}{R}\right] \\ &= -P\left\{\frac{\partial V_R}{\partial R} + \frac{1}{R}\frac{\partial V_\Theta}{\partial \Theta} + \frac{2V_R}{R} + \frac{V_\Theta}{R}\cot\Theta\right\} + Q\left\{\left[\frac{\partial V_R}{\partial R} - \frac{1}{R}\frac{\partial V_\Theta}{\partial \Theta} - \frac{V_R}{R}\right]\cos 2\Psi + \right. \\ &\quad \left. + \left[\frac{1}{R}\frac{\partial V_R}{\partial \Theta} + \frac{\partial V_\Theta}{\partial R} - \frac{V_\Theta}{R}\right]\sin 2\Psi + \frac{V_\Theta}{R}\cot\Theta + \frac{V_R}{R}\right\} \geq 0,\end{aligned}$$

which, after substituting $(2.32)_1$ and (2.34), becomes

$$\sin 2\Psi \frac{\mathrm{d}U}{\mathrm{d}\Theta} + U(1 - 3\cos 2\Psi) \geq 0, \tag{5.16}$$

noting that $Q \geq 0$. Further, from (2.35), (5.16) simplifies to give the rate of work inequality

$$\frac{3 - \sin\phi + (3\sin\phi - 1)\cos 2\Psi}{\sin\phi + \cos 2\Psi} \geq 0. \tag{5.17}$$

We again note that for the special case of $\sin\phi = 1$, the rate of work inequality (5.17) is always satisfied. For all other values of ϕ, we must have

$$|\Psi| < \frac{\pi}{4} + \frac{1}{2}\phi. \tag{5.18}$$

This condition has been checked for a variety of parameter values for the problem considered in Section 4, and was found to hold true in each case. We conclude that the choice of hoop stress (2.26) is not inconsistent with the important requirement (5.11).

We mention that if the alternate condition (2.27) for the hoop stress was chosen, then (5.17) would be replaced with the similar condition

$$\frac{3+\sin\phi+(3\sin\phi+1)\cos 2\Psi}{\sin\phi+\cos 2\Psi}\geq 0.$$

However, this inequality also implies (5.18), so the necessary condition for non-negative energy dissipation would remain unchanged.

6. Conclusions

We have considered the problem of gravity flow of granular materials both in symmetrical hoppers with inserts and asymmetrical hoppers. These inserts are primarily used to influence the flow of the material to ensure that mass-flow occurs, rather than funnel-flow. For flow near the outlet of the hopper, we have applied the radial field stress similarity solutions, which give rise to highly non-linear coupled ordinary differential equations, while the flow fields have been determined using the non-dilatant double-shearing theory. We have analytically calculated the first two terms of a perturbation scheme valid for $1-\sin\phi \ll 1$, for both the stress and velocity fields, which are then applied to the problems considered. The perturbation results are in excellent agreement with a full numerical solution for large values of ϕ, as expected, and also gives reasonable predictions for values of ϕ as low as 45°.

We note here that the dynamic equations for Spencer's double-shearing theory are known to be linearly ill-posed (as are many other plasticity-type theories for granular flow), in the sense that small perturbations to existing solutions may grow exponentially in time (see [38], for example). This property puts some doubt on whether the steady solutions examined in this study actually describe real granular flows, although, as mentioned in the Introduction, there is experimental evidence (see [30], for example) that suggests the double-shearing theory is indeed appropriate for flows in the neighbourhood of a hopper outlet. This topic is still a matter for debate, and we have not pursued it here.

One of the goals of this study was to ascertain whether or not the perturbation solutions for highly frictional granular materials could be utilized for asymmetrical hoppers and symmetrical hoppers with inserts, in order to calculate the granular flow patterns near the outlet of the hopper. This has been achieved, and with the use of the non-dilatant double-shearing theory we have confirmed that the solutions presented satisfy the physically necessary requirement that the rate of working be non-negative. At present it is not precisely clear under what conditions an insert will convert a funnel-flow hopper into a mass-flow hopper (a simple idea involves the use of an envelope curve as discussed in Subsection 2.3) page 12, but with any future developments the results presented here may prove useful.

Appendix A. Correction terms in two-dimensions

Here we consider the ordinary differential Equation (3.9). With use of the change of variables (3.4), we can transform this equation into

$$(1+h^2)(h+\xi)\frac{\mathrm{d}^2\psi_1}{\mathrm{d}\xi^2}+2[1+h^2+2h(h+\xi)h']\frac{\mathrm{d}\psi_1}{\mathrm{d}\xi}+2\left[(h+\xi)h'^2-\frac{1+h^2}{h+\xi}h'\right]\psi_1$$
$$=h'^2\left[\frac{(1+\xi^2)^2}{(h+\xi)^2}-\xi(h+\xi)\right]+\frac{h(1+h^2)}{h+\xi}h',$$

which, after substituting the parametric solutions for h and ξ (3.6), can be transformed into

$$2\omega[C_2^2+I^2(\omega)]\frac{\mathrm{d}^2\psi_1}{\mathrm{d}\omega^2}+\left\{8\omega^{1/2}\mathrm{e}^{\omega/2}I(\omega)+(1-\omega)[C_2^2+I^2(\omega)]\right\}\frac{\mathrm{d}\psi_1}{\mathrm{d}\omega}+\{4\mathrm{e}^{\omega}+I^2(\omega)+C_2^2\}\psi_1$$
$$=-\frac{1}{2C_2}I(\omega)[C_2^2+I^2(\omega)]-\frac{2}{C_2}\mathrm{e}^{\omega}[2\omega^{-1/2}\mathrm{e}^{\omega/2}-I(\omega)]+\frac{1}{4C_2}\omega^{3/2}\mathrm{e}^{-\omega/2}[C_2^2+(2\omega^{-1/2}\mathrm{e}^{\omega/2}-I(\omega))^2]^2,$$

where ω is the parameter. Progress can be made by noting the left-hand side of this equation is simply

$$2\omega\frac{\mathrm{d}^2\Phi}{\mathrm{d}\omega^2}+(1-\omega)\frac{\mathrm{d}\Phi}{\mathrm{d}\omega}+\Phi,$$

where the variable Φ is defined by $\Phi=[C_2^2+I^2(\omega)]\psi_1$, so that we can integrate twice to find

$$\Phi=C_4(1-\omega)+\tfrac{1}{4}[2\omega^{1/2}\mathrm{e}^{\omega/2}+(1-\omega)I(\omega)]\left[\int_0^{\omega}(1-t)K(t)\mathrm{d}t+C_3\right]-$$
$$-\tfrac{1}{4}(1-\omega)\int_0^{\omega}[2t^{1/2}\mathrm{e}^{t/2}+(1-t)I(t)]K(t)\mathrm{d}t,$$

where C_3 and C_4 are constants and $K(t)$ is given by (3.11).

Now, by substituting the perturbation expansions $(3.1)_1$ and $(3.1)_2$ in (2.13), we find that

$$F_1=\frac{F_0}{2(1+\psi_0')}\{-2\psi_1'+2\psi_1(1+\psi_0')[\tan\psi_0+\cot(\theta+\psi_0)]+$$
$$+(1+\psi_0')\tan\psi_0[\tan\psi_0-\cot(\theta+\psi_0)]+\sec^2\psi_0\},$$

where F_0 is given by $(3.3)_1$. After making the transformations in (3.4), we arrive at the expression

$$F_1=\frac{F_0}{2(h+\xi)^2}\left\{2\frac{(1+h^2)(h+\xi)^2}{h'}\frac{\mathrm{d}\psi_1}{\mathrm{d}\xi}+2(1+h^2)(h+\xi)\psi_1+\right.$$
$$\left.+(1-h\xi)(1-2h\xi-h^2)-\frac{(1+h^2)^2}{h'}\right\},$$

which can be simplified further to $(3.10)_2$ with use of the parametric solutions (3.6).

Finally, from (2.18), $(3.1)_1$, $(3.1)_3$, (3.4) and (3.6) we find that u_1 is

$$u_1=u_0\int_{\theta}^{\theta_c}(2\psi_1+\tan\psi_0)\sec^2\psi_0\mathrm{d}\theta$$
$$=\frac{u_0}{8C_2^2}\left\{\mathrm{e}^{-\omega_c}[C_2^2+I^2(\omega_c)]^2-\mathrm{e}^{-\omega}[C_2^2+I^2(\omega)]^2+\right.$$
$$\left.+2\int_{\omega_c}^{\omega}t^{-1/2}\mathrm{e}^{-t/2}[C_2^2+I^2(t)][2C_2\psi_1+I(t)]\mathrm{d}t\right\},$$

where u_0 is the leading-order solution $(3.8)_2$, and ψ_1 is the correction term $(3.10)_1$. This expression may be simplified to give $(3.10)_3$.

Appendix B. General solution for asymmetrical wedge hopper

For an asymmetrical wedge hopper, the general solutions for the leading-order and correction terms in the region $\pi/2\le\theta\le\gamma_2$ are given here. The leading-order solution for ψ_0 is

$$\cot(\theta+\psi_0)=\frac{I(\omega)}{D_2},\quad \tan\theta=\frac{2s^{-1/2}\mathrm{e}^{s/2}-I(s)}{D_2},\tag{B.1}$$

where s acts as a parameter, $I(s) = J(s) + D_1$, and D_1 and D_2 are arbitrary constants. The solutions for F_0 and u_0 are given parametrically by

$$F_0 = \frac{1}{4}\frac{s^{-1/2}\mathrm{e}^{-s/2}[D_2^2 + I^2(s)]}{\left\{D_2^2 + [2s^{-1/2}\mathrm{e}^{s/2} - I(s)]^2\right\}^{1/2}}, \qquad u_0 = \frac{\bar{u}\, s\left\{D_2^2 + [2s^{-1/2}\mathrm{e}^{s/2} - I(s)]^2\right\}}{s_c\left\{D_2^2 + [2s_c^{-1/2}\mathrm{e}^{s_c/2} - I(s_c)]^2\right\}}, \tag{B.2}$$

where the parameter value $s = s_c$ corresponds to $\theta = \theta_c$. The correction terms are given by

$$\begin{aligned}
\psi_1 &= \frac{1}{D_2^2 + I^2(s)}\left\{D_4(1-s) + \frac{1}{4}[2s^{1/2}\mathrm{e}^{s/2} + (1-s)I(s)]\left[\int_0^s (1-t)K(t)\mathrm{d}t + D_3\right] - \right. \\
&\qquad \left. -\frac{1}{4}(1-s)\int_0^s [2t^{1/2}\mathrm{e}^{t/2} + (1-t)I(t)]K(t)\mathrm{d}t\right\}, \\
F_1 &= F_0\frac{s^{1/2}\mathrm{e}^{-s/2}}{8D_2^2}[D_2^2 + I^2(s)]\left\{8D_2\frac{\mathrm{d}\psi_1}{\mathrm{d}s} + 4D_2\psi_1 + (1+s)s^{-1/2}\mathrm{e}^{-s/2}[D_2^2 + I^2(s)]\right. \\
&\qquad \left. -6s^{-1/2}\mathrm{e}^{s/2}I(s) + \frac{8s^{-1}\mathrm{e}^s I^2(s)}{D_2^2 + I^2(s)}\right\}, \\
u_1 &= \frac{u_0}{8D_2^2}\left\{\mathrm{e}^{-s_c}[D_2^2 + I^2(s_c)]^2 - \mathrm{e}^{-s}[D_2^2 + I^2(s)]^2 + 2\int_{s_c}^s t^{-1/2}\mathrm{e}^{-t/2}[D_2^2 + I^2(t)][2D_2\psi_1 + I(t)]\mathrm{d}t\right\},
\end{aligned} \tag{B.3}$$

where D_3 and D_4 are constants of integration, F_0 given by $(\mathrm{B}.2)_1$, u_0 given by $(\mathrm{B}.2)_2$, and the function K given by the expression

$$\begin{aligned}
K(s) = &-\frac{1}{2D_2}s^{-1/2}\mathrm{e}^{-s/2}I(s)[D_2^2 + I^2(s)]^2 - \frac{2}{D_2}s^{-1/2}\mathrm{e}^{s/2}[2s^{-1/2}\mathrm{e}^{s/2} - I(s)] \\
&+ \frac{1}{4D_2}s\mathrm{e}^{-s}[D_2^2 + (2s^{-1/2}\mathrm{e}^{s/2} - I(s))^2]^2.
\end{aligned} \tag{B.4}$$

These expressions are equivalent to (3.6)–$(3.8)_2$ and $(3.10)_1$–(3.11), except we have assigned different labels to the parameter and the constants of integration.

Appendix C. Correction terms in axial symmetry

Here we consider the ordinary differential equation (4.7). With use of the change of variables (4.4), we can transform this equation into

$$\begin{aligned}
&(1+H^2)(\xi+H)\frac{\mathrm{d}^2\Psi_1}{\mathrm{d}\xi^2} + [4H(\xi+H)H' + 3(1+H^2)]\frac{\mathrm{d}\Psi_1}{\mathrm{d}\xi} + \left[2(\xi+H)H'^2 - \frac{3(1+H^2)}{\xi+H}H'\right]\Psi_1 \\
&\quad = \left[\xi(\xi+H) - \frac{(1+\xi^2)^2}{(\xi+H)^2}\right]H'^2 - \left[1 + (H-\xi)^2 - \frac{(1+\xi^2)^2}{(\xi+H)^2}\right]H' + \frac{(1+H^2)(H\xi - 1)}{(\xi+H)^2},
\end{aligned}$$

which, after substituting the parametric solutions for H and ξ (4.4), can be transformed into

$$\begin{aligned}
&3\omega[C_2^2 + I^2(\omega)]\frac{\mathrm{d}^2\Psi_1}{\mathrm{d}\omega^2} + \{12\omega^{2/3}\mathrm{e}^{\omega/3}I(\omega) + (1-\omega)[C_2^2 + I^2(\omega)]\}\frac{\mathrm{d}\Psi_1}{\mathrm{d}\omega} + \{6\omega^{1/3}\mathrm{e}^{2\omega/3} + C_2^2 + I^2(\omega)\}\Psi_1 \\
&\quad = C_2\omega^{-1/3}\mathrm{e}^{\omega/3} + \frac{3}{C_2}\omega^{1/3}\mathrm{e}^{2\omega/3}[3\omega^{-1/3}\mathrm{e}^{\omega/3} - I(\omega)] - \frac{1}{9C_2}\omega^{1/3}\mathrm{e}^{-\omega/3}(1+\omega)[C_2^2 + (3\omega^{-1/3}\mathrm{e}^{\omega/3} - I(\omega))^2]^2 + \\
&\qquad + \frac{1}{C_2}\omega^{-1/3}\mathrm{e}^{\omega/3}[2I(\omega) - 3\omega^{-1/3}\mathrm{e}^{\omega/3}]^2 + \frac{1}{9C_2}\omega^{-2/3}\mathrm{e}^{-\omega/3}[C_2^2 + I^2(\omega)][3\omega^{-1/3}\mathrm{e}^{\omega/3}I(\omega) - C_2^2 - I^2(\omega)],
\end{aligned}$$

where ω is the parameter. Progress can be made by noting that the left-hand side of this equation is simply

$$3\omega\frac{\mathrm{d}^2\Phi}{\mathrm{d}\omega^2}+(1-\omega)\frac{\mathrm{d}\Phi}{\mathrm{d}\omega}+\Phi,$$

where the variable Φ is defined to be $\Phi=[C_2^2+I^2(\omega)]\Psi_1$, so that we can integrate twice to find

$$\Phi=C_4(1-\omega)+\frac{1}{9}[3\omega^{2/3}\mathrm{e}^{\omega/3}+(1-\omega)I(\omega)]\left[\int_0^\omega(1-t)K(t)\mathrm{d}t+C_3\right]$$
$$-\frac{1}{9}(1-\omega)\int_0^\omega[3t^{2/3}\mathrm{e}^{t/3}+(1-t)I(t)]K(t)\mathrm{d}t,$$

where C_3 and C_4 are constants and $K(t)$ is given by (4.9).

Now, by substituting the perturbation expansions $(4.1)_1$ and $(4.1)_2$ in (2.31), we find that

$$G_1=-\frac{G_0}{2}\left\{\frac{2\Psi_1'+\cot\Theta\tan\Psi_0-2\tan^2\Psi_0-1}{1+\Psi_0'}+2\Psi_1[\tan(\Theta+\Psi_0)-\tan\Psi_0]\right.$$
$$\left.-\tan\Psi_0[\tan(\Theta+\Psi_0)+\tan\Psi_0]\right\},$$

where G_0 is given by $(4.3)_1$, and can be simplified to $(4.8)_2$ with the use of the parametric solutions (4.4).

Finally, from (2.36), $(4.1)_1$, $(4.1)_3$ and (4.4), we find that U_1 is

$$U_1=-\frac{3U_0}{2}\int_{\Theta_c}^{\Theta}(2\Psi_1+\tan\Psi_0)\sec^2\Psi_0\ \mathrm{d}\Theta,$$
$$=-\frac{U_0}{6C_2^2}\int_{\omega_c}^{\omega}t^{-2/3}\mathrm{e}^{-t/3}[C_2^2+I^2(t)][2C_2\Psi_1+I(t)]\mathrm{d}t+\frac{U_0}{18C_2^2}\int_{\omega_c}^{\omega}t^{-1/3}\mathrm{e}^{-2t/3}[C_2^2+I^2(t)]^2\mathrm{d}t,$$

where U_0 is the leading-order solution $(4.6)_2$, and Ψ_1 is the correction term $(4.8)_1$. This expression may be simplified to give $(4.8)_3$.

Appendix D. Constants of integration for axial symmetry

In the following, we will use the integral

$$J(\omega)=\int_0^\omega t^{-1/3}\mathrm{e}^{t/3}\mathrm{d}t, \tag{D.1}$$

so that from (4.5) we have $I(\omega)=J(\omega)+C_1$. Now, for the case of a cone-in-cone insert (see Figure 3) we need only consider the domain $\alpha_2\le\Theta\le\alpha_1$, since the flow is symmetrical around $\Theta=0$. We suppose that the parameter ω in (4.4) takes the values $\omega=\omega_1$ and $\omega=\omega_2$ at $\Theta=\Theta_1$ and $\Theta=\Theta_2$, respectively. By applying the boundary conditions $(4.10)_1$ and $(4.10)_3$, we find the constants C_1 and C_2 are given in terms of these parameter values by

$$C_1=C_2\tan(\alpha_1+\mu_1)-J(\omega_1),\ C_2=\frac{3\omega_2^{-1/3}\mathrm{e}^{\omega_2/3}-J(\omega_2)+J(\omega_1)}{\cot\alpha_2+\tan(\alpha_1+\mu_1)}, \tag{D.2}$$

while the values of ω_1 and ω_2 are determined by the pair of transcendental equations

$$\frac{1}{3}\omega_2^{1/3}\mathrm{e}^{-\omega_2/3}[J(\omega_1)-J(\omega_2)]=\frac{\tan(\alpha_1+\mu_1)-\tan(\alpha_2-\mu_2)}{\cot\alpha_2+\tan(\alpha_2-\mu_2)},$$
$$\left(\frac{\omega_2}{\omega_1}\right)^{1/3}\mathrm{e}^{(\omega_1-\omega_2)/3}=\frac{\cot\alpha_1+\tan(\alpha_1+\mu_1)}{\cot\alpha_2+\tan(\alpha_2-\mu_2)}.$$

The other two boundary conditions $(4.10)_2$ and $(4.10)_4$ provide the following equations for C_3 and C_4

$$C_3 = \frac{1}{M(\omega_1) - mM(\omega_2)} \left\{ \frac{9}{2}(m \tan \mu_2 [C_2^2 + I^2(\omega_2)] + \tan \mu_1 [C_2^2 + I^2(\omega_1)]) + \right.$$
$$\left. + (1-\omega_1) \int_{\omega_2}^{\omega_1} M(t)K(t)\mathrm{d}t + mM(\omega_2) \int_0^{\omega_2} (1-t)K(t)\mathrm{d}t - M(\omega_1) \int_0^{\omega_1} (1-t)K(t)\mathrm{d}t \right\},$$
$$C_4 = \frac{1}{M(\omega_1) - mM(\omega_2)} \left\{ -\frac{1}{2(1-\omega_2)} (\tan \mu_1 M(\omega_2)[C_2^2 + I^2(\omega_1)] + \tan \mu_2 M(\omega_1)[C_2^2 + I^2(\omega_2)]) + \right.$$
$$\left. + \frac{M(\omega_1)M(\omega_2)}{9(1-\omega_2)} \int_{\omega_2}^{\omega_1} (1-t)K(t)\mathrm{d}t - \frac{1}{9} \left[mM(\omega_2) \int_0^{\omega_1} M(t)K(t)\mathrm{d}t - M(\omega_1) \int_0^{\omega_2} M(t)K(t)\mathrm{d}t \right] \right\},$$

where $M(\omega)$ and m are defined by

$$M(\omega) = 3\omega^{2/3} \mathrm{e}^{\omega/3} + (1-\omega)I(\omega), \qquad m = \frac{1-\omega_1}{1-\omega_2},$$

respectively.

Acknowledgements

This work is supported by the Australian Research Council through an Australian Post-Doctoral Fellowship (GMC), a Discovery Project Grant (SWM) and an Australian Professorial Fellowship (JMH). NT gratefully acknowledges the Institute for the Promotion of Teaching Science and Technology, Thailand, for financial support. The authors are also grateful to one of the referees for a number of helpful comments, and in particular for recommending that the authors confirm the physical applicability of the solutions presented here by establishing the non-negative rate of working as described in Section 5.

References

1. A. Drescher, *Analytical Methods in Bin-load Analysis*. Amsterdam: Elsevier (1991) 255pp.
2. J.M. Hill and G.M. Cox, Cylindrical cavities and classical rat-hole theory occuring in bulk materials. *Int. J. Num. Anal. Meth. Geomech.* 24 (2000) 971–990.
3. J.M. Hill and G.M. Cox, Stress profiles for tapered cylindrical cavities in granular media. *Int. J. Solids Struct.* 38 (2001) 3795–3811.
4. A.J.M. Spencer and J.M. Hill, Non-dilatant double-shearing theory applied to granular funnel-flow in hoppers. *J. Engng. Maths.* 41 (2001) 55–73.
5. J.R. Johanson, Controlling flow patterns in bins by use of an insert. *Bulk Solids Handling* 2 (1982) 495–498.
6. J.R. Johanson and W.K. Kleysteuber, Flow-corrective inserts in bins. *Chem. Engng. Progr.* 62 (1966) 79–83.
7. U. Tüzün and R.M. Nedderman, Gravity flow of granular materials round obstacles – I. Investigation of the effects of inserts on flow patterns inside a silo. *Chem. Engng. Sci.* 40 (1985) 325–336.
8. U. Tüzün and R.M. Nedderman, Gravity flow of granular materials round obstacles – II. Investigation of the stress profiles at the wall of a silo with inserts. *Chem. Engng. Sci.* 40 (1985) 337–351.
9. S. Ding, S.R. De Silva and G.G. Enstad, Effect of passive inserts on the granular flow from silos using numerical solutions. *Particulate Sci. Tech.* 21 (2003) 211–226.
10. S-C. Yang, S-S. Hsiau, The simulation and experimental study of granular materials discharged from a silo with the placement of inserts. *Powder Technol.* 120 (2001) 244–255.
11. B. Dantoin, R. Hossfeld and K. McAtee, Converting from funnel-flow to mass-flow. *Power* 147 (2003) 61.
12. A.W. Roberts, Bulk solids handling: Recent developments and future directions. *Bulk Solids Handling* 11 (1991) 17–35.
13. P.A. Gremaud and J.V. Matthews, On the computation of steady hopper flows I. Stress determination for Coulomb materials. *J. Comp. Phys.* 166 (2001) 63–83.

14. P.A. Gremaud, J.V. Matthews and D.G. Schaeffer, Secondary circulation in granular flow through nonaxisymmetric hoppers. *SIAM J. Appl. Math.* 64 (2003) 583–600.
15. A.W. Jenike, Steady gravity flow of frictional-cohesive solids in converging channels. *J. Appl. Mech.* 31 (1964) 5–11.
16. A.W. Jenike, Gravity flow of frictional-cohesive solids - Convergence to radial stress fields. *J. Appl. Mech.* 32 (1965) 205–207.
17. J.R. Johanson, Stress and velocity fields in the gravity flow of bulk solids. *J. Appl. Mech.* 31 (1964) 499–506.
18. V.V. Sokolovskii, *Statics of Granular Media*. Oxford: Pergamon Press (1965) 270pp.
19. G.M. Cox and J.M. Hill, Some exact mathematical solutions for granular stock piles and granular flow in hoppers. *Math. Mech. Solids* 8 (2003) 21–50.
20. J.M. Hill and G.M. Cox, An exact parametric solution for granular flow in a converging hopper. *Z. Angew Math. Phys. (ZAMP)* 52 (2001) 657–668.
21. N. Thamwattana and J.M. Hill, Perturbation solutions for highly frictional granular media. *Proc. R. Soc. London A* 461 (2005) 21–42.
22. Australian Standards, *Loads on Bulk Solids Containers*. Australian Standards 3774 (1996) 77pp.
23. S.W. Perkins, Non-linear limit analysis for the bearing capacity of highly frictional soils. In: J.P. Mohsen (ed), *Proc. 2nd Congress on Computing in Civil Engineering, Atlanta*, 4 June 1995, ASCE (1995) pp. 629–636.
24. S.W. Perkins, Bearing capacity of highly frictional material. *ASTM Geotech. Testing J.* 18 (1995) 450–462.
25. S. Sture, Constitutive issues in soil liquefaction. In: P.V. Lade and J.A. Yanamuro (eds.), *Proc. Physics and Mechanics of Soil Liquefaction, Baltimore*, 10–11 September 1998, Balkema (1999) pp. 133–143.
26. A.J.M. Spencer, A theory of the kinematics of ideal soils under plane strain conditions. *J. Mech. Phys. Solids* 21 (1964) 337–351.
27. A.J.M. Spencer, Deformation of ideal granular materials. In: H.G. Hopkins and M.J. Sewell (eds.), *Mechanics of Solids: The Rodney Hill 60th Anniversary Volume*. Oxford: Pergamon (1982) pp. 607–652.
28. A.J.M. Spencer, Remarks on coaxiality in fully developed gravity flows of dry granular materials. In: N.A. Fleck and A.C.F. Cocks (eds.), *Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) pp. 227–238.
29. A.J.M. Spencer and N.J. Bradley, Gravity flow of a granular material in compression between vertical walls and through a tapering vertical channel. *Q. J. Mech. Appl. Math.* 45 (1992) 733–746.
30. P.A. Gremaud, Numerical issues in plasticity models for granular materials. *J. Volcanol. Geotherm. Res.* 137 (2004) 1–9.
31. G.M. Cox and J.M. Hill, Some exact velocity profiles for granular flow in converging hoppers, *Z. Angew Math. Phys. (ZAMP)* 56 (2005) pp. 92–106.
32. S.W. McCue and J.M. Hill, Free surface problems for static Coulomb–Mohr granular solids. *Mathematics and Mechanics of Solids* (2005) to appear.
33. A.D. Cox, G. Eason and H.G. Hopkins, Axially symmetric plastic deformations in soils. *Phil. Trans. R. Soc. London A* A254 (1961) 1–45.
34. A.J.M. Spencer and N.J. Bradley, Gravity flow of granular materials in converging wedges and cones. In: K.Z. Markov (ed.), *Proc. 8th Int. Symp. Continuum Models and Discrete Systems, Varna, Bulgaria*, 11–16 June 1995, Singapore: World Scientific (1996) pp. 581–590.
35. S.W. McCue, I.K. Johnpillai and J.M. Hill, New stress and velocity fields for highly frictional granular materials. *IMA J. Appl. Math.* 70 (2005) pp. 92–118.
36. R.L. Burden and J.D. Faires, *Numerical Analysis*. Boston: PWS Publishing (1993) 841pp.
37. S.C. Hunter, *Mechanics of Continuous Media*. Chichester: Ellis Horwood Limited (1983) 640pp.
38. D. Harris, Ill- and well-posed models of granular flow. *Acta Mech.* 146 (2001) 199–225.

Journal of Engineering Mathematics (2005) 52: 93–106
DOI 10.1007/s10665-004-2260-9

Micromechanical constitutive modelling of granular media: evolution and loss of contact in particle clusters

B.S. GARDINER and A. TORDESILLAS*
Department of Mathematics and Statistics, The University of Melbourne, Victoria 3010, Australia (atordesi@ms.unimelb.edu.au)

Received 1 May 2003; accepted in revised form 16 August 2004

Abstract. Micromechanical constitutive equations are developed which allow for the broad range of interparticle interactions observed in a real deforming granular assembly: microslip contact, gross slip contact, loss of contact and an evolution in these modes of contact as the deformation proceeds. This was accomplished through a synergetic use of contact laws, which account for interparticle resistance to both sliding and rolling, together with *strain-dependent* anisotropies in contacts and the normal contact force. By applying the constitutive model to the bi-axial test it is demonstrated that the model can correctly predict the evolution of various anisotropies as well as the formation of a distinct shear band. Moreover, the predicted shear-band properties (*e.g.* thickness, prolonged localisation, void ratio) are an even better fit with experimental observations than were previously found by use of previously developed micromechanical models.

Key words: anisotropy, fabric evolution, granular media

1. Introduction

Tordesillas and Walsh [1] developed a framework for constructing micropolar constitutive models for dry monodisperse granular materials. Their method is based on averaging the interactions within a particle cluster, which consists of a single particle and its immediate neighbours (see Figure 1). By considering only a small number of particles in its representative volume element, their method diverges from many previous micromechanical models, with the subsequent advantage that this high-resolution technique should be capable of capturing microstructures consisting of only a few particles *e.g.* shear bands. The generalised constitutive relations derived in their paper were as follows:

$$\sigma_{\alpha\beta} = \frac{1-\upsilon}{\pi R}\int_{\Omega} f_\alpha n_\beta \Phi \,\mathrm{d}\boldsymbol{n}, \tag{1.1}$$

$$\mu_\alpha = \frac{1-\upsilon}{\pi R}\int_{\Omega} \left[M n_\alpha + R e_{\beta\Phi 3} f_\beta n_\alpha n_\phi\right] \Phi \,\mathrm{d}\boldsymbol{n}, \tag{1.2}$$

where $\sigma_{\alpha\beta}$ is the $\alpha\beta$-component of stress, μ_α is the α-component of couple stress, R is the particle radius, υ is the void ratio of the Voronoi cell (ratio of the void volume to the total volume of the cell), f_α is the α-component of the contact force, n_β is the β-component of the outward unit vector $\boldsymbol{n}$ from the particle centre in the direction of contact, M is the rolling resistance, $e_{\beta\phi 3}$ is the permutation symbol, and Ω represents all possible orientations in space. A repeated subscript signifies a summation over the range of the repeated subscript (in a two-dimensional model the summation is from 1 to 2). The weighting function Φ represents

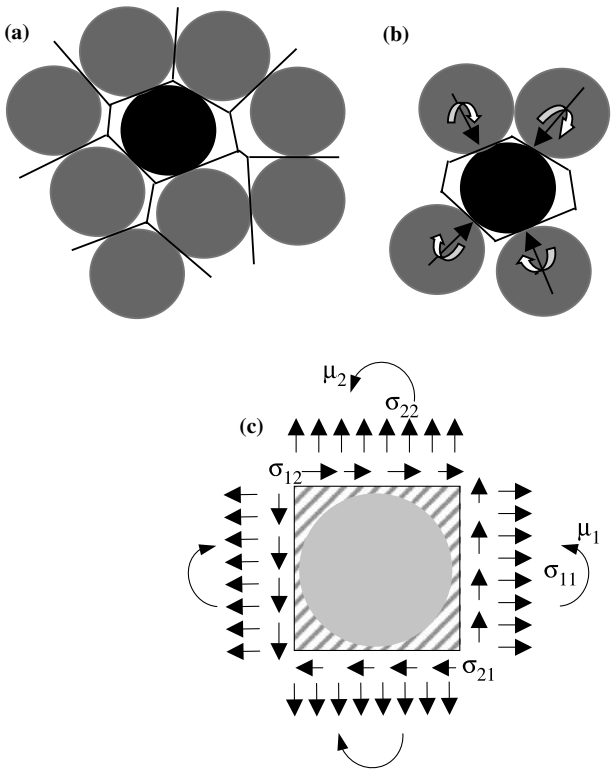


Figure 1. The homogenisation process employed by Tordesillas and Walsh [1]: (a) a Voronoi tessellation of the discrete assembly yields a Voronoi cell for each particle and its immediate void space, (b) interactions between a particle and its contacting neighbours (*i.e.*, forces and moments at contacts) are averaged over the particle's Voronoi cell, (c) corresponding micropolar continuum element of the same area as that of the Voronoi cell.

the angular-contact-density distribution function, and describes the probability of finding a contacting particle for a given direction. Hence, for a particle with N neighbouring contacts, it follows that

$$\int_{\Omega} \Phi \, \mathrm{d}\boldsymbol{n} = N.$$

The function Φ may be used to introduce contact anisotropy. The form of this angular-contact-density distribution function, however, remains an open problem and is an area in need of further investigation. It has been established experimentally that particle arrangements (fabric) and rearrangements (fabric evolution) govern bulk behaviour of granular media. A summary of key experimental findings on fabric evolution is given in [2, Chapter 4]. In particular, it has been observed that contacts within a granular material tend to align themselves in the direction of maximum compressive strain [3]. To account for this experimental result, Walsh and Tordesillas [4] adopted a bipolar contact-density distribution function in which the highest probability of finding a contact was aligned with the direction of maximum compressive strain. The form of their contact-density distribution function is shown in Figure 2 and defined by

$$\Phi\left(\boldsymbol{n}, \boldsymbol{m}, \chi, \nu\right) = \frac{N}{2\pi}\left(1 + \chi \cos\left(2\theta\right)\right), \tag{1.3}$$

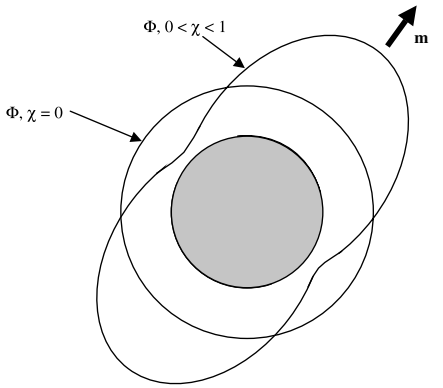


Figure 2. The contact density distribution function Φ, in polar coordinates, for the isotropic case $\chi = 0$, and the anisotropic case $0 < \chi \leq 1$.

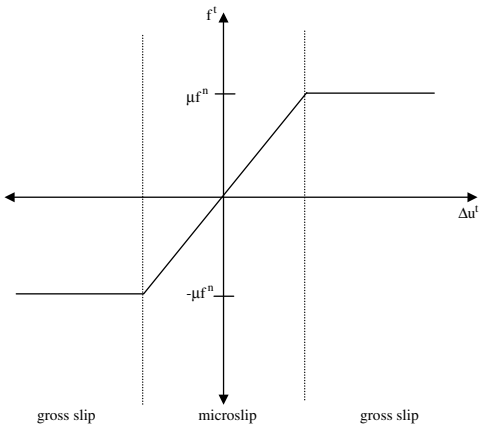


Figure 3. Graphical representation of Coulomb's law (Equation (2.5)).

where χ, the degree of the initial contact anisotropy, can take values in the range $0 < \chi \leq 1$. The symbol θ is defined by $\cos\theta = \boldsymbol{m} \cdot \boldsymbol{n}$, where the 'dot' denotes the scalar product and $\boldsymbol{m}$ is the unit vector pointing in the direction of greatest probability of finding a contacting neighbour. The drawback of this approach is that it limits the fabric evolution to a predefined mode of deformation, which may compromise a model's ability to predict the development of certain anisotropies arising in the formation and evolution of highly localised structures like shear bands (*e.g.* see [2, Chapter 4], [5]). In contrast, Tordesillas and Walsh [1] use an angular contact-density distribution function in which the maximum probability of finding a contact is aligned with the direction of highest density increase. This direction is given by the negative of the gradient of the void-ratio distribution function. In such a case, the contact anisotropy is able to evolve as the assembly deforms, since the void ratio is linked with strain; however, in practice the model was difficult to use without resorting to further simplifying assumptions regarding the evolution of the strain distribution throughout the whole assembly. Apart from these differences in the form for Φ, the two constitutive models [1,4], are identical. In particular, they both neglect: (i) gross slip and loss of contact, (ii) force anisotropy, and (iii) strain-dependent evolution in the force and contact anisotropy. In contrast with these previous models, this paper presents constitutive equations that allow for the possibility of microslip, gross slip and loss of contact. Contact laws, which describe these different modes of contact, are used in conjunction with strain-dependent contact forces, thereby allowing the current model to capture the evolution of these contact forces, the various modes of contact and contact anisotropy.

2. Incorporating gross slip, microslip and loss of contact

Forces and moments between contacting particles were not explicitly expressed in the previous discussion. Expressions for these contact forces and moments need to be inserted into the generalised constitutive laws, Equations (1.1) and (1.2), before the constitutive laws can be used. In this section, expressions are developed for contact forces and moments, which account for the broad range of interactions arising in real deforming granular media.

The contact force, appearing in (1.1) and (1.2), can be separated into normal f^n and tangential f^t components,

$$f_\alpha = f^n n_\alpha + f^t t_\alpha, \tag{2.1}$$

where t_α is the α-component of the unit vector tangential to the contact. The normal force at a contact can be thought of as having two parts: (i) the initial normal force at a contact, and (ii) the normal force resulting from the relative normal displacement of contacting particles. If it is assumed that the particle deformation is linear elastic and that contacts are cohesionless, then the normal force at a contact can be written as

$$f^n = \begin{cases} f^{\text{initial}} + k^n \Delta u^n & \text{if } f^{\text{initial}} + k^n \Delta u^n < 0, \\ 0 & \text{if } f^{\text{initial}} + k^n \Delta u^n \geq 0, \end{cases} \tag{2.2}$$

where f^{initial} is the initial normal contact force, k^n is the particle's normal stiffness coefficient, and Δu^n is the relative normal displacement of two contacting particles. The previous models of Tordesillas and Walsh [1] and Walsh and Tordesillas [4] assumed a constant f^n and so did not allow an evolution or loss of contacts.

It is well-accepted that the normal contact force (and for that matter, the tangential contact force) is not generally equal at all contacts, even prior to deformation. That is, particles rarely share forces equally amongst all neighbours. A striking demonstration of this normal contact force anisotropy can be seen in the photo-elastic disk experiments of Howell *et al.* [6]. In these experiments, it is clearly seen that particles tend to transmit forces through only 2 or 3 of their contacts, resulting in the so-called force chains. The other contacting particles can be considered to be spectator particles, as they have little or no contribution to the force transmission. Preferential force transmission occurs even prior to deformation of a granular material (*e.g.* static granular pile). To account for this potential normal force anisotropy, a weighting function, similar to that used for contact anisotropy (Equation (1.3)), can be combined with f^{initial}, such that the angular dependent normal force at a contact is expressed as

$$f^n(\theta) = \begin{cases} f^{\text{initial}}(1 + \chi^* \cos(2\theta + \phi)) + k^n \Delta u^n & \text{if } f^{\text{initial}}(1 + \chi^* \cos(2\theta + \phi)) + k^n \Delta u^n < 0, \\ 0 & \text{if } f^{\text{initial}}(1 + \chi^* \cos(2\theta + \phi)) + k^n \Delta u^n \geq 0. \end{cases} \tag{2.3}$$

In (2.3), χ^* represents the degree of the initial normal contact-force anisotropy and can take values in the range $0 \leq \chi^* \leq 1$, with $\chi^* = 0$ corresponding to the isotropic case. Again, θ is defined by $\cos\theta = \boldsymbol{m} \cdot \boldsymbol{n}$ and ϕ is defined such that $\cos\frac{\phi}{2}$ represents the dot product between $\boldsymbol{m}$ and a unit vector in the direction of the maximum normal contact force. Note that the integrals, which, respectively, define the stress and couple stress in (1.1) and (1.2), are to be evaluated with respect to θ. By defining the normal contact force according to (2.3), not only has a potential initial normal contact-force anisotropy been introduced, but also the normal contact-force anisotropy is now able to evolve naturally with strain, as the relative normal displacement can be related to strain via:

$$\Delta u^n = 2R\varepsilon_{\alpha\beta} n_\alpha n_\beta, \tag{2.4}$$

where $\varepsilon_{\alpha\beta}$ is the $\alpha\beta$-component of strain. Furthermore, an evolving anisotropy in the normal contact force results in an evolving angular-contact distribution, as contacts are lost in directions for which $f^{\text{initial}}(1 + \chi^* \cos(2\theta + \phi)) + k^n \Delta u^n \geq 0$.

In Equation (2.1), the tangential contact force can be related to the normal contact force by following Coulomb's contact law (see Figure 3) for the various modes of contact (*i.e.*, microslip and gross slip) and no contact:

$$f^t = \begin{cases} \text{sign}(\Delta u^t)\mu|f^n| & \text{if } |k^t\Delta u^t| \geq |\mu f^n| \text{ and } f^n < 0 \Rightarrow \text{gross slip}, \\ k^t\Delta u^t & \text{if } |k^t\Delta u^t| < |\mu f^n| \text{ and } f^n < 0 \Rightarrow \text{microslip}, \\ 0 & \text{if } f^n \geq 0 \Rightarrow \text{no contact}, \end{cases} \tag{2.5}$$

where μ is the inter-particle friction coefficient, k^t is the particle's tangential stiffness coefficient, and Δu^t is the tangential component of the relative displacement of the contact points and is related to strain and curvature (gradient in rotation) κ_φ according to

$$\Delta u^t = 2R\left(\varepsilon_{qr} n_r t_q - R n_\varphi \kappa_\varphi\right), \tag{2.6}$$

such that

$$\text{sign}\left(\Delta u^t\right) = \begin{cases} 1 & \text{if } \varepsilon_{qr} n_r t_q - R n_\varphi \kappa_\varphi > 0, \\ -1 & \text{if } \varepsilon_{qr} n_r t_q - R n_\varphi \kappa_\varphi < 0. \end{cases} \tag{2.7}$$

Strictly speaking, the tangential contact force should have an additional term to represent the initial tangential contact force acting on a particle. However, the application of the current model is restricted to cases in which the granular specimen is prepared such that the expected initial tangential contact force is zero. Unlike the initial normal contact force, the initial tangential contact force may be either positive or negative, such that the expected initial tangential contact force in many circumstances will be close to zero. Applications for which the expected initial normal force may not be zero are, for now, not considered.

Since f^n has an evolving anisotropy, and f^t depends on f^n, it follows that f^t also has an evolving anisotropy. More specifically, not only do these contact laws (Equations (2.3) and (2.5)) allow the contact mode to be identified, but they also allow the direction of these contact modes to change with deformation, as the inequality constraints in Equation (2.5) contain an angular dependence due to the angular dependence of f^n and Δu^t. In other words, the mode of contact will vary with the direction of contact. At any stage in a granular media's deformation, a neighbouring particle pair will be sharing *only one* of the following: a microslip contact, a gross slip contact, or no contact. Therefore, the integration over the angle in Equations (1.1) and (1.2) can be separated into a sum of integrals, with each integral representing a contact mode (*i.e.*, gross slip or microslip) or no contact. That is, Equations (1.1) and (1.2) may be written as

$$\sigma_{\alpha\beta} = \frac{1-\upsilon}{\pi R}\sum_{i=1}^{x}\left[\underbrace{\int_{a_i}^{c_i} f_\alpha^{\text{G.S.}} n_\beta \Phi \,\mathrm{d}\theta}_{\text{gross slip}} + \underbrace{\int_{c_i}^{d_i} f_\alpha^{\text{M.S.}} n_\beta \Phi \,\mathrm{d}\theta}_{\text{microslip}}\right], \tag{2.8}$$

$$\mu_\alpha = \frac{1-\upsilon}{\pi R}\sum_{i=1}^{x}\left[\underbrace{\int_{a_i}^{c_i}\left[M^{\text{G.S.}} n_\alpha + R e_{\beta\phi 3} f_\beta^{\text{G.S.}} n_\alpha n_\phi\right]\Phi \,\mathrm{d}\theta}_{\text{gross slip}} + \underbrace{\int_{c_i}^{d_i}\left[M^{\text{M.S.}} n_\alpha + R e_{\beta\phi 3} f_\beta^{\text{M.S.}} n_\alpha n_\phi\right]\Phi \,\mathrm{d}\theta}_{\text{microslip}}\right], \tag{2.9}$$

where the superscripts 'G.S.' and 'M.S.' denote the gross slip and microslip terms, respectively. The integrals corresponding to regions of no contact are not shown, as regions of no contact

do not contribute (directly) to the stress or couple stress. The integration limits and the upper limit of the summation index x obey the following relationships:

$$\sum_{i=1}^{x}[(c_i-a_i)+(d_i-c_i)+(e_i-d_i)]=2\pi, \quad \text{and} \quad a_1=0;\, a_{i+1}=e_i;\, e_x=2\pi. \tag{2.10}$$

Equation (2.10) is a condition that ensures that all possible contact directions ($0\to 2\pi$) have been assigned a contact mode. In (2.10), the quantity (e_i-d_i) corresponds to the range of angles for which there are no contacts. The gross slip regions (c_i-a_i) are further separated into two regions, (b_i-a_i) and (c_i-b_i), which are related to sign $\left(\Delta u^t\right)$ being positive or negative, respectively. The limits of integration in (2.8) and (2.9) need to be determined from the inequality constraints in (2.5) and (2.7). For example, the limits a_i, c_i and d_i, demarcating the regions of gross slip and microslip, must satisfy $k^t\left|\Delta u^t\right|-\mu\left|f^n\right|=0$. The integration limits a_i, d_i and e_i satisfy $f^n=0$, and the gross-slip integration limit b_i must satisfy $\left(\varepsilon_{qr}n_r t_q - Rn_\varphi\kappa_\varphi\right)=0$. If, for example, with increasing angle a gross slip region (c_3-b_3) is immediately followed by a region of no contact (e_3-d_3), the microslip integration limits are set to have the property $d_3=c_3=0$. These integration limits will now be referred to as the transition angles. Due to the form of the constraints in (2.5) and (2.7), involving absolute values of cosine-squared functions, there exists an upper bound on the value that the summation index x may take.

A linear contact law is used to approximate the rolling resistance for both the microslip and gross slip contacts, such that

$$M^{\text{G.S.}}=2k'^{\text{G.S.}}Rk_\psi n_\psi, \tag{2.11}$$

$$M^{\text{M.S.}}=2k'^{\text{M.S.}}Rk_\psi n_\psi, \tag{2.12}$$

where $k'^{\text{G.S.}}$ and $k'^{\text{M.S.}}$ are rolling-stiffness coefficients.

In this paper, the bipolar form for the contact-density distribution function Φ used by Walsh and Tordesillas [4] is adopted (see Equation (1.3) and Figure 2). However, unlike in [4], no assumption is imposed on the direction of maximum probability of finding a contact $\boldsymbol{m}$ (direction of contact anisotropy) at this stage of the analysis. Hence, the direction of $\boldsymbol{m}$ can be defined at the point of implementation of the resulting constitutive law. Furthermore, in the present analysis, the form of Φ and the direction of $\boldsymbol{m}$ serve only as *initial* conditions for the contact anisotropy. Since the contact laws introduced earlier incorporate loss of contacts, then, as deformation proceeds, the contact-density distribution function will be, in effect, modified (evolving) as various angles are "switched off" for regions in which particle contacts are broken.

We note that the bipolar form of Φ is based on a numerical simulation of the direction of contacts averaged over a large number of particles [3]. Furthermore, the generalised constitutive equations (1.1) and (1.2) are not restricted to a specific particle-cluster configuration, which was the original reason for Tordesillas and Walsh [1] to introduce the weighting function Φ. Therefore, when interpreting the contact mode and force anisotropy results in Section 4 of this paper, we emphasize that the results will be an *expectation* rather than an exact prediction of the contact behaviour within a granular specimen.

The form of Φ in Equation (1.3) is now substituted in Equations (2.8) and (2.9), along with the expressions for the contact forces and moments (2.3), (2.5), (2.11) and (2.12). With these substitutions, Equations (2.8) and (2.9) can be integrated to give the following micromechanical constitutive laws:

$$\sigma_{\alpha\beta}=a\,(\upsilon)\left[P_{\alpha\beta}+Q_{\alpha\beta mn}\varepsilon_{mn}+R_{\alpha\beta j}\varepsilon_{\alpha j}+S_{\alpha\beta sj}\varepsilon_{sj}\right], \tag{2.13}$$

$$\mu_\alpha=a\,(\upsilon)\,T_{\alpha\psi}\kappa_\psi, \tag{2.14}$$

where

$$a(\upsilon) = \frac{1}{2\pi R}\left[\frac{(1-\upsilon)\left[36-4\pi\sqrt{3}\right]-\pi^2}{6-\pi\sqrt{3}}\right]. \tag{2.15}$$

The coefficients $P_{\alpha\beta}$, $Q_{\alpha\beta mn}$, $R_{\alpha\beta j}$, $S_{\alpha\beta sj}$ and $T_{\alpha\psi}$ depend on the transition angles, the various anisotropy parameters and the physical properties of particles, and are defined in the Appendix. To obtain (2.15) it has been assumed, following Hinrichsen *et al.* [7], that

$$\mathrm{N} = \frac{1}{6-\pi\sqrt{3}}\left[36-4\pi\sqrt{3}-\frac{\pi^2}{1-\upsilon}\right], \tag{2.16}$$

for $0 \leq N \leq 6$. Equation (2.16) reflects an expected relationship between the number of contacts per particle and the void ratio and, in a similar manner to (1.3), it was obtained by averaging over a large number of particles. As such, any results obtained using (2.16) will be considered to reflect an average or expectation rather than an exact prediction for a specific particle cluster.

3. Fabric and force evolution in a bi-axial compression test

To demonstrate the capabilities of the micromechanically-based constitutive model, we examine the formation of shear bands in a bi-axial test, using the method provided by Mühlhaus and Vardoulakis [8] for micropolar continua. As illustrated in Figure 4, the specimen is compressed in the X_2-direction at a constant rate. The vertical boundaries are allowed to deform such that the normal stress σ_{11} remains constant along these boundaries. Prior to shear-band formation, the deformation is assumed to be homogeneous and $\varepsilon_{12}=\varepsilon_{21}=\kappa_1=\kappa_2=0$. In addition to predicting shear-band evolution, specific attention will be given to the model's predictions of microstructural development (*e.g.* contact and contact-force evolution).

Briefly, the method originally proposed by Mühlhaus and Vardoulakis [8] involves solving the rate form of the stress equilibrium equations for a Cosserat/micropolar material, and then looking for special solutions that may exist within a narrow region (shear band). These special solutions correspond to an inhomogeneous deformation that is distinct from the homogeneous deformation occurring both prior to the onset of a shear band and outside the shear band after the band has formed. The onset of the inhomogeneous solution is referred to as the bifurcation point.

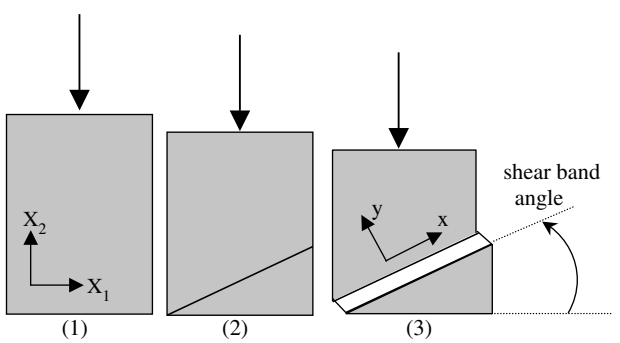


Figure 4. Stages in shear-band formation within a bi-axial test: (a) initial specimen, (1) homogeneous deformation prior to shear band formation, (2) onset of shear band or bifurcation point, and (3) evolution of shear band.

Adopting the nomenclature employed by Tordesillas *et al.* [9], we can write the stress-rate equilibrium equations in shear-band coordinates (x, y) in terms of time rates of displacement V and rotation $\frac{\mathrm{d}\omega}{\mathrm{d}t}$ as:

$$A_{\alpha\eta} V_\eta'' + A_{\alpha 31} \frac{\mathrm{d}\omega''}{\mathrm{d}t} + A_{\alpha 32} \frac{\mathrm{d}\omega'}{\mathrm{d}t} = 0, \tag{3.1}$$

$$A_{3\eta 1} V_\eta'' + A_{3\eta 2} V_\eta' + A_{331} \frac{\mathrm{d}\omega''}{\mathrm{d}t} + A_{332} \frac{\mathrm{d}\omega'}{\mathrm{d}t} + A_{333} \frac{\mathrm{d}\omega}{\mathrm{d}t} = 0, \tag{3.2}$$

where only the gradients across the shear band (the y-coordinate) are nonzero with $(\ )' = \frac{\mathrm{d}(\ )}{\mathrm{d}y}$; $\alpha = 1, 2$; $\eta = 1, 2$; and the various coefficients (*i.e.*, the subscripted 'A's) come from the rate form of the constitutive law. That is, by writing

$$\left\langle \frac{\mathrm{d}\sigma_{\alpha\beta}}{\mathrm{d}t} \right\rangle = L_{\alpha\beta\eta\lambda} \frac{\mathrm{d}\varepsilon_{\eta\lambda}}{\mathrm{d}t}, \tag{3.3}$$

$$\left\langle \frac{\mathrm{d}\mu_\lambda}{\mathrm{d}t} \right\rangle = G_{\lambda\alpha\beta} \frac{\mathrm{d}\varepsilon_{\alpha\beta}}{\mathrm{d}t} + H_{\lambda\alpha} \frac{\mathrm{d}\kappa_\alpha}{\mathrm{d}t}, \tag{3.4}$$

where $\langle\ \rangle$ represent the Jaumann time derivative, we can be represent the various coefficients in (3.1) and (3.2) as

$$A_{11} = s_\lambda s_1 L_{111\lambda} + s_\lambda s_2 L_{121\lambda} - \frac{1}{2} s_2 s_2 \sigma, \quad A_{12} = s_\lambda s_1 L_{112\lambda} + s_\lambda s_2 L_{122\lambda} + \frac{1}{2} s_1 s_2 \sigma,$$

$$A_{21} = s_\lambda s_1 L_{211\lambda} + s_\lambda s_2 L_{221\lambda} - \frac{1}{2} s_1 s_2 \sigma, \quad A_{22} = s_\lambda s_1 L_{212\lambda} + s_\lambda s_2 L_{222\lambda} + \frac{1}{2} s_1 s_1 \sigma,$$

$$A_{132} = e_{\eta\lambda 3} s_1 L_{11\eta\lambda} + e_{\eta\lambda 3} s_2 L_{12\eta\lambda}, \quad A_{232} = e_{\eta\lambda 3} s_2 L_{22\eta\lambda} + e_{\eta\lambda 3} s_1 L_{21\eta\lambda},$$

$$A_{131} = A_{231} = 0, \quad A_{311} = s_\lambda s_1 G_{11\lambda} + s_\lambda s_2 G_{21\lambda},$$

$$A_{321} = s_\lambda s_1 G_{12\lambda} + s_\lambda s_2 G_{22\lambda}, \quad A_{312} = s_\lambda L_{211\lambda} - s_\lambda L_{121\lambda},$$

$$A_{322} = s_\lambda L_{212\lambda} - s_\lambda L_{122\lambda}, \quad A_{331} = H_{1\lambda} s_\lambda s_1 + H_{2\lambda} s_\lambda s_2,$$

$$A_{332} = e_{\eta\lambda 3} s_1 G_{1\eta\lambda} + e_{\eta\lambda 3} s_2 G_{2\eta\lambda}, \quad A_{333} = e_{\eta\lambda 3} L_{21\eta\lambda} - e_{\eta\lambda 3} L_{12\eta\lambda}. \tag{3.5}$$

In Equation (3.5), $\sigma = \sigma_{11} - \sigma_{22}$, $s_1 = -\sin\theta_b$ and $s_2 = \cos\theta_b$, where θ_b is the shear-band angle (see Figure 4).

Solutions are then sought for (3.1) which satisfy the following boundary conditions (in accordance with experimental observations and DEM simulations of Iwashita and Oda [5] and Oda and Kazama [10]:

(i) across a shear band (*i.e.*, $-d_b \geq y \geq d_b$ where $2d_b$ is the shear-band thickness), the displacement rate is an even function of y and the rotation rate is an odd function of y;

(ii) continuity of traction and couple stress at the shear-band boundary;

(iii) $V_1 = \frac{V_1^*}{2}$ and $V_2 = \frac{V_2^*}{2}$ at $y = d_b$ and $V_1 = -\frac{V_1^*}{2}$ and $V_2 = -\frac{V_2^*}{2}$ at $y = -d_b$. Since it is assumed that regions outside the shear band act rigidly after the bifurcation strain is reached, and that the bi-axial test is strain-controlled, the rate of vertical displacement of the rigid outer regions of the specimen V_2^* is prescribed. The horizontal rate of displacement V_1^*, on the other hand, needs to be determined from the strain within the shear band.

Once the displacement and rotation rates are determined, from solving (3.1) and (3.2) subject to the boundary conditions, an incremental procedure can be adopted to determine the strain and rotation within a shear band, and therefore the evolution of the shear band. For a more detailed discussion of the shear-banding problem and the method outlined here, see [9].

4. Results and discussion

Prior to deformation, we assume that the particles are randomly packed under conditions in which gravity is negligible. Therefore, the expected contact distribution and the expected contact-force distribution are isotropic, *i.e.*, $\chi = \chi^* = 0$. Only after deformation occurs would one expect the contact distribution and the contact-force distribution to become anisotropic and take on a preferential direction. Model predictions are now presented for the following model parameters: $\chi = \chi^* = 0$; $m_2 = 1$; $v_0 = 0{\cdot}2$; $\phi = 0$; $f^{\text{initial}} = -(1/20)Rk^n$; $k^n = k^t = 4 \times 10^7$ N/m; $R = 0{\cdot}005$ m. These parameters reflect a moderate particle-packing density with an initially isotropic contact and contact-force distribution.

Figure 5 shows the change in thickness of a shear band inclined at an angle of 65 degrees to the X_1-axis. The thickness rapidly approaches a value of 4–5 particle diameters. In real sands for which there is a distribution of particle sizes and shapes, reported shear-band widths range from around 7–8 particle diameters [10] up to 17 particle diameters [11]. However, two-dimensional Schneebeli systems, consisting of circular rods, display shear bands that involve fewer particles (1–4 particle diameters) than those found in real sands [12]. Note that the thickness predicted by the current model is also consistent with the shear-band analysis based on the deformation theory of plasticity for frictional materials with internal constraint [8], [13]. Although not shown, other shear-band inclinations display similar behaviour.

In Figure 6, it can be seen that the expected contact distribution evolves with increasing strain. Stage 1, the innermost (smallest) ring, corresponds to the initial contact mode distribution. Stage 2, the middle ring, corresponds to the contact-mode distribution at the bifurcation point (onset of shear banding). Stage 3, the outside (largest) ring, corresponds to the contact mode distribution within a shear band at a shear strain of 0·8. Initially, all contacts are microslip contacts (stage 1). With compression, however, contacts are lost in the direction perpendicular to the applied compressive strain, and the regions of microslip shrink to a narrow angle in the direction of the maximum principal strain (stage 2). This pattern of loss and retention of contacts is consistent with the formation of particle columns at the onset

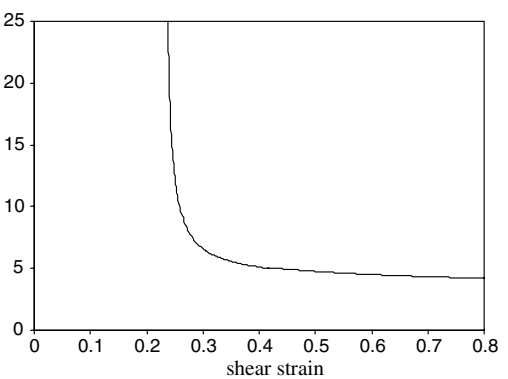


Figure 5. The variation of the shear-band thickness with shear strain for a shear band inclined at 65 degrees to the X_1-axis.

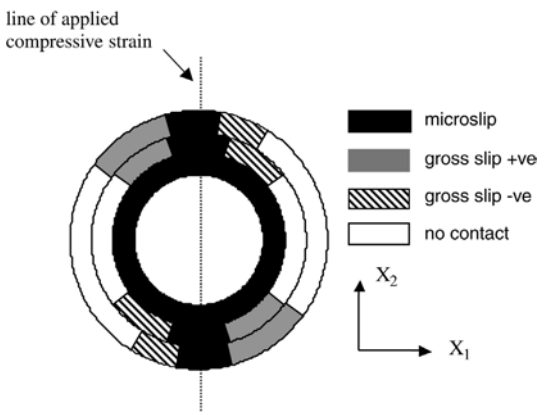


Figure 6. Evolution of contact modes for the three stages of the bi-axial test.

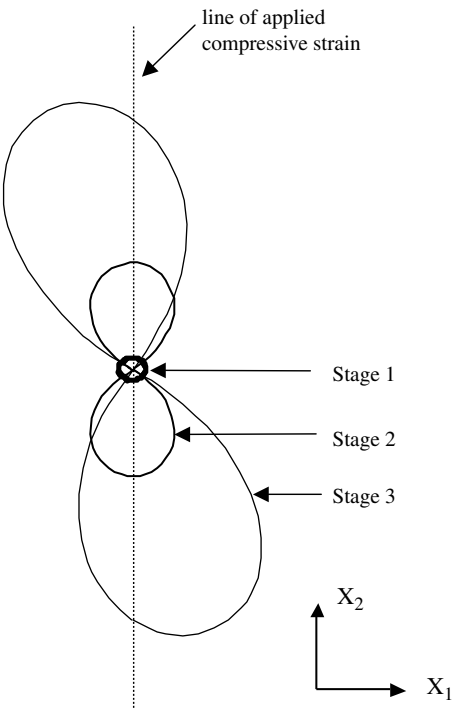


Figure 7. A polar plot showing the evolution of the normal contact force for the three stages of the bi-axial test.

of shear banding. It is thought that the buckling of these particle columns is the underlying mechanism responsible for the softening observed during shear banding [10]. With a further increase in shear strain it is seen that there is a rotation of the direction of contacts (and no contacts) and the contact modes, which is also consistent with observed microstructural development inside shear bands (stage 3).

Figure 7 displays the evolution of the normal contact force for the three stages in shear-band development presented in Figure 6. Firstly, it is confirmed that normal contact-force anisotropy can develop from an initially isotropic distribution using this constitutive model. At the bifurcation point the normal contact force has become highly anisotropic, with a maximum direction aligned with the maximum principal strain. Secondly, within the shear band, the direction of the maximum normal force begins to rotate away from the direction of maximum principal strain. The direction and rate of rotation of the normal contact force is the same as the direction of rotation of the contact modes shown in Figure 6, and the combined contact mode and normal contact force behaviour is consistent with the rotation of particle columns and force chains in a shear band.

Figure 8 shows the change in the *global* void ratio (void ratio averaged across the entire specimen) for stage 1, and the change in the *local* void ratio (void ratio averaged across the band) during stage 3. The initial overall compaction by an applied compressive strain is expected, given that the initial void ratio is 0·2. The observed decrease in void ratio of 0·05 corresponds to an average gain of only one contact per particle. At first glance this result would seem to contradict the loss of contacts observed in Figure 6. However, one should recall that: (i) the contact modes are only expectations, and (ii) a void ratio of 0·2 only cor-

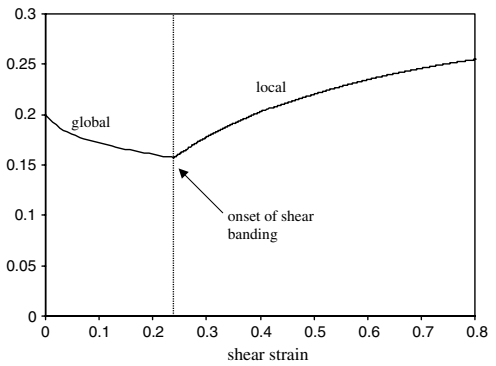


Figure 8. The variation of void ratio with deformation.

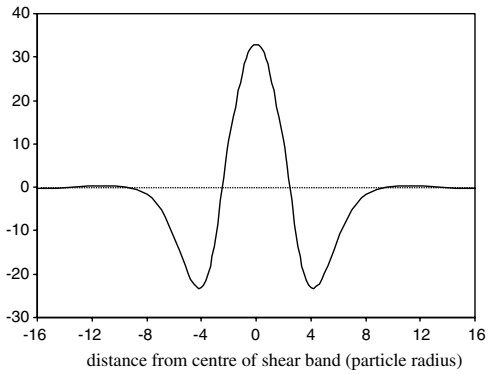


Figure 9. Total rotation in the region of the shear band at a shear strain of 0·8.

responds to approximately three contacts per particle, such that a particle can readily accommodate an additional contact within the region predicted to have a contact in Figure 6. The increase in void ratio within a shear band, as shown in Figure 8, is consistent with the experimental observations of dilation and the appearance of large voids within shear bands in sands (*e.g.* see [2, Chapter 4]) and assemblies Schneebeli rods [12].

The total rotation in the region of the shear band, as shown in Figure 9, refers to the net accumulation of rotation ω (see Equation (3.1)) from the start of deformation up to a shear strain of 0·8. Note that although the shear band thickness at a shear strain of 0·8 is only 4–5 particle diameters (Figure 5), the microstructure beyond 5 particle diameters has been disturbed (finite rotations) due to prior shear strains for which the shear band thickness was greater than 5 particle diameters. The shear-band thickness shown in Figure 5 should then be considered an instantaneous thickness, whereas a shear-band thickness implied by rotations should resemble that seen in experiments.

Note that the current model predicts rotations to change sign within a shear band. This would seem to contradict the findings from DEM simulations of Iwashita and Oda [5], in which particles were seen to rotate in one direction on average. It could be argued that, as the shear-band thickness in Figure 5 is in the range of 4–5 particle diameters, beyond a shear strain of 0·8, further rotations will only occur within 4–5 particle radii from the centre of the shear band, such that the central, unidirectional, peak will begin to dominate the observed rotations. However, larger shear strains are beyond the small strain limitations of the current constitutive model.

5. Conclusions

A high-resolution, micromechanically-based, Cosserat constitutive model for the quasi-static deformation of a dry granular material has been presented. The model incorporates inter-particle contacts undergoing microslip, gross slip, rolling resistance and loss of contact. Fabric and contact force anisotropies are also introduced, and the evolution of these anisotropies is naturally accounted for. When applied to the shear-banding problem, the micromechanically-based constitutive laws successfully reproduces expected shear-band thicknesses and the contact and force anisotropies evolve in a way consistent with the formation and evolution of particle columns and force chains within a shear band. Ongoing work is focused on understanding the evolution of the anisotropies, through a DEM analysis, in preparation for the extension of these models to the large deformations encountered in engineering applications.

Appendix A. Coefficients appearing in Equation (2.13) and (2.14)

The various coefficients in the constitutive equations (2.13) and (2.14) are

$$P_{\alpha\beta} = \frac{f^{\text{initial}}}{\pi} \sum_{i=1}^{x} \left\lfloor \begin{array}{l} w_1(a_i, d_i) W_{\alpha\beta 1} + w_2(a_i, d_i) W_{\alpha\beta 2} + w_3(a_i, d_i) W_{\alpha\beta 3} \\ +\mu e_{s\alpha 3} \left[\begin{array}{l} (w_1(b_i, c_i) - w_1(a_i, b_i)) W_{s\beta 1} + (w_2(b_i, c_i) - w_2(a_i, b_i)) W_{s\beta 2} \\ + (w_3(b_i, c_i) - w_3(a_i, b_i)) W_{s\beta 3} \end{array} \right] \end{array} \right\rfloor,$$

$$Q_{\alpha\beta mn} = \frac{2Rk^n}{\pi} \sum_{i=1}^{x} \left\lfloor \begin{array}{l} w_4(a_i, d_i) W_{\alpha\beta 1} W_{mn1} + w_5(a_i, d_i) \left(W_{\alpha\beta 1} W_{mn2} + W_{\alpha\beta 2} W_{mn1} \right) \\ +w_6(a_i, d_i) \left(W_{\alpha\beta 1} W_{mn3} + W_{\alpha\beta 2} W_{mn2} + W_{\alpha\beta 3} W_{mn1} \right) \\ +w_7(a_i, d_i) \left(W_{\alpha\beta 2} W_{mn3} + W_{\alpha\beta 3} W_{mn2} \right) + w_8(a_i, d_i) W_{\alpha\beta 3} W_{mn3} \\ +\mu e_{s\alpha 3} \left[\begin{array}{l} (w_4(b_i, c_i) - w_4(a_i, b_i)) W_{s\beta 1} W_{mn1} \\ + (w_5(b_i, c_i) - w_5(a_i, b_i)) \left(W_{s\beta 1} W_{mn2} + W_{s\beta 2} W_{mn1} \right) \\ + (w_6(b_i, c_i) - w_6(a_i, b_i)) \left(W_{s\beta 1} W_{mn3} + W_{s\beta 2} W_{mn2} + W_{s\beta 3} W_{mn1} \right) \\ + (w_7(b_i, c_i) - w_7(a_i, b_i)) \left(W_{s\beta 2} W_{mn3} + W_{s\beta 3} W_{mn2} \right) \\ + (w_8(b_i, c_i) - w_8(a_i, b_i)) W_{s\beta 3} W_{mn3} \end{array} \right] \end{array} \right\rfloor,$$

$$R_{\alpha\beta j} = \frac{2Rk^t}{\pi} \sum_{i=1}^{x} \left[w_9(c_i, d_i) W_{j\beta 1} + w_{10}(c_i, d_i) W_{j\beta 2} + w_{11}(c_i, d_i) W_{j\beta 3} \right]$$

$$S_{\alpha\beta sj} = -\frac{2Rk^t}{\pi} \sum_{i=1}^{x} \left\lfloor \begin{array}{l} w_{12}(c_i, d_i) W_{\alpha\beta 1} W_{sj1} + w_{13}(c_i, d_i) \left(\begin{array}{l} e_{h\alpha 3} W_{h\beta 1} W_{sj1} + e_{hs3} W_{\alpha\beta 1} W_{hj1} \\ + e_{h\beta 3} W_{\alpha h1} W_{sj1} + e_{hj3} W_{\alpha\beta 1} W_{sh1} \end{array} \right) \\ +w_{14}(c_i, d_i) \left(\begin{array}{l} W_{j\beta 1} \delta_{s\alpha} + W_{js1} \delta_{\beta\alpha} + W_{s\beta 1} \delta_{j\alpha} + W_{\alpha\beta 1} \delta_{js} \\ + W_{\alpha j1} \delta_{\beta s} + W_{\alpha s1} \delta_{\beta j} - 6 W_{\alpha\beta 1} W_{sj1} \end{array} \right) \\ +w_{15}(c_i, d_i) \left(\begin{array}{l} e_{h\beta 3} W_{j_{h1}} W_{s\alpha 3} + e_{hj3} W_{h_{\beta 1}} W_{s_{\alpha 3}} \\ + e_{h\alpha 3} W_{s_{h1}} W_{j_{\beta 3}} + e_{hs3} W_{h\alpha 1} W_{j_{\beta 3}} \end{array} \right) \\ +w_{16}(c_i, d_i) \left(W_{\alpha\beta 1} W_{sj1} + \delta_{j_\beta} \delta_{s_\alpha} - W_{j\beta 1} \delta_{s_\alpha} - \delta_{j_\beta} W_{\alpha s1} \right) \end{array} \right\rfloor,$$

$$T_{\alpha\psi} = \frac{2R}{\pi} \sum_{i=1}^{x} \left\lfloor \begin{array}{l} k^{G.S.} \left[w_9(a_i, c_i) W_{\alpha\psi 1} + w_{10}(a_i, c_i) W_{\alpha\psi 2} + w_{11}(a_i, c_i) W_{\alpha\psi 3} \right] \\ + \left(k^{M.S.} + R^2 k^t \right) \left[w_9(c_i, d_i) W_{\alpha\psi 1} + w_{10}(c_i, d_i) W_{\alpha\psi 2} + w_{11}(c_i, d_i) W_{\alpha\psi 3} \right] \end{array} \right\rfloor,$$

where

$$\delta_{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta, \end{cases},$$

$$W_{\alpha\beta 1} = m_\alpha m_\beta,$$

$$W_{\alpha\beta 2} = \left[e_{\phi\beta 3} W_{\alpha\phi 1} + e_{\phi\alpha 3} W_{\phi\beta 1} \right],$$

$$W_{\alpha\beta 3} = \left[\delta_{\alpha\beta} - W_{\alpha\beta 1} \right],$$

$$w_1(x, y) = \frac{1}{48} \begin{pmatrix} 12(2 + \chi + (1 + \chi) \chi^* \cos\phi)(y - x) + 3(\chi + 4) \chi^* \sin\phi (\cos 2y - \cos 2x) \\ +3(\chi + 1) \chi^* \sin\phi (\cos 4y - \cos 4x) + \chi \chi^* \sin\phi (\cos 6y - \cos 6x) \\ +3(4(\chi + 1) + (3\chi + 4) \chi^* \cos\phi)(\sin 2y - \sin 2x) \\ +3(\chi + (\chi + 1) \chi^* \cos\phi)(\sin 4y - \sin 4x) + \chi \chi^* \cos\phi (\sin 6y - \sin 6x) \end{pmatrix},$$

$$w_2(x, y) = \frac{1}{48} \begin{pmatrix} 12\chi^* \sin\phi (x - y) + 3\chi \chi^* (\cos(\phi - 2x) - \cos(\phi - 2y)) \\ +12(\cos 2x - \cos 2y) + 3\chi (\cos 4x - \cos 4y) \\ +3\chi^* (\cos(\phi + 4x) - \cos(\phi + 4y)) + \chi \chi^* (\cos(\phi + 6x) - \cos(\phi + 6y)) \end{pmatrix},$$

$$w_3(x,y)=\frac{1}{48}\begin{bmatrix}12(2-\chi+(\chi-1)\chi^*\cos\phi)(y-x)\\+12(\chi-1)(\sin 2y-\sin 2x)+3\chi\chi^*(\sin(\phi-2y)-\sin(\pi-2x))\\+6\chi^*(2-\chi)(\sin(\phi+2y)-\sin(\phi+2x))-3\chi(\sin 4y-\sin 4x)\\+3\chi^*(\chi-1)(\sin(\phi+4y)-\sin(\phi+4x))-\chi\chi^*(\sin(\phi+6y)-\sin(\phi+6x))\end{bmatrix},$$

$$w_4(x,y)=\frac{1}{96}\begin{pmatrix}3(8\chi+12)(y-x)+3(7\chi+8)(\sin 2y-\sin 2x)\\+3(2\chi+1)(\sin 4y-\sin 4x)+\chi(\sin 6y-\sin 6x)\end{pmatrix},$$

$$w_5(x,y)=\frac{1}{96}(3(\chi+4)(\cos 2x-\cos 2y)+3(\chi+1)(\cos 4x-\cos 4y)+\chi(\cos 6x-\cos 6y)),$$

$$w_6(x,y)=\frac{1}{96}(12(y-x)+3\chi(\sin 2y-\sin 2x)-3(\sin 4y-\sin 4x)-\chi(\sin 6y-\sin 6x)),$$

$$w_7(x,y)=\frac{1}{96}(3(\chi-4)(\cos 2y-\cos 2x)-3(\chi-1)(\cos 4y-\cos 4x)+\chi(\cos 6y-\cos 6x)),$$

$$w_8(x,y)=\frac{1}{96}\begin{pmatrix}3(-8\chi+12)(y-x)-3(-7\chi+8)(\sin 2y-\sin 2x)\\+3(-2\chi+1)(\sin 4y-\sin 4x)+\chi(\sin 6y-\sin 6x)\end{pmatrix},$$

$$w_9(x,y)=\frac{1}{16}(4(2+\chi)(y-x)+4(1+\chi)(\sin 2y-\sin 2x)+\chi(\sin 4y-\sin 4x)),$$

$$w_{10}(x,y)=\frac{1}{8}(\cos 2x-\cos 2y)(2+\chi(\cos 2x+\cos 2y)),$$

$$w_{11}(x,y)=\frac{1}{16}(4(2-\chi)(y-x)+4(\chi-1)(\sin 2y-\sin 2x)-\chi(\sin 4y-\sin 4x)),$$

$$w_{12}(x,y)=\frac{1}{96}\begin{pmatrix}(36+24\chi)(y-x)+3(8+7\chi)(\sin 2y-\sin 2x)\\+3(1+2\chi)(\sin 4y-\sin 4x)+\chi(\sin 6y-\sin 6x)\end{pmatrix},$$

$$w_{13}(x,y)=\frac{1}{48}(\cos 2x-\cos 2y)\begin{pmatrix}6+2\chi+3(1+\chi)(\cos 2x+\cos 2y)+\chi(\cos 4x+\cos 4y)\\+\chi\cos[2(x-y)]+\chi\cos[2(x+y)]\end{pmatrix},$$

$$w_{14}(x,y)=\frac{1}{96}(12(y-x)+3\chi(\sin 2y-\sin 2x)-3(\sin 4y-\sin 4x)-\chi(\sin 6y-\sin 6x)),$$

$$w_{15}(x,y)=\frac{1}{96}\begin{pmatrix}6\chi\left(\cos^2 2x-\cos^2 2y\right)+3(\chi-4)(\cos 2y-\cos 2x)\\+3(\cos 4y-\cos 4x)+\chi(\cos 6y-\cos 6x)\end{pmatrix},$$

$$w_{16}(x,y)=\frac{1}{96}\begin{pmatrix}(36-24\chi)(y-x)+3(7\chi-8)(\sin 2y-\sin 2x)\\-3(2\chi-1)(\sin 4y-\sin 4x)+\chi(\sin 6y-\sin 6x)\end{pmatrix}.$$

For the incremental procedure outlined in Section 3, the change in the void ratio from an initial void ratio ν_0 is given by

$$\begin{aligned}\nu-\nu_0&=\frac{\text{Area}-\text{Area of particles}}{\text{Area}}-\frac{\text{Initial Area}-\text{Area of particles}}{\text{Initial Area}}\\&=(1-\nu)\left[\frac{\text{Area}-\text{Initial Area}}{\text{Initial Area}}\right],\end{aligned}$$

if ε_{11} and ε_{22} are much larger than ε_{12} and ε_{21} yet still much smaller than 1, $\nu_0 << 1$ and $\nu << 1$, then the change in void ratio can be approximated by $\nu-\nu_0=(1-\nu_0)[\varepsilon_{11}+\varepsilon_{22}]$.

Acknowledgements

The authors gratefully acknowledge the support of the US Army Research Office under grant number DAAD19-02-1-0216 and the Melbourne Research Development Grant Scheme.

References

1. A. Tordesillas and S.D.C. Walsh, Incorporating rolling resistance and contact anisotropy in micromechanical models of granular media, *Powder Technol.* 124 (2002) 106–111.
2. M. Oda and K. Iwashita (eds.), *Mechanics of Granular Materials: an Introduction*. A.A. Rotterdam: Balkema (1999) 383 pp.
3. L. Rothenburg, R.J. Bathurst and M.B. Dusseault, *Micromechanical Ideas in Constitutive Modelling of Granular Materials*, In: J. Biarez and R. Gourves, (eds.), *Powders and Grains*. Rotterdam: Balkema, (1989) 355–363.
4. S.D.C. Walsh and A. Tordesillas, Stranger than friction – micromechanics of granular media. In: S. Shoop and G. Blaisdell (eds.), *Proceedings, 14th International Conference of the ISTVS*. Vicksburg, Mississippi, USA (2002) pp. 1–10.
5. K. Iwashita and M. Oda, Micro-deformation mechanism of shear banding process based on modified distinct element method. *Powder Technol.* 109 (2000) 192–205.
6. D. Howell, R.P. Behringer and C. Veje, Stress fluctuations in a 2D granular Couette experiment: A continuous transition. *Phys. Rev. Lett.* 82 (1999) 5241–5244.
7. H.J. Hinrichsen, J. Feder and T. Jøssang, Random packing of disks in two dimensions. *Phys. Rev. A* 41 (1990) 4199–4209.
8. H.B. Mühlhaus and I. Vardoulakis, The thickness of shear bands in granular materials. *Geotechnique* 37 (1987) 271–283.
9. A. Tordesillas, J.F. Peters and B.S. Gardiner, Shear band evolution and accumulated microstructural development in Cosserat Media. *Int. J. Num. Analyt. Methods Geomech.* 28 (2004) 981–1010.
10. M. Oda and H. Kazama, Microstructure of shear bands and its relation to the mechanisms of dilatancy and failure of dense granular soils. *Geotechnique* 48 (1998) 465–481.
11. W.W. Harris, G. Viggiani, M.A. Mooney and R.J. Finno, Use of stereophotogrammetry to analyze the development of shear bands in sand. *Geotechn. Testing J.* 18 (1995) 405–420.
12. F. Calvetti, G. Combe and J. Lanier, Experimental micro-mechanical analysis of a 2D-granular material: Relation between structure evolution and loading path. *Mech. Cohesive-Frictional Mater.* 2 (1997) 121–163.
13. J.P. Bardet and J. Proubet, Shear-band analysis in idealized granular material. *J. Engng. Mech.* 118 (1992) 397–415.

Journal of Engineering Mathematics (2005) 52: 107–135
DOI 10.1007/s10665-004-3717-6

A hyperbolic well-posed model for the flow of granular materials

D. HARRIS and E.F. GREKOVA[1]
Department of Mathematics, UMIST, P.O. Box 88, Manchester, M60 1QD, UK. (E-mail: david.harris@umist.ac.uk). [1]*Laboratory of Dynamical Systems, IPME, St. Petersburg, Russia*

Received 12 July 2004; accepted in revised form 17 September 2004

Abstract. A plasticity model for the flow of granular materials is presented which is derived from a physically based kinematic rule and which is closely related to the double-shearing model, the double-sliding free-rotating model and also to the plastic-potential model. All of these models incorporate various notions of the concept of rotation-rate and the crucial idea behind the model presented here is that it identifies this rotation-rate with a property associated with a Cosserat continuum, namely, the intrinsic spin. As a consequence of this identification, the stress tensor may become asymmetric. For simplicity, in the analysis presented here, the material parameters are assumed to be constant. The central results of the paper are that (a) the model is hyperbolic for two-dimensional steady-state flows in the inertial regime and (b) the model possesses a domain of linear well-posedness. Specifically, it is proved that incompressible flows are well-posed.

Key words: granular materials, hyperbolic, rigid-plastic, well-posed

1. Introduction

The last 10 to 15 years have seen a rapid increase in research effort into the study of granular materials. Such materials, and systems involving them, exhibit complex and interesting behaviour. Civil engineers have long had a need to model soils, in which the densely packed grains, with liquid and gas occupying the interstices, exhibit solid-like behaviour, while chemical engineers have needed to model fluid suspensions, liquid and gas, in which granular material is dispersed in the fluid. In recent years, these traditional areas involving practical engineering problems, have attracted the attention of physicists, who have developed an interest in such systems both to understand their fundamental behaviour and also for the purpose of analogy when investigating other complex systems. In addition to this, many industries (for example, the chemical and food industries) handle granular solids and both storage and bulk flow give rise to problems which have a direct adverse economic effect. One method of trying to solve these problems is by way of obtaining a better theoretical understanding of the underlying principles of the physics and mechanics of granular materials.

In this paper we shall consider a mass of granular material occupying a region of space and undergoing a flow or deformation. The grains are assumed to be densely packed, *i.e.*, each grain is in contact with several of its neighbours and contact with a neighbouring grain is of finite duration and non-impulsive (*i.e.*, contact between grains is not modelled as instantaneous impact). The kinematics of such systems have been remarkably resistant to successful mathematical modelling. There are a number of different types of models which may be summarised as (1) discrete, (2) statistical mechanical and (3) continuum, in which the first and third appear to be more appropriate for the very densely packed systems considered here. The statistical–mechanical theories are associated with the work of Jenkins, see for example [1]. Despite the popularity of discrete modelling based upon Newtonian mechanics in which

the grains are modelled as small bodies and in which grain interactions may be modelled by including frictional, elastic and viscous effects in various ways, for large-scale systems computing power is still not adequate to solve those boundary- and initial-value problems of practical interest. For large systems the convenience of a continuum approach is a tempting goal, even though granular systems may appear to be at the borderline, or even beyond, that at which a continuum model may be considered applicable. Continuum models are also a useful framework within which to propose simple theories which may capture some aspect of the behaviour of the real material. And herein lies the problem: no continuum model has gained wide acceptance for its applicability to granular materials. Geotechnical and civil engineers usually use the so-called plastic-potential model, in which a yield condition (an algebraic inequality satisfied by the stress components) is assumed to hold in conjunction with the stress equilibrium equations and the strain increment or deformation-rate tensor is obtained by differentiation of the plastic potential with respect to the components of stress. If the yield condition and plastic potential are determined from the same function, the flow rule is called associated, otherwise it is called non-associated. Such plasticity models are now almost universal in soil mechanics; see for example the classic paper by Drucker and Prager [2], or, for a realistic plastic potential based upon careful experimentation, see [3], but they have not been adopted by researchers in any other field. It should be noted that there are dissenting voices even in geotechnical engineering, in particular there have been significant contributions due to G. Gudehus [4] and D. Kolymbas in the development of hypoplastic models; see [5]. An alternative class of models has been based directly upon physical arguments concerning the manner of flow of granular materials, but these again have not found widespread acceptance, finding support mainly from applied mathematicians. Historically, chemical engineers have studied systems in which the grains are in suspension, either in liquid or in a gas, and have thus treated the problem as one of fluid mechanics or rheology. Recently, physicists have conducted experiments on small-scale systems and are attempting to construct new types of model for such systems. It is also becoming more and more common to conduct computer simulations in lieu of performing real experiments and although there is something to be said in favour of such simulations, there is a danger that the distinction between simulation results and the results of real experiments is becoming blurred, particularly in view of the current fashion of referring to simulations as experiments! Discrete models, and the attendant simulations, have proponents in all disciplines, a key reference here is the classic paper by Cundall and Strack [6].

In this paper we wish to consider a unified plasticity theory based upon physical arguments. Such theories have had a number of theoretical problems associated with them, of which perhaps the most important is the loss of hyperbolicity and ill-posedness. Perhaps this latter term requires some explanation. The traditional definition of a well-posed problem is that it is one for which the solution exists, is unique and depends continuously on the boundary and initial data. An ill-posed problem, then, is merely the negation of this, one in which at least one of the above conditions fails to hold. Understood in this way, the phrase ill-posed problem can mean any one of a number of things, and in this sense is imprecise and not itself well (*i.e.*, uniquely) defined. However, we shall use the term in the following specific way. A set of partial differential equations is linearly ill-posed with respect to initial conditions if, given a solution to the initial-value problem, a sinusoidal perturbation of the given solution grows without bound in the limit of vanishingly short wavelengths. The basic facts concerning ill-posedness with regard to the models of granular materials were established in a number of important papers by Schaeffer and co-workers, [7–9], in which the method of frozen coefficients was used; see also, for a related paper, [10]. The implementation of the method

of frozen coefficients used here is based upon that used in [11]. It may happen that the mathematical ill-posedness reflects some strong physical instability and in such a case the ill-posedness may be acceptable. But in the absence of a strong physical instability it seems more likely that equations admitting such a strong mathematical instability cannot be an acceptable mathematical model of the physical process. In this case we must look to some property in the physics or the mechanics of the problem that will regularise the equations. In the case of granular materials there appears to be no consensus as to whether ill-posed equations form an acceptable model. On the one hand, there is reluctance among many researchers to accept ill-posed theories, on the other hand the fact that granular materials do exhibit unstable behaviour gives credence to the possibility that they be modelled in some sense by an ill-posed theory. There also appear to be many researchers, particularly those working on quasi-static problems, who are oblivious to, or even ignore, the fact that the model they work with is ill-posed.

We shall construct a theory which incorporates the following physical properties and attributes. We regard these as the minimum necessary requirements to construct an acceptable theory from the point of view of both theoretical and physical properties. These are:

(a) a pressure-dependent yield condition which introduces the concept of internal friction and frictionally generated resistive stresses. A yield condition is an essential part of many plasticity models and seems well-founded physically, taking an analogy with the Coulomb law of dry friction from Newtonian particle and rigid-body mechanics;

(b) packed discrete grains that may deform, but certainly cannot overlap, can only flow by one grain overtaking another or by being overtaken, and this suggests that the basic continuum deformation is by shear;

(c) the impenetrability condition for the grains means that the grains have to re-arrange themselves, riding up and down over each other, to flow past one another. This rearrangement gives rise to dilatation, *i.e.*, to volume and bulk density changes;

(d) individual grains may rotate and the rotations combine to affect the macroscopic flow. This gives rise to a continuum rate of rotation phenomenon distinct from that of the anti-symmetric part of the velocity-gradient tensor (or vorticity).

Whether or not the above conditions are sufficient for a workable continuum model for granular materials is an open question but they appear all to be necessary for a complete theory capable of explaining sufficient physical facts. We regard the ill-posedness of the plastic-potential model as a sign that anything less is insufficient. The ill-posedness of the double-shearing model is due to the choice of the rate of rotation of the principal axes of stress as a measure of the rate of rotation described in (d). In this paper we replace this quantity by a physical angular velocity (intrinsic spin), a primitive kinematic quantity that has a rotational inertia and stress associated with it: the stress tensor, in general, may be non-symmetric.

As mentioned above, it is sometimes stated that ill-posedness of the governing equations is merely a reflection of the unstable nature of granular flow. But there has, as yet, been no convincing demonstration that the mathematical ill-posedness is associated with a particular physical instability. The nature of the ill-posedness is too strong, too all-pervading to be a reflection, say, of the growth of a single shear-band. In the authors' opinion, a more likely hypothesis is that the models are mathematically ill-posed due to a missing or incorrect physical law or attribute or due to the various physical laws being combined in an inconsistent way. We have stated above the physical properties that we wish to introduce into the model; all that remains is to express these laws in such a manner, and in such a combination, that the model is well-posed.

It is, perhaps, necessary to emphasise the following point. Plasticity models are mainly used in the quasi-static regime. But for an ill-posed plasticity model there is, strictly speaking, no such thing as a quasi-static regime. The inertia terms may be negligible in many practical problems, but they are not identically zero. For an ill-posed model, no matter how infinitesimally small the inertia terms are, they will grow increasingly large with time. For this reason, in this paper, where we demonstrate that our plasticity model is well-posed, it is essential that we retain the inertia terms in both the translational and rotational equations of motion.

Since no model within the confines of classical continuum mechanics has found general acceptance, the possibility must be faced that perhaps no such model exists. Indeed, many researchers have turned to the framework of a Cosserat continuum in order to regularise the equations, see for example [12]. However, such models are greatly complicated in comparison to standard models due to the presence of couple-stress. Further, the standard plastic-potential model and the double-shearing model both appear to capture successfully some aspects of the behaviour of granular materials. In this paper we wish to preserve as much as possible the relative simplicity of these latter models and hence wish to use a framework which extends the classical continuum as little as possible and to obtain equations as similar as possible to the plastic-potential and double-shearing models. As far as the authors are aware there is no conclusive experimental evidence for the existence of couple stress in granular materials and its introduction greatly complicates the equations. Moreover, the existence of grain rotation in granular materials is very much self-evident and yet researchers seem to have overlooked the possibility that incorporating some measure of grain rotation may itself be sufficient to both regularise the equations and to produce a model capable of explaining the major features of the bulk flow of granular materials. For this reason we introduce a Cosserat continuum in which there is no couple-stress but there is both rotation and rotational inertia and refer to this as a reduced Cosserat continuum.

For simplicity we shall consider a perfectly plastic model, *i.e.*, one in which the material parameters are constants and the conclusions presented here have been proved in this context. But it should be pointed out that the model may easily be extended to include density-hardening or softening, or indeed any other type of hardening or softening. In fact, since the material cannot undergo dilatation indefinitely, it is necessary for the parameter governing dilatation to reduce in magnitude to zero as the deformation or flow proceeds in order to obtain a realistic model. In this paper we shall ultimately avoid the question of the evolution of dilatancy and its effect on well-posedness by considering the important special case of incompressible flows.

2. The mathematical model

Consider a body $\mathcal{B}$ comprising a mass of granular material occupying a region $\mathcal{R}$ at time t in three-dimensional space and let $\mathcal{F}$ denote an inertial frame from which to observe $\mathcal{B}$. In the standard continuum model, the following field variables are defined at each spatial point P of $\mathcal{R}$. Relative to $\mathcal{F}$, let $\mathbf{v}$, σ, ρ denote the velocity, Cauchy stress and bulk density of the granular material at the material point of $\mathcal{B}$ instantaneously occupying the point P. We shall consider an enhanced continuum model, a type of Cosserat continuum, in which what is usually termed a material particle, in addition to the above, also possesses some of the attributes of a rigid body, namely an intrinsic spin ω and a moment of inertia $\mathbf{I}$, and for this reason is referred to as a material-point body. A standard Cosserat continuum possesses a further physical quantity, namely the couple-stress. However, couple-stress will not appear in our equations and so we refrain from introducing it into our continuum, and for this reason we call

it a reduced Cosserat continuum. The quantity ω is an extra field variable and $\mathbf{I}$ represents a material property. As a result of their introduction the stress tensor will not, in general, be symmetric. In order to give an intuitive meaning to ω and $\mathbf{I}$, we may note that $\mathbf{I}$ is a second-order symmetric tensor and, as such, there exist three mutually perpendicular directions (which will not be unique if the principal values of inertia are not distinct) relative to which the matrix representation of $\mathbf{I}$ is diagonal. This triad of directions is fixed in the material-point body. During the deformation or flow the orientation of this triad will, in general, vary and ω is the spin or angular velocity of this triad.

In this paper we shall restrict consideration to planar flows of $\mathcal{B}$. Let $\mathbf{\Gamma}$ denote the velocity-gradient tensor with components defined by

$$\Gamma_{ij} = \frac{\partial v_i}{\partial x_j} \tag{1}$$

and define the deformation-rate tensor, $\mathbf{d}$, and the spin tensor, $\mathbf{s}$, with components

$$d_{ij} = \tfrac{1}{2}\left(\Gamma_{ij} + \Gamma_{ji}\right), \quad s_{ij} = \tfrac{1}{2}\left(\Gamma_{ij} - \Gamma_{ji}\right) \tag{2}$$

as its symmetric and anti-symmetric parts, respectively. Let Ox_i, Ox_i' denote two sets of rectangular Cartesian co-ordinate axes, in the plane of flow, with the latter inclined at an angle ϑ (measured anti-clockwise positive) to the former and denote the components of σ by σ_{ij}, σ_{ij}' relative to each set of axes, respectively, where the subscripts take the values 1 and 2. The stress components transform as

$$\begin{aligned}
\sigma_{11}' &= \tfrac{1}{2}\left[\sigma_{11} + \sigma_{22} + (\sigma_{11} - \sigma_{22})\cos 2\vartheta + (\sigma_{21} + \sigma_{12})\sin 2\vartheta\right], \\
\sigma_{12}' &= \tfrac{1}{2}\left[\sigma_{12} - \sigma_{21} - (\sigma_{11} - \sigma_{22})\sin 2\vartheta + (\sigma_{21} + \sigma_{12})\cos 2\vartheta\right], \\
\sigma_{21}' &= \tfrac{1}{2}\left[\sigma_{21} - \sigma_{12} - (\sigma_{11} - \sigma_{22})\sin 2\vartheta + (\sigma_{21} + \sigma_{12})\cos 2\vartheta\right], \\
\sigma_{22}' &= \tfrac{1}{2}\left[\sigma_{11} + \sigma_{22} - (\sigma_{11} - \sigma_{22})\cos 2\vartheta - (\sigma_{21} + \sigma_{12})\sin 2\vartheta\right].
\end{aligned} \tag{3}$$

It is convenient to define the invariant quantities

$$p_\sigma = -\tfrac{1}{2}\left(\sigma_{11} + \sigma_{22}\right), \quad r_\sigma = \tfrac{1}{2}\left(\sigma_{12} - \sigma_{21}\right), \quad q_\sigma = \tfrac{1}{2}\left[(\sigma_{11} - \sigma_{22})^2 + (\sigma_{21} + \sigma_{12})^2\right]^{1/2} \tag{4}$$

and to define ψ_σ, the angle that the greater principal direction of the *symmetric* part of the stress makes with the x_1-axis, by

$$\tan 2\psi_\sigma = \frac{\sigma_{21} + \sigma_{12}}{\sigma_{11} - \sigma_{22}} \tag{5}$$

and then

$$\begin{aligned}
\sigma_{11} &= -p_\sigma + q_\sigma \cos 2\psi_\sigma, \qquad \sigma_{22} = -p_\sigma - q_\sigma \cos 2\psi_\sigma, \\
\sigma_{12} &= r_\sigma + q_\sigma \sin 2\psi_\sigma, \qquad \sigma_{21} = -r_\sigma + q_\sigma \sin 2\psi_\sigma.
\end{aligned} \tag{6}$$

2.1. The stress equations of motion

Let ∂_t, ∂_i denote partial differentiation with respect to time t, x_i, $i = 1,2$, respectively; then the stress equations of motion are

$$\begin{aligned}
&\rho\partial_t v_1 + \rho v_1 \partial_1 v_1 + \rho v_2 \partial_2 v_1 - \partial_1 \sigma_{11} - \partial_2 \sigma_{21} - \rho F_1 = 0, \\
&\rho\partial_t v_2 + \rho v_1 \partial_1 v_2 + \rho v_2 \partial_2 v_2 - \partial_1 \sigma_{12} - \partial_2 \sigma_{22} - \rho F_2 = 0,
\end{aligned} \tag{7}$$

where F_1, F_2 denote the components of the body force per unit mass. From Equations (4) and (7) we obtain

$$\begin{aligned}
&\rho\partial_t v_1 + \rho v_1\partial_1 v_1 + \rho v_2\partial_2 v_1 + \partial_1 p_\sigma - \cos 2\psi_\sigma\partial_1 q_\sigma + 2q_\sigma \sin 2\psi_\sigma\partial_1\psi_\sigma \\
&\qquad + \partial_2 r_\sigma - \partial_2 q_\sigma \sin 2\psi_\sigma - 2q_\sigma \cos 2\psi_\sigma\partial_2\psi_\sigma - \rho F_1 = 0, \\
&\rho\partial_t v_2 + \rho v_1\partial_1 v_2 + \rho v_2\partial_2 v_2 - \partial_1 r_\sigma - \sin 2\psi_\sigma\partial_1 q_\sigma - 2q_\sigma \cos 2\psi_\sigma\partial_1\psi_\sigma \\
&\qquad + \partial_2 p_\sigma + \partial_2 q_\sigma \cos 2\psi_\sigma - 2q_\sigma \sin 2\psi_\sigma\partial_2\psi_\sigma - \rho F_2 = 0.
\end{aligned} \tag{8}$$

The equation of rotational motion for a reduced Cosserat continuum may be written

$$\rho I\,(\partial_t\omega + v_1\partial_1\omega + v_2\partial_2\omega) + \rho\,(\partial_t I + v_1\partial_1 I + v_2\partial_2 I)\,\omega - 2r_\sigma - \rho G = 0, \tag{9}$$

where G is the body moment per unit mass. In this paper we assume that the moment-of-inertia tensor $\mathbf{I}$ is prescribed. At each point P, it represents, effectively, the moment of inertia of a representative volume element in the real material. A lower bound for this volume is the volume occupied by a single grain, but, in general, it will be a mesoscopic domain containing several grains and the attendant void space. In a standard Cosserat continuum, the equation of rotational motion would also contain the couple-stress.

The constitutive equation must be such as to allow the possibility of a non-symmetric state of stress. In this paper our intention is to propose a theory which is as similar as possible to standard plasticity models. Here, we shall adopt the simplest method of incorporating a possibly asymmetric stress tensor into the theory, namely, we shall assume that the symmetric and anti-symmetric parts of the stress may be treated separately, *i.e.*, they are independent of each other. We shall assume that the symmetric part of the stress satisfies a yield condition of the form

$$f\,(p_\sigma, q_\sigma, \rho) \le 0, \tag{10}$$

where the function f is such that an angle of internal friction

$$\sin\phi = -f_p/f_q \tag{11}$$

may be defined, and where subscripts p, q indicate partial differentiation with respect to p_σ, q_σ, respectively. The simplest and most important yield condition of this type is the Mohr–Coulomb criterion, where if τ_n^s denotes the tangential traction associated with the symmetric part of the stress; then

$$\left|\tau_n^s\right| \le -\sigma_n \tan\phi + k, \tag{12}$$

where σ_n denotes the normal component of traction across a surface with normal direction $\mathbf{n}$ and ϕ, k denote the angle of internal friction and cohesion, respectively. In invariant form the Coulomb yield condition may be written

$$q_\sigma \le p_\sigma \sin\phi + k\cos\phi. \tag{13}$$

In addition, we introduce a rotational yield condition of the form

$$|r_\sigma| \le m, \tag{14}$$

where $m > 0$ is a material parameter, which we will call the rotational yield strength. We shall regard m as a new material parameter, independent of both ϕ and k. Note that (14) is not

differentiable at $r_\sigma = 0$. It is possible for the stress tensor to be symmetric, $r_\sigma = 0$, and so we incorporate this case together with the case of equality in the inequality (14) by writing

$$r_\sigma = \epsilon m, \tag{15}$$

where ϵ may take one of the values $+1, 0, -1$ according as to whether r_σ is positive, zero or negative. Introducing a rotational yield condition in a rigid–plastic context brings with it the usual problem of indeterminacy, albeit in a slightly altered form. In classical plasticity the stress is indeterminate in the so-called rigid region, *i.e.*, in the region where $f(p_\sigma, q_\sigma, \rho) < 0$. In the case of a rotational yield condition, the asymmetric part of the stress is zero in the rotationally quasi-static regime. Whereas the linear momentum equations involve the divergence of the stress, allowing a multiplicity of solutions to the equilibrium equations, the balance of angular momentum involves the asymmetric stress directly, so a rotationally quasi-static flow leads to a symmetric stress tensor and to an indeterminacy, up to a constant, in the intrinsic spin field and hence also in the velocity field. We may call this the de Josselin de Jong indeterminacy, since it was first discovered by him in his double-sliding free-rotating model, [13]. The indeterminacy is removed in cases where there is a boundary condition to be satisfied by ω, or if physical arguments may be deployed to show that $\omega = 0$. An alternative way to remove the indeterminacy is to include rotational elasticity (presumably in the context of a full elastic-plastic model for the symmetric part of the stress as well) or by using rotational viscosity instead of the rotational yield condition. We will not pursue this idea in this paper and shall consider only the case of a rotational yield condition. Also, we wish to construct a generalisation of classical plasticity that is well-posed, but which is as similar to the classical case as possible, so we shall concentrate here on regions in which the translational yield condition is satisfied. Thus, the following regions are considered:

1. the indeterminate region $\mathcal{R}_i$, where one or both of $q_\sigma < p_\sigma \sin\phi + k\cos\phi$ and $r_\sigma \neq \epsilon m$ hold;
2. the fully deforming region $\mathcal{R}_d$, where $q_\sigma = p_\sigma \sin\phi + k\cos\phi$ and $r_\sigma = \epsilon m$.

Henceforth in this paper we shall consider only the region $\mathcal{R}_d$. In this region we may eliminate q_σ, r_σ from the stress equations of motion using the yield conditions (13) and (14) with equality holding, to obtain the equations

$$\begin{aligned}&\rho\partial_t v_1 + \rho v_1\partial_1 v_1 + \rho v_2\partial_2 v_1 + (1 - \sin\phi\cos 2\psi_\sigma)\,\partial_1 p_\sigma - \sin\phi\sin 2\psi_\sigma\,\partial_2 p_\sigma \\ &\quad + 2q_\sigma \sin 2\psi_\sigma\,\partial_1\psi_\sigma - 2q_\sigma\cos 2\psi_\sigma\,\partial_2\psi_\sigma - \rho F_1 = 0,\end{aligned} \tag{16}$$

$$\begin{aligned}&\rho\partial_t v_2 + \rho v_1\partial_1 v_2 + \rho v_2\partial_2 v_2 - \sin\phi\sin 2\psi_\sigma\,\partial_1 p_\sigma + (1 + \sin\phi\cos 2\psi_\sigma)\,\partial_2 p_\sigma \\ &\quad - 2q_\sigma\cos 2\psi_\sigma\,\partial_1\psi_\sigma - 2q_\sigma\sin 2\psi_\sigma\,\partial_2\psi_\sigma - \rho F_2 = 0.\end{aligned} \tag{17}$$

Also, the continuity equation governing the evolution of the bulk density reads

$$\partial_t\rho + v_1\partial_1\rho + v_2\partial_2\rho + \rho\partial_1 v_1 + \rho\partial_2 v_2 = 0. \tag{18}$$

3. The kinematic equations

In this section we derive a pair of kinematic equations governing the flow of the ideal material, which are formally identical to the double-shearing equations due to Spencer [14] for incompressible materials and extended by Mehrabadi and Cowin [15] to dilatant materials. See also [16] for further developments of the model. The double-sliding free-rotating model, [17], was also extended to dilatant materials in a similar way to that of Mehrabadi and Cowin

[15]. However, one quantity occurring in the equations is here given a different interpretation to the above models. For a unified derivation of the double-shearing, double-sliding free-rotating and plastic-potential models, see [18]. It should be noted that single-shearing models have also been proposed; see for example [19–21], but it is difficult with such models to admit a sufficiently wide variety of flows. The derivation of the kinematic equations governing the flow for the model presented here is based on that of Harris [18].

Define the quantity ε by

$$\varepsilon = \pi/4 + \phi/2; \tag{19}$$

then at each point of $\mathcal{R}_d$, *i.e.*, the region in which the Coulomb yield condition is satisfied, it is well known that the inequality (12) is satisfied across line segments directed at angles of $\psi_\sigma + \varepsilon$, $\psi_\sigma - \varepsilon$ to the x_1-axis. We shall refer to these two directions, symmetric on either side of the major principal stress direction, as the α_1- and α_2-directions, respectively, or as the Coulomb yield directions associated with the inequality (12). Let $\mathbf{t}_{\alpha_i}$ denote a unit vector in the α_i-direction.

We shall also define a second pair of (non-coincident) directions which characterise the dilatancy of the material and which make angles of $\psi_\sigma - \varepsilon + \nu$, $\psi_\sigma + \varepsilon - \nu$ with the positive x_1-direction and will be referred to as the β_1-, β_2-directions, respectively. Thus, the β-directions are defined in terms of the α-directions and ν, where ν is called the *angle of dilatancy* and is to be regarded as a material parameter. In an incompressible material, $\nu = 0$ and in this case the β_1-(β_2-)direction coincides with the α_2-(α_1-)direction. Let $\mathbf{t}_{\beta_i}$ denote a unit vector in the β_i-direction.

Let P denote a material-point body in $\mathcal{R}_d$ and let Q denote a material-point body in an infinitesimal neighbourhood of P, with Q distinct from P. Let $\mathbf{ds}_P^Q$ denote the infinitesimal position vector of Q relative to P. Let Q_{α_i} denote the point of intersection of the α_i-direction through P with the α_j-direction through Q, where, here and in the sequel, $j = 2$ when $i = 1$ and $j = 1$ when $i = 2$. Consider the material body comprising the set of all material-point bodies in an infinitesimal neighbourhood of P and also, for fixed but arbitrary Q, consider the material body instantaneously occupying the infinitesimal parallelogram $PQ_{\alpha_1}Q_{\alpha_2}Q$. Let $\mathbf{dv}_P^{Q_{\alpha_i}}$ denote the velocity of the material point body at Q_{α_i} relative to that at P, $\mathbf{ds}_P^{Q_{\alpha_i}}$ denote the position vector of Q_{α_i} relative to P and let $\mathbf{n}_{\alpha_i}$ denote the normal to the α_i-direction (measured anti-clockwise positive from the direction $\mathbf{t}_{\alpha_i}$).

The kinematic model comprises two separate postulates. Firstly, the manner of the instantaneous rate of deformation of the material occupying the parallelogram $PQ_{\alpha_1}Q_{\alpha_2}Q$ in terms of the two adjacent sides PQ_{α_1}, PQ_{α_2} is specified by prescribing the velocities of the material-point bodies at Q_{α_1} and Q_{α_2} relative to P. Secondly, the manner in which the intrinsic spin gives rise to a velocity of Q relative to P is prescribed. It should be noted that this is a purely continuum hypothesis, no attempt being made to average the micro-mechanical velocities and rotations. The model is to be validated, or refuted, by comparison of its predictions with the properties of real granular materials.

3.1. The proposed kinematic hypothesis

The first postulate is that, relative to the inertial frame $\mathcal{F}$, the velocity of the material-point body at Q_{α_i} relative to P, $\mathbf{dv}_P^{Q_{\alpha_i}}$, is given by

$$\mathbf{dv}_P^{Q_{\alpha_i}} = k_{\beta_i} \left| \mathbf{ds}_P^{Q_{\alpha_i}} \right| \mathbf{t}_{\beta_i} \cos\phi, \tag{20}$$

where k_{β_i} is a proportionality factor called the *shear strength* in the β_i-direction and i takes the values 1 and 2. We shall call Equation (20) a β_i-direction dilatant shear on the α_i-direction, since both material-point bodies P and Q_{α_i} lie on the same α_i-line segment and their relative velocity is directed in the β_i-direction. For an arbitrary point Q in the neighbourhood of P the velocity of the material-point body at Q relative to that at P due to the two dilatant shears of Equation (20) is then postulated to be given by their sum

$$\mathbf{dv}_P^Q\Big|_{\mathcal{S}} = \mathbf{dv}_P^{Q_{\alpha_1}} + \mathbf{dv}_P^{Q_{\alpha_2}}, \tag{21}$$

relative to the inertial frame $\mathcal{F}$. The α_1- and α_2-directions are the shear directions and we shall also refer to them as the *slip directions*. A fundamental postulate of the model is that the slip directions coincide with the Coulomb yield directions. It should be carefully noted that Equations (20) and (21) hold at each instant of time. As the flow proceeds, the vectors $\mathbf{t}_{\alpha_i}$ and $\mathbf{t}_{\beta_i}$, $i=1,2$, will, in general, vary their orientation in space and, at each instant, the dilatant shears (20) are defined relative to the current orientations of the $\mathbf{t}_{\alpha_i}$, $\mathbf{t}_{\beta_i}$.

The second postulate is that, again relative to the inertial frame $\mathcal{F}$, the velocity of the material-point body at Q relative to that at P due to the intrinsic spin is the local rigid rotation

$$\mathbf{dv}_P^Q\Big|_{\mathcal{R}} = \omega \times \mathbf{ds}_P^Q. \tag{22}$$

Thus, the resultant relative velocity of Q relative to P, relative to the inertial frame $\mathcal{F}$, is

$$\mathbf{dv}_P^Q = \mathbf{dv}_P^Q\Big|_{\mathcal{S}} + \omega \times \mathbf{ds}_P^Q\big|_{\mathcal{R}}, \tag{23}$$

where ω is evaluated at P.

This is the complete kinematic rule for the local rate of deformation. Thus, for any point Q in the neighbourhood of P the velocity of the material point body at Q relative to that at P may be written

$$\mathbf{dv}_P^Q = k_{\beta_1}\left|\mathbf{ds}_P^{Q_{\alpha_1}}\right|\mathbf{t}_{\beta_1}\cos\phi + k_{\beta_2}\left|\mathbf{ds}_P^{Q_{\alpha_2}}\right|\mathbf{t}_{\beta_2}\cos\phi + \omega\times\mathbf{ds}_P^Q, \tag{24}$$

relative to $\mathcal{F}$.

3.2. The standard double-shearing kinematic hypothesis

It is instructive to compare the kinematic hypothesis given above, with that of the standard double-shearing model in terms of the method and notation of this paper. In order to state the hypothesis, a second frame of reference, this time non-inertial, is required. Let $\mathcal{G}$ denote a frame of reference fixed with respect to the principal axes of stress. Since the principal axes of stress may rotate, these axes form another spinning triad. Let ω_ψ denote the spin (angular velocity) of this triad relative to the inertial frame $\mathcal{F}$. The standard double-shearing model is also based upon Equations (20) and (21) but now we take these equations to be true relative to the frame $\mathcal{G}$. Since we require our constitutive equations to be expressed relative to the inertial frame $\mathcal{F}$, we must add the relative velocity $\mathbf{dv}_P^Q\Big|_{\mathcal{F}} = \omega_\psi\times\mathbf{ds}_P^Q$ due to the rigid spin ω_ψ of the frame $\mathcal{G}$ relative to $\mathcal{F}$. The complete kinematic hypothesis for the standard double-shearing model is then

$$\mathbf{dv}_P^Q = \mathbf{dv}_P^{Q_{\alpha_1}} + \mathbf{dv}_P^{Q_{\alpha_2}} + \omega_\psi\times\mathbf{ds}_P^Q \tag{25}$$

where the $\mathbf{dv}_P^{Q_{\alpha_i}}$ are again given by Equation (20). We see that the two kinematic hypotheses are formally identical, and will hence give rise to formally identical equations. In the planar

case, where the orientation of the principal axes of stress in the plane of deformation is determined by ψ_σ, it is clear that ω_ψ is expressible in terms of the material time derivative of ψ_σ (material derivative since the material spins at this rate). Since the double-shearing equations have been derived from an equation essentially equivalent to Equation (25) in [18], we shall refrain from deriving them again here.

Now, the assertion that we observe two dilatant shears relative to axes fixed relative to the principal axes of the Cauchy stress tensor is equivalent to the assumption that, locally, the material rotates with these axes. This is an unusual, if not unique, assumption in continuum mechanics. It is, ultimately, this assumption that renders the double-shearing equations ill-posed, that causes the instability of a time-dependent simple shear, and is the cause of the discrepancy between the theory and certain experimental data; see [22] and [23]. The postulate that (20) and (21) are true relative to the frame $\mathcal{G}$ implies that the material rotates locally with the principal axes of stress and this is tantamount to postulating a physical law. It is this physical law that the authors identify as causing both the theoretical difficulties of ill-posedness and instability and also the discrepancy with the available experimental data. It is this physical law that the authors seek to replace with another law, the proposed kinematic hypothesis (25), which does not suffer from these disadvantages.

3.3. The proposed kinematic equations

We see then, that the essential difference between the standard double-shearing kinematic hypothesis and the one proposed here is that one spin of a triad, the spin of the principal axes of stress, is replaced by the spin of another triad, the spin of the principal axes of inertia. The proposed kinematic hypothesis is more in keeping with the double-sliding free-rotating model due to de Josselin de Jong [2,17]. However, this latter model is indeterminate, consisting, in one formulation, of a system of equations with more unknowns than equations (the unknown not matched with an equation being ω) and in a second formulation, of a set of inequalities. The standard double-shearing model represents one method of closing the double-sliding free-rotating equations, by prescribing the rotation to coincide with the rotation of the principal axes of stress. We may view the model presented here as an alternative closure of the double-sliding free-rotating equations, in which the rotation is identified with the spin of the material-point body and the system is closed by the equation of rotational motion. We turn now to the derivation of the equations arising from the proposed kinematic hypothesis (24). Now the angle between the α_1-direction and the normal to the α_2-direction is

$$\psi_\sigma + \varepsilon - (\psi_\sigma - \varepsilon + \pi/2) = \phi, \tag{26}$$

while the angle between the α_2-direction and the normal to the α_1-direction is

$$\psi_\sigma - \varepsilon - (\psi_\sigma + \varepsilon + \pi/2) = -\phi - \pi. \tag{27}$$

Hence the projections of $\mathbf{ds}_P^{Q\alpha_i}$ and $\mathbf{ds}_P^Q$ onto the direction $\mathbf{n}_{\alpha_j}$ are equal and may be written as

$$\left|\mathbf{ds}_P^{Q\alpha_1}\right| \cos\phi = \mathbf{n}_{\alpha_2} \cdot \mathbf{ds}_P^Q, \quad \left|\mathbf{ds}_P^{Q\alpha_2}\right| \cos\phi = -\mathbf{n}_{\alpha_1} \cdot \mathbf{ds}_P^Q. \tag{28}$$

Hence

$$\mathbf{dv}_P^Q = k_{\beta_1} \left(\mathbf{n}_{\alpha_2} \cdot \mathbf{ds}_P^Q\right) \mathbf{t}_{\beta_1} - k_{\beta_2} \left(\mathbf{n}_{\alpha_1} \cdot \mathbf{ds}_P^Q\right) \mathbf{t}_{\beta_2} + \omega \times \mathbf{ds}_P^Q, \tag{29}$$

i.e.,

$$\mathbf{dv}_P^Q = \left[k_{\beta_1} \left(\mathbf{t}_{\beta_1} \otimes \mathbf{n}_{\alpha_2}\right) - k_{\beta_2} \left(\mathbf{t}_{\beta_2} \otimes \mathbf{n}_{\alpha_1}\right) + \Omega\right] \cdot \mathbf{ds}_P^Q, \tag{30}$$

where Ω denotes the anti-symmetric tensor dual to ω. We now suppose that the velocity and rotational fields are sufficiently smooth in terms of the velocity-gradient tensor

$$\mathbf{dv}_P^Q = \boldsymbol{\Gamma} \cdot \boldsymbol{ds}_P^Q \tag{31}$$

and so

$$\boldsymbol{\Gamma} = k_{\beta_1} \left(\mathbf{t}_{\beta_1} \otimes \mathbf{n}_{\alpha_2}\right) - k_{\beta_2} \left(\mathbf{t}_{\beta_2} \otimes \mathbf{n}_{\alpha_1}\right) + \Omega, \tag{32}$$

where

$$\Omega = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}. \tag{33}$$

Now,

$$\begin{aligned} &\mathbf{t}_{\beta_1} = (\cos(\psi_\sigma - \varepsilon + \nu), \sin(\psi_\sigma - \varepsilon + \nu)), \quad \mathbf{t}_{\beta_2} = (\cos(\psi_\sigma + \varepsilon - \nu), \sin(\psi_\sigma + \varepsilon - \nu)), \\ &\mathbf{n}_{\alpha_1} = (-\sin(\psi_\sigma + \varepsilon), \cos(\psi_\sigma + \varepsilon)), \quad \mathbf{n}_{\alpha_2} = (-\sin(\psi_\sigma - \varepsilon), \cos(\psi_\sigma - \varepsilon)) \end{aligned} \tag{34}$$

and

$$\mathbf{t}_{\beta_1} \otimes \mathbf{n}_{\alpha_2} = \begin{bmatrix} -\cos(\psi_\sigma - \varepsilon + \nu)\sin(\psi_\sigma - \varepsilon) & \cos(\psi_\sigma - \varepsilon + \nu)\cos(\psi_\sigma - \varepsilon) \\ -\sin(\psi_\sigma - \varepsilon + \nu)\sin(\psi_\sigma - \varepsilon) & \sin(\psi_\sigma - \varepsilon + \nu)\cos(\psi_\sigma - \varepsilon) \end{bmatrix}, \tag{35}$$

$$\mathbf{t}_{\beta_2} \otimes \mathbf{n}_{\alpha_1} = \begin{bmatrix} -\cos(\psi_\sigma + \varepsilon - \nu)\sin(\psi_\sigma + \varepsilon) & \cos(\psi_\sigma + \varepsilon - \nu)\cos(\psi_\sigma + \varepsilon) \\ -\sin(\psi_\sigma + \varepsilon - \nu)\sin(\psi_\sigma + \varepsilon) & \sin(\psi_\sigma + \varepsilon - \nu)\cos(\psi_\sigma + \varepsilon) \end{bmatrix}. \tag{36}$$

Hence

$$\begin{aligned} \partial_1 v_1 &= -k_{\beta_1}\cos(\psi_\sigma - \varepsilon + \nu)\sin(\psi_\sigma - \varepsilon) + k_{\beta_2}\cos(\psi_\sigma + \varepsilon - \nu)\sin(\psi_\sigma + \varepsilon), \\ \partial_2 v_1 &= k_{\beta_1}\cos(\psi_\sigma - \varepsilon + \nu)\cos(\psi_\sigma - \varepsilon) - k_{\beta_2}\cos(\psi_\sigma + \varepsilon - \nu)\cos(\psi_\sigma + \varepsilon) - \omega, \\ \partial_1 v_2 &= -k_{\beta_1}\sin(\psi_\sigma - \varepsilon + \nu)\sin(\psi_\sigma - \varepsilon) + k_{\beta_2}\sin(\psi_\sigma + \varepsilon - \nu)\sin(\psi_\sigma + \varepsilon) + \omega, \\ \partial_2 v_2 &= k_{\beta_1}\sin(\psi_\sigma - \varepsilon + \nu)\cos(\psi_\sigma - \varepsilon) - k_{\beta_2}\sin(\psi_\sigma + \varepsilon - \nu)\cos(\psi_\sigma + \varepsilon), \end{aligned}$$

from which we obtain

$$\begin{aligned} \partial_1 v_1 + \partial_2 v_2 &= \left(k_{\beta_1} + k_{\beta_2}\right)\sin\nu, \\ \partial_1 v_1 - \partial_2 v_2 &= -k_{\beta_1}\sin(2\psi_\sigma - 2\varepsilon + \nu) + k_{\beta_2}\sin(2\psi_\sigma + 2\varepsilon - \nu) \\ &= k_{\beta_1}\cos(2\psi_\sigma - \phi + \nu) + k_{\beta_2}\cos(2\psi_\sigma + \phi - \nu), \\ \partial_2 v_1 + \partial_1 v_2 &= k_{\beta_1}\cos(2\psi_\sigma - 2\varepsilon + \nu) - k_{\beta_2}\cos(2\psi_\sigma + 2\varepsilon - \nu) \\ &= k_{\beta_1}\sin(2\psi_\sigma - \phi + \nu) + k_{\beta_2}\sin(2\psi_\sigma + \phi - \nu), \\ \partial_2 v_1 - \partial_1 v_2 &= \left(k_{\beta_1} - k_{\beta_2}\right)\cos\nu - 2\omega. \end{aligned} \tag{37}$$

From Equations $(37)_{\text{ii}}$ and $(37)_{\text{iii}}$ we may solve for the quantities k_{β_1} and k_{β_2} in terms of the components of the deformation-rate tensor to obtain

$$\begin{aligned} (d_{11} - d_{22})\sin(2\psi_\sigma + \phi - \nu) - 2d_{12}\cos(2\psi_\sigma + \phi - \nu) &= k_{\beta_1}\sin 2(\phi - \nu), \\ (d_{11} - d_{22})\sin(2\psi_\sigma - \phi + \nu) - 2d_{12}\cos(2\psi_\sigma - \phi + \nu) &= -k_{\beta_2}\sin 2(\phi - \nu), \end{aligned} \tag{38}$$

provided $\nu \neq \phi$. Subtracting and adding Equations (38) gives

$$\begin{aligned} \left(k_{\beta_1} + k_{\beta_2}\right)\sin 2(\phi - \nu) &= (d_{11} - d_{22})\left[\sin(2\psi_\sigma + \phi - \nu) - \sin(2\psi_\sigma - \phi + \nu)\right] \\ &\quad + 2d_{12}\left[\cos(2\psi_\sigma - \phi + \nu) - \cos(2\psi_\sigma + \phi - \nu)\right] \\ &= 2\left[(d_{11} - d_{22})\cos 2\psi_\sigma + 2d_{12}\sin 2\psi_\sigma\right]\sin(\phi - \nu), \end{aligned}$$

i.e.,

$$\left(k_{\beta_1}+k_{\beta_2}\right)\cos\left(\phi-\nu\right)=\left(d_{11}-d_{22}\right)\cos 2\psi_\sigma+2d_{12}\sin 2\psi_\sigma \tag{39}$$

and

$$\begin{aligned}\left(k_{\beta_1}-k_{\beta_2}\right)\sin 2\left(\phi-\nu\right)&=\left(d_{11}-d_{22}\right)\left[\sin\left(2\psi_\sigma+\phi-\nu\right)+\sin\left(2\psi_\sigma-\phi+\nu\right)\right]\\&\quad-2d_{12}\left[\cos\left(2\psi_\sigma+\phi-\nu\right)+\cos\left(2\psi_\sigma-\phi+\nu\right)\right]\\&=2\left[\left(d_{11}-d_{22}\right)\sin 2\psi_\sigma-2d_{12}\cos 2\psi_\sigma\right]\cos\left(\phi-\nu\right),\end{aligned}$$

i.e.,

$$\left(k_{\beta_1}-k_{\beta_2}\right)\sin\left(\phi-\nu\right)=\left(d_{11}-d_{22}\right)\sin 2\psi_\sigma-2d_{12}\cos 2\psi_\sigma. \tag{40}$$

Eliminating the quantities $k_{\beta_1}+k_{\beta_2}$ and $k_{\beta_1}-k_{\beta_2}$, between Equations $(37)_{\mathrm{i}}$, $(37)_{\mathrm{iv}}$, (39) and (40) gives the following pair of equations

$$d_{11}+d_{22}=\frac{\sin\nu}{\cos\left(\phi-\nu\right)}\left[\left(d_{11}-d_{22}\right)\cos 2\psi_\sigma+2d_{12}\sin 2\psi_\sigma\right], \tag{41}$$

$$2\left(\omega-s_{21}\right)=\frac{\cos\nu}{\sin\left(\phi-\nu\right)}\left[\left(d_{11}-d_{22}\right)\sin 2\psi_\sigma-2d_{12}\cos 2\psi_\sigma\right]. \tag{42}$$

Equations (41) and (42) are the required constitutive equations governing the flow. They are formally identical to the Mehrabadi–Cowin equations [15]; however, the quantity ω is here interpreted as the intrinsic spin, whereas in the Mehrabadi–Cowin equations it is interpreted as the material derivative of the quantity ψ_σ.

Equations (41) and (42) are frame-indifferent. To see this, we demonstrate that they are unchanged in form under a superposed rigid-body motion. Consider two velocity and intrinsic spin fields (v_1, v_2, ω), $(v_1^{(1)}, v_2^{(1)}, \omega^{(1)})$, differing only by a rigid-body spin $(0, 0, \Omega)$, measured anti-clockwise positive, then substituting

$$v_1=v_1^{(1)}-\Omega x_2,\quad v_2=v_2^{(1)}+\Omega x_1,\quad \omega=\omega^{(1)}+\Omega$$

in Equations (41) and (42) gives the required result.

In the case where $\phi=\nu$, the first three of Equations (37) are equivalent to those for an associated flow rule, see [2],

$$d_{11}+d_{22}=\sin\phi\left[\left(d_{11}-d_{22}\right)\cos 2\psi_\sigma+2d_{12}\sin 2\psi_\sigma\right], \tag{43}$$

$$\left(d_{11}-d_{22}\right)\sin 2\psi_\sigma-2d_{12}\cos 2\psi_\sigma=0. \tag{44}$$

Equation (44) is the statement of coaxiality of the stress and deformation-rate tensors. Equations (41) and (42) also reduce to Equations (43) and (44) when $\phi=\nu$. In this case the velocity field is independent of ω and the quantity $k_{\beta_1}-k_{\beta_2}$ is indeterminate. Equation $(37)_{\mathrm{iv}}$ then also becomes indeterminate unless we make the additional assumption that $k_{\beta_1}=k_{\beta_2}$ and then $(37)_{\mathrm{iv}}$ reduces to

$$\omega=s_{21}, \tag{45}$$

i.e., the intrinsic spin ω is determined by the velocity field and is equal to half the vorticity. The anti-symmetric part of the stress required to ensure satisfaction of this kinematic constraint is obtained from the equation of rotational motion; Equation (9) and, in this case, the yield condition (14) must be omitted from the model, since the material must be

able to sustain the anti-symmetric stress required in order to ensure satisfaction of Equation (45). In this sense, the model links, inextricably, the concepts of dilatancy, coaxiality, intrinsic spin and vorticity. Since it is an experimental fact that the magnitude of the angle of dilatancy is less than the angle of internal friction, it follows necessarily that the model predicts both non-coaxiality and non-coincidence of the intrinsic spin with half the vorticity, the rotational yield condition limiting the magnitude of the anti-symmetric part of the stress.

In summary, the model proposed in this section, expressed mathematically in Equations (41) and (42), may be described in physical terms by saying that the flow consists of a local intrinsic spin together with simultaneous dilatant shears on two slip directions and, further, these slip directions coincide with the Coulomb yield directions. The derivation is essentially algebraic, depending only on the yield condition and is independent of the equations of motion. When the full equations of the model are considered as a set of first-order partial differential equations, the question naturally arises as to the relationship, if any, between the slip and yield directions on the one hand and the spatial characteristic directions for a steady-state motion, on the other. We show in the next section that for this model there are three distinct spatial characteristic directions, two of which coincide with the coincident slip and yield directions, while the remaining characteristic direction corresponds to the direction of the streamlines.

It turns out that the well- or ill-posedness of this system of first-order partial differential equations is dependent upon certain properties of the characteristic directions. In this way, the mechanical and kinematic concepts of yield and slip directions are directly related to the mathematical concept of characteristic direction, and hence to the well- or ill-posedness of the model.

4. The steady-state equations are hyperbolic

For steady-state flows in the Ox_1x_2-plane, the equations governing the model (9), (16–18), (41), (42) become,

$$\begin{aligned}\rho v_1 \partial_1 v_1 + \rho v_2 \partial_2 v_1 + (1 - \sin\phi \cos 2\psi_\sigma)\, \partial_1 p_\sigma - \sin\phi \sin 2\psi_\sigma\, \partial_2 p_\sigma \\ + 2q_\sigma \sin 2\psi_\sigma\, \partial_1 \psi_\sigma - 2q_\sigma \cos 2\psi_\sigma\, \partial_2 \psi_\sigma - \rho F_1 = 0,\end{aligned} \tag{46}$$

$$\begin{aligned}\rho v_1 \partial_1 v_2 + \rho v_2 \partial_2 v_2 - \sin\phi \sin 2\psi_\sigma\, \partial_1 p_\sigma + (1 + \sin\phi \cos 2\psi_\sigma)\, \partial_2 p_\sigma \\ - 2q_\sigma \cos 2\psi_\sigma\, \partial_1 \psi_\sigma - 2q_\sigma \sin 2\psi_\sigma\, \partial_2 \psi_\sigma - \rho F_2 = 0,\end{aligned} \tag{47}$$

$$\rho I\, (v_1 \partial_1 \omega + v_2 \partial_2 \omega) + \rho\, (v_1 \partial_1 I + v_2 \partial_2 I)\, \omega - 2r_\sigma - \rho G = 0, \tag{48}$$

$$v_1 \partial_1 \rho + v_2 \partial_2 \rho + \rho \partial_1 v_1 + \rho \partial_2 v_2 = 0, \tag{49}$$

$$\begin{aligned}[\cos(\phi - \nu) - \sin\nu \cos 2\psi_\sigma]\, \partial_1 v_1 - (\sin\nu \sin 2\psi_\sigma)\, \partial_2 v_1 - (\sin\nu \sin 2\psi_\sigma)\, \partial_1 v_2 \\ + [\cos(\phi - \nu) + \sin\nu \cos 2\psi_\sigma]\, \partial_2 v_2 = 0,\end{aligned} \tag{50}$$

$$\begin{aligned}-(\cos\nu \sin 2\psi_\sigma)\, \partial_1 v_1 + [\sin(\phi - \nu) + \cos\nu \cos 2\psi_\sigma]\, \partial_2 v_1 \\ + [-\sin(\phi - \nu) + \cos\nu \cos 2\psi_\sigma]\, \partial_1 v_2 + (\cos\nu \sin 2\psi_\sigma)\, \partial_2 v_2 + 2\omega \sin(\phi - \nu) = 0.\end{aligned} \tag{51}$$

Let

$$\mathbf{z}^t = (z_1, z_2, z_3, z_4, z_5, z_6) = (v_1, v_2, \omega, \rho, p_\sigma, \psi_\sigma), \tag{52}$$

where the superscript t denotes transpose; then Equations (46–51) may be written in matrix form,

$$B_1(\mathbf{z})\,\partial_1\mathbf{z} + B_2(\mathbf{z})\,\partial_2\mathbf{z} + \mathbf{c}(\mathbf{z}) = 0, \tag{53}$$

where

$$B_1(\mathbf{z}) = \begin{bmatrix} \rho v_1 & 0 & 0 & 0 & b_{15}^1 & b_{16}^1 \\ 0 & \rho v_1 & 0 & 0 & b_{25}^1 & b_{26}^1 \\ 0 & 0 & \rho I v_1 & 0 & 0 & 0 \\ \rho & 0 & 0 & v_1 & 0 & 0 \\ b_{51}^1 & b_{52}^1 & 0 & 0 & 0 & 0 \\ b_{61}^1 & b_{62}^1 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{54}$$

$$B_2(\mathbf{z}) = \begin{bmatrix} \rho v_2 & 0 & 0 & 0 & b_{15}^2 & b_{16}^2 \\ 0 & \rho v_2 & 0 & 0 & b_{25}^2 & b_{26}^2 \\ 0 & 0 & \rho I v_2 & 0 & 0 & 0 \\ 0 & \rho & 0 & v_2 & 0 & 0 \\ b_{51}^2 & b_{52}^2 & 0 & 0 & 0 & 0 \\ b_{61}^2 & b_{62}^2 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{55}$$

$$\mathbf{c}(\mathbf{z}) = \begin{bmatrix} -\rho F_1 \\ -\rho F_2 \\ \rho \dot{I}\omega - 2r_\sigma - \rho G \\ 0 \\ 0 \\ 2\omega \sin(\phi - \nu) \end{bmatrix}, \tag{56}$$

where the superposed dot denotes the material derivative and where

$$\begin{aligned} b_{15}^1 &= 1 - \sin\phi \cos 2\psi_\sigma, & b_{16}^1 &= 2q_\sigma \sin 2\psi_\sigma, \\ b_{25}^1 &= -\sin\phi \sin 2\psi_\sigma, & b_{26}^1 &= -2q_\sigma \cos 2\psi_\sigma, \\ b_{51}^1 &= \cos(\phi - \nu) - \sin\nu \cos 2\psi_\sigma, & b_{52}^1 &= -\sin\nu \sin 2\psi_\sigma, \\ b_{61}^1 &= -\cos\nu \sin 2\psi_\sigma, & b_{62}^1 &= -\sin(\phi - \nu) + \cos\nu \cos 2\psi_\sigma, \end{aligned} \tag{57}$$

$$\begin{aligned} b_{15}^2 &= -\sin\phi \sin 2\psi_\sigma, & b_{16}^2 &= -2q_\sigma \cos 2\psi_\sigma, \\ b_{25}^2 &= 1 + \sin\phi \cos 2\psi_\sigma, & b_{26}^2 &= -2q_\sigma \sin 2\psi_\sigma, \\ b_{51}^2 &= -\sin\nu \sin 2\psi_\sigma, & b_{52}^2 &= \cos(\phi - \nu) + \sin\nu \cos 2\psi_\sigma, \\ b_{61}^2 &= \sin(\phi - \nu) + \cos\nu \cos 2\psi_\sigma & b_{62}^2 &= \cos\nu \sin 2\psi_\sigma. \end{aligned} \tag{58}$$

Let $u = u(x_1, x_2) = c$, where c is a constant, denote a curve in the Ox_1x_2-plane on which the solution $\mathbf{z}$ is known and define

$$\xi_1 = \partial_1 u, \quad \xi_2 = \partial_2 u. \tag{59}$$

Then $\mathbf{z} = \mathbf{z}(u) = \mathbf{z}(x_1, x_2)$ and, letting $d_u\mathbf{z}$ denote differentiation with respect to u, we have

$$\partial_1\mathbf{z} = d_u\mathbf{z}\,\partial_1 u = \xi_1 d_u\mathbf{z}, \quad \partial_2\mathbf{z} = d_u\mathbf{z}\,\partial_2 u = \xi_2 d_u\mathbf{z}. \tag{60}$$

Since $d_u\mathbf{z}=0$ on the curve $u(x_1, x_2)=c$, we may regard $d_u\mathbf{z}$ as an exterior derivative. Let

$$E = B_1\xi_1 + B_2\xi_2; \tag{61}$$

then

$$E = \begin{bmatrix} \rho e & 0 & 0 & 0 & e_{15} & e_{16} \\ 0 & \rho e & 0 & 0 & e_{25} & e_{26} \\ 0 & 0 & \rho I e & 0 & 0 & 0 \\ \rho\xi_1 & \rho\xi_2 & 0 & e & 0 & 0 \\ e_{51} & e_{52} & 0 & 0 & 0 & 0 \\ e_{61} & e_{62} & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{62}$$

where

$$e = v_1\xi_1 + v_2\xi_2, \quad e_{ij} = b^1_{ij}\xi_1 + b^2_{ij}\xi_2. \tag{63}$$

Using Equations (60) and (61), we may write Equation (53) as

$$E d_u\mathbf{z} + \mathbf{c} = \mathbf{0}. \tag{64}$$

This is a set of equations for the exterior derivative $d_u\mathbf{z}$ which enables the solution $\mathbf{z}$ to be continued into the Ox_1x_2-plane away from the curve $u(x_1, x_2)=c$. Equation (64) fails to have a solution, *i.e.*, $\mathbf{z}$ cannot be continued off the curve $u(x_1, x_2)=c$, if

$$\det E = 0. \tag{65}$$

This is the condition that $u(x_1, x_2)=c$ be a characteristic curve. Expanding the determinant gives

$$\det E = \rho I e^2 (e_{26}e_{15} - e_{16}e_{25})(e_{51}e_{62} - e_{61}e_{52}), \tag{66}$$

where

$$e_{26}e_{15} - e_{16}e_{25} = -2q_\sigma A(\xi_1, \xi_2), \tag{67}$$

$$e_{51}e_{62} - e_{61}e_{52} = \cos(\phi - 2\nu) A(\xi_1, \xi_2) \tag{68}$$

and

$$A = A(\xi_1, \xi_2) = (\cos 2\psi_\sigma - \sin\phi)\xi_1^2 + 2\sin 2\psi_\sigma \xi_1\xi_2 - (\cos 2\psi_\sigma + \sin\phi)\xi_2^2. \tag{69}$$

Thus,

$$\det E = -2q_\sigma \rho I \cos(\phi - 2\nu) e^2 A^2 \tag{70}$$

and we note that the contributions to $\det E$ from the stress and kinematic equations uncouple and so we may refer to the two pairs of characteristic curves arising from Equations (67) and (68) as the stress and velocity characteristic curves, respectively. Further, their contributions are identical up to a multiplicative factor, and hence these curves coincide. All the characteristic curves are given by the condition (65) and hence one of the following must hold

$$v_1\xi_1 + v_2\xi_2 = 0, \tag{71}$$

$$(\cos 2\psi_\sigma - \sin\phi)\xi_1^2 + (2\sin 2\psi_\sigma)\xi_1\xi_2 - (\cos 2\psi_\sigma + \sin\phi)\xi_2^2 = 0. \tag{72}$$

Thus, the system has

1. a repeated characteristic linear in ξ_1, ξ_2,
2. a repeated pair of characteristic curves, quadratic in ξ_1 and ξ_2,

and, the system is hyperbolic in the sense that all characteristic directions are real, albeit degenerate in that each characteristic is repeated. Let $u = u(x_1, x_2) = c$ be a characteristic curve, then the condition

$$\mathrm{d}u = \xi_1 \mathrm{d}x_1 + \xi_2 \mathrm{d}x_2 = 0 \tag{73}$$

gives

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1} = -\frac{\xi_1}{\xi_2} = m_i, \tag{74}$$

say, where i takes the values 1,2,3. Hence, the root of the linear equation may be written $m_3 = v_2/v_1$, and the characteristic direction determined by this equation corresponds to the streamlines of the flow, while the roots of the quadratic equation

$$(\cos 2\psi_\sigma - \sin\phi)\, m_i^2 - (2\sin 2\psi_\sigma)\, m_i - (\cos 2\psi_\sigma + \sin\phi) = 0 \tag{75}$$

determine the characteristic directions in the Ox_1x_2 plane given by

$$m_1 = \tan(\psi_\sigma + \varepsilon), \quad m_2 = \tan(\psi_\sigma - \varepsilon). \tag{76}$$

These are the α_1- and α_2-directions defined after Equation (19) and the angle between them is $\frac{1}{2}\pi + \phi$. Thus, the characteristic directions of the system of governing partial differential equations coincide with the Coulomb yield directions and the slip directions. In fact, for equations of the form considered here, the condition that the stress and velocity characteristic directions coincide is a necessary, but not sufficient condition, for the linear well-posedness of the model. The plastic-potential model for incompressible materials does not have coincident stress and velocity characteristics and is ill-posed. On the other hand, the incompressible double-shearing model does have coincident stress and velocity characteristics but is also ill-posed. In the next two sections we demonstrate that the model proposed here, which is closely related to both the plastic potential and to the double-shearing models does admit a class of flows for which the model is linearly well-posed, namely the class of flows for an incompressible material.

5. Linearisation of the equations in the incompressible case

We now begin the proof that the model contains a domain of well-posedness. The calculation for the full model is extremely lengthy and requires specification of the evolution of the angle of dilatancy ν, since the standard assumption of perfect plasticity in which ν is considered constant is not adequate here. The total amount of dilatancy or compressibility must be limited, for otherwise the model becomes invalid, either as the density reduces below the level at which the grains can remain in contact, or as it increases above the level which requires grain overlap. Unlimited dilatancy, in particular, may give rise to a mathematical ill-posedness, in this case valid and caused by a physical instability, namely the phase change from solid or liquid-like behaviour to gaseous-like behaviour. The model proposed here, of course, becomes invalid for such dilute flows. Consequently, we consider the special case of incompressible flows and put $\nu = 0$ and omit ρ from the set of dependent variables. In this case, Equation (41) becomes identical with the continuity equation for an incompressible material and so we omit the latter. The full time-dependent equations for incompressible flows will be written as

$$A\,\partial_t \mathbf{z} + B_1(\mathbf{z})\,\partial_1 \mathbf{z} + B_2(\mathbf{z})\,\partial_2 \mathbf{z} + \mathbf{c}(\mathbf{z}) = 0, \tag{77}$$

where

$$A=\begin{bmatrix} \rho & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & \rho I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{78}$$

and B_1, B_2 are as in the previous section but with the fourth row and column omitted. We now linearise this set of equations. Let $\mathbf{Z}^{\mathbf{t}}=(Z_1, Z_2, Z_3, Z_4, Z_5)=(V_1, V_2, \Omega, P_\sigma, \Psi_\sigma)$ denote a known solution of Equation (77) and consider a perturbation $\mathbf{z}'$ of $\mathbf{Z}$ such that

$$\mathbf{z}=\mathbf{Z}+\mathbf{z}'. \tag{79}$$

Also let Q_σ, R_σ denote the known q_σ, r_σ fields for this solution, q'_σ, r'_σ the corresponding perturbations of the q_σ, r_σ fields, then

$$\begin{aligned} &v_1=V_1+v'_1, \quad v_2=V_2+v'_2, \quad \omega=\Omega+\omega', \quad p_\sigma=P_\sigma+p'_\sigma, \\ &\psi_\sigma=\Psi_\sigma+\psi'_\sigma, \quad q_\sigma=Q_\sigma+q'_\sigma, \quad r_\sigma=R_\sigma+r'_\sigma, \end{aligned} \tag{80}$$

Finally, let D_{ij}, Σ_{ij} denote the components of the deformation-rate and stress tensors in the underlying unperturbed solution. Substituting Equation (79) in Equation (77), linearizing the resulting equations and using the fact that $\mathbf{Z}$ is a solution of (77), we have

$$A\partial_t\mathbf{z}'+B_1(\mathbf{Z})\,\partial_1\mathbf{z}'+B_2(\mathbf{Z})\,\partial_2\mathbf{z}'+\mathbf{C}(\mathbf{Z})\,\mathbf{z}'=0, \tag{81}$$

where

$$\mathbf{C}(\mathbf{Z})=\begin{bmatrix} c_{11} & c_{12} & 0 & c_{14} & c_{15} \\ c_{21} & c_{22} & 0 & c_{24} & c_{25} \\ c_{31} & c_{32} & c_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_{53} & 0 & c_{55} \end{bmatrix} \tag{82}$$

and

$$\begin{aligned} &c_{11}=\rho_0\partial_1 V_1, \quad c_{12}=\rho_0\partial_2 V_1, \\ &c_{14}=2\sin\phi\,(\partial_1\Psi_\sigma\sin 2\Psi_\sigma-\partial_2\Psi_\sigma\cos 2\Psi_\sigma), \\ &c_{15}=2[\sin\phi\,(\partial_1 P_\sigma\sin 2\Psi_\sigma-\partial_2 P_\sigma\cos 2\Psi_\sigma) \\ &\qquad +2Q_\sigma\,(\partial_1\Psi_\sigma\cos 2\Psi_\sigma+\partial_2\Psi_\sigma\sin 2\Psi_\sigma)], \\ &c_{21}=\rho_0\partial_1 V_2, \quad c_{22}=\rho_0\partial_2 V_2, \\ &c_{24}=-2\sin\phi\,(\partial_1\Psi_\sigma\cos 2\Psi_\sigma+\partial_2\Psi_\sigma\sin 2\Psi_\sigma), \\ &c_{25}=-2[\sin\phi\,(\partial_1 P_\sigma\cos 2\Psi_\sigma+\partial_2 P_\sigma\sin 2\Psi_\sigma) \\ &\qquad -2Q_\sigma\,(\partial_1\Psi_\sigma\sin 2\Psi_\sigma-\partial_2\Psi_\sigma\cos 2\Psi_\sigma)], \\ &c_{31}=\rho I\partial_1\Omega, \quad c_{32}=\rho I\partial_2\Omega, \quad c_{33}=\rho\,(\partial_t I+V_1\partial_1 I+V_2\partial_2 I), \\ &c_{53}=2\sin\phi, \quad c_{55}=-2Q_D\cos 2\,(\Psi_D-\Psi_\Sigma), \end{aligned} \tag{83}$$

where

$$Q_D=\frac{1}{2}\left[(D_{11}-D_{22})^2+4D_{12}^2\right]^{1/2}, \tag{84}$$

$$\tan 2\Psi_D=\frac{2D_{12}}{D_{11}-D_{22}}, \tag{85}$$

i.e., Q_D and Ψ_D are the maximum shear-rate and angle that the greater principal direction of the deformation-rate tensor makes with the x_1-axis, in the underlying unperturbed field, respectively. Let the entries of $B_1(\mathbf{Z})$, $B_2(\mathbf{Z})$ be denoted by B^1_{ij}, B^2_{ij}, respectively; then, since $\nu = 0$, and renumbering the rows and columns as necessary, we obtain from Equations (57), (58)

$$\begin{aligned}
&B^1_{14} = 1 - \sin\phi\cos 2\Psi_\Sigma, \quad B^1_{15} = 2Q_\Sigma \sin 2\Psi_\Sigma,\\
&B^1_{24} = -\sin\phi\sin 2\Psi_\sigma, \quad B^1_{25} = -2Q_\Sigma\cos 2\Psi_\Sigma,\\
&B^1_{41} = 1, \quad B^1_{42} = 0,\\
&B^1_{51} = -\sin 2\Psi_\Sigma, \quad B^1_{52} = -\sin\phi + \cos 2\Psi_\Sigma,\\
&B^2_{14} = -\sin\phi\sin 2\Psi_\Sigma, \quad B^2_{15} = -2Q_\Sigma\cos 2\Psi_\Sigma,\\
&B^2_{24} = 1 + \sin\phi\cos 2\Psi_\Sigma, \quad B^2_{25} = -2Q_\Sigma\sin 2\Psi_\Sigma,\\
&B^2_{41} = 0, \quad B^2_{42} = 1,\\
&B^2_{51} = \sin\phi, \quad B^2_{52} = \sin 2\Psi_\Sigma,
\end{aligned} \tag{86}$$

where

$$Q_\Sigma = \frac{1}{2}\left[(\Sigma_{11} - \Sigma_{22})^2 + 4\Sigma_{12}^2\right]^{1/2}, \tag{87}$$

$$\tan 2\Psi_\Sigma = \frac{\Sigma_{21} + \Sigma_{12}}{\Sigma_{11} - \Sigma_{22}}.$$

6. Method of frozen coefficients for the incompressible case

We now apply the method of frozen coefficients to show that the model is linearly well-posed for incompressible flows. Consider a perturbation $\mathbf{z}'$ of the original solution $\mathbf{Z}$, with initial time t_0 in the neighbourhood of the point $\mathbf{x}_0$ in which $\mathbf{z}'$ is a normal mode solution of the linearised Equations (81), *i.e.*,

$$\mathbf{z}' = \mathbf{z_0}\exp\left[\varsigma\,(t - t_0) + \mathrm{i}\mathbf{k}\cdot(\mathbf{x} - \mathbf{x}_0)\right] \tag{88}$$

where

$$\mathbf{z_0^t} = \left(v_1^0, v_2^0, \omega^0, p_\sigma^0, \psi_\sigma^0\right) \tag{89}$$

denotes the initial amplitude of the perturbation, ς denotes the (possibly complex) frequency, $\mathrm{i} = \sqrt{-1}$, $\mathbf{k} = (k_1, k_2)$ the (real) vector wave-number, $\mathbf{x} = (x_1, x_2)$, t_0 denotes the initial time and $\mathbf{x}_0$ denotes a fixed spatial point in $\mathcal{R}$. Now,

$$\partial_t \mathbf{z}' = \varsigma\mathbf{z}', \qquad \partial_1\mathbf{z}' = \mathrm{i}k_1\mathbf{z}', \qquad \partial_2\mathbf{z}' = \mathrm{i}k_2\mathbf{z}', \tag{90}$$

and substituting in Equation (81) gives the homogeneous set of algebraic linear equations for $\mathbf{z_0}$,

$$\left[A\varsigma + E(\mathbf{k}, \mathbf{Z})\right]\mathbf{z_0} = 0, \tag{91}$$

where E denotes the matrix

$$E(\mathbf{k}, \mathbf{Z}) = \mathrm{i}k_1 B_1(\mathbf{Z}) + \mathrm{i}k_2 B_2(\mathbf{Z}) + C(\mathbf{Z}) \tag{92}$$

$$= \begin{bmatrix} e_{11} & c_{12} & 0 & e_{14} & e_{15} \\ c_{21} & e_{22} & 0 & e_{24} & e_{25} \\ c_{31} & c_{32} & e_{33} & 0 & 0 \\ e_{41} & e_{42} & 0 & 0 & 0 \\ e_{51} & e_{52} & c_{53} & 0 & c_{55} \end{bmatrix} \tag{93}$$

and where

$$e_{11} = \rho_0 e + c_{11}, \quad e_{22} = \rho_0 e + c_{22}, \quad e_{33} = \rho_0 I e + c_{33}, \tag{94}$$

$$e = \mathrm{i}k_1 V_1 + \mathrm{i}k_2 V_2, \tag{95}$$

$$\begin{aligned} e_{ij} &= \mathrm{i}k_1 B_{ij}^1 + \mathrm{i}k_2 B_{ij}^2 + c_{ij}, \quad i = 1, 2,\ j = 4, 5 \\ e_{ij} &= \mathrm{i}k_1 B_{ij}^1 + \mathrm{i}k_2 B_{ij}^2, \quad i = 4, 5,\ j = 1, 2. \end{aligned} \tag{96}$$

Thus, each e_{ij} is linear in $\mathrm{i}\mathbf{k}$, while c_{ij} is independent of $\mathbf{k}$. Recall that the symbol E was used for a related, but different, matrix in the section demonstrating hyperbolicity of the steady state equations. The similarities and differences between the two matrices are worth noting. The condition that Equation (91) gives rise to non-trivial solutions for $\mathbf{z_0}$, is

$$\det[A\varsigma + E(\mathbf{k}, \mathbf{Z})] = 0. \tag{97}$$

Now, the matrix A is singular and Equation (91) represents a generalised eigenvalue problem for ς,[1] and care must be taken to ensure that all possibilities for non-trivial solutions are found. Accordingly, we first reduce Equation (91) to a standard eigenvalue problem by using the fourth and fifth equations to eliminate the unknowns $p_\sigma^0, \psi_\sigma^0$. However, it should be pointed out that a direct expansion of the determinant in Equation (97) gives the same results concerning well-posedness as those given below, and, moreover, the calculation is shorter. Now, the fourth row of Equation (91) states

$$e_{41} v_1^0 + e_{42} v_2^0 = 0. \tag{98}$$

Multiplying the first equation by e_{41}, the second by e_{42} and adding gives

$$\begin{aligned} &(e_{11}e_{41} + e_{42}c_{21})\, v_1^0 + (e_{41}c_{12} + e_{42}e_{22})\, v_2^0 \\ &\quad + (e_{41}e_{14} + e_{42}e_{24})\, p_\sigma^0 + (e_{41}e_{15} + e_{42}e_{25})\, \psi_\sigma^0 = 0. \end{aligned} \tag{99}$$

The fifth row of Equation (91) states

$$e_{51} v_1^0 + e_{52} v_2^0 + c_{53}\omega^0 + c_{55}\psi_\sigma^0 = 0.$$

Multiplying the first equation by e_{51}, the second by e_{52} and adding gives

$$\begin{aligned} &(e_{11}e_{51} + c_{21}e_{52})\, v_1^0 + (c_{12}e_{51} + e_{22}e_{52})\, v_2^0 - \rho_0 \varsigma c_{53}\omega^0 \\ &\quad + (e_{15}e_{51} + e_{25}e_{52})\, p_\sigma^0 + (e_{15}e_{51} + e_{25}e_{52} - c_{55}\rho_0\varsigma)\, \psi_\sigma^0 = 0. \end{aligned} \tag{100}$$

[1]One of the authors (DH) is indebted for remarks by Prof. D.G. Schaeffer on this point.

Equations (99) and (100) may be solved for $p_\sigma^0,\ \psi_\sigma^0$,

$$p_\sigma^0 = \frac{(f_{41}+g_{41}\rho_0\varsigma)\,v_1^0+(f_{42}+g_{42}\rho_0\varsigma)\,v_2^0+g_{43}\rho_0\varsigma\omega^0}{f+g\rho_0\varsigma}, \tag{101}$$

$$\psi_\sigma^0 = \frac{f_{51}v_1^0+f_{52}v_2^0+g_{53}\rho_0\varsigma\omega^0}{f+g\rho_0\varsigma}, \tag{102}$$

where

$$f=(e_{14}e_{25}-e_{15}e_{24})\,(e_{42}e_{51}-e_{41}e_{52})\,,\quad g=c_{55}\,(e_{41}e_{14}+e_{42}e_{24}) \tag{103}$$

and these are the key quantities in determining the linear well-posedness of the system. Also

$$\begin{aligned} f_{41}&=(c_{21}e_{15}-e_{11}e_{25})\,(e_{42}e_{51}-e_{41}e_{52})\,,\\ f_{42}&=(e_{15}e_{22}-e_{25}c_{12})\,(e_{42}e_{51}-e_{41}e_{52})\,,\\ f_{51}&=(e_{15}c_{21}-e_{25}e_{11})\,(e_{41}e_{52}-e_{42}e_{51})\,,\\ f_{52}&=(e_{14}e_{22}-e_{24}c_{12})\,(e_{41}e_{52}-e_{42}e_{51})\,, \end{aligned} \tag{104}$$

$$\begin{aligned} g_{41}&=-c_{55}\,(e_{11}e_{41}+e_{42}c_{21})\,,\quad & g_{42}&=-c_{55}\,(e_{41}c_{12}+e_{42}e_{22})\,,\\ g_{43}&=c_{53}\,(e_{41}e_{14}+e_{42}e_{24})\,, & g_{43}&=-c_{53}\,(e_{41}e_{14}+e_{42}e_{24})\,. \end{aligned} \tag{105}$$

Elimination of $p_\sigma^0, \psi_\sigma^0$ from the equations and defining $\mathbf{z}_r^0=\left(v_1^0, v_2^0, \omega^0\right)$ reduces the system of equations to

$$\begin{bmatrix} g\rho_0^2\varsigma^2+h_{11}\rho_0\varsigma+i_{11} & h_{12}\rho_0\varsigma+i_{12} & h_{13}\rho_0\varsigma \\ h_{21}\rho_0\varsigma+i_{21} & g\rho_0^2\varsigma^2+h_{22}\rho_0\varsigma+i_{22} & h_{23}\rho_0\varsigma \\ c_{31} & c_{32} & \rho_0 I\varsigma+e_{33} \end{bmatrix} \begin{bmatrix} v_1^0 \\ v_2^0 \\ \omega^0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{106}$$

where

$$\begin{aligned} h_{11}&=f+e_{11}g+e_{15}g_{41}, & h_{12}&=c_{12}g+e_{15}g_{42},\\ h_{13}&=e_{15}g_{43}+e_{16}g_{53}, & h_{21}&=c_{21}g+e_{25}g_{41},\\ h_{22}&=f+e_{22}g+e_{25}g_{42}, & h_{23}&=e_{25}g_{43}+e_{26}g_{53},\\ i_{11}&=e_{11}f+e_{15}f_{41}+e_{16}f_{51}, & i_{12}&=c_{12}f+e_{15}f_{42}+e_{16}f_{52},\\ i_{21}&=c_{21}f+e_{26}f_{51}+e_{25}f_{41}, & i_{22}&=e_{22}f+e_{25}f_{42}+e_{26}f_{52}. \end{aligned} \tag{107}$$

A direct calculation shows that

$$i_{11}=i_{12}=i_{21}=i_{22}=0 \tag{108}$$

and so the equations reduce to

$$F(\varsigma,\mathbf{k},\mathbf{Z})\mathbf{z}_r^0=0, \tag{109}$$

where

$$F(\varsigma,\mathbf{k},\mathbf{Z})=\begin{bmatrix} g\rho_0^2\varsigma^2+h_{11}\rho_0\varsigma & h_{12}\rho_0\varsigma & h_{13}\rho_0\varsigma \\ h_{21}\rho_0\varsigma & g\rho_0^2\varsigma^2+h_{22}\rho_0\varsigma & h_{23}\rho_0\varsigma \\ c_{31} & c_{32} & \rho_0 I\varsigma+e_{33} \end{bmatrix}. \tag{110}$$

This is now a standard eigenvalue problem and Equation (109) has non-trivial solutions for $\mathbf{z}_r^0$ provided

$$\det F(\varsigma,\mathbf{k},\mathbf{Z})=0 \tag{111}$$

and equation (111), the *dispersion relation* for the model, determines ς in terms of $\mathbf{k}$ and $\mathbf{Z}$,

$$\varsigma = \varsigma(\mathbf{k}, \mathbf{Z}). \tag{112}$$

Bearing in mind Equation (88), the *growth rate* of the model, for given $\mathbf{k}$, $\mathbf{Z}$ is defined to be the quantity $\mathfrak{Re}(\varsigma)$ where $\mathfrak{Re}$ denotes real part. It is convenient to define the *wave number tensor* $\mathbf{K}$ by

$$\mathbf{K} = \mathbf{k} \otimes \mathbf{k} = \begin{bmatrix} k_1^2 & k_1 k_2 \\ k_1 k_2 & k_2^2 \end{bmatrix} \tag{113}$$

with invariant k, where

$$k = \sqrt{k_1^2 + k_2^2}, \tag{114}$$

and the *wave tensor angle* ψ_K by

$$\tan \psi_K = k_2 / k_1. \tag{115}$$

Then

$$k_1 = k \cos \psi_K, \quad k_2 = k \sin \psi_K. \tag{116}$$

For $k \neq 0$ define

$$\bar{k}_i = k_i / k. \tag{117}$$

The relations

$$\sin 2\psi_K = \frac{2k_1 k_2}{k_1^2 + k_2^2} = 2\bar{k}_1 \bar{k}_2, \tag{118}$$

$$\cos 2\psi_K = \frac{k_1^2 - k_2^2}{k_1^2 + k_2^2} = \bar{k}_1^2 - \bar{k}_2^2. \tag{119}$$

will be particularly useful. Then

$$\det F(\varsigma, k, \psi_K, \mathbf{Z}) = 0 \tag{120}$$

and expanding the dispersion relation (120) using Equation (110) gives the following quintic equation in ς

$$A_3 (\rho_0 \varsigma)^5 + A_2 (\rho_0 \varsigma)^4 + A_1 (\rho_0 \varsigma)^3 + A_0 (\rho_0 \varsigma)^2 = 0, \tag{121}$$

in which the coefficients $A_i = A_i(k, \psi_K, \mathbf{Z})$ are

$$\begin{aligned} A_3 &= I g^2, \\ A_2 &= g [e_{33} g + I (h_{11} + h_{22})], \\ A_1 &= I (h_{11} h_{22} - h_{12} h_{21}) + e_{33} g (h_{11} + h_{22}) - g (c_{31} h_{13} + c_{32} h_{23}), \\ A_0 &= e_{33} (h_{11} h_{22} - h_{12} h_{21}) + c_{32} (h_{13} h_{21} - h_{11} h_{23}) + c_{31} (h_{12} h_{23} - h_{13} h_{22}), \end{aligned} \tag{122}$$

where, from Equations (107)

$$h_{11}+h_{22}=2f+c_{55}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right], \tag{123}$$

$$h_{11}h_{22}-h_{12}h_{21}=f^2+c_{55}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]f, \tag{124}$$

$$c_{31}h_{13}+c_{32}h_{23}=c_{53}\left(e_{14}e_{25}-e_{15}e_{24}\right)\left(c_{31}e_{42}-c_{32}e_{41}\right), \tag{125}$$

$$\begin{aligned} &c_{32}\left(h_{13}h_{21}-h_{11}h_{23}\right)+c_{31}\left(h_{12}h_{23}-h_{13}h_{22}\right)\\ &\quad=c_{53}\left(c_{32}b_{41}-c_{31}b_{42}\right)\left(e_{14}e_{25}-e_{15}e_{24}\right)f. \end{aligned} \tag{126}$$

Thus,

$$A_2=2Ifg+e_{33}g^2+Ic_{55}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]g, \tag{127}$$

$$\begin{aligned} A_1=If^2&+\left\{Ic_{55}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]+2e_{33}g\right\}f\\ &+\left\{c_{55}e_{33}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]\right.\\ &\left.-c_{53}\left(c_{31}e_{42}-c_{32}e_{41}\right)\left(e_{14}e_{25}-e_{15}e_{24}\right)\right\}g, \end{aligned} \tag{128}$$

$$\begin{aligned} A_0=\{&e_{33}\left\{f+c_{55}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]\right\}\\ &+c_{53}\left(c_{32}e_{41}-c_{31}e_{42}\right)\left(e_{14}e_{25}-e_{15}e_{24}\right)\}f. \end{aligned} \tag{129}$$

Now, $\varsigma=\varsigma\left(k,\psi_K,\mathbf{Z}\right)$ and we define

$$\varsigma_\infty=\varsigma_\infty\left(\psi_K,\mathbf{Z}\right)=\lim_{k\to\infty}\varsigma\left(k,\psi_K,\mathbf{Z}\right). \tag{130}$$

The quantity $\mathfrak{Re}\left(\varsigma_\infty\right)$ is called the *asymptotic growth rate* of the model. The model is called *linearly well-posed* if $\mathfrak{Re}\left(\varsigma_\infty\right)$ is finite for all values of ψ_K. On the other hand, it is called *linearly ill-posed* if $\mathfrak{Re}\left(\varsigma_\infty\right)\to+\infty$ as $k\to\infty$ for some value of ψ_K. Thus, a well-posed model may admit unstable solutions but the growth rate ς must be bounded in its dependence on k. But then, in a linearly ill-posed model, ς is unbounded as a function of k, so, the shorter the wavelength of the perturbation, the larger its growth rate, pointing to a particularly strong kind of instability. Of course, these growing perturbations are solutions of only the linearised equations of the model, which become invalid as the perturbations grow. In the full model, the strength of the growth of perturbations may be mitigated by the nonlinearities. However, for the quasi-linear models considered here, a (nonlinearly) ill-posed model cannot give rise to a well-posed linearisation, and a linearly ill-posed model cannot be the linearisation of a (nonlinearly) well-posed model; see Strang [24]. It is in this sense that the linearised analysis presented here gives a valid deduction for the full quasi-linear model. Our goal is to evaluate $\mathfrak{Re}\left(\varsigma_\infty\right)$ for all values of ψ_K and this we will do by finding the asymptotic growth rate of the roots of Equation (121).

7. Linear well-posedness for normal directions

From Equation (121) either $\varsigma=0$ is a repeated root, corresponding to a neutral affect on growth or decay, or

$$A_3\left(\rho_0\varsigma\right)^3+A_2\left(\rho_0\varsigma\right)^2+A_1\rho_0\varsigma+A_0=0. \tag{131}$$

By inspection, from Equations $(122)_1$, (127–129), we have

$$\begin{aligned} &A_3\left(\mathbf{k},\mathbf{Z}\right)=O\left(k^4\right),\quad A_2\left(\mathbf{k},\mathbf{Z}\right)=O\left(k^6\right),\\ &A_1\left(\mathbf{k},\mathbf{Z}\right)=O\left(k^8\right),\quad A_0\left(\mathbf{k},\mathbf{Z}\right)=O\left(k^9\right). \end{aligned} \tag{132}$$

Let the roots of the dispersion relation, Equation (131), be denoted by α, β, γ; then

$$\alpha+\beta+\gamma=O\left(k^2\right), \quad \alpha\beta+\alpha\gamma+\beta\gamma=O\left(k^4\right), \quad \alpha\beta\gamma=O\left(k^5\right). \tag{133}$$

Hence, there are two roots $O\left(k^2\right)$ and one root $O(k)$. We now investigate the asymptotic behaviour of the roots α, β, γ.

7.1. The roots $O\left(k^2\right)$

To investigate the $O\left(k^2\right)$ roots, we define

$$\bar{\varsigma}=\bar{\varsigma}\left(k, \psi_K, \mathbf{Z}\right)=\varsigma/k^2, \quad \bar{e}_{ij}=e_{ij}/k, \quad \bar{c}_{ij}=c_{ij}/k. \tag{134}$$

Also let

$$\bar{f}=\left(\bar{e}_{14}\bar{e}_{25}-\bar{e}_{15}\bar{e}_{24}\right)\left(\bar{e}_{42}\bar{e}_{51}-\bar{e}_{41}\bar{e}_{52}\right), \qquad \bar{g}=c_{55}\left(\bar{e}_{41}\bar{e}_{14}+\bar{e}_{42}\bar{e}_{24}\right). \tag{135}$$

We now consider the behaviour of $\mathfrak{Re}\left(\bar{\varsigma}\right)$ in the limit $k\to\infty$ with ψ_K held constant. Let $\bar{\varsigma}_\infty=\lim_{k\to\infty}\bar{\varsigma}$ then $\mathfrak{Re}\left(\bar{\varsigma}_\infty\right)$ determines the asymptotic growth rate of the model since

$$\bar{\varsigma}_\infty=\lim_{k\to\infty}\varsigma_\infty/k^2. \tag{136}$$

Clearly, for a $O\left(k^2\right)$ root, the model is linearly well-posed if, and only if, $\bar{\varsigma}_\infty\le 0$. We re-scale the dispersion relation to render the $O\left(k^2\right)$ root finite in the limit $k\to\infty$ and use the scaled dispersion relation to obtain the value of $\mathfrak{Re}\left(\varsigma_\infty\right)$. Dividing Equation (131) by k^{10}, we may write the dispersion relation as follows:

$$\bar{A}_3\left(\rho_0\bar{\varsigma}\right)^3+\bar{A}_2\left(\rho_0\bar{\varsigma}\right)^2+\bar{A}_1\rho_0\bar{\varsigma}+\bar{A}_0/k=0, \tag{137}$$

where $\bar{A}_3=\bar{A}_3\left(\psi_K, \mathbf{Z}\right)$, all other $\bar{A}_i=\bar{A}_i\left(k, \psi_K, \mathbf{Z}\right)$ and

$$\begin{aligned}
\bar{A}_3&=I\bar{g}^2,\\
\bar{A}_2&=2I\bar{f}\bar{g}+\tfrac{1}{k}\bar{e}_{33}\bar{g}^2+I\bar{c}_{55}\left[\bar{e}_{42}\left(\bar{e}_{11}\bar{e}_{24}-\bar{e}_{14}\bar{c}_{21}\right)+\bar{e}_{41}\left(\bar{e}_{22}\bar{e}_{14}-\bar{e}_{24}\bar{c}_{12}\right)\right]\bar{g},\\
\bar{A}_1&=I\bar{f}^2+\left\{I\bar{c}_{55}\left[\bar{e}_{52}\left(\bar{e}_{11}\bar{e}_{24}-\bar{e}_{14}\bar{c}_{21}\right)+\bar{e}_{41}\left(\bar{e}_{22}\bar{e}_{14}-\bar{e}_{24}\bar{c}_{12}\right)\right]+\tfrac{2}{k}\bar{e}_{33}\bar{g}\right\}\bar{f}\\
&\quad+\tfrac{1}{k}\left\{\bar{c}_{55}\bar{e}_{33}\left[\bar{e}_{42}\left(\bar{e}_{11}\bar{e}_{24}-\bar{e}_{14}\bar{c}_{21}\right)+\bar{e}_{41}\left(\bar{e}_{22}\bar{e}_{14}-\bar{e}_{25}\bar{c}_{12}\right)\right]\right.\\
&\quad\left.-\bar{c}_{53}\left(\bar{c}_{31}\bar{e}_{42}-\bar{c}_{32}\bar{e}_{41}\right)\left(\bar{e}_{14}\bar{e}_{25}-\bar{e}_{15}\bar{e}_{24}\right)\right\}\bar{g},\\
\bar{A}_0&=\left\{\bar{e}_{33}\left\{\bar{f}+\bar{c}_{55}\left[\bar{e}_{42}\left(\bar{e}_{11}\bar{e}_{24}-\bar{e}_{14}\bar{c}_{21}\right)+\bar{e}_{41}\left(\bar{e}_{22}\bar{e}_{14}-\bar{e}_{24}\bar{c}_{12}\right)\right]\right\}\right.\\
&\quad\left.+\bar{c}_{53}\left(\bar{c}_{32}\bar{e}_{41}-\bar{c}_{31}\bar{e}_{42}\right)\left(\bar{e}_{14}\bar{e}_{25}-\bar{e}_{15}\bar{e}_{24}\right)\right\}\bar{f}.
\end{aligned} \tag{138}$$

Let $k\to\infty$ with ψ_K held constant and define

$$\bar{e}_{ij}^{\infty}=\lim_{k\to\infty}e_{ij}/k, \tag{139}$$

$$\bar{f}_\infty=\left(\bar{e}_{14}^{\infty}\bar{e}_{25}^{\infty}-\bar{e}_{15}^{\infty}\bar{e}_{24}\right)\left(\bar{e}_{42}^{\infty}\bar{e}_{51}^{\infty}-\bar{e}_{41}^{\infty}\bar{e}_{52}^{\infty}\right), \quad \bar{g}_\infty=c_{55}\left(\bar{e}_{41}^{\infty}\bar{e}_{14}^{\infty}+\bar{e}_{42}^{\infty}\bar{e}_{24}^{\infty}\right), \tag{140}$$

noting that

$$\lim_{k\to\infty}\bar{c}_{ij}=0. \tag{141}$$

Then the dispersion relation reduces to the following cubic equation, the *first reduced asymptotic dispersion relation* for $\bar{\varsigma}_\infty$,

$$\bar{A}_3^{\infty}\left(\rho_0\bar{\varsigma}_\infty\right)^3+\bar{A}_2^{\infty}\left(\rho_0\bar{\varsigma}_\infty\right)^2+\bar{A}_1^{\infty}\rho_0\bar{\varsigma}_\infty=0, \tag{142}$$

where

$$\bar{A}_i^{\infty} = \bar{A}_i^{\infty}(\psi_K, \mathbf{Z}) = \lim_{k\to\infty} \bar{A}_i \tag{143}$$

and

$$\bar{A}_3^{\infty} = I\bar{g}_{\infty}^2, \quad \bar{A}_2^{\infty} = 2I\bar{f}_{\infty}\bar{g}_{\infty}, \quad \bar{A}_1^{\infty} = I\bar{f}_{\infty}^2, \quad \bar{A}_0^{\infty} = \bar{e}_{33}\bar{f}_{\infty}^2. \tag{144}$$

From Equation (142), either $\bar{\varsigma}_{\infty} = 0$ (which corresponds to the $O(k)$ bounded root of Equation (131)) or

$$\left(\rho_0 \bar{g}_{\infty}\bar{\varsigma}_{\infty} + \bar{f}_{\infty}\right)^2 = 0, \tag{145}$$

i.e.,

$$\bar{\varsigma}_{\infty} = -\frac{\left(\bar{e}_{14}^{\infty}\bar{e}_{25}^{\infty} - \bar{e}_{15}^{\infty}\bar{e}_{24}\right)\left(\bar{e}_{42}^{\infty}\bar{e}_{51}^{\infty} - \bar{e}_{41}^{\infty}\bar{e}_{52}^{\infty}\right)}{\rho_0 c_{55}\left(\bar{e}_{41}^{\infty}\bar{e}_{14}^{\infty} + \bar{e}_{42}^{\infty}\bar{e}_{24}^{\infty}\right)}. \tag{146}$$

Now,

$$\bar{e}_{ij}^{\infty} = \mathrm{i}\bar{k}_1 B_{ij}^1 + \mathrm{i}\bar{k}_2 B_{ij}^2, \tag{147}$$

and hence, on using Equations (147), (86), (118), (119) and (19), we have

$$\begin{aligned}
\bar{e}_{14}^{\infty}\bar{e}_{25}^{\infty} - \bar{e}_{15}^{\infty}\bar{e}_{24} &= 2Q_{\Sigma}\left[\cos 2\left(\Psi_{\Sigma} - \psi_K\right) - \sin\phi\right], \\
&= 4Q_{\Sigma}\cos\left(\Psi_{\Sigma} + \varepsilon - \psi_K\right)\cos\left(\Psi_{\Sigma} - \varepsilon - \psi_K\right), \\
\bar{e}_{42}^{\infty}\bar{e}_{51}^{\infty} - \bar{e}_{41}^{\infty}\bar{e}_{52}^{\infty} &= \left[\cos 2\left(\Psi_{\Sigma} - \psi_K\right) - \sin\phi\right], \\
&= 2\cos\left(\Psi_{\Sigma} + \varepsilon - \psi_K\right)\cos\left(\Psi_{\Sigma} - \varepsilon - \psi_K\right), \\
\bar{e}_{41}^{\infty}\bar{e}_{14}^{\infty} + \bar{e}_{42}^{\infty}\bar{e}_{24}^{\infty} &= -1 + \sin\phi\cos 2\left(\Psi_{\Sigma} - \psi_K\right).
\end{aligned} \tag{148}$$

Thus, using Equation $(83)_{\mathrm{ix}}$ yields

$$\bar{\varsigma}_{\infty} = -\frac{Q_{\Sigma}\left[\cos 2\left(\Psi_{\Sigma} - \psi_K\right) - \sin\phi\right]^2}{\rho_0 Q_D \cos 2\left(\Psi_D - \Psi_{\Sigma}\right)\left[1 - \sin\phi\cos 2\left(\Psi_{\Sigma} - \psi_K\right)\right]} \tag{149}$$

$$= -\frac{4Q_{\Sigma}\sin^2\left(\Psi_{\Sigma} - \psi_K + \frac{1}{4}\pi - \frac{1}{2}\phi\right)\sin^2\left(\Psi_{\Sigma} - \psi_K - \frac{1}{4}\pi + \frac{1}{2}\phi\right)^2}{\rho_0 Q_D \cos 2\left(\Psi_D - \Psi_{\Sigma}\right)\left[1 - \sin\phi\cos 2\left(\Psi_{\Sigma} - \psi_K\right)\right]}. \tag{150}$$

Now, for all values of ψ_K

$$1 - \sin\phi\cos 2\left(\Psi_{\Sigma} - \psi_K\right) > 0, \tag{151}$$

provided $0 < \phi < \frac{1}{2}\pi$ and so $\bar{\varsigma}_{\infty} \le 0$ provided the following bound on non-coaxiality in the underlying prescribed solution is satisfied:

$$-\pi/4 < \Psi_D - \Psi_{\Sigma} < \pi/4. \tag{152}$$

We note from its derivation, that the asymptotic equation (142) fails to be valid when

$$\left(\bar{e}_{14}\bar{e}_{25} - \bar{e}_{15}\bar{e}_{24}\right)\left(\bar{e}_{42}\bar{e}_{51} - \bar{e}_{41}\bar{e}_{52}\right) = 0, \tag{153}$$

since terms in Equation (131) which have been omitted from (142) on the condition that they may be neglected in comparison with the term on the left-hand side of Equation (153) can no

longer be so when this term is identically zero. Equation (153) holds when ψ_K takes one of the values ψ_K^1, ψ_K^2, where

$$\psi_K^1 = \Psi_\Sigma - \pi/4 - \phi/2 \pm \pi/2 = \Psi_\Sigma - \varepsilon \pm \pi/2,$$
$$\psi_K^2 = \Psi_\Sigma + \pi/4 + \phi/2 \pm \pi/2 = \Psi_\Sigma + \varepsilon \pm \pi/2, \tag{154}$$

i.e., the directions normal to the coincident stress and velocity characteristic directions of the underlying prescribed solution. A *degenerate direction in the* (k_1,k_2)*-plane* corresponds to one of the values ψ_K^1, ψ_K^2 of ψ_K given by Equation (154). All other directions in the (k_1,k_2)-plane are called *normal directions*. Thus, we have shown that the two roots $O(k^2)$ give a well-posed contribution in the normal directions.

7.2. The root $O(k)$

We now consider the remaining root of Equation (131) subject to Equations (132). To investigate this root we again re-scale the dispersion relation, this time to render the $O(k)$ root finite in the limit $k \to \infty$ and use the scaled dispersion relation to obtain the value of ς_∞. We now divide Equation (131) by k^9 to obtain the dispersion relation

$$\frac{1}{k^2}\bar{A}_3\,(\rho_0\bar{\varsigma})^3 + \frac{1}{k}\bar{A}_2\,(\rho_0\bar{\varsigma})^2 + \bar{A}_1\rho_0\bar{\varsigma} + \bar{A}_0 = 0, \tag{155}$$

where the $\bar{A}_i$ are given by Equation (138). Let $k \to \infty$ with ψ_K held constant then the dispersion relation reduces to the following linear equation, the *second reduced asymptotic dispersion relation* for $\bar{\varsigma}_\infty$,

$$\bar{A}_1^\infty \rho_0 \bar{\varsigma}_\infty + \bar{A}_0^\infty = 0, \tag{156}$$

where the $\bar{A}_i^\infty$ are given by Equation (144). Thus

$$\bar{\varsigma}_\infty = -\bar{e}_\infty, \tag{157}$$

where

$$\bar{e}_\infty = \mathrm{i}\,(V_1 \cos\psi_k + V_2 \sin\psi_k)\,, \tag{158}$$

which is purely imaginary, $\mathfrak{Re}\,(\bar{\varsigma}_\infty) = 0$, and hence this root cannot contribute to ill-posedness. Again we see that the reduced dispersion relation ceases to be valid in a direction perpendicular to the corresponding characteristic, in this case the streamline characteristic, *i.e.*, the second reduced asymptotic dispersion relation ceases to be valid in the direction

$$\psi_K^3 = \tan^{-1}(-V_1/V_2)\,. \tag{159}$$

7.3. Degenerate directions

We now turn to the values of ψ_K that correspond to the degenerate directions, ψ_K^1, ψ_K^2, and also to the direction ψ_K^3, *i.e.*, the directions normal to each of the three distinct characteristic directions. As stated above, the asymptotic equations (142) break down for $\psi_k = \psi_K^1, \psi_K^2$. In this case

$$k_1 = \pm k \sin(\Psi_\sigma \pm \varepsilon)\,, \quad k_2 = \pm k \cos(\Psi_\sigma \pm \varepsilon)\,, \tag{160}$$

and we now derive asymptotic dispersion equations applicable to the degenerate directions. It seems intuitively clear that in the degenerate directions the terms which give rise to the roots

$O(k^2)$ are absent, thus leaving only the root $O(k)$ to determine the growth rate. We verify here that this is indeed the case. Now, let ψ_K take one of the values given by Equations (154); then

$$\begin{aligned}&\bar{e}^{\infty}_{14}\bar{e}^{\infty}_{25}-\bar{e}^{\infty}_{15}\bar{e}_{24}=\bar{e}^{\infty}_{42}\bar{e}^{\infty}_{51}-\bar{e}^{\infty}_{41}\bar{e}^{\infty}_{52}=0,\\&\bar{e}^{\infty}_{41}\bar{e}^{\infty}_{14}+\bar{e}^{\infty}_{42}\bar{e}^{\infty}_{24}=-\cos^2\phi.\end{aligned}\tag{161}$$

since $\cos 2(\Psi_\sigma-\psi_K)=\sin\phi$. Also, Equations $(122)_1$, (127–129) reduce to

$$\begin{aligned}&A_3=Ig^2\\&A_2=e_{33}g^2+Ic_{55}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]g,\\&A_1=\{c_{55}e_{33}\left[e_{42}\left(e_{11}e_{24}-e_{14}c_{21}\right)+e_{41}\left(e_{22}e_{14}-e_{24}c_{12}\right)\right]\\&\qquad -c_{53}\left(c_{31}e_{42}-c_{32}e_{41}\right)\left(e_{14}e_{25}-e_{15}e_{24}\right)\}g,\\&A_0=0.\end{aligned}\tag{162}$$

Hence

$$\left[A_3(\rho_0\varsigma)^2+A_2(\rho_0\varsigma)+A_1\right](\rho_0\varsigma)^3=0,\tag{163}$$

i.e., either $\varsigma=0$, a triple root, or

$$A_3(\rho_0\varsigma)^2+A_2(\rho_0\varsigma)+A_1=0,\tag{164}$$

where

$$A_3=O\left(k^4\right),\quad A_2=O\left(k^5\right),\quad A_1=O\left(k^6\right).\tag{165}$$

Denoting the roots of the quadratic equation (164) by α,β,

$$\alpha+\beta=O(k),\quad \alpha\beta=O\left(k^2\right)\tag{166}$$

i.e., there are two roots $O(k)$. Defining

$$\bar{\varsigma}=\varsigma/k\tag{167}$$

and dividing the equation by k^6 gives

$$\bar{A}_3(\rho_0\bar{\varsigma})^2+\bar{A}_2(\rho_0\bar{\varsigma})+\bar{A}_1=0,\tag{168}$$

where

$$\begin{aligned}&\bar{A}_2=\bar{e}_{33}\bar{g}^2+Ic_{55}\left[\bar{e}_{42}\left(\bar{e}_{11}\bar{e}_{24}-\bar{e}_{14}\bar{c}_{21}\right)+\bar{e}_{41}\left(\bar{e}_{22}\bar{e}_{14}-\bar{e}_{24}\bar{c}_{12}\right)\right]\bar{g},\\&\bar{A}_1=\{c_{55}\bar{e}_{33}\left[\bar{e}_{42}\left(\bar{e}_{11}\bar{e}_{24}-\bar{e}_{14}\bar{c}_{21}\right)+\bar{e}_{41}\left(\bar{e}_{22}\bar{e}_{14}-\bar{e}_{24}\bar{c}_{12}\right)\right]\\&\qquad -c_{53}\left(\bar{c}_{31}\bar{e}_{42}-\bar{c}_{32}\bar{e}_{41}\right)\left(\bar{e}_{14}\bar{e}_{25}-\bar{e}_{15}\bar{e}_{24}\right)\}\bar{g}.\end{aligned}\tag{169}$$

Letting $k\to\infty$ gives

$$\bar{A}^{\infty}_3=I\bar{g}^2_{\infty},\quad \bar{A}^{\infty}_2=2\rho_0 I\bar{e}_{\infty}\bar{g}^2_{\infty},\quad \bar{A}^{\infty}_1=I\rho_0^2\bar{e}^2_{\infty}\bar{g}^2_{\infty}$$

and the *third reduced asymptotic dispersion relation* is

$$\bar{\varsigma}^2_{\infty}+2\bar{e}_{\infty}\bar{\varsigma}_{\infty}+\bar{e}^2_{\infty}=0\tag{170}$$

with repeated root

$$\bar{\varsigma}=-\bar{e}_{\infty}=-\mathrm{i}\left(V_1\cos\psi^j_K+V_2\sin\psi^j_K\right),$$

where j may take the values 1, 2 and again this purely imaginary root cannot cause ill-posedness.

Finally, we turn to the direction, $\psi_K^3 = \tan^{-1}(-V_1/V_2)$. If ψ_K^3 does not coincide with one of ψ_K^1, ψ_K^2 then ψ_K^3 is a normal direction and need not be considered further. If ψ_K^3 does coincide with one of ψ_K^1, ψ_K^2 then both $f = e = 0$, and so Equations $(122)_i$, (127–129) reduce further to

$$\begin{aligned} &A_3 = Ig^2, \quad A_2 = c_{33}g^2 + Ic_{55}[e_{42}(c_{11}e_{24} - e_{14}c_{21}) + e_{41}(c_{22}e_{14} - e_{24}c_{12})]g, \\ &A_1 = c_{55}c_{33}[e_{42}(c_{11}e_{24} - e_{14}c_{21}) + e_{41}(c_{22}e_{14} - e_{24}c_{12})]g, \quad A_0 = 0, \end{aligned} \tag{171}$$

but then

$$A_3 = A_2 = A_1 = O(k^4) \tag{172}$$

and all the roots are bounded and hence cannot contribute to ill-posedness.

8. Conclusions and discussion

We have presented a rigid/perfectly plastic model for the flow of granular materials which is closely related to the double-shearing model, the double-sliding free-rotating model and the associated flow rule. The two essential results of the paper are that (a) the model has been shown to be hyperbolic for steady-state flows in two space dimensions, irrespective of whether the flow is quasi-static or dynamic, and (b) incompressible flows are well-posed.

We make some further remarks concerning the significance of the model and its properties. One way to regard the model is that the double-sliding free-rotating model is indeterminate, containing one more unknown (ω) than equations. The double-shearing model closes this set of equations by taking ω to be $\dot{\psi}_\sigma$, the material derivative of the angle, that the algebraically greater principal stress direction makes with the x_1-axis. The present model is an alternative method of closure, in which the model is augmented by a further unknown (r_σ) and further equations (the equation of rotational motion and the rotational yield condition).

For the double-shearing model, many of the analytic solutions that have been found are such that the quantity $\dot{\psi}_\sigma$, is zero. All of these solutions are also solutions of the present model in which the intrinsic spin ω is zero. It may be anticipated that these solutions will be of use in applications of the present model, and work is currently in progress on this. The model has been constructed purposefully to demonstrate the existence of a model which contains sufficient mechanical and kinematic properties to describe the major bulk properties of granular materials, which also contains a domain of well-posedness and which retains the property of remaining hyperbolic in the inertial regime. This is in contrast to both the double-shearing model and the plastic-potential model when a non-associated flow rule is used.

Hyperbolicity in both the quasi-static and inertial regimes is a desirable property. In a real granular material, the inertial terms can never be identically zero, even though they may be very small. A model which changes type from hyperbolic to elliptic in the presence of inertia has the following difficulty. Solutions for a hyperbolic model need not be smooth, discontinuities in the field variables or their spatial derivatives are very common, indeed one of the standard methods for obtaining solutions to the quasi-static stress field involves patching together solutions which are continuous in the stresses, but in which the spatial derivatives tangential to one family of characteristics, are discontinuous. Solutions to elliptic models exhibit much more smoothness, typically, they must be analytic functions. So for a model which changes type from hyperbolic to elliptic, it appears that solutions in the inertial regime must be of a completely different character to solutions in the quasi-static regime, even in the case where

the inertial terms are arbitrarily small. This difference also holds true for boundary conditions. Solutions of hyperbolic models often exhibit discontinuities in tangential components of velocity and this describes well the slipping of a granular material over a bounding surface, or an internal slip surface. Elliptic models, on the other hand, cannot readily incorporate such discontinuities due to the required smoothness of the solution. Hence an elliptic model must resort to velocity boundary conditions such as the no-slip condition. However, it may be argued that the no-slip condition can never be assured to apply to granular materials. If a bounding surface be perfectly rough, so that no relative velocity is allowed (this does not, of course, preclude rolling) between the material comprising the boundary and the granular material itself, then the layer of grains in direct contact with the boundary will surely stick to it, but the next layer of grains further out *may indeed* slip on the inner layer. In this case it is often said that the material slips on itself. Since, in a continuum model, the grain size may be taken as zero, this slip line cannot be distinguished from the boundary itself and a tangential velocity discontinuity at the boundary is required.

References

1. J.T. Jenkins and M.W. Richman, Kinetic theory for plane flows of a dense gas of idealised, rough, inelastic, circular disks. *Phys. Fluids* 28 (1985) 3485–3495.
2. D.C. Drucker and W. Prager, Soil mechanics and plastic analysis or limit design. *Q. Appl. Math.* 10 (1952) 157–165.
3. P.V. Lade, Elasto-plastic stress–strain theory for cohesionless soil with curved yield surfaces. *Int. J. Solids Structs.* 13 (1977) 1019–1035.
4. G. Gudehus, A comprehensive constitutive equation for granular materials. *Soils Found.* 36 (1996) 1–12.
5. D. Kolymbas, I. Herle and P.A. von Wolffersdorff, Hypoplastic constitutive equation with internal variables. *Int. J. Numer Anal. Meth. Geomech.* 19 (1995) 415–436.
6. P.A. Cundall and O.D.L. Strack, A discrete numerical model for granular assemblies. *Geotechnique* 37 (1979) 47–65.
7. D.G. Schaeffer, Mathematical issues in the continuum formulation of slow granular flow. In: D.D. Joseph and D.G. Schaeffer (eds), *Two Phase Waves in Fluidized Beds, Sedimentation and Granular Flows.* Minneapolis: Institute of Mathematics and its Applications, University of Minnesota (1990) pp. 118–129.
8. E.B. Pitman and D.G. Schaeffer, Stability of time dependent compressible granular flow in two dimensions. *Comm. Pure Appl. Math.* 40 (1987) 421–447.
9. D.G. Schaeffer and E.B. Pitman, Ill-posedness in three-dimensional plastic flow. *Comm. Pure Appl. Math.* XLI (1988) 879–890.
10. D.G. Schaeffer, A mathematical model for localization in granular flow. *Proc. R. Soc. London* A 436 (1992) 217–250.
11. D. Harris, Ill- and well-posed models of granular flow. *Acta Mech.* 146 (2001) 199–225.
12. S. Diebels and W. Ehlers, On basic equations of multiphase micropolar materials. *Tech. Mech.* 16 (1996) 77–88.
13. G. de Josselin de Jong, *Statics and Kinematics of the Failable Zone of a Granular Material*. Doctoral thesis. Delft: Uitgeverij Waltmann (1959) 119 pp.
14. A.J.M. Spencer, A theory of the kinematics of ideal soils under plane strain conditions. *J. Mech. Phys. Solids* 12 (1964) 337–351.
15. M.M. Mehrabadi and S.C. Cowin, Initial planar deformation of dilatant granular materials. *J. Mech. Phys. Solids* 26 (1978) 269–284.
16. A.J.M. Spencer, Deformation of ideal granular materials. In: H.G. Hopkins and M.J. Sewell (eds), *The Rodney Hill 60th. Anniversary Volume.* Oxford: Pergamon Press (1981) pp. 607–652.
17. G. de Josselin de Jong, Mathematical elaboration of the double-sliding, free-rotating. *Arch. Mech.* 29 (1977) 561–591.
18. D. Harris, A unified formulation for plasticity models of granular and other materials. *Proc. R. Soc. London* A450 (1995) 37–49.

19. G.A. Geniev, Problems of the dynamics of a granular medium (in Russian). *Akad. Stroit Archit. SSSR, Moscow* (1958) 120 pp.
20. D. Harris, Modelling mathematically the flow of granular materials. In: N.A. Fleck and A.C.F. Cocks (eds), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) pp. 239–250.
21. D. Harris, Characteristic relations for a model for the flow of granular materials. *Proc. R. Soc. London* A457 (2001) 349–370.
22. A. Drescher and G. de Josselin de Jong, Photoelastic verification of a mechanical model for the flow of granular material. *J. Mech. Phys. Solids* 29 (1972) 337–351.
23. J.C. Savage and D.C. Lockner, A test of the double-shearing model of flow of granular materials. *J. Geophys. Res.* 102 (1997) 12287–12294.
24. G. Strang, Necessary and insufficient conditions for well-posed Cauchy problems. *J. Diff. Eq.* 2 (1966) 107–114.

Journal of Engineering Mathematics (2005) 52: 137–146
DOI 10.1007/s10665-004-6010-9

Towards a theory of granular plasticity

SHAUN C. HENDY
Applied Mathematics, Industrial Research Ltd, Lower Hutt, New Zealand

Received 1 July 2003; accepted in revised form 19 July 2004

Abstract. A theory of granular plasticity based on the time-averaged rigid-plastic flow equations is presented. Slow granular flows in hoppers are often modeled as rigid-plastic flows with frictional yield conditions. However, such constitutive relations lead to systems of partial differential equations that are ill-posed: they possess instabilities in the short-wavelength limit. In addition, features of these flows clearly depend on microstructure in a way not modeled by such continuum models. Here an attempt is made to address both short-comings by splitting variables into 'fluctuating' plus 'average' parts and time-averaging the rigid-plastic flow equations to produce effective equations which depend on the 'average' variables and variances of the 'fluctuating' variables. Microstructural physics can be introduced by appealing to the kinetic theory of inelastic hard-spheres to develop a constitutive relation for the new 'fluctuating' variables. The equations can then be closed by a suitable consitutive equation, requiring that this system of equations be stable in the short-wavelength limit. In this way a granular length-scale is introduced to the rigid-plastic flow equations.

Key words: granular flow, granular temperature, plasticity

1. Introduction

The flow and handling of granular materials is of major importance to many industries. Yet, despite efforts over several decades, the modeling of such flows has achieved only modest success. Dense gravity-driven flows in hoppers have been often modeled as elastic–plastic continua, for example. In this picture, the granular material flows as a plastic with a frictional yield condition, and deforms as an elastic solid otherwise. This model has been used to analyze mass flows, where the material is flowing throughout the hopper, but has failed to provide adequate agreement with experiment in the prediction of quantities such as discharge rate, for example [1]. Despite such shortcomings, Jenike's [2] construction of steady-state incompressible rigid-plastic radial solutions in hoppers with simple geometries has been of considerable importance in hopper design [3, Chapter 10]. These are solutions for quasistatic flow (inertial effects are neglected) where grains travel radially in conical or wedge-shaped hoppers. Only recently have numerical solutions of steady-state quasistatic flows in more complicated geometries been produced [4].

However, it is now clear that there are serious mathematical difficulties with the equations for time-dependent incompressible rigid-plastic flow (IRPF). In many instances, the equations for such flows have been shown to be ill-posed, *i.e.*, they possess instabilities with arbitrarily short wavelengths [5,6]. Hence, it is problematic to interpret the steady-state rigid-plastic flow equations as long-time solutions of the time-dependent equations. Additionally, both the steady-state and time-dependent equations have physical shortcomings. There are several features of hopper flows where the particle size may be important, such as chute flows [7] and shear-banding [8] indicating that it may not be appropriate to model such dense granular flows as continua.

Nonetheless, attempts have been made to describe such flow features in continuum theories in some average sense [9]. Indeed, the development of the shear-banding instability may be related to ill-posedness in continuum plastic flow models [10]. One approach to dealing with this in a continuum theory is to model materials with internal structure by the inclusion of extra terms motivated by physics at the granular scale. These extra terms can damp the instabilities that lead to ill-posedness and can be used to predict shear-band thickness Muhlhaus [9]. Including such damping terms which act at a granular length-scale and 'thickening' the shear-bands is a practical approach to studying granular flows using continuum equations.

Indeed, such an approach is analogous to including Reynolds stresses in the Navier–Stokes equations to model turbulence. In granular systems the analog of the Reynolds stresses is granular temperature. Granular temperature arises naturally in the theory of inelastic gases [11]. Savage [12] introduced granular temperature into plastic flow equations in order to facilitate the introduction of the particle size into the continuum equations for plastic flow. We will introduce granular temperature for a similar purpose here, appealing to the kinetic theory of inelastic gases to introduce grain size into the equations.

We begin by examining the IRPF equations in two-dimensions and review the proof that these equations are ill-posed. We then introduce a granular temperature in a general way, before building it specifically into the plastic flow equations. We then show that with an appropriate choice of constitutive relations, the modified IRPF equations are well-posed.

2. Equations for an incompressible rigid-plastic flow

We begin by considering the flow under gravity of an incompressible granular material in a wedge-shaped hopper under plane strain. We model the flow of this material as a continuum rigid-plastic flow. The equations for such a flow consist of the incompressibility condition,

$$\nabla \cdot \boldsymbol{u} = 0, \tag{1}$$

where u_i are the components of the velocity field, and the momentum equations:

$$\rho \left(\frac{\partial u_i}{\partial t} + \boldsymbol{u} \cdot \nabla u_i \right) + \nabla_j \sigma_{ij} = \rho g_i, \tag{2}$$

where σ_{ij} is the symmetric stress tensor, g_i is the acceleration due to gravity and ρ is the density of the flowing granular material. Note that we define σ to be positive when forces are compressive.

Here we will consider flows where plastic deformation is occurring everywhere in the hopper, *i.e.*, the material is at plastic yield throughout the hopper. We will use an extended von Mises yield condition. In terms of the principal stresses σ_i, this condition is written as

$$(\sigma_1 - \sigma)^2 + (\sigma_2 - \sigma)^2 + (\sigma_3 - \sigma)^2 \leq k^2 \sigma^2, \tag{3}$$

where $k = \sqrt{2} \sin \varphi$, φ is the internal angle of friction of the material and $\sigma = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3)$ is the average trace of the stress tensor (we will refer to σ as the average stress). If this inequality is satisfied exactly then the material is deforming plastically. Under plane strain $\sigma_2 = \sigma = \frac{1}{2}(\sigma_1 + \sigma_3)$, so in this case we may consider a strictly two-dimensional yield condition:

$$(\sigma_1 - \sigma)^2 + (\sigma_3 - \sigma)^2 = k^2 \sigma^2. \tag{4}$$

We now assume a *non-associated flow* rule of the form

$$\sigma_{ij} = \sigma \delta_{ij} + \mu V_{ij}, \tag{5}$$

where $V_{ij} = \nabla_{(i} u_{j)} = (\nabla_i u_j + \nabla_j u_i)/2$ and μ is some, as yet unspecified, scalar function of the normal stress and strain rates. If we compare the flow rule (5) to the yield condition (3), then we see such a flow will satisfy the yield condition exactly if we choose the function μ to be

$$\mu = \frac{k\sigma}{\|V\|}, \tag{6}$$

where $\|V\| = (V_{ij} V_{ji})^{1/2}$.

Equations (1), (2), (5) and (6) form a closed system for incompressible rigid-plastic flow in plane strain. For granular flows in hoppers, these equations are only valid for so-called mass flows where the material is flowing throughout the hopper. When the hopper is not sufficiently steep, funnel flows can develop where material flows down a central funnel leaving a stagnant region adjacent to the walls. Indeed, radial solutions have been used to study the transition between mass and funnel flow which is thought to occur when the rigid-plastic equations become singular as the rate of deformation vanishes [13]. We will restrict our attention to mass flows where rigid-plastic flow occurs everywhere in a given domain.

Combining Equations (1), (2), (5) and (6) together we obtain the equations:

$$\left(\frac{\partial u_i}{\partial t} + \boldsymbol{u} \cdot \nabla u_i\right) = \rho g_i - \nabla_j \left(\sigma \left(\delta_{ij} - k A_{ij}\right)\right), \tag{7}$$

$$\nabla_i u_i = 0, \tag{8}$$

where $A_{ij} = V_{ij}/\|V\|$. We note that $\mathrm{Tr}(A_{ij}) = 0$ and that $\mathrm{Tr}(A_{ik} A_{kj}) = 1$.

The Equations (7) and (8) have been shown to be linearly ill-posed in certain geometries and for certain parameter values [5]. Specifically, the linearized equations of motion that describe the propagation of a small disturbance in the flow, possess unstable plane-wave solutions in the short wavelength limit. In what follows, we will consider the linearized equations of motion for a plane-wave disturbances. We restrict ourselves here to 2D flows under plane-strain (4).

We will begin by writing the two-dimensional rigid-plastic equations ((7) and (8)) in non-dimensional form as follows:

$$\hat{u}_i = u_i/u_0, \quad \hat{\sigma} = \sigma/\rho g L, \quad \hat{x}_i = x_i/L, \quad \hat{t} = t\, g/u_0, \tag{9}$$

where u_0 is some characteristic velocity and L is some characteristic length-scale of the problem. The equations for the rigid-plastic flow then become

$$\frac{\partial \hat{u}_i}{\partial \hat{t}} + \mathrm{Fr}^2 \hat{\boldsymbol{u}} \cdot \hat{\nabla} u_i = g_i/g - \hat{\nabla}_j \left(\hat{\sigma} \left(\delta_{ij} - k A_{ij}\right)\right), \tag{10}$$

$$\hat{\nabla} \cdot \hat{\boldsymbol{u}} = 0, \tag{11}$$

where $\mathrm{Fr} = u_0/\sqrt{gL}$ is the Froude number. For the remainder of this section we will drop the $\hat{}$ notation and assume that we are dealing with dimensionless quantities.

The linearized equations of motion for a small disturbance $(\delta\mathbf{u}, \delta\sigma)$ propagating on a smooth background solution $(\mathbf{u}, \sigma)$ to Equations (10) and (11) can be shown to be

$$\frac{\partial \delta u_i}{\partial t} + \mathrm{Fr}^2 \left(\boldsymbol{u} \cdot \nabla \delta u_i + \delta \boldsymbol{u} \cdot \nabla u_i\right) = \nabla_j \left(\left(k A_{ij} - \delta_{ij}\right) \delta\sigma + \sigma\, \delta A_{ij}\right), \tag{12}$$

$$\nabla \cdot \delta \boldsymbol{u} = 0, \tag{13}$$

where

$$\delta A_{ij} = \hat{A}_{ijkl} \frac{\nabla_k \delta u_l}{\|V\|} = \left(\delta_{i(k}\delta_{l)j} - A_{ij}A_{kl}\right) \frac{\nabla_k \delta u_l}{\|V\|}.$$

We now consider plane-wave disturbances $(\delta \boldsymbol{u}, \delta p) = \exp(\lambda t + \mathrm{i}\boldsymbol{\xi} \cdot \boldsymbol{x})(\boldsymbol{a}, \alpha)$ propagating with wavevector ξ. In general α and $\boldsymbol{a}$ will be complex quantities. From the linearized equations, we obtain the following relations for λ, $\boldsymbol{a}$ and α:

$$\lambda a_i = B_i \alpha + C_{ij} a_j, \tag{14}$$

$$\boldsymbol{\xi} \cdot \boldsymbol{a} = 0, \tag{15}$$

where

$$B_i = \left(k \nabla_j A_{ij} + \mathrm{i}\left(k A_{ij} - \delta_{ij}\right) \xi_j\right), \tag{16}$$

$$C_{ij} = -\mathrm{Fr}^2 \left(\nabla_j u_i + \mathrm{i}(\boldsymbol{\xi} \cdot \boldsymbol{u})\delta_{ij}\right) + \xi_l\left(-\mu \hat{A}_{ijkl}\xi_k + \mathrm{i}\nabla_k\left(\mu \hat{A}_{ijkl}\right)\right). \tag{17}$$

One can solve (14) and (15) for λ for every wavevector ξ. The real part of λ determines the growth or decay of the plane-wave disturbance with wavevector ξ, and the imaginary part determines the speed of propagation of the disturbance. If the real part of λ is positive for any ξ, we refer to this mode as *linearly unstable*, as this mode will grow rapidly in time. If, for a unique background solution of the equations, there are unstable modes with growth rates $\mathfrak{Re}(\lambda) > 0$ that diverge in the short-wavelength limit ($\xi \to \infty$) then we will call these equations and the corresponding solutions *linearly ill-posed*.

Using the condition $\boldsymbol{\xi} \cdot \boldsymbol{a} = 0$ we can eliminate α from the equation for λ:

$$\lambda a_i = P_{ij} a_j, \tag{18}$$

$$P_{ij} = \left(\delta_{ik} - \frac{B_i \chi_k}{\mathbf{B} \cdot \boldsymbol{\chi}}\right) C_{kj}, \tag{19}$$

$$\alpha = -\frac{C_{ij}}{\mathbf{B} \cdot \boldsymbol{\chi}} \chi_i a_j, \tag{20}$$

where $\boldsymbol{\chi} = \boldsymbol{\xi}/|\xi|$. The eigenvalues of the matrix P (19) determine the growth and propagation of the infinitesimal plane-wave disturbance.

Since $\left(\delta_{ij} - \frac{B_i \chi_j}{\mathbf{B} \cdot \boldsymbol{\chi}}\right) B_j = 0$, it follows that P_{ij} has at least one zero eigenvalue $\lambda_1 = 0$. Since we are working in two dimensions, the remaining eigenvalue is equal to the trace of P: $\lambda_2 = \mathrm{Tr}(P)$. This eigenvalue is thus given by

$$\lambda_2 = \mathrm{Tr}(C) - \frac{C_{ij}\chi_i B_j}{\mathbf{B} \cdot \boldsymbol{\chi}}. \tag{21}$$

We now consider the short wavelength ($|\xi| \to \infty$) limit of (21). The leading-order term in ξ on the right-hand side goes as $\mathrm{O}(|\xi|^2)$ and is real with coefficient

$$-\mu \frac{A(A - k/2)}{1 - kA}, \tag{22}$$

where $A = \chi_i A_{ij} \chi_j$. This leading order term was considered by Schaeffer [5] in his analysis of (7) and (8). In two dimensions it is straight-forward to show that $|A_{ij}\chi_j|^2 = 1/2$ so that $|A|^2 = |\chi_i A_{ij} \chi_j|^2 \leq 1/2$. Thus the denominator is always positive for angles of friction $\varphi < \pi/2$. However we see that the numerator is negative for $0 < A < \frac{k}{2}$. This occurs when the direction

of $\boldsymbol{\chi}$ lies between the direction of the velocity characteristics (these lie at $\pm\pi/4$ to the principal stress directions) and the direction of the stress characteristics (these lie at angles of $(\pm\varphi + \pi/2)/2$ to the principal stress directions) of the background solution.

We conclude that plane wave disturbances with wavevectors ξ that lie in directions between the stress and velocity characteristic directions will be unstable in the short wavelength limit as the real part of λ_2 will be positive and is $O(\xi^2)$ in this limit. Thus we conclude that the two-dimensional granular flow equations are linearly ill-posed [5].

3. Granular temperature

Our aim is to introduce microstructural terms in order to regularize the ill-posedness of the 2D IRPF equations. To do this we now introduce a granular temperature by decomposing the velocity field into a mean, slowly varying part $\mathbf{U}(\boldsymbol{x}, t)$ and a rapidly fluctuating part $\boldsymbol{u}'(\boldsymbol{x}, t)$:

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{U}(\boldsymbol{x}, t) + \boldsymbol{u}'(\boldsymbol{x}, t). \tag{23}$$

Taking the time-average of $\boldsymbol{u}(\boldsymbol{x}, t)$ then yields

$$\langle \boldsymbol{u}(\boldsymbol{x}, t) \rangle = \boldsymbol{U}(\boldsymbol{x}, t), \tag{24}$$

so that $\langle \boldsymbol{u}' \rangle = 0$. If we then let

$$T_{ij} = \langle u_i' u_j' \rangle, \tag{25}$$

then the granular temperature is defined by $T = \mathrm{Tr}(T_{ij})$. The fluctuating velocity field introduces a new scale which we write as $T_0^{1/2}$. Thus, the granular temperature scales as T_0. In what follows we will assume that $T_{ij} = \frac{1}{n} T \delta_{ij}$ where n is the number of spatial dimensions.

The incompressibility condition can now be split into a mean and fluctuating part. By 'fluctuating part' we mean the difference between the full Equation (1) and the equation obtained by averaging (1):

$$\nabla \cdot \boldsymbol{U} = 0, \tag{26}$$

giving the fluctuating part as

$$\nabla \cdot \boldsymbol{u}' = 0. \tag{27}$$

Now consider the fluctuating part of the momentum equation (2). This averaged momentum equation is given by

$$\rho \left(\frac{\partial U_i}{\partial t} + \boldsymbol{U} \cdot \boldsymbol{\nabla} U_i \right) + \nabla_j \left(\langle \sigma_{ij} \rangle + \rho T_{ij} \right) = \rho g_i. \tag{28}$$

Subtracting this from (2) gives the so-called fluctuating part of the momentum equation:

$$\rho \left(\frac{\partial u_i'}{\partial t} + \boldsymbol{U} \cdot \boldsymbol{\nabla} u_i' + u_j' \nabla_j U_i + u_j' \nabla_j u_i' \right) + \nabla_j \sigma_{ij} = \nabla_j \left(\langle \sigma_{ij} \rangle + \rho T_{ij} \right), \tag{29}$$

where we have made use of (1) and (23). Now multiplying (29) by u_k' and contracting over the indices i and k, we obtain an equation for the granular temperature:

$$\frac{\rho}{2} \left(\frac{\partial T}{\partial t} + \boldsymbol{U} \cdot \boldsymbol{\nabla} T + \nabla_j \langle u_i' u_i' u_j' \rangle \right) + \rho\, T_{ij} \nabla_j U_i = -\langle u_i' \nabla_j \sigma_{ij} \rangle, \tag{30}$$

A conservation equation for the granular temperature can be developed on general grounds [12]:

$$\frac{\rho}{2}\left(\frac{\partial T}{\partial t}+\boldsymbol{U}\cdot\boldsymbol{\nabla}T+U_iT_{ij}\nabla_j+\frac{1}{2}\nabla_j\langle u_i'u_i'u_j'\rangle\right)$$
$$=-\nabla_jQ_j-\langle\sigma_{ij}\rangle\nabla_iU_j-\gamma, \tag{31}$$

where $\boldsymbol{Q}$ is the energy flux vector and γ is the rate of energy dissipation per unit volume. This form of conservation equation for granular temperature can also be obtained from the kinetic theory of inelastic spheres [11]. Bocquet *et al.* [14] applied this kinetic theory to dense Couette flows, deriving the following expressions for Q_i and γ in the dense limit of the kinetic theory:

$$Q_i=-\kappa\frac{\sigma d}{T^{1/2}}\nabla_iT, \tag{32}$$

$$\gamma=\epsilon\frac{\sigma T^{1/2}}{d}, \tag{33}$$

where $\kappa>0$ and $\epsilon>0$ are dimensionless material constants, and d is a representative grain diameter. In what follows we will use the conservation equation for temperature (31) with the energy flux and dissipation terms given by (32) and (33).

Comparing Equation (30) then with (31), we arrive at the following identity:

$$\langle\sigma_{ij}V_{ij}\rangle+\epsilon\frac{\sigma T^{1/2}}{d}=\nabla_i\left(\langle\sigma_{ij}u_j'\rangle+\kappa\frac{\sigma d}{T^{1/2}}\nabla_iT\right). \tag{34}$$

It is useful now to work with non-dimensional quantities. Writing $D=d/L$, we find that

$$\langle\sigma_{ij}V_{ij}\rangle+\frac{1}{D}\left(\frac{T_0^{1/2}}{u_0}\right)\epsilon\sigma T^{1/2}=\left(\frac{T_0^{1/2}}{u_0}\right)\nabla_i\left(D\kappa\frac{\sigma}{T^{1/2}}\nabla_iT+\left(\frac{T_0^{1/2}}{u_0}\right)\langle\sigma_{ij}u_j'\rangle\right). \tag{35}$$

Now, for smooth flows we expect $D\ll1$. In this case, since the first term on the lefthand side is O(1), we find the following relation for the temperature scale T_0 must hold:

$$T_0^{1/2}\sim Du_0, \tag{36}$$

and thus $T_0\ll u_0^2$.

We now consider the fluctuating part of the stress tensor. Returning to the fluctuating part of the momentum equation we can estimate the relative magnitude of the stress fluctuations, $\Delta\sim\sigma'/\rho gL$, that correspond to the velocity fluctuations $u'\sim T_0^{1/2}$. Rewriting (29) in terms of non-dimensional quantities and rearranging in terms of the stress fluctuations $\sigma_{ij}-\langle\sigma_{ij}\rangle$ we find:

$$\frac{\Delta}{D}\nabla_j\left(\sigma_{ij}-\langle\sigma_{ij}\rangle\right)=\frac{\partial u_i'}{\partial t}+\mathrm{Fr}^2\left(U_j\nabla_ju_i'+u_j'\nabla_jU_i\right)+D\mathrm{Fr}\nabla_j\left(u_j'u_i'-T_{ij}\right), \tag{37}$$

where we have made use of (36) to eliminate T_0. We will only consider slow, smooth flows here where $\mathrm{Fr}\ll1$ and $D\ll1$, so we see that in this case $\Delta\sim D$. Thus for slow smooth flows, we expect stress fluctuations of order ρgd to couple to the fluctuating part of the velocity.

This analysis has given us some indication of how the granular temperature and stress fluctuations may scale in smooth slow flows by appealing to the dense-limit of the kinetic theory of inelastic gases. As yet we have not specified the stress tensor nor the nature of the stress fluctuations. This is our goal in the next section.

4. Mean stress

To relate the velocity and stress fluctuations, we now assume that the flow rule (5) and constitutive law (6) hold for the full stress and velocity fields:

$$\sigma_{ij} = \sigma\left(\delta_{ij} - kA_{ij}\right). \tag{38}$$

We can decompose this stress tensor into average and fluctuating parts by expanding σ and $\mathbf{u}$ about their means, $\langle\sigma\rangle = \bar{\sigma}$ and $\langle\mathbf{u}\rangle = \mathbf{U}$ respectively. To facilitate this we introduce the following notation:

$$\begin{aligned}
&\bar{V}_{ij} = V_{ij}|_{\boldsymbol{u}=\boldsymbol{U}} = \nabla_{(i}U_{j)},\\
&\bar{\sigma}_{ij} = \bar{\sigma}\left(\delta_{ij} - k\bar{A}_{ij}\right),\\
&\qquad \ldots \text{etc.},
\end{aligned}$$

so that the bar denotes evaluation at $\boldsymbol{u} = \boldsymbol{U}$. Now, expanding σ_{ij} as indicated, where σ' is the fluctuating part of σ, and continuing to work with non-dimensional quantities, we obtain

$$\begin{aligned}
\sigma_{ij} = \bar{\sigma}_{ij} + D&\left(\sigma'\left(\delta_{ij} - k\frac{\bar{V}_{ij}}{\|\bar{V}\|}\right) - \frac{k\sigma}{\|\bar{V}\|}\hat{A}_{ijkl}\nabla_k u'_l\right)\\
&- D^2\frac{k}{\|\bar{V}\|}\left(\hat{A}_{ijkl}\sigma' - \frac{\bar{\sigma}}{\|\bar{V}\|}\hat{C}_{ijklmn}\nabla_m u'_n\right)\nabla_k u'_l + \mathrm{O}(D^3),
\end{aligned} \tag{39}$$

where

$$\hat{A}_{ijkl} = \delta_{i(k}\delta_{l)j} - \bar{A}_{ij}\bar{A}_{kl}, \tag{40}$$

and

$$\hat{C}_{ijklmn} = \frac{1}{2}\left(\hat{A}_{ijkl}\bar{A}_{mn} + \hat{A}_{klmn}\bar{A}_{ij} + \hat{A}_{mnij}\bar{A}_{kl}\right). \tag{41}$$

Note that we have assumed that we are dealing with a smooth, slow flow so that (36) holds. Consequently this expansion produces a power series in D. Further, we can now compute the mean stress

$$\langle\sigma_{ij}\rangle = \bar{\sigma}_{ij} - \frac{kD^2}{\|\bar{V}\|}\left(\hat{A}_{ijkl}\langle\sigma'\nabla_k u'_l\rangle - \frac{\bar{\sigma}}{\|\bar{V}\|}\hat{C}_{ijklmn}\langle\nabla_k u'_l\nabla_m u'_n\rangle\right), \tag{42}$$

to order D^2. Note that the $\mathrm{O}(D)$ terms have vanished as they are linear in the (zero mean) fluctuating variables. Thus the mean stress $\langle\sigma_{ij}\rangle$, which appears in the averaged momentum equation and the temperature conservation equation, receives contributions of order D^2 due to variances in velocity and stress fluctuations. Specifically these $\mathrm{O}(D^2)$ contributions to the average stress involve the variances $\langle\sigma'\nabla_k u'_l\rangle$ and $\langle\nabla_k u'_l\nabla_m u'_n\rangle$. To close this system, we wish to write these variances as functions of the mean variables $(\boldsymbol{U}, \bar{\sigma}, T)$. Here we propose the following simple relations:

$$\langle\sigma'\nabla_k u'_l\rangle = \psi\nabla_k\left(T^{1/2}\nabla_l\bar{\sigma}\right), \tag{43}$$

$$\langle\nabla_k u'_l\nabla_m u'_n\rangle = 2\phi\nabla_k\nabla_m T_{ln} = \phi\delta_{ln}\nabla_k\nabla_m T, \tag{44}$$

where ψ and ϕ are dimensionless constants. In the next section we will motivate this choice of closure relations by showing they lead to a well-posed system of equations for smooth, slow two-dimensional flows.

The average stress now takes the form:

$$\langle\sigma_{ij}\rangle = \bar{\sigma}_{ij} - D^2\frac{k}{\|\bar{V}\|}\left(\psi\hat{A}_{ijkl}\nabla_k(T^{1/2}\nabla_l\bar{\sigma}) - \frac{\phi\bar{\sigma}}{\|\bar{V}\|}\delta_{k(i}\bar{A}_{j)m}\nabla_k\nabla_m T\right). \tag{45}$$

We can now write down a set of equations for incompressible rigid-plastic flow involving only averaged quantities, correct to $O(D^2)$:

$$\nabla\cdot\boldsymbol{U} = 0, \tag{46}$$

$$\left(\frac{\partial U_i}{\partial t} + \mathrm{Fr}^2\left(\boldsymbol{U}\cdot\nabla U_i + D^2\nabla_j T_{ij}\right)\right) + \nabla_j\left(\bar{\sigma}\left(\delta_{ij} - k\bar{A}_{ij}\right)\right)$$
$$-kD^2\nabla_j\left(\frac{\psi}{\|\bar{V}\|}\hat{A}_{ijkl}\nabla_k(T^{1/2}\nabla_l\bar{\sigma}) - \frac{\phi\bar{\sigma}}{\|\bar{V}\|^2}\delta_{k(i}\bar{A}_{j)m}\nabla_k\nabla_m T\right) = g_i/g, \tag{47}$$

$$\frac{D^2}{2}\left(\frac{\partial T}{\partial t} + \mathrm{Fr}^2\left(U_j\nabla_j T + T_{ij}\nabla_j U_i\right)\right)$$
$$= D^2\left(\kappa\nabla_j\left(\frac{\bar{\sigma}}{T^{1/2}}\nabla_j T\right) - \frac{k\phi\bar{\sigma}}{\|\bar{V}\|}\nabla^2 T\right) + \bar{\sigma}\left(k\|\bar{V}\| - \epsilon T^{1/2}\right). \tag{48}$$

These constitute a closed set of equations for the averaged variables $\boldsymbol{U}$, $\bar{\sigma}$ and T.

5. Stability

In this section we examine the stability of slow, smooth solutions (*i.e.,* $\mathrm{Fr}\ll 1$, $D\ll 1$) to Equations (48–50). Again we consider plane-wave disturbances $(\delta\boldsymbol{u}, \delta p, \delta T) = \exp(\lambda t + \mathrm{i}\boldsymbol{\xi}\cdot\boldsymbol{x})$ $(\boldsymbol{a}, \alpha, \tau)$ propagating with wavevector $\boldsymbol{\xi}$. From the linearized equations, we obtain the following equations for $\boldsymbol{a}$, α and τ:

$$\begin{pmatrix} \xi^2\mu Q_{ij} + \lambda\delta_{ij} & \mathrm{i}\xi^3 D^2 R r_i & -\mathrm{i}\xi^3 D^2 S s_i \\ \chi_j & 0 & 0 \\ -\mathrm{i}\xi k\sigma\chi_i A_{ij} & -\mathrm{i}\xi D^2 G & D^2(\xi^2 H + \lambda) \end{pmatrix}\begin{pmatrix} a_j \\ \alpha \\ \tau \end{pmatrix} = 0, \tag{49}$$

where

$$Q_{ij} = \hat{A}_{ijkl}\chi_k\chi_l + O(1/\xi), \quad R = \frac{k\psi}{\|V\|}, \quad S = \mu\phi,$$
$$r_i = \hat{A}_{ijkl}\chi_j\chi_k\chi_l + O(1/\xi), \quad s_i = \delta_{k(i}A_{j)m}\chi_j\chi_k\chi_m + O(1/\xi),$$
$$G = \frac{\boldsymbol{\chi}\cdot\nabla T}{T^{1/2}} + O(1/\xi), \quad H = \frac{\kappa\bar{\sigma}}{T^{1/2}} - \frac{k\phi}{\|\bar{V}\|} + O(1/\xi). \tag{50}$$

It is straightforward to eliminate α and τ to obtain the following eigenvalue problem:

$$\left(W\lambda\delta_{ik} + P_{ij}Z_{jk}\right)a_k = 0, \tag{51}$$

where

$$W = D^2(R(\lambda + \chi^2 H)r_i - \mathrm{i}\xi G S s_i)\chi_i, \tag{52}$$

$$P_{ij} = \delta_{ij} - \frac{r_i\chi_j}{(\boldsymbol{r}\cdot\boldsymbol{\chi})}, \tag{53}$$

$$Z_{jk} = \xi^2\mu W Q_{jk} - \mathrm{i}\xi^3 D^2 S s_j\chi_l(\mu G Q_{lk} - \mathrm{i}\xi k\sigma R\boldsymbol{r}\cdot\boldsymbol{\chi}A_{lk}). \tag{54}$$

Again one can show that the matrix $P_{ij}Z_{jk}$ has at least one zero eigenvalue (since $P_{ij}r_j = 0$). Thus, in two dimensions the remaining eigenvalue is equal to the trace of $P_{ij}Z_{jk}$. This identity results in a quadratic equation for this eigenvalue λ_2, giving two solutions to leading order in ξ (provided $\psi \neq 0$):

$$\lambda_2 = \frac{\xi^2 \bar{\sigma}}{2}\left(-\Omega_1 \pm \sqrt{\Omega_1^2 - 4\Omega_2} + \mathrm{O}\left(\frac{1}{\xi}\right)\right), \tag{55}$$

where

$$\Omega_1 = \frac{\kappa}{T^{1/2}} + \frac{k}{\|\bar{V}\|}\left(\frac{A^2}{1-A^2} - \phi\right), \tag{56}$$

$$\Omega_2 = \frac{1}{2(1-A^2)\|\bar{V}\|}\left(\frac{\kappa}{T^{1/2}} - \frac{k\phi}{2\|\bar{V}\|}(1 + A^2 - 2A^4)\right). \tag{57}$$

If we choose $\phi < 0$, then we see that both Ω_1 and Ω_2 are positive for $k, \kappa > 0$ (recall that $|A|^2 \leq 1/2$). There are then two cases: if $\Omega_1^2 - 4\Omega_2 > 0$, then the $\mathrm{O}(\xi^2)$ contribution to λ_2 is real and since $\Omega_2 > 0$, both of the values for λ_2 are negative. However, if $\Omega_1^2 - 4\Omega_2 \leq 0$, then the real part of the $\mathrm{O}(\xi^2)$ contribution to λ_2 is $-\Omega_1\xi^2/2 < 0$. We conclude that the real part of the leading order contribution in ξ to λ_2 is negative for $\phi < 0$. Thus, smooth, slow flows governed by Equations (48)–(50) are well-posed provided we choose $\phi < 0$ and $\psi \neq 0$.

6. Discussion

Equations (48)–(50) give a set of equations for the average variables $\mathbf{U}$, $\bar{\sigma}$ and T. In the limit where $d \to 0$, these equations reduce to the equations for incompressible rigid-plastic flow plus an algebraic relation linking granular temperature and the deformation rate $\|\bar{V}\|$. We have shown that the choice of closure relations (45) and (46) lead to a well-posed set of equations for smooth slow flows under the conditions that $\phi < 0$ and $\psi \neq 0$. Physically, this first requirement corresponds to the condition that the $\nabla^2 T$ term in Equation (50) be dissipative. Thus the effect of this term in the temperature conservation equation (50) is to conduct granular temperature from hot regions to cool regions. In addition, the coefficient of the first closure relation (45) must be non-zero to preserve the coupling between high-order derivatives of the pressure and temperature in the momentum equation and guarantee well-posedness. This term in the momentum equation conserves temperature intrinsically and so does not appear in the temperature conservation equation.

Although we have not provided a clear physical motivation of the closure relations, they are dimensionally and tensorially consistent. A detailed theory of velocity and stress fluctuations could be developed to extend or replace equations (45) and (46). Another approach might be to expand the variances in all dimensionally and tensorially consistent terms involving T and $\bar{\sigma}$ and their derivatives. In this case, we suggest that demanding that the resulting equations be well-posed would be a useful discrimant to apply to such closure relations. We have worked in two-dimensional plane-strain conditions here for simplicity. A full theory would need to address these issues in three dimensions for compressible granular flows.

We have also not examined the question of whether solutions to (48–50) reduce to solutions of the IRPF in this limit. Note that the equations (48–50) involve T and derivatives of T, which require new boundary conditions over and above those required by the IRPF equations. Boundary conditions that are placed on the granular temperature will effectively place new conditions on the velocity derivatives through Equation (50) in the limit $d \to 0$. The

nature of these boundary conditions on the temperature will determine whether solutions to (48)–(50) reduce to solutions of the IRPF as $d \to 0$. For instance, one could demand that any boundary conditions on T preserve solutions of the IRPF equations in this limit. Whether this is physically reasonable or not, requires a more detailed examination of granular temperature than we have undertaken here.

An important assumption in deriving the expression for the average stress (41), was that the full stress tensor obeyed the constitutive laws (5) and (6). This allowed us to relate velocity fluctuations to the stress fluctuations. Savage [12] relates velocity and stress fluctuations using a constitutive law developed directly from the kinetic theory of inelastic gases. Savage also suggests a relationship between the temperature scale and the particle size that is very similar to (36). Indeed, our analysis in Section 3 can be directly applied to the equations of Savage.

References

1. R.B. Thorpe, An experimental clue to the importance of dilation in determining the flow rate of a granular material from a hopper or bin. *Chem. Engng. Sci.* 47 (1992) 4295–4303.
2. A. Jenike, Gravity flow of bulk solids. Technical Report, Utah Engineering Experimental Station, University of Utah, Salt Lake City (1964).
3. R.M. Nedderman, *Statics and Kinematics of Granular Materials*. London: Cambridge University Press (1992) 368 pp.
4. P.-A. Gremaud and J.V. Matthews, On the computation of steady hopper flows: I, stress determination for coulomb materials. Technical Report CRSC TR99-35, Center for Research in Scientific Computing, North Carolina State University, Raleigh (1999).
5. D.G. Schaeffer, Instability in the evolution equations describing incompressible granular flow. *J. Diff. Equ.* 66 (1987) 19–50.
6. K.C. Valanis and J.F. Peters, Ill-posedness of the initial and boundary value problems in non-associative plasticity. *Acta Mech.* 114 (1996) 1–25.
7. S.B. Savage, The mechanics of rapid granular flows. *Adv. Appl. Mech.* 24 (1984) 289–366.
8. R.M. Nedderman and C. Laohakul, The thickness of the shear zone of flowing granular materials. *Powder Technol.* 25 (1980) 91–100.
9. H.-B. Muhlhaus and F. Oka, Dispersion and wave propagation in discrete and continuous models for granular materials. *J. Solids Structs.* 33 (1996) 2841–2858.
10. D.G. Schaeffer, Instability and ill-posedness in the deformation of granular materials. *Int. J. Num. Anal. Methods Geomech.* 14 (1990) 253–278.
11. J.T. Jenkins and S.B. Savage, A theory for the rapid flow of identical, smooth, nearly elastic, spherical particles. *J. Fluid Mech.* 130 (1983) 187–202.
12. S.B. Savage, Analysis of slow high-concentration flows of granular materials. *J. Fluid Mech.* 377 (1998) 1–26.
13. A. Drescher, On the criteria for mass flow in hoppers. *Powder Technol.* 73 (1992) 251–260.
14. L. Bocquet, W. Losert, D. Schalk, T.C. Lubensky and J.P. Gollub, Granular shear flow dynamics and forces: experiment and continuum theory. *Phys. Rev. E* 65 (2001) 011307.

Journal of Engineering Mathematics (2005) 52: 147–165
DOI 10.1007/s10665-004-7768-5

Formulation of non-standard dissipative behavior of geomaterials

MOHAMMED HJIAJ*, WENXIONG HUANG, KRISTIAN KRABBENHØFT and SCOTT W. SLOAN

Department of Civil, Surveying and Environmental Engineering, The University of Newcastle, NSW 2308, Australia
**Author for corresspondence: (mohammed.hjiaj@insa-rennes.fr)*

Received 10 October 2004; accepted in revised form 13 December 2004

Abstract. In this paper, fundamental mathematical concepts for modeling the dissipative behavior of geomaterials are recalled. These concepts are illustrated on two basic models and applied to derive a new form of the evolution law of the modified Cam-clay model. The aim is to discuss the mathematical structure of the constitutive relationships and its consequences on the structural level. It is recalled that non-differentiable potentials provide an appropriate means of modeling rate-independent behavior. The Cam-clay model is revisited and a standard version is presented. It is seen that this standard version is non-dissipative, which at the same time explains why a non-standard version is needed. The partial normality is exploited and an implicit variational formulation of the modified Cam-clay model is derived. As a result, the solution of boundary-value problems can be replaced by seeking stationary points of a functional.

Key words: convex analysis, Fenchel transform, internal variables, modified Cam-clay model, variational inequality

1. Introduction

Developing a model for the inelastic behavior of geomaterials is usually carried out in a rather empirical fashion where experimental data are curve-fitted to derive the constitutive relations. The relevance of the model is then assessed on the basis of its capability to reproduce several key characteristics of material behavior. Nowadays, modeling has attained a high degree of sophistication allowing a fine description of the behavior of materials. As a result complex constitutive relations have emerged that have to be implemented in finite-element codes. Lately much attention has been paid to thermodynamic consistency of constitutive models (see [1] for an application to geotechnical materials). The two principles of thermodynamics are used to validate or reject models if they fail to pass the thermodynamic test. This aspect is essential to ensure that constitutive models are physically consistent. However, the mathematical structure of the constitutive relations is crucial when questions such as existence and uniqueness of the boundary-value problems (BVP) need to be answered. Furthermore, convergence of numerical algorithms is closely related to properties of the constitutive operator. Unfortunately this aspect of the problem is rarely addressed, particularly in geomechanics.

Often constitutive relations are provided in a rather explicit form where the stress-like variables are given as a function of the strain-like variables. For instance, the viscoelastic strain rates are related to the stresses through a linear operator. This defines a mapping between the stress space and the strain rate space called constitutive operator. If the constitutive operator is invertible then the inverse relationship exists. Another example is the flow rule in associated plasticity where the plastic strain rates are related to the stresses through the so-called normality rule and the complementarity conditions. This formulation of the flow rule does not provide an explicit expression of the constitutive operator. The main difficulty

stems from the non-smooth (multi-valued) character of this constitutive model. Using Convex Analysis tools, Moreau [2,3] has shown that, under some conditions, the constitutive operator can be derived from a scalar-valued function that acts like a potential for the flow rule. The work of Moreau [2,3] on the mathematical structure of mechanical laws is an important step in material modeling. The main contribution is probably a unified framework proposed for mechanical models including the multi-valued ones. The variational structure revealed by the "potential form" of the constitutive relationship prove to be useful regarding to the numerical and mathematical aspects of boundary-value problems. This property ensures the existence of stationary principles that becomes minimum principles if the functional is convex. Another key-step has been accomplished by Nguyen Quoc Son [4,5] who extended Moreau's work to more complex multivalued laws ("visco"-plasticity with hardening, damage,$\cdots$) using the phenomenological approach with internal variables. However, geomaterials exhibit frictional behavior, they undergo plastic changes of volume and need to be modeled by plasticity theories with 'non-associated' (or non-standard) flow rule. It is shown that a non-standard model for the modified Cam-clay is not avoidable, otherwise it will lead to a non-dissipative model which contradicts experimental investigations (see [6]). A convenient formulation of this model based on implicit normality is discussed.

2. The constitutive operator

The phenomenological approach with internal variables provides a unified framework for developing various models arising in engineering applications. It consists of supplementing the deformation $\boldsymbol{\varepsilon}$ by a set of internal (strain-like) variables $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_i,\ i = 1, \ldots, n)$ which account for the internal restructuring taking place during the dissipative process. The number and the mathematical nature (tensor, vector or scalar) of the internal variables depend on the model under consideration. The notation used here will be one in which symmetric second-order tensors are represented as six-dimensional vectors and denoted by bold letters. More complex operator are capital doubled (*e.g.* $\mathbb{D}$ for Hooke's tensor). For the sake of a compact representation, internal variables $\boldsymbol{\kappa}_i$ are grouped together in a unique vector $\boldsymbol{\kappa} \in \mathbb{R}^m$ made by the following ordered n-tuples:

$$\boldsymbol{\kappa}^{\mathrm{t}} = \left[\boldsymbol{\kappa}_1^{\mathrm{t}}, \ldots, \boldsymbol{\kappa}_i^{\mathrm{t}}, \ldots, \boldsymbol{\kappa}_n^{\mathrm{t}}\right]$$

where "t" stands for the usual transposition, $\mathbb{R}^m$ is a m-dimensional vector space and $\boldsymbol{\kappa}_i^{\mathrm{t}}$ can be either a vectorial representation of a tensor, a vector or a scalar. The rate of an internal variable, also called velocity, is denoted by a superimposed dot. A set of generalized stresses $\boldsymbol{\pi} = (\boldsymbol{\pi}_i,\ i = 1, \ldots, n)$, responsible for the internal modifications, are defined such that (generalized stresses) $\times$ (rate of change of internal variables) gives the rate of dissipation. Grouping together the generalized stresses in the vector $\boldsymbol{\pi} \in \mathbb{R}^m$, one may give the rate of dissipation as a scalar product in $\mathbb{R}^m$

$$\boldsymbol{\pi} \cdot \dot{\boldsymbol{\kappa}} = \boldsymbol{\pi}_1^{\mathrm{t}} \cdot \dot{\boldsymbol{\kappa}}_1 + \cdots + \boldsymbol{\pi}_i \cdot \dot{\boldsymbol{\kappa}}_i + \cdots + \boldsymbol{\pi}_n \cdot \dot{\boldsymbol{\kappa}}_n,$$

where a dot "$\circ \cdot \circ$" represents the usual scalar product. The m-dimensional linear space $\mathbb{R}^m$, whose elements are the velocities, is called the velocity space and denoted by $\mathcal{V}$. The bilinear form generated by the rate of dissipation puts the velocity space $\mathcal{V}$ in duality with the force space $\mathcal{F}$ comprising the generalized stresses $\boldsymbol{\pi}$:

$$\forall (\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}) \in \mathcal{V} \times \mathcal{F} \ \mapsto\ \dot{\boldsymbol{\kappa}} \cdot \boldsymbol{\pi} \in \mathbb{R}. \tag{1}$$

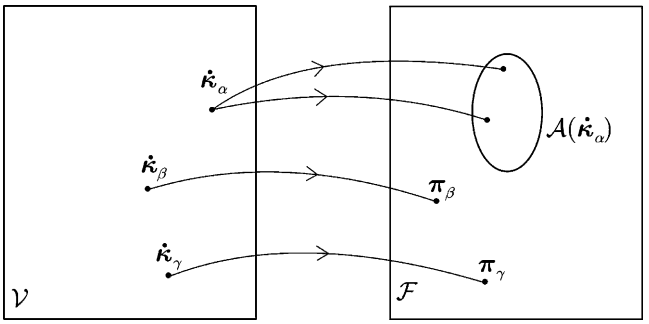


Figure 1. Dissipative mapping.

It is said that $\boldsymbol{\pi}$ and $\dot{\boldsymbol{\kappa}}$ are conjugated with respect to the dissipation. While the evolution of the strain can be controlled externally, the internal variables evolve according to some additional laws called *evolution laws* which complement the state laws (*e.g.* elastic law for an elastic perfectly plastic model). These laws, which describe the evolution of the internal modifications, establish relationships between the rate of change of each $\boldsymbol{\kappa}_i$ and each generalized stress $\boldsymbol{\pi}_i$. From a mathematical viewpoint, the global evolution law defines a certain mapping between $\mathcal{V}$ and $\mathcal{F}$, denoted by $\mathcal{A}$, which maps each $\dot{\boldsymbol{\kappa}} \in \mathcal{V}$ to the set, possibly empty, $\mathcal{A}(\dot{\boldsymbol{\kappa}}) \subset \mathcal{F}$ (see Figure 1).

The relationship between $\dot{\boldsymbol{\kappa}}$ and $\boldsymbol{\pi}$ can be expressed in the following explicit form

$$\mathcal{A}: \mathcal{V} \to \mathcal{F}: \dot{\boldsymbol{\kappa}} \mapsto \boldsymbol{\pi} \in \mathcal{A}(\dot{\boldsymbol{\kappa}}). \tag{2}$$

The operator $\mathcal{A}(\cdot)$ transforms $\dot{\boldsymbol{\kappa}}$ defined in the domain $\mathscr{D}(\mathcal{A}) \subset \mathscr{V}$ into $\boldsymbol{\pi}$ defined in the range of the operator $\mathscr{R}(\mathcal{A}) \subset \mathcal{F}$. In the most general case, the dissipative mapping $\mathcal{A}$ will be nonlinear. Further, the operator $\mathcal{A}$ is said to be single-valued or multi-valued at $\dot{\boldsymbol{\kappa}}$ according to whether $\mathcal{A}(\dot{\boldsymbol{\kappa}})$ is a singleton or a set containing more than one element (see Figure 1). The multi-valuedness is a desirable feature for a dissipative law like the flow rule in plasticity. In what follows, only invertible operators are considered. If the map $\mathcal{A}$ is invertible, the inverse law has the form:

$$\mathcal{A}^{-1}: \mathcal{F} \to \mathcal{V}: \boldsymbol{\pi} \mapsto \dot{\boldsymbol{\kappa}} \in \mathcal{A}^{-1}(\boldsymbol{\pi}). \tag{3}$$

A class of dissipative materials, interesting from both a mathematical and a computational point of view, are those for which the dissipative operator can be obtained as a gradient or a subgradient[1] of a function for all its elements of its domain. For a single-valued operator, the condition ensuring that such a function does, in fact, exist, is the conservativity of the operator, *i.e.*, the vanishing of the related integral along every closed curve in the domain of the operator. If the constitutive operator is differentiable, this condition is ensured by the symmetry of its first Fréchet derivative D

$$[D\mathcal{A}(\dot{\boldsymbol{\kappa}})\mathrm{d}\dot{\boldsymbol{\kappa}}]\delta\dot{\boldsymbol{\kappa}} = [D\mathcal{A}(\dot{\boldsymbol{\kappa}})\delta\dot{\boldsymbol{\kappa}}]\mathrm{d}\dot{\boldsymbol{\kappa}} \tag{4}$$

for any vectors $\mathrm{d}\dot{\boldsymbol{\kappa}}$ and $\delta\dot{\boldsymbol{\kappa}}$ in $\mathcal{V}$. In that case, the constitutive operator can be obtained as the gradient of a scalar-valued function $\phi(\dot{\boldsymbol{\kappa}})$ for all its elements of its domain:

$$\boldsymbol{\pi} = D\phi(\dot{\boldsymbol{\kappa}}). \tag{5}$$

[1]The subgradient is a generalization of the gradient to non-differentiable functions.

In order for the potential $\phi(\dot{\boldsymbol{\kappa}})$ to be convex the operator $\mathcal{A}$ must be positive definite. The "potentiality" condition (4) and the convexity condition must be checked separately before we can claim that an operator derives from a convex potential. There is no complete answer as to whether a multi-valued operator corresponds to the subgradient operator of a scalar-valued function which is not necessarily convex. Some partial results exist which make use of the prox-regularity concept [7, pp. 609–618]. However, if the operator satisfies the maximal cyclic monotonicity condition, it is proven that this operator can be derived as the subgradient of a *convex* scalar-valued function. The operator $\mathcal{A}$ is *cyclically monotonic* if for any family of pairs $(\dot{\boldsymbol{\kappa}}_i, \boldsymbol{\pi}_i) \in \mathcal{V} \times \mathcal{F}$, $i = 0, 1, \ldots, n$ such that $\dot{\boldsymbol{\kappa}}_i \in \mathcal{A}^{-1}(\boldsymbol{\pi}_i)$, the following inequality holds

$$\sum_{i=0}^{n} (\boldsymbol{\pi}_{i+1} - \boldsymbol{\pi}_i) \cdot \dot{\boldsymbol{\kappa}}_i \leq 0, \quad \text{with } n+1 \equiv 0. \tag{6}$$

Therefore the relationship between $\dot{\boldsymbol{\kappa}}$ and $\boldsymbol{\pi}$ takes the following potential structure

$$\dot{\boldsymbol{\kappa}} \in \partial \phi(\boldsymbol{\pi}), \tag{7}$$

where $\phi(\boldsymbol{\pi})$ is a convex scalar-valued function satisfying

$$\phi(\boldsymbol{\pi}') - \phi(\boldsymbol{\pi}) \geq \dot{\boldsymbol{\kappa}} \cdot (\boldsymbol{\pi}' - \boldsymbol{\pi}). \tag{8}$$

The relation (7) represents a *differential inclusion* and the function $\phi(\boldsymbol{\pi})$ is called a *pseudo-potential* where the term "pseudo" is used to emphasize that this function is non-differentiable. The multi-valued character of the relationship $\dot{\boldsymbol{\kappa}}(\boldsymbol{\pi})$ lies in the non-differentiability of $\phi(\boldsymbol{\pi})$ which requires the use of the mathematical operator "$\in$". The condition (6) seems to be quite complicated to use in practice and it is preferred to consider two pairs $(\dot{\boldsymbol{\kappa}}_0, \boldsymbol{\pi}_0)$ and $(\dot{\boldsymbol{\kappa}}_1, \boldsymbol{\pi}_1)$ to obtains the inequality

$$(\boldsymbol{\pi}_0 - \boldsymbol{\pi}_1) \cdot (\dot{\boldsymbol{\kappa}}_0 - \dot{\boldsymbol{\kappa}}_1) \geq 0, \tag{9}$$

which means that the mapping is *monotonically increasing*. If we can find two pairs that violate inequality (9), then the mapping is not monotonically increasing. Obviously, if the mapping is not monotonically increasing, it is not cyclically monotonic. Practically, the condition (9) is used as an necessary test. To rule out the existence of a convex pseudo-potential, it is enough to show that there exist two pairs violating the inequality (9). If we assume the existence of a convex subset $K \subset \mathcal{F}$

$$K \subset \mathcal{F} = \{\boldsymbol{\pi} \in \mathcal{F} \mid f(\boldsymbol{\pi}) \leq 0\}, \tag{10}$$

such that $\dot{\boldsymbol{\kappa}} \neq 0$ if $f(\boldsymbol{\pi}) = 0$, the condition (9) is equivalent to the normality rule. Accordingly, if the evolution law does not satisfy the normality then there is no convex pseudo-potential. In the next section more details will be given about rate-independent models.

The primary advantage of having a potential structure for the constitutive relations is that both analytical and physical insights may be gained. At least for geometrically linear solid mechanics boundary-value problems, it will result in the possibility of applying the calculus of variations. This technique consists of replacing the problem with a system of differential or integro-differential equations (BVP) by the equivalent problem of seeking the stationary points of a proper functional. The extremal formulations make the qualitative study of the problem easier, *i.e.*, the study of the existence, uniqueness and regularity of the solution through the so-called direct methods of variational calculus.

A broad range of dissipative materials present in engineering have more complex dissipative laws which can not take the convenient form of a potential law. One of most illustrative example is the Coulomb frictional-contact law. Other examples are typically those provided by dissipative laws of geomaterials and cyclic (visco)-plasticity models. In this context, the following question arises naturally: how can one preserve all the benefits of a formulation based on the definition of a scalar-valued function, *i.e.*, a potential structure of the dissipative law? An answer to this question is to relax the explicit relation introduced by the potential form by admitting an implicit one.

3. Potential and pseudo-potential

In this section, fundamental aspects of the material modeling discussed in the introduction are illustrated by considering two classical models. The first one is the linear viscous model which under usual symmetry conditions leads to a pair of dual differentiable potentials. The second example is the rigid-perfectly-plastic model where the associated flow rule (with complementarity relations) defines a multi-valued operator that can be derived from a non-differentiable potential. Further details on potential in constitutive modeling are given by Mróz [8, pp. 1–37].

3.1. Potential

Among classical dissipative models, the linear viscous material provides probably the most elementary dissipative law. There is only one internal variable which corresponds to the viscous deformation $\boldsymbol{\varepsilon}^v$, conjugated to the Cauchy stress tensor $\boldsymbol{\sigma}$. The evolution law takes the following simple form

$$\mathcal{E} \mapsto \mathcal{S} : \dot{\boldsymbol{\varepsilon}}^v \mapsto \boldsymbol{\sigma} = \mathcal{L}\,\dot{\boldsymbol{\varepsilon}}^v, \tag{11}$$

where $\mathcal{L}$ is a linear mapping represented by a 6×6 matrix whose elements are constant. The spaces $\mathcal{E}$ and its dual $\mathcal{S}$ merely correspond the six-dimensional space of symmetric second-order tensors. We suppose that $\mathcal{L}$ is symmetric and invertible. Trivially, the scalar-valued function $\psi(\dot{\boldsymbol{\varepsilon}}^v)$

$$\psi : \mathcal{E} \mapsto \mathbb{R} : \dot{\boldsymbol{\varepsilon}}^v \mapsto \frac{1}{2}(\dot{\boldsymbol{\varepsilon}}^v)^{\mathrm{t}}\mathcal{L}\,\dot{\boldsymbol{\varepsilon}}^v \tag{12}$$

is a quadratic form on the velocity space $\mathcal{E}$ and $\mathcal{L}\,\dot{\boldsymbol{\varepsilon}}^v$ is its gradient at $\dot{\boldsymbol{\varepsilon}}^v$. The potential ψ is convex only if the operator $\mathcal{L}$ is positive definite

$$(\dot{\boldsymbol{\varepsilon}}^v)^{\mathrm{t}}\mathcal{L}\,\dot{\boldsymbol{\varepsilon}}^v \geq 0. \tag{13}$$

Thus, the relation (11) may be equivalently written as

$$\boldsymbol{\sigma} = \operatorname{grad}\psi(\dot{\boldsymbol{\varepsilon}}^v). \tag{14}$$

A nice consequence of the *normality rule* (14) is the possibility to make use of the Legendre transform to invert the law (11). If $\boldsymbol{\sigma}$ is related to $\dot{\boldsymbol{\varepsilon}}^v$ by means of a potential $\psi(\dot{\boldsymbol{\varepsilon}}^v)$, Legendre showed that $\dot{\boldsymbol{\varepsilon}}^v$ is, in turn, related to $\boldsymbol{\sigma}$ through a potential ψ^* such that

$$\dot{\boldsymbol{\varepsilon}}^v = \operatorname{grad}\psi^*(\boldsymbol{\sigma}). \tag{15}$$

The potential $\psi^*(\boldsymbol{\sigma})$ is equal to

$$\psi^*(\boldsymbol{\sigma}) = \dot{\boldsymbol{\varepsilon}}^v \cdot \operatorname{grad}\psi(\dot{\boldsymbol{\varepsilon}}^v) - \psi(\dot{\boldsymbol{\varepsilon}}^v), \tag{16}$$

and its expression in terms of $\boldsymbol{\sigma}$ is

$$\psi^*(\boldsymbol{\sigma}) = (\mathcal{L}^{-1}\boldsymbol{\sigma}) \cdot \operatorname{grad} \psi(\mathcal{L}^{-1}\boldsymbol{\sigma}) - \psi(\mathcal{L}^{-1}\boldsymbol{\sigma}). \tag{17}$$

If the operator $\mathcal{L}$ is linear then ψ^* is given by

$$\psi^* : \mathcal{S} \mapsto \mathbb{R} : \boldsymbol{\sigma} \mapsto \frac{1}{2}\boldsymbol{\sigma}^{\mathrm{t}}\mathcal{L}^{-1}\boldsymbol{\sigma}. \tag{18}$$

The functions $\psi(\dot{\boldsymbol{\varepsilon}}^v)$ and $\psi^*(\boldsymbol{\sigma})$ are conjugate (or dual) and related by the following equality

$$\psi(\dot{\boldsymbol{\varepsilon}}^v) + \psi^*(\boldsymbol{\sigma}) = \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^v. \tag{19}$$

For any pair $(\dot{\boldsymbol{\varepsilon}}^{v\prime}, \boldsymbol{\sigma}') \in \mathcal{E} \times \mathcal{S}$ not related by the constitutive relation, we have

$$\psi(\dot{\boldsymbol{\varepsilon}}^{v\prime}) + \psi^*(\boldsymbol{\sigma}') \neq \boldsymbol{\sigma}' \cdot \dot{\boldsymbol{\varepsilon}}^{v\prime}. \tag{20}$$

It is worth mentioning that the Legendre transform does not require that the potential $\psi(\dot{\boldsymbol{\varepsilon}}^v)$ is convex. However, if the constitutive operator satisfies (13), the potential $\psi(\dot{\boldsymbol{\varepsilon}}^v)$ is convex and $\psi^*(\boldsymbol{\sigma})$, which is also convex, can be obtained using the following maximization procedure

$$\psi^*(\boldsymbol{\sigma}) = \sup_{\dot{\boldsymbol{\varepsilon}}^v} [\boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^v - \psi(\dot{\boldsymbol{\varepsilon}}^v)]. \tag{21}$$

When ψ is concave, the same formula is used where "sup" is replaced by a "inf". We can remark that, with these definitions, there is no need for the potential to be differentiable. As a consequence of (21), we have

$$\psi(\dot{\boldsymbol{\varepsilon}}^{v\prime}) + \psi^*(\boldsymbol{\sigma}') \geq \boldsymbol{\sigma}' \cdot \dot{\boldsymbol{\varepsilon}}^{v\prime}, \quad \forall (\dot{\boldsymbol{\varepsilon}}^{v\prime}, \boldsymbol{\sigma}') \in \mathcal{E} \times \mathcal{S}. \tag{22}$$

The mathematical properties of the potentials reflect the nature of the behavior. The one-to-one relation is related to the differentiability of the potential. The convexity of the potentials is a consequence of the monotonic nature of the behavior. Finally, quadratic form implies a linear relation between static and kinematic quantities. All information about the behavior is contained in the function $\psi(\dot{\boldsymbol{\varepsilon}}^v)$. Probably the most attractive property is the existence of variational principles. Weak formulations of a boundary-value problem involving such materials lead to differentiable minimum principles. It is found that the term variational may be understood in different ways. This term can mean: weak formulation (*i.e.*, principle of virtual work), stationary principle or extremum principle. Although a weak formulation can be associated with most of the physical problems, a few of them admit a stationary principle and even fewer admit an extremum principle. Convex or concave potentials lead to extremum principles which are particularly attractive from both a mathematical and a computational point of view. In fact, mathematicians have used them to prove existence and eventually uniqueness of the solution to the corresponding boundary-value problem. Further, the possibility of searching for the solution of a physical problem as a minimum point of a convex functional on a convex set is especially relevant from a computational point view.

Consider a perfectly viscous body occupying a bounded domain $\Omega \subset \mathbb{R}^3$ with boundary Γ, subjected to imposed body force $\bar{\boldsymbol{b}}$, imposed surface traction $\bar{\boldsymbol{t}}$ on the part Γ_t of Γ, and imposed velocity $\bar{\dot{\boldsymbol{u}}}$ on the remaining part of the boundary $\Gamma_u = \Gamma - \Gamma_t$. We assume that the body is fixed on Γ_u, *i.e.*, $\bar{\dot{\boldsymbol{u}}} = \boldsymbol{0}$. Velocities and strain rates are assumed to be small, so that

geometry changes can be neglected and the analysis is performed on the reference configuration. The weak form of equilibrium equations yields to the following functional

$$\int_{\Omega} \boldsymbol{\sigma} \cdot \delta \dot{\boldsymbol{\varepsilon}}^{v} \, \mathrm{d}\Omega - \int_{\Omega} \bar{\boldsymbol{b}} \cdot \delta \dot{\boldsymbol{u}} \, \mathrm{d}\Omega - \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \delta \dot{\boldsymbol{u}} \, \mathrm{d}\Gamma = 0, \tag{23}$$

where $\delta \dot{\boldsymbol{u}} \in \mathscr{V}_{ad}$ is defined by

$$\mathscr{V}_{ad} = \left\{ \delta \dot{\boldsymbol{u}} \in \mathscr{V} \mid \delta \dot{\boldsymbol{\varepsilon}}^{v} = \nabla_S \, \delta \dot{\boldsymbol{u}} \quad \text{in}, \Omega \, , \; \delta \dot{\boldsymbol{u}} = \boldsymbol{0} \quad \text{on} \, \Gamma_u \right\};$$

here ∇_S is the symmetric part of the gradient operator. As a result of the potential structure of the constitutive relation, a solution of the virtual work equation (23) corresponds to a stationary point

$$\mathscr{J}(\dot{\boldsymbol{u}}) \rightarrow \text{stationary} \quad \text{over} \quad \mathscr{V}_{ad}, \tag{24}$$

where

$$\mathscr{J}(\dot{\boldsymbol{u}}) = \int_{\Omega} \psi(\dot{\boldsymbol{\varepsilon}}^{v}(\dot{\boldsymbol{u}})) \, \mathrm{d}\Omega - \int_{\Omega} \bar{\boldsymbol{b}} \cdot \dot{\boldsymbol{u}} \, \mathrm{d}\Omega - \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \dot{\boldsymbol{u}} \, \mathrm{d}\Gamma, \tag{25}$$

which is analogous to the energy functional for elastic bodies. The functional $\mathrm{J}(\dot{\boldsymbol{u}})$ is convex if $\psi(\dot{\boldsymbol{\varepsilon}}^{v})$ is convex/concave and the stationary principle becomes a minimum/maximum principle:

$$\mathscr{J}(\dot{\boldsymbol{u}}) \longrightarrow \min \quad \text{over} \quad \mathscr{V}. \tag{26}$$

In solid mechanics, the weak form (23) always exists but is equivalent to (24) only if the symmetry conditions are satisfied. The second advantage of having a potential is that properties of the solutions appear more explicitly.

Indeed, if J is convex/concave the BVP (functional) has a unique solution but if $\mathscr{J}$ is non-convex it has more than one solution (Figure 2). The extremal formulations are particularly suitable for finding numerical solutions of the problem through direct solution procedures based on combining finite-element and optimization procedures.

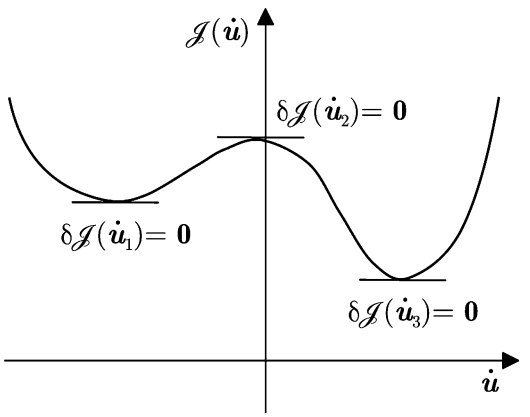


Figure 2. Non-convex functional.

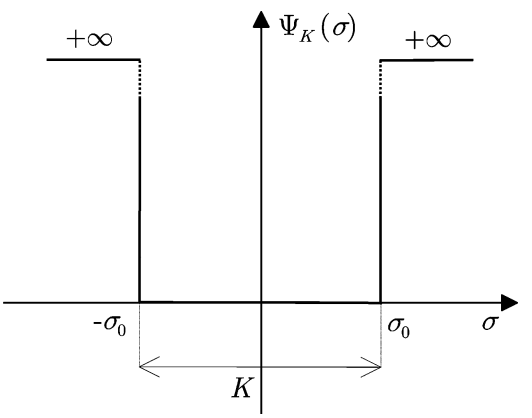


Figure 3. Indicator function for the uniaxial plastic model.

3.2. Pseudo-potential

The concept of potential, although attractive, is not relevant to describe all dissipative laws. The most simple counter-example is given by the classical Mises metal plasticity. The internal variable is the plastic strain tensor $\boldsymbol{\varepsilon}^p \in \mathcal{E}$ and the corresponding associated variable is the Cauchy stress tensor $\boldsymbol{\sigma} \in \mathcal{S}$. For convenience, the plastic strain is decomposed into the volumetric plastic strain e_m^p (belonging to the one-dimensional vector space $\mathcal{E}_m$) and the plastic strain deviator $\mathbf{e}^p$ (belonging to the five-dimensional vector space $\mathcal{E}_d$). The corresponding dual variables are the mean stress s_m (belonging to the space $\mathcal{S}_m$, dual of $\mathcal{E}_m$) and the stress deviator $\mathbf{s}$ (belonging to the space $\mathcal{S}_d$, dual of $\mathcal{E}_d$). Clearly, we have the following decomposition

$$\mathcal{E} = \mathcal{E}_m \oplus \mathcal{E}_d \quad \text{and} \quad \mathcal{S} = \mathcal{S}_m \oplus \mathcal{S}_d$$

and the dual pairing is defined by the bilinear form:

$$\dot{\boldsymbol{\varepsilon}}^p \cdot \boldsymbol{\sigma} = \dot{e}_m^p \, s_m + \dot{\mathbf{e}}^p \cdot \mathbf{s}. \tag{27}$$

The space $\mathcal{S}_d$ is equipped with the von Mises norm

$$\|\mathbf{s}\|_{eq} = \left(\frac{3}{2} \|\mathbf{s}\|^2 \right)^{\frac{1}{2}}, \tag{28}$$

which is used to define a closed convex set of admissible stresses, denoted K

$$K = \left\{ (s_m, \mathbf{s}) \in \mathcal{S} \mid \|\mathbf{s}\|_{eq} - \sigma_0 \leq 0 \right\}, \tag{29}$$

where σ_0 is the yield stress. For an associated flow rule, the direction of the plastic strain rate is given by the gradient to the yield function and its magnitude by the plastic multiplier:

$$\dot{\boldsymbol{\varepsilon}}^p = \dot{\lambda} \frac{\partial f}{\partial \boldsymbol{\sigma}}. \tag{30}$$

The latter is required to satisfy the complementarity relations

$$f(\boldsymbol{\sigma}) \leq 0, \quad \dot{\lambda} \geq 0, \quad \dot{\lambda} \, f(\boldsymbol{\sigma}) = 0. \tag{31}$$

The previous relations refer to the rate formulation of the plastic flow and is probably the most popular form of the flow rule. To gain more insight into the nature of the plastic behavior, the complete flow rule can be written analytically, using a '*if ... then ... else*' statement, as follows

If $\boldsymbol{\sigma} \in \operatorname{int} K$ *then*

$\quad \dot{\boldsymbol{\varepsilon}}^p = \mathbf{0}$! elastic loading/unloading

else

$\quad \left\{ \boldsymbol{\sigma} \in \operatorname{bd} K \text{ and } \exists \, \dot{\lambda} > 0 \text{ such that } \dot{\boldsymbol{\varepsilon}}^p = \dot{\lambda} \frac{\partial f}{\partial \boldsymbol{\sigma}} \right\}$! plastic loading

endif

where "int K" and "bd K" denotes the interior and the boundary of K, repectively. The first part of the statement highlights an important feature of the flow rule, namely its multi-valued nature. Indeed, the zero plastic strain rate can be related to an infinite number of stress states which correspond to the whole elastic domain. As a consequence the constitutive operator $\mathcal{A}$

is multi-valued and therefore it cannot be obtained as a gradient of a differentiable potential. We will see that a potential form of the flow rule can still be derived but by considering a non-differentiable potential, the non-differentiability being required by the multi-valued character of the flow rule. An alternative form of the flow rule (30) and the complementarity relations (31) is the one given by the maximum dissipation inequality, known as Hill's principle:

$$\boldsymbol{\sigma} \in K : \dot{\boldsymbol{\varepsilon}}^p \cdot (\boldsymbol{\sigma} - \boldsymbol{\sigma}') \geq 0, \quad \forall \boldsymbol{\sigma}' \in K. \tag{32}$$

The variational inequality is appreciated by mathematicians because it is a suitable tool for proving the existence of solutions. By transforming adequately the inequality (32), one can obtain the set-valued mapping relating the stress $\boldsymbol{\sigma}$ and the plastic strain rate $\dot{\boldsymbol{\varepsilon}}^p$, *i.e.,* the relationship $\dot{\boldsymbol{\varepsilon}}^p(\boldsymbol{\sigma})$. The idea, due to Moreau [2,3], is to make use of the indicator function [7, p. 6] of the set K to which the stresses $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are required to belong. This particular function, frequently used in Convex Analysis, is defined by

$$\Psi_K(\boldsymbol{\sigma}) = \begin{cases} 0 & \text{if } \boldsymbol{\sigma} \in K, \\ +\infty & \text{otherwise.} \end{cases} \tag{33}$$

This function has a zero ground level within the elastic domain and infinite walls along the yield surface. Figure 3 gives a schematic interpretation of the indicator function for the one-dimensional case.

The function $\Psi_K(\boldsymbol{\sigma})$ is not differentiable in the classical sense. However, the indicator function is convex if the set to which it refers is convex. Having at hand this tool, a key step is to rewrite the variational inequality (32) in the following manner

$$\dot{\boldsymbol{\varepsilon}}^p \cdot (\boldsymbol{\sigma} - \boldsymbol{\sigma}') + \Psi_K(\boldsymbol{\sigma}') \geq \Psi_K(\boldsymbol{\sigma}), \tag{34}$$

where the member function "$\in$" in (32) has been replaced by the value of the indicator function at the corresponding stress. Both inequalities (32) and (34) are equivalent. In fact, we remark that, if inequality (32) is satisfied, it follows that inequality (34) is also satisfied. Conversely, if inequality (34) holds, by taking $\boldsymbol{\sigma}'$ in K, we see that $\Psi_K(\boldsymbol{\sigma}')$ has a finite value (zero) and $\Psi_K(\boldsymbol{\sigma})$ must be equal to zero which means that $\boldsymbol{\sigma}$ is in K and therefore inequality (32) is fulfilled. The inequality (34) corresponds to the *convexity inequality* applied to a non-differentiable function [7, p. 301]. It means that $\dot{\boldsymbol{\varepsilon}}^p$ belongs to the subdifferential of $\Psi_K(\boldsymbol{\sigma})$ at $\boldsymbol{\sigma}$ or equivalently, $\dot{\boldsymbol{\varepsilon}}^p$ and $\boldsymbol{\sigma}$ are related by the differential inclusion:

$$\dot{\boldsymbol{\varepsilon}}^p \in \partial \Psi_K(\boldsymbol{\sigma}). \tag{35}$$

The subdifferential of $\Psi_K(\boldsymbol{\sigma})$ at $\boldsymbol{\sigma}$ corresponds to the set of all subgradients of $\Psi_K(\boldsymbol{\sigma})$ at $\boldsymbol{\sigma}$. It defines a multi-valued operator that maps each point in the domain $\mathcal{S}$ of the function to the closed convex set of its subgradient:

$$\partial \Psi_K(\boldsymbol{\sigma}) : \mathcal{S} \mapsto \mathcal{E} : \boldsymbol{\sigma} \mapsto \partial \Psi_K(\boldsymbol{\sigma}). \tag{36}$$

In particular, for a differentiable function, the subdifferential is reduced to a singleton which corresponds to the classical gradient. The function $\Psi_K(\boldsymbol{\sigma})$ is called complementary dissipation pseudo-potential and denoted by $\psi^*(\boldsymbol{\sigma})$. The relation (35) is equivalent to the flow rule and the complementarity conditions. The above considerations show that by simply allowing the potential to be non-differentiable, we can produce a "potential structure" of the relationship between $\dot{\boldsymbol{\varepsilon}}^p$ and $\boldsymbol{\sigma}$. Accordingly, we can say that differentiable potentials are suited only for

single-valued laws and non-differentiable potentials provide an effective means of representing multi-valued constitutive laws. With the setting (35), the relationship may be inverted by applying the Fenchel transform

$$\psi(\dot{\boldsymbol{\varepsilon}}^p) = \sup_{\boldsymbol{\sigma}} \left[\boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^p - \Psi_K(\boldsymbol{\sigma})\right] = \sup_{\boldsymbol{\sigma} \in K} \left[\boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^p\right], \tag{37}$$

where $\psi(\dot{\boldsymbol{\varepsilon}}^p)$ is the dissipation pseudo-potential, which is convex by construction. The inverse flow rule is then

$$\boldsymbol{\sigma} \in \partial \psi(\dot{\boldsymbol{\varepsilon}}^p), \tag{38}$$

which is equivalent to

$$\boldsymbol{\sigma} \cdot (\dot{\boldsymbol{\varepsilon}}^p - \dot{\boldsymbol{\varepsilon}}^{p\prime}) + \psi(\dot{\boldsymbol{\varepsilon}}^{p\prime}) \geq \psi(\dot{\boldsymbol{\varepsilon}}^p). \tag{39}$$

The formulations (35) and (38) of the flow rule and its inverse are particularly useful for associating dual extremum principles to boundary-value problems involving rigid plastic materials with associated flow rules. The dissipated power is exactly equal to $\psi(\dot{\boldsymbol{\varepsilon}}^p)$; hence, in the present case, $\psi(\dot{\boldsymbol{\varepsilon}}^p)$ may be called the dissipation function of the material. The previous developments are illustrated by considering the deviatoric plastic model. The expression of $\psi^*(\boldsymbol{\sigma})$ does not pose any particular difficulty as it always coincides with the indicator function of the elastic domain but we need to derive the expression of the dissipation pseudo-potential. Using the decomposition (27), the scalar product can be decomposed as

$$\sup_{\boldsymbol{\sigma} \in K} \left[\boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^p\right] = \sup_{\boldsymbol{\sigma} \in K} \left[\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \, \dot{e}_m^p\right]. \tag{40}$$

It is clear that the supremum will be achieved for a vector $\boldsymbol{\sigma}$ colinear to $\dot{\boldsymbol{\varepsilon}}^p$:

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \, \dot{e}_m^p \leq \|\mathbf{s}\| \, \|\dot{\mathbf{e}}^p\| + s_m \, \dot{e}_m^p. \tag{41}$$

To be able to use the yield criterion, we replace the Euclidean norm of $\mathbf{s}$ by the von Mises norm (28)

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \, \dot{e}_m^p \leq \|\mathbf{s}\|_{eq} \, \|\dot{\mathbf{e}}^p\|_{eq}^* + s_m \, \dot{e}_m^p, \tag{42}$$

where $\| \bullet \|_{eq}^*$ corresponds to the dual norm of $\| \bullet \|_{eq}$ and is defined on the space $\mathcal{E}_d$

$$\|\dot{\mathbf{e}}^p\|_{eq}^* = \left(\frac{2}{3} \|\dot{\mathbf{e}}^p\|^2\right)^{\frac{1}{2}}. \tag{43}$$

Taking into account that $\|\mathbf{s}\|_{eq}$ is bounded by σ_0, we have

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \, \dot{e}_m^p \leq \sigma_0 \, \|\dot{\mathbf{e}}^p\|_{eq}^* + s_m \, \dot{e}_m^p. \tag{44}$$

Two distinct possibilities emerge: if $\dot{e}_m^p = 0$, the supremum (40) is $\sigma_0 \, \|\dot{\mathbf{e}}^p\|_{eq}^*$. If on the other hand, $\dot{e}_m^p \neq 0$ then, since the value of s_m is unbounded, so is the supremum (40). Thus, the expression of the dissipation pseudo-potential is given by

$$\psi(\dot{\boldsymbol{\varepsilon}}^p) = \begin{cases} \sigma_0 \, \|\dot{\mathbf{e}}^p\|_{eq}^* & \text{if } \dot{e}_m^p = 0 \\ +\infty & \text{if } \dot{e}_m^p \neq 0 \end{cases}. \tag{45}$$

Introducing the indicator function $\Psi_{\{0\}}(\dot{e}_m^p)$ defined by

$$\Psi_{\{0\}}(\dot{e}_m^p) = \begin{cases} 0 & \text{if } \dot{e}_m^p = 0 \\ +\infty & \text{if } \dot{e}_m^p \neq 0 \end{cases}, \tag{46}$$

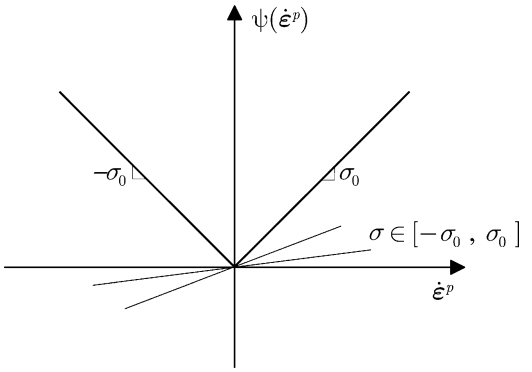


Figure 4. One-dimensional dissipation pseudo-potential.

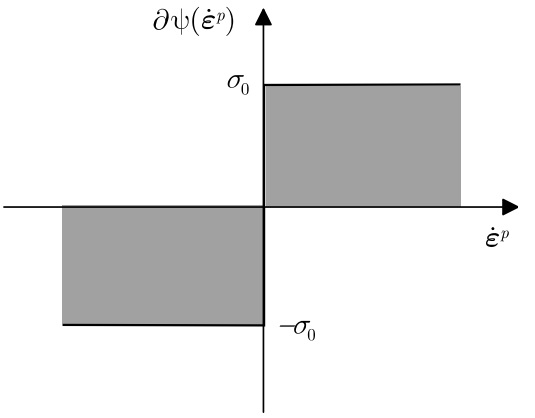


Figure 5. Graph of the rigid-plastic law.

we may write the expression of the dissipation pseudo-potential as

$$\psi(\dot{\boldsymbol{\varepsilon}}^p) = \sigma_0 \, \|\dot{\boldsymbol{\varepsilon}}^p\|^*_{eq} + \Psi_{\{0\}}(\dot{e}^p_m). \tag{47}$$

As a consequence of the Fenchel transform property, the pseudo-potential of dissipation is convex. Moreover, it is a positive function, homogeneous of degree one which takes a zero value at the origin where the function is non-differentiable. In most papers, the dissipation function is defined without taking into account the internal constraints on the plastic strain rate (or internal variable rate in more general models). We decide to take into account these internal constraints by adding to the mechanical dissipation an indicator function, so that the dissipation has the value $+\infty$ for non-physical states. We call this function the "extended dissipation pseudo-potential". As a result, the dissipation is defined for any plastic strain rate vector $\dot{\boldsymbol{\varepsilon}}^p \in \mathcal{E}$ including those which are actually physically impossible. In one-dimension, the expression of the dissipation pseudo-potential (see Figure 4) becomes

$$\psi(\dot{\varepsilon}^p) = \sigma_0 \, |\dot{\varepsilon}^p|. \tag{48}$$

The pseudo-potential of dissipation is differentiable everywhere except at the origin (see on the graph). The subdifferential of $\psi(\dot{\varepsilon}^p)$ is given by

$$\partial\psi(\dot{\varepsilon}^p) = \begin{cases} -\sigma_0, & \sigma < 0 \\ (-\sigma_0, \sigma_0), & \sigma = 0 \\ \sigma_0, & \sigma > 0 \end{cases}. \tag{49}$$

It corresponds to the multi-valued rigid plastic model, geometrically represented in Figure 5. As can be seen from Figure 4, the subdifferential at the origin corresponds the elastic domain.

$$K = \partial\psi(\mathbf{0}).$$

Again, the set-valued relation is related to the non-differentiability of the potential and the convexity of the potentials is a consequence of the monotonic nature of the behavior. Finally, the fact that the pseudo-potential is homogeneous of degree one implies a rate-independent behavior. All relevant information about the behavior is contained in the function $\psi(\dot{\boldsymbol{\varepsilon}}^p)$. Furthermore, the convexity of the pseudo-potential of dissipation is a consequence of the maximum of dissipation principle. The functions $\psi(\dot{\boldsymbol{\varepsilon}}^p)$ and $\psi^*(\boldsymbol{\sigma})$ satisfy the following relation:

$$\psi(\dot{\boldsymbol{\varepsilon}}^{p\prime}) + \psi^*(\boldsymbol{\sigma}') \geq \dot{\boldsymbol{\varepsilon}}^{p\prime} \cdot \boldsymbol{\sigma}', \quad \forall \left(\dot{\boldsymbol{\varepsilon}}^{p\prime}, \boldsymbol{\sigma}'\right) \in \mathcal{E} \times \mathcal{S}. \tag{50}$$

Equality is reached when a couple $(\dot{\boldsymbol{\varepsilon}}^p, \boldsymbol{\sigma})$ is linked by the evolution law. Finally, the evolution law can be expressed equivalently by

$$\dot{\boldsymbol{\varepsilon}}^p \in \partial \Psi_K(\boldsymbol{\sigma}) \Leftrightarrow \boldsymbol{\sigma} \in \partial \psi(\dot{\boldsymbol{\varepsilon}}^p) \Leftrightarrow \psi(\dot{\boldsymbol{\varepsilon}}^p) + \psi^*(\boldsymbol{\sigma}) = \dot{\boldsymbol{\varepsilon}}^p \cdot \boldsymbol{\sigma}. \tag{51}$$

Although the pseudo-potential was introduced here for the simple model of rigid-perfect plasticity, the previous approach can be easily extended to take into account hardening effects by providing additional (to the plastic strain) internal variables. The elastic domain is then expressed in the generalized stress space and the rate form of the evolution law takes the form of a generalized normality rule. This leads to the class of Generalized Standard Materials (GSM) of Bernard Halphen and Nguyen Quoc Son [4, 5] who encompass a large class of materials (plasticity, viscoplasticity and damage models). An abundant literature can be found about evolution laws of plastic materials. In particular, contributions have been devoted to the analysis of the connections existing among Hill's principle of maximum dissipation, the existence of a pseudo-potential of dissipation and the normality rule to a convex elastic locus (see [9, pp. 71–83]).

4. The Cam-clay model

It is well known that geomaterials have a very complicated behavior compared to metals, even if only monotonic loading is considered. It is therefore a challenge in geomechanics to develop relatively simple mathematical models that are able to predict, at least qualitatively, a great number of fundamental aspects of soil behavior. The success of the modified Cam-clay model lies in its ability to capture many of the characteristics of clay behavior by using only a limited number of material parameters. This model belongs to the class of critical-state models which originated from the work of Roscoe and his co-workers at the University of Cambridge [6]. Recent work on the modified Cam-clay model using thermodynamics has been carried out by Collins [10]. Commonly observed features such as hardening/softening, contractancy/dilatancy and the tendency to eventually reach a state in which the stress state and the volume change become stationary are all captured by the modified Cam-clay model. Even at present, the modified Cam-clay model remains widely used for computational applications as further evidence of its success. In this section, we first recall the relations governing the dissipation of the model. The elasticity relations are not discussed (Figure 6).

5. Classical formulation

The modified Cam-clay yield surface (see Figure 6) is defined by

$$f(\mathbf{s}, s_m, p_c) = \|\mathbf{s}\|_{eq}^2 + M^2 s_m^2 - 2\, M^2 s_m\, p_c, \tag{52}$$

where p_c is the "*critical state pressure*" and M a material constant defined by

$$M = \frac{6 \sin\phi}{3 - \sin\phi}$$

with ϕ the internal friction angle. The solid-mechanics convention of strains and stresses is used. In the plane $(s_m, \|\mathbf{s}\|_{eq})$, the yield surface is represented by a family of ellipses (Figure 6) passing through the origin, taking a maximum value for $s_m = p_c$ and intercepting the mean

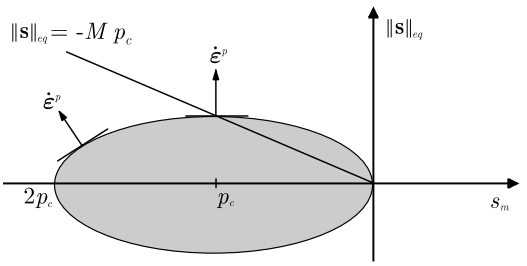


Figure 6. The Cam-clay Model.

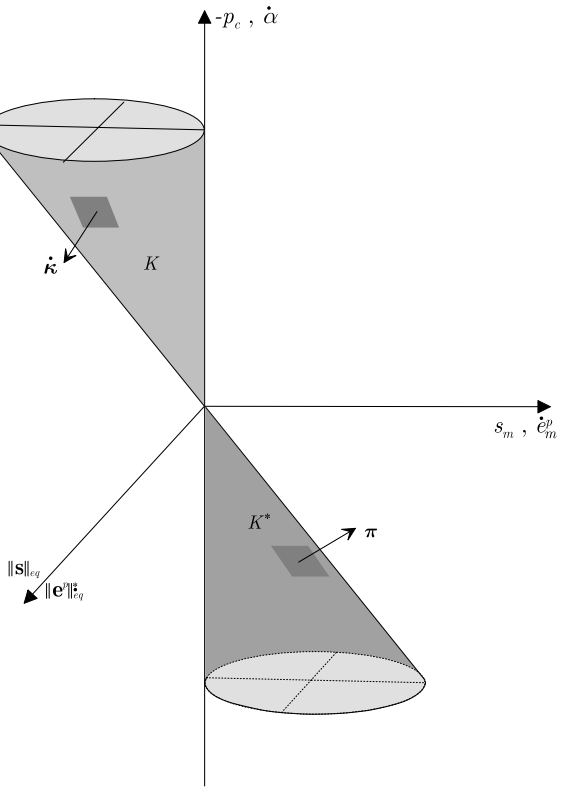


Figure 7. Modified Cam-clay in the generalized stress space.

stress axis at $s_m = 2\,p_c$. This point corresponds to the elastic limit under a hydrostatic loading and is called the "*preconsolidation pressure*". The plastic flow obeys the normality rule

$$\dot{\mathbf{e}}^p = \dot{\lambda}\frac{\partial f}{\partial \mathbf{s}} = 3\,\dot{\lambda}\,\mathbf{s}, \tag{53}$$

$$\dot{e}^p_m = \dot{\lambda}\frac{\partial f}{\partial\, s_m} = 2\,\dot{\lambda}\,M^2(s_m - p_c) \tag{54}$$

and the evolution of the elastic domain is governed by the relation

$$\dot{p}_c = \upsilon\, p_c\, \dot{e}^p_m \tag{55}$$

with

$$\upsilon = \frac{1+e}{\eta - \varsigma},$$

where e is the void ratio of the soil mass, η is the virgin compression index and ς the swell/recompression index. Equation (55) shows that the contractancy leads to a decrease of p_c and therefore the ellipse expands so that the elastic domain is enlarged. On the contrary, dilatancy leads to an increase of p_c (softening phase) which corresponds to a reduction of the elastic domain. When the plastic volumetric strain is zero, p_c becomes constant and the elastic domain stationary. During hardening/softening, the top of the ellipse moves along the straight line $\|\mathbf{s}\|_{eq} = -M\,p_c$ called the "*critical state line*". The critical state line divides the

stress space into a contractant and a dilatant region. For a stress state situated on the part of the yield function where $s_m > p_c$, the behavior will be plastically dilatant. On the contrary, for any stress state situated on the part of the yield function where $s_m < p_c$, the behavior will be plastically contractant. The point situated on the yield curve at the intersection with the critical-state line strains at constant plastic volume.

6. Internal-variable formulation

As mentioned before, our aim is to discuss the dissipative behavior of the model. In addition to the plastic strain $\dot{\boldsymbol{\varepsilon}}^p$, a scalar internal variable α is introduced. This variable accounts for hardening/softening. The conjugated variables are $\boldsymbol{\sigma}$ and p_c, respectively, and the dissipation is given by

$$\boldsymbol{\pi} \cdot \dot{\boldsymbol{\kappa}} = \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^p - p_c \dot{\alpha} = \dot{\mathbf{e}}^p \cdot \mathbf{s} + \dot{e}_m^p s_m - p_c \dot{\alpha}, \tag{56}$$

where $\boldsymbol{\pi}$ and $\dot{\boldsymbol{\kappa}}$ are given by

$$\boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\sigma} \\ p_c \end{bmatrix} \quad \text{and} \quad \dot{\boldsymbol{\kappa}} = \begin{bmatrix} \dot{\boldsymbol{\varepsilon}}^p \\ -\dot{\alpha} \end{bmatrix}. \tag{57}$$

The elastic domain K is now defined in the generalized stress space

$$K = \left\{ \boldsymbol{\pi} \in \mathcal{S} \mid \|\mathbf{s}\|_{eq}^2 + M^2 s_m^2 - 2 M^2 s_m \, p_c \leq 0 \right\}. \tag{58}$$

This corresponds to a cone as shown in Figure 7. Relation (55), which is referred to as the rate form of the state equation, becomes

$$\dot{p}_c = \upsilon \, p_c \, \dot{\alpha}. \tag{59}$$

Comparing (55) and (59), we deduce the so-called hardening rule

$$\dot{e}_m^p - \dot{\alpha} = 0. \tag{60}$$

The evolution law for α does not satisfy the normal rule since we have

$$\dot{\lambda} \frac{\partial f}{\partial p_c} = -2 \dot{\lambda} M^2 s_m \neq -\dot{\alpha}. \tag{61}$$

Therefore, the model is not standard generalized, since we do not have generalized normality (normality for each component of $\dot{\boldsymbol{\kappa}}$). Therefore, we need to introduce another scalar function, called plastic potential, from which the evolution laws can be deduced by applying the normality rule. The expression of the plastic potential is

$$g(\mathbf{s}, s_m, p_c) = \|\mathbf{s}\|_{eq}^2 + M^2 (p_c - s_m) \tag{62}$$

and the evolution laws are given by

$$\dot{\mathbf{e}}^p = \dot{\lambda} \frac{\partial g}{\partial \mathbf{s}} = 3 \dot{\lambda} \, \mathbf{s}, \quad \dot{e}_m^p = \dot{\lambda} \frac{\partial g}{\partial s_m} = 2 \dot{\lambda} M^2 (s_m - p_c), \quad -\dot{\alpha} = \dot{\lambda} \frac{\partial g}{\partial p_c} = 2 \dot{\lambda} M^2 (p_c - s_m).$$

Before continuing, let us derive an equivalent expression for the yield function which will have the property of being homogeneous of degree one. The square in (52) is completed to get

$$\|\mathbf{s}\|_{eq}^2 + M^2 (s_m - p_c)^2 \leq M^2 p_c^2. \tag{63}$$

Taking into account that p_c is always negative, we obtain an alternative expression of the elastic domain given by

$$\sqrt{\|\mathbf{s}\|_{eq}^2 + M^2(s_m - p_c)^2} \leq -M\, p_c \tag{64}$$

and the expression of the yield function is now

$$f(\boldsymbol{\sigma}, p_c) = \sqrt{\|\mathbf{s}\|_{eq}^2 + M^2(s_m - p_c)^2} + M\, p_c, \tag{65}$$

which is homogenous of degree one:

$$f(\beta\,\boldsymbol{\sigma}, \beta\, p_c) = \beta\, f(\boldsymbol{\sigma}, p_c). \tag{66}$$

By introducing the following notations

$$\boldsymbol{\sigma} = \begin{bmatrix} \mathbf{s} \\ s_m \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} \mathbf{0} \\ p_c \end{bmatrix},$$

we obtain the yield function (65) as follows:

$$f(\boldsymbol{\sigma}, p_c) = \|\boldsymbol{\sigma} - \mathbf{X}\|_{cc} + M\, p_c,$$

where

$$\|\boldsymbol{\sigma} - \mathbf{X}\|_{cc} = \left(\|\mathbf{s}\|_{eq}^2 + M^2(s_m - p_c)^2\right)^{\frac{1}{2}}. \tag{67}$$

With this new expression of the yield function, the flow rule becomes

$$\dot{\mathbf{e}}^p = \frac{3\,\dot{\lambda}}{2}\,\frac{\mathbf{s}}{\|\boldsymbol{\sigma} - \mathbf{X}\|_{cc}}, \tag{68}$$

$$\dot{e}_m^p = \dot{\lambda}\, M^2\,\frac{(s_m - p_c)}{\|\boldsymbol{\sigma} - \mathbf{X}\|_{cc}}, \tag{69}$$

which leads to the following expression for the plastic multiplier $\dot{\lambda}$

$$\dot{\lambda} = \|\dot{\boldsymbol{\varepsilon}}^p\|_{cc}^* = \left(\|\dot{\mathbf{e}}^p\|_{eq}^2 + \frac{(\dot{e}_m^p)^2}{M^2}\right)^{\frac{1}{2}}, \tag{70}$$

where the norm $\|\bullet\|_{cc}^*$ defined on the velocity space is dual to the norm (67) in the sense that

$$(\boldsymbol{\sigma} - \mathbf{X}) \cdot \dot{\boldsymbol{\varepsilon}}^p \leq \|\boldsymbol{\sigma} - \mathbf{X}\|_{cc}\, \|\dot{\boldsymbol{\varepsilon}}^p\|_{cc}^*. \tag{71}$$

7. Standard version of the modified Cam-clay model

Suppose that we would like to have a generalized standard model. The hardening rule (60) must be different in order to satisfy the generalized normality. Indeed, by applying the normality rule for the internal variable α, we find the following relationship for the hardening rule

$$-\dot{\alpha} = \dot{\lambda}\frac{\partial f}{\partial p_c} = -\dot{e}_m^p + M\, \|\dot{\boldsymbol{\varepsilon}}^p\|_{cc}^*. \tag{72}$$

So, to have a standard model the relation, Equation (72) should be used instead of (60). It is worth mentioning that now $-\alpha$ is a non-decreasing variable and therefore softening cannot

occur. With this expression of $-\alpha$, the evolution rule for all internal variables can be written in the following compact relation:

$$\dot{\boldsymbol{\kappa}} \in \partial \psi^*(\boldsymbol{\pi}), \tag{73}$$

where $\psi^*(\boldsymbol{\pi})$ is the complementary dissipation pseudo-potential which corresponds to the indicator function of the elastic domain K expressed in the generalized stress space

$$\Psi_K(\boldsymbol{\pi}) = \begin{cases} 0 & \text{if } \boldsymbol{\pi} \in K, \\ +\infty & \text{otherwise} \end{cases}. \tag{74}$$

A standard model for clay seems not to be appropriate, since it does not reproduce the softening behavior observed experimentally. Furthermore, the dissipation is always equal to zero. Indeed, the dissipation is obtained as follows

$$\psi(\dot{\boldsymbol{\kappa}}) = \sup_{\boldsymbol{\pi} \in K} [\boldsymbol{\pi} \cdot \dot{\boldsymbol{\kappa}}] = \sup_{\boldsymbol{\pi} \in K} \left[\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \dot{e}_m^p - p_c \dot{\alpha}\right]. \tag{75}$$

It is clear that the supremum will be achieved for a vector π colinear to $\dot{\kappa}$:

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \dot{e}_m^p - p_c \dot{\alpha} \leq \|\mathbf{s}\| \, \|\dot{\mathbf{e}}^p\| + s_m \dot{e}_m^p - p_c \dot{\alpha}. \tag{76}$$

Adding and subtracting $p_c \dot{e}_m^p$, we have

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \dot{e}_m^p - p_c \dot{\alpha} \leq \|\mathbf{s}\| \, \|\dot{\mathbf{e}}^p\| + (s_m - p_c) \dot{e}_m^p + p_c (\dot{e}_m^p - \dot{\alpha}). \tag{77}$$

Now, we make use of the Cauchy–Schwarz inequality (71), to obtain

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \dot{e}_m^p - p_c \dot{\alpha} \leq \|\boldsymbol{\sigma} - \mathbf{X}\|_{cc} \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + p_c (\dot{e}_m^p - \dot{\alpha}). \tag{78}$$

Taking into account that $\|\boldsymbol{\sigma} - \mathbf{X}\|_{cc}$ is bounded by $-M\, p_c$, we have

$$\mathbf{s} \cdot \dot{\mathbf{e}}^p + s_m \dot{e}_m^p - p_c \dot{\alpha} \leq -p_c \left(M \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + \dot{\alpha} - \dot{e}_m^p\right). \tag{79}$$

Two distinct possibilities emerge. If we have

$$M \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + \dot{\alpha} - \dot{e}_m^p \leq 0, \tag{80}$$

then because $-p_c$ is positive, the dissipation is equal to zero. On the other hand, if we have

$$M \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + \dot{\alpha} - \dot{e}_m^p \geq 0, \tag{81}$$

then, since the value of $-p_c$ is unbounded, so is the supremum (75). Thus, we have

$$\psi(\dot{\boldsymbol{\kappa}}) = \begin{cases} 0 & \text{if } M \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + \dot{\alpha} - \dot{e}_m^p \leq 0 \\ +\infty & \text{if } M \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + \dot{\alpha} - \dot{e}_m^p \geq 0 \end{cases}, \tag{82}$$

in short

$$\psi(\dot{\boldsymbol{\kappa}}) = \Psi_{K^*}(\dot{\boldsymbol{\kappa}}), \tag{83}$$

where K^* is the cone dual to the cone K (see Figure 7) and defined by

$$K^* = \{\dot{\boldsymbol{\kappa}} \in \mathcal{V} \mid M \, \|\dot{\boldsymbol{e}}^p\|_{cc}^* + \dot{\alpha} - \dot{e}_m^p \leq 0\}. \tag{84}$$

These results do not come as a surprise. Indeed, it is well known that, if the yield surface does not strictly contain the origin and the generalized normality rule applies, then the dissipation

is zero and the dissipation pseudo-potential is the cone dual to the cone of admissible stresses K (see Figure 7). A typical example where the dissipation is also zero is given by the cohesionless Mohr–Coulomb model with an associated flow rule. Accordingly, to have a non-zero dissipation, we need a non-standard law as the one provided by the modified Cam-clay model itself. The functions $\psi(\dot{\boldsymbol{\kappa}})$ and $\psi^*(\boldsymbol{\pi})$ satisfy the following relation:

$$\psi(\dot{\boldsymbol{\kappa}}') + \psi^*(\boldsymbol{\pi}') \geq \boldsymbol{\pi}' \cdot \dot{\boldsymbol{\kappa}}', \quad \forall \left(\boldsymbol{\pi}', \dot{\boldsymbol{\kappa}}'\right) \in \mathcal{V} \times \mathcal{F}. \tag{85}$$

A pair $(\boldsymbol{\pi}, \dot{\boldsymbol{\kappa}})$ related by the generalized normality satisfies

$$\dot{\boldsymbol{\kappa}} \in \partial\psi^*(\boldsymbol{\pi}) \Leftrightarrow \boldsymbol{\pi} \in \partial\psi(\dot{\boldsymbol{\kappa}}) \Leftrightarrow \psi(\dot{\boldsymbol{\kappa}}) + \psi^*(\boldsymbol{\pi}) = \boldsymbol{\pi} \cdot \dot{\boldsymbol{\kappa}}. \tag{86}$$

Using the yield function as given by (65), which is homogeneous of degree one, we can see that the dissipation is zero by simply applying the generalized normality and using the Euler identity

$$\mathcal{D} = \boldsymbol{\pi} \cdot \dot{\boldsymbol{\kappa}} = \boldsymbol{\pi} \cdot \dot{\lambda} \frac{\partial f}{\partial \boldsymbol{\pi}} = \dot{\lambda}\, f(\boldsymbol{\pi}) = 0,$$

since $f(\boldsymbol{\pi}) = 0$. If the origin is inside the convex domain K in the generalized stresses space, it is possible to express the region K as a level set $\{\boldsymbol{\pi} : \gamma_K(\boldsymbol{\pi}) \leq 1\}$ where γ_K is a nonnegative, positively homogeneous convex function called the *gauge*. The mathematical definition of the gauge is

$$\gamma_K(\boldsymbol{\pi}) = \inf\{\mu > 0 : \boldsymbol{\pi} \in \mu K\}$$

and the dissipation is now given by

$$\mathcal{D} = \dot{\lambda},$$

where the plastic multiplier $\dot{\lambda}$ is obtained using $\gamma_K(\boldsymbol{\pi})$.

8. Implicit normality rule

The modified Cam-clay model is non-standard, but we will see below that it is still possible to obtain a variational formulation of the evolution law. In the original model, the non-normality is partial and concerns only the internal variable α. Applying the following change of variables

$$-\dot{\vartheta} = -\left(\dot{\alpha} - M \left\|\dot{\boldsymbol{e}}\right\|^p\right), \tag{87}$$

we can recover the normality rule

$$\dot{\boldsymbol{e}}^p \in \partial_{\boldsymbol{\sigma}} \psi^*(\boldsymbol{\sigma}, p_c) \quad \text{and} \quad -\dot{\vartheta} \in \partial_{p_c} \psi^*(\boldsymbol{\sigma}, p_c) \tag{88}$$

or in more compact form

$$\dot{\boldsymbol{\xi}} \in \partial\psi^*(\boldsymbol{\pi}), \tag{89}$$

where the vector $\dot{\boldsymbol{\xi}}$ is given by

$$\dot{\boldsymbol{\xi}} = \begin{bmatrix} \dot{\boldsymbol{e}}^p \\ -\dot{\vartheta} \end{bmatrix}. \tag{90}$$

The inverse law is

$$\boldsymbol{\pi} \in \partial\psi(\dot{\boldsymbol{\xi}}), \tag{91}$$

where $\psi(\dot{\boldsymbol{\xi}})$ is the indicator function of K^* which now depends on $\dot{\boldsymbol{\varepsilon}}^p$ and $-\dot{\vartheta}$

$$K^* = \{\dot{\boldsymbol{\xi}} \in \mathcal{V} \mid M \, \|\dot{\boldsymbol{\varepsilon}}^p\|_{cc}^* + \dot{\vartheta} - \dot{e}_m^p \leq 0\}. \tag{92}$$

Here $\psi(\dot{\boldsymbol{\xi}})$ and $\psi^*(\boldsymbol{\pi})$ satisfy the Fenchel inequality

$$\psi(\dot{\boldsymbol{\xi}}') + \psi^*(\boldsymbol{\pi}') \geq \boldsymbol{\pi}' \cdot \dot{\boldsymbol{\xi}}', \quad \forall \left(\boldsymbol{\pi}', \dot{\boldsymbol{\xi}}'\right) \in \mathcal{V} \times \mathcal{F}. \tag{93}$$

Although this relation provides further insight into such a plastic model, additional developments can still be made to establish a relationship between $\dot{\boldsymbol{\kappa}}$ and $\boldsymbol{\pi}$ based on a normality rule. To recover a relation between the dual variables $\dot{\boldsymbol{\kappa}}$ and $\boldsymbol{\pi}$, we add $\boldsymbol{\pi}' \cdot \dot{\boldsymbol{\kappa}}'$ to both sides of (93),

$$\psi(\dot{\boldsymbol{\xi}}') + \psi^*(\boldsymbol{\pi}') + \boldsymbol{\pi}' \cdot (\dot{\boldsymbol{\kappa}}' - \dot{\boldsymbol{\xi}}') \geq \boldsymbol{\pi}' \cdot \dot{\boldsymbol{\kappa}}', \quad \forall (\boldsymbol{\pi}', \dot{\boldsymbol{\xi}}') \in \mathcal{V} \times \mathcal{F}. \tag{94}$$

The left-hand side of (94) is a function of both $\dot{\boldsymbol{\kappa}}'$ and $\boldsymbol{\pi}'$, which cannot be represented as the sum of two functions, one of $\dot{\boldsymbol{\kappa}}'$ and another of $\boldsymbol{\pi}'$. We call this function a *bi-potential* and its general expression is given by

$$b_p(\dot{\boldsymbol{\xi}}', \boldsymbol{\pi}') := \psi(\dot{\boldsymbol{\xi}}') + \psi^*(\boldsymbol{\pi}') + \boldsymbol{\pi}' \cdot (\dot{\boldsymbol{\kappa}}' - \dot{\boldsymbol{\xi}}'). \tag{95}$$

The right-hand side of (95) is developed using the change of variables (87). The cone K^* is not the dual of K anymore. Its expression is given by

$$K^* = \{\dot{\boldsymbol{\kappa}} \in \mathcal{V} \mid \dot{\alpha} - \dot{e}_m^p \leq 0\}. \tag{96}$$

Developing the scalar product in (95), we obtain the bi-potential for the modified Cam-clay model:

$$b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}) = \Psi_K(\boldsymbol{\pi}) + \Psi_{K^*}(\boldsymbol{\kappa}) - M \, p_c \|\dot{\boldsymbol{\varepsilon}}^p\|_{cc}^*. \tag{97}$$

The bi-potential is positive function and satisfies the fundamental inequality

$$b_p(\dot{\boldsymbol{\kappa}}', \boldsymbol{\pi}') \geq \dot{\boldsymbol{\kappa}}' \cdot \boldsymbol{\pi}'. \tag{98}$$

A strict equality is obtained in (98) for any pair $(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi})$ related by the evolution law:

$$b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}) = \dot{\boldsymbol{\kappa}} \cdot \boldsymbol{\pi}. \tag{99}$$

The relations (98) and (99) can be combined to give

$$\forall \boldsymbol{\pi}' \in \mathcal{F}: \quad b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}') - b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}) \geq \dot{\boldsymbol{\kappa}} \cdot (\boldsymbol{\pi}' - \boldsymbol{\pi}), \tag{100}$$

$$\forall \dot{\boldsymbol{\kappa}}' \in \mathcal{V}: \quad b_p(\dot{\boldsymbol{\kappa}}', \boldsymbol{\pi}) - b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}) \geq \boldsymbol{\pi} \cdot (\dot{\boldsymbol{\kappa}}' - \dot{\boldsymbol{\kappa}}), \tag{101}$$

which means that

- the bi-potential is bi-convex that is $b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi})$ is a convex function of $\dot{\boldsymbol{\kappa}} \in \mathcal{V}$ for each $\boldsymbol{\pi} \in \mathcal{F}$ and a convex function of $\boldsymbol{\pi} \in \mathcal{F}$ for each $\dot{\boldsymbol{\kappa}} \in \mathcal{V}$;
- the evolution law and its inverse derive from the bi-potential $b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi})$

$$\dot{\boldsymbol{\kappa}} \in \partial_{\boldsymbol{\pi}} b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}) \quad \text{and} \quad \boldsymbol{\pi} \in \partial_{\dot{\boldsymbol{\kappa}}} b_p(\dot{\boldsymbol{\kappa}}, \boldsymbol{\pi}). \tag{102}$$

The advantage of the present formulation results in a compact form of the evolution law formulated with one variational inequality. This formulation of the evolution law can be advantageously exploited to derive a robust algorithm. The relations (102) are essential for the derivation of stationary principles involving a functional that depends now on both the velocities and the stresses.

9. Conclusions

An important task for an engineer is to assess the stability of a geotechnical structure. Nowadays this task is carried out using finite-element codes in conjunction with complex constitutive models. For most practical cases (small displacements and small deformations), the nature of the constitutive operator has a significant influence on the convergence of numerical algorithms. It has been recognized that a variational formulation has several advantages, among them, the possibility to associate extremum (or at least stationary) principles to weak formulations of the initial/boundary-value problems. It also permits to express the behavior in a succinct manner. Indeed the stress-strain relationship derives from a scalar-valued function which acts as a potential. For plastic models such a property exists if the maximum dissipation principle holds. However, geomaterial models do not exhibit such a strong variational structure. By allowing an implicit form of the evolution rule, one may recover a weaker variational formulation and the pseudo-potential concept (introduced by Moreau) can be extended to cover non-standard behaviors. The pseudo-potential is replaced by the bi-potential, which depends on both the generalized stresses and the velocities. The bi-potential is not convex but bi-convex, which means convex with respect to the generalized stresses and the plastic strain rates when considered separately. The partial sub-derivatives of the bi-potential yield the evolution law and its inverse. It has been shown that the evolution law of the modified Cam-clay model can be derived from a bi-potential which serves as a "potential" for both the generalized stresses and the velocities. As a consequence, coupled extremum principles exist. These principles are not as strong as the usual extremum (or stationary) principles since they involve static and kinematic variables, but at least they provide new insights into this difficult problem. Further research has to be carried out to design new algorithms using, for instance, mixed formulations.

References

1. I.F. Collins and G.T. Houlsby, Application of thermomechanical principles to the modelling of geotechnical materials. *Proc. R. Soc. London, Series A* 453 (1997) 1975–2001.
2. J.J. Moreau, La notion de sur-potentiel et les liaisons en élastostatique. *Comptes Rendus de l'Académie des Sciences* 267 (1968) 954–957.
3. J.J. Moreau, Sur les lois de frottement, de plasticité et de viscosité. *Comptes Rendus de l'Académie des Sciences* 271 (1970) 608–611.
4. Quoc Son. Nguyen, Matériaux élasto-visco-plastiques à potentiel généralisé. *Comptes Rendus de l'Académie des Sciences* 277 (1973) 915–918.
5. B. Halphen and Quoc Son. Nguyen, Sur les matériaux standard généralisés. *J. Mécanique* 14 (1975) 39–63.
6. K.H. Roscoe and J.B. Burland, On the generalized stress–strain behaviour of "wet" clays. In: J. Heyman, & F.A. Leckie (eds.), *Engineering Plasticity*. Cambridge: Cambridge University Press (1968) pp. 535–609.
7. R.T. Rockafellar and J.B. Wets, *Variational Analysis*. Berlin: Springer (1998) 733 pp.
8. Z. Mróz, *Mathematical Models of Inelastic Behavior*. Solid Mechanics Division, University of Waterloo, Canada (1973) 160 pp.
9. W. Han and B.D. Reddy, *Plasticity. Mathematical Theory and Numerical Analysis*. Berlin: Springer (1999) 371 pp.
10. I.F. Collins, A systematic procedure for constructing critical state models in three dimensions. *Int. J. Solids Struct.* 40 (2003) 4379–4397.

Journal of Engineering Mathematics (2005) 52: 167–184
DOI 10.1007/s10665-004-6451-1

Bifurcation analysis for shear localization in non-polar and micro-polar hypoplastic continua

WENXIONG HUANG, MOHAMMED HJIAJ and SCOTT W. SLOAN
School of Engineering, The University of Newcastle, Callaghan NSW 2308, Australia
(wh670@alinga.newcastle.edu.au)

Received 16 December 2003; accepted in revised form 15 November 2004

Abstract. In this paper, shear localization in granular materials is studied as a bifurcation problem based on a conventional (non-polar) and a micro-polar continuum description. General bifurcation conditions are formulated for a non-polar hypoplastic model and its micro-polar continuum extension. These conditions define stress, couple stress and density states at which weak discontinuity bifurcation may occur. The stress states for bifurcation are then compared with the peak stress states, which define a bounding surface for the accessible stress domain in the principal stress space. The results show that, in a micro-polar continuum description, the constitutive model may no longer be associated with weak discontinuity bifurcation.

Key words: bifurcation analysis, granular materials, hypoplasticity, micro-polar continuum, non-polar continuum

1. Introduction

Localized deformation, in the form of narrow shear zones known as shear bands, is a widely observed phenomenon in granular materials when a certain limit state is approached. Analysis of shear localization can give some insight into the failure mechanism of granular materials and may provide further understanding for constitutive modelling of inelastic behavior.

Following the pioneering work of Rice [1] and Rudnicki and Rice [2], the phenomenon of shear localization has often been directly associated with the weak discontinuity bifurcation properties of constitutive models. The onset of shear localization was understood "in the sense that the constitutive relations may allow the homogeneous deformation of an initially uniform material to lead to a bifurcation point, at which non-uniform deformation can be incipient in a planar band under conditions of continuing equilibrium and continuing deformation outside the zone of localization" [2]. The technique of bifurcation analysis has been widely used to predict the occurrence of shear localization [3,4].

It is well known that constitutive models developed within the framework of conventional continuum theory are limited to modelling pre-bifurcation behavior. Due to the lack of an internal length, these models cannot capture the thickness of the localized zones. As a result, numerical modelling of post-bifurcation behavior suffers from severe mesh-dependency.

Consideration of the relevant micro-mechanics leads to the application of enhanced continuum theory, including non-local theory, higher-gradient theory and micro-polar continuum theory, in the macro-description of granular materials. These theories introduce a characteristic length, which regularizes the failure process with non-local effects. In particular, recent work has established links between inter-granular contacts and the continuum quantities such as stress and couple stress [5,6]. Thus micro-polar continuum theory has become a suitable framework for macro-description of granular media. The theory has been used to study shear

localization in granular materials by Mühlhaus [7], de Borst [8], Dietsche *et al.* [9] and Ehlers and Volk [10] who adopted an elastoplastic approach, and by Tejchman [11], Tejchman and Bauer [12], Bauer and Huang [13], Tejchman and Gudehus [14] and Huang and Bauer [15] who adopted a hypoplastic approach. It is widely accepted that post-bifurcation behavior can be well captured by micro-polar continuum models. Numerical results for shearing of a granular layer between two parallel plates show that a localized shear zone of finite thickness can be obtained, which is mesh-independent provided that the element size is small enough [14, 15].

An open question with the micro-polar continuum description of granular materials is whether shear localization will occur in a homogeneous deforming specimen in the form of weak discontinuity bifurcation (as is the case in a conventional continuum description). This paper focuses on bifurcation analysis of constitutive models developed for cohesionless granular materials using a conventional (non-polar) continuum approach and a micro-polar continuum approach. The non-polar hypoplastic model proposed by Gudehus [16] and Bauer [17], and its micro-polar continuum extension formulated by Huang *et al.* [18], are employed for this purpose. In these non-polar and micro-polar models, the void ratio is incorporated as a measurement of density. This allows the stationary state (the so-called critical state) to be described and the pressure- and density-dependent behavior to be captured for a wide range of stress and density levels with a single set of constitutive constants.

Hypoplastic constitutive models belong to a category known as incrementally nonlinear models. Bifurcation analysis of non-polar models of this type was first discussed by Kolymbas [19] and Chambon and Desrues [20]. A more general analysis can be found in [21,22]. Bauer and Huang [23] and Bauer [24] discussed the pressure and density effects in bifurcation and shear localization. Wu and Sikora [25] and Wu [26] presented a bifurcation surface in principal-stress space for a density-independent hypoplastic model. Fewer results are known for bifurcation analysis of micro-polar continuum models. Some early results can be found in, for instance, [9,27].

In this paper, a general criterion for bifurcation is first derived for the non-polar hypoplastic continuum. By extending the concept of a weak discontinuity in the micro-polar continuum, a similar bifurcation criterion is formulated for the micro-polar hypoplastic continuum. In order to assess the possibility of shear bifurcation in a general way, a condition for the peak stress state is provided. Geometric representations of the bifurcation states and peak stress states in the deviatoric stress plane are then presented, from which the accessibility of bifurcation points can be determined.

Symbolic notation is used for vectors and tensors in this paper. Vectors and second-order tensors are distinguished by bold-faced font and fourth-order tensors by calligraphic font. Index notation is used by referring to a fixed orthogonal Cartesian coordinate system. The second-order unit tensor and the fourth-order unit tensor are denoted by $\mathbf{I}$ and $\mathcal{I}$, respectively. Their components read $(\mathbf{I})_{ij} = \delta_{ij}$ and $(\mathcal{I})_{ijkl} = \delta_{ik}\delta_{jl}$ with δ_{ij} being the Kronecker delta. The permutation symbol $\boldsymbol{\epsilon}$ is used with $(\boldsymbol{\epsilon})_{ijk} = 1$ for $ijk \in \{\{1,2,3\},\{2,3,1\},\{3,1,2\}\}$, $(\boldsymbol{\epsilon})_{ijk} = -1$ for $ijk \in \{\{1,3,2\},\{2,1,3\},\{3,2,1\}\}$ and $(\boldsymbol{\epsilon})_{ijk} = 0$ otherwise. The dot-product operation is used for $\mathbf{a}\cdot\mathbf{b} = a_i b_i$, $(\mathbf{A}\cdot\mathbf{b})_i = A_{ij}b_j$, $\mathbf{A}:\mathbf{B} = A_{ij}B_{ij}$ and $(\mathcal{A}:\mathbf{B})_{ij} = A_{ijkl}B_{kl}$. Here, the usual summation convention for dummy indices is adopted. Dyadic multiplication is denoted by the symbol $\otimes$. For instance $(\mathbf{a}\otimes\mathbf{b})_{ij} = a_i b_j$ and $(\mathbf{A}\otimes\mathbf{B})_{ijkl} = A_{ij}B_{kl}$. The Euclidean norm is used for all vectors so that $\|\mathbf{a}\| = \sqrt{\mathbf{a}\cdot\mathbf{a}}$ and all tensors so that $\|\mathbf{A}\| = \sqrt{\mathbf{A}:\mathbf{A}}$. The Nabla operator ∇ is so defined that $(\nabla\mathbf{a})_{ij} = \partial a_i/\partial x_j$ and $(\nabla\cdot\mathbf{a}) = \partial a_i/\partial x_i$.

2. Hypoplastic description of granular materials

2.1. Outline of the micro-polar continuum

A micro-polar continuum is characterized by additional rotational degrees of freedom and the presence of a couple stress which is work-conjugated with the micro-curvature (associated with the rotational degrees of freedom). Material particles in a micro-polar continuum can translate and rotate independently. Here we use the vectors $\dot{\mathbf{u}}$ and $\dot{\mathbf{w}}^c$ to denote the rate of translation (velocity) and the rate of rotation (angular velocity) of material particles. Then the deformation rate of a micro-polar (Cosserat) continuum can be measured by the strain rate $\dot{\boldsymbol{\varepsilon}}^c$ and the micro-curvature rate $\dot{\boldsymbol{\kappa}}$, which are defined by the following kinematic relations:

$$\dot{\boldsymbol{\varepsilon}}^c = \nabla\dot{\mathbf{u}} + \boldsymbol{\epsilon}\cdot\dot{\mathbf{w}}^c, \quad \dot{\varepsilon}^c_{ij} = \partial\dot{u}_i/\partial x_j + \epsilon_{ijk}\dot{w}^c_k, \tag{1a}$$

$$\dot{\boldsymbol{\kappa}} = \nabla\dot{\mathbf{w}}^c, \qquad \dot{\kappa}_{ij} = \partial\dot{w}^c_i/\partial x_j. \tag{1b}$$

By introducing the notion of micro-spin $\dot{\boldsymbol{\omega}}^c = -\boldsymbol{\epsilon}\cdot\dot{\mathbf{w}}^c$, Equation (1a) can also be written in the form of:

$$\dot{\boldsymbol{\varepsilon}}^c = \dot{\boldsymbol{\varepsilon}} + \dot{\boldsymbol{\omega}} - \dot{\boldsymbol{\omega}}^c,$$

where $\dot{\boldsymbol{\varepsilon}} = \mathbf{d} = \frac{1}{2}[\nabla\dot{\mathbf{u}} + (\nabla\dot{\mathbf{u}})^T]$ and $\dot{\boldsymbol{\omega}} = \frac{1}{2}[\nabla\dot{\mathbf{u}} - (\nabla\dot{\mathbf{u}})^T]$, the symmetric part and the skew symmetric part of the velocity gradient, are the strain rate and the spin (termed the macro spin in this paper) for the non-polar continuum. The strain rate for the micro-polar continuum $\dot{\boldsymbol{\varepsilon}}^c$ is generally non-symmetric. In the case where the micro-spin coincides with the macro-spin, it becomes symmetric and coincides with the strain rate for the non-polar continuum.

The equilibrium equations for a micro-polar continuum must take into account the presence of the couple stress $\boldsymbol{\mu}$. In cases where body forces and body couples are absent, the *local* form of the equilibrium equations read:

$$\nabla\cdot\boldsymbol{\sigma}^T = \mathbf{0}, \qquad \partial\sigma_{ij}/\partial x_j = 0, \tag{2a}$$

$$\nabla\cdot\boldsymbol{\mu}^T - \boldsymbol{\epsilon}:\boldsymbol{\sigma} = \mathbf{0}, \quad \partial\mu_{ij}/\partial x_j - \epsilon_{ijk}\sigma_{jk} = 0. \tag{2b}$$

From the second equation, we see that the stress tensor is generally non-symmetric for a micro-polar continuum. When the couple stress vanishes, the stress tensor becomes symmetric.

2.2. Hypoplastic model for non-polar and micro-polar continuum

In hypoplasticity, the granular material is described as a continuum in terms of material state and material properties. The latter are represented by constitutive constants which do not change during loading. In a non-polar (conventional) continuum approach, the material state is characterized by the current stress $\boldsymbol{\sigma}$ and the void ratio e, while in a micro-polar continuum description the couple stress $\boldsymbol{\mu}$ is also included.

In the nonpolar continuum description, a practical hypoplastic constitutive equation for the cohesionless granular materials has the following general form [16]:

$$\mathring{\boldsymbol{\sigma}} = f_s(\mathcal{L}(\hat{\boldsymbol{\sigma}}):\dot{\boldsymbol{\varepsilon}} + f_d\mathbf{N}(\hat{\boldsymbol{\sigma}})\|\dot{\boldsymbol{\varepsilon}}\|). \tag{3}$$

The objective time rate of the Cauchy stress, $\mathring{\boldsymbol{\sigma}}$, is a nonlinear function of the strain rate $\dot{\boldsymbol{\varepsilon}}$. Herein the fourth-order tensor $\mathcal{L}$ and the second-order tensor $\mathbf{N}$ are functions of the normalized stress $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}/\mathrm{tr}\ \boldsymbol{\sigma}$, and the scalar factors f_s and f_d (known as the stiffness factor and the density factor, respectively) are functions of the mean pressure $p = -\mathrm{tr}\ \boldsymbol{\sigma}/3$ and the void

ratio e. The elaborated nonlinear tensor-valued function allows the description of different stiffness for loading and unloading without employing the elastoplastic concept of decomposition of strain rate into an elastic part and a plastic part. As $\mathring{\boldsymbol{\sigma}}$ is a positively homogeneous of order one in $\dot{\boldsymbol{\varepsilon}}$, the rate-independent behavior is described by (3).

In a micropolar continuum extension [18] of constitutive equation (3), the objective time rate of the stress, $\mathring{\boldsymbol{\sigma}}$, and the objective time rate of couple stress, $\mathring{\boldsymbol{\mu}}$, are expressed as

$$\mathring{\boldsymbol{\sigma}} = f_{\mathrm{s}}[\mathcal{L}^{\sigma\sigma} : \dot{\boldsymbol{\varepsilon}}^{c} + \mathcal{L}^{\sigma\mu} : \dot{\boldsymbol{\kappa}}^{*} + f_{\mathrm{d}}\mathbf{N}^{\sigma} R], \tag{4a}$$

$$\mathring{\boldsymbol{\mu}} = d_{50} f_{\mathrm{s}}[\mathcal{L}^{\mu\mu} : \dot{\boldsymbol{\kappa}}^{*} + \mathcal{L}^{\mu\sigma} : \dot{\boldsymbol{\varepsilon}}^{c} + f_{\mathrm{d}}\mathbf{N}^{\mu} R], \tag{4b}$$

where d_{50} denotes the mean grain diameter, which is a natural length scale for granular materials, $\dot{\boldsymbol{\kappa}}^{*} = d_{50}\dot{\boldsymbol{\kappa}}$ is the scaled micro-curvature rate, and R is a combination of the norms of the strain rate and the micro-curvature rate in the form of

$$R = \sqrt{\dot{\boldsymbol{\varepsilon}} : \dot{\boldsymbol{\varepsilon}} + \delta^{2}\dot{\boldsymbol{\kappa}} : \dot{\boldsymbol{\kappa}}}, \tag{5}$$

where

$$\delta = (a_{\mu}/\hat{a}) d_{50} \tag{6}$$

is a characteristic length which scales the thickness of shear bands. The fourth-order tensors and the second-order tensors depend on the normalized stress tensor, $\hat{\boldsymbol{\sigma}}$, and the normalized couple-stress tensor, $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}/(d_{50}\mathrm{tr}\ \boldsymbol{\sigma})$, with the following representations:

$$\begin{aligned} \mathcal{L}^{\sigma\sigma} &= \hat{a}^{2}\mathcal{I} + \hat{\boldsymbol{\sigma}} \otimes \hat{\boldsymbol{\sigma}}, \quad \mathcal{L}^{\sigma\mu} = \hat{\boldsymbol{\sigma}} \otimes \hat{\boldsymbol{\mu}}, \quad \mathbf{N}^{\sigma} = \hat{a}(\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\sigma}}^{d}), \\ \mathcal{L}^{\mu\mu} &= a_{\mu}^{2}\mathcal{I} + \hat{\boldsymbol{\mu}} \otimes \hat{\boldsymbol{\mu}}, \quad \mathcal{L}^{\mu\sigma} = \hat{\boldsymbol{\mu}} \otimes \hat{\boldsymbol{\sigma}}, \quad \mathbf{N}^{\mu} = 2\hat{a}\hat{\boldsymbol{\mu}}. \end{aligned} \tag{7}$$

Herein $\hat{\boldsymbol{\sigma}}^{d} = \hat{\boldsymbol{\sigma}} - \mathbf{I}/3$ denotes the deviator of the normalized stress, and $\hat{a}$ and a_{μ} are two parameters related to the limit value of stress and couple stress at stationary states.

A stationary state is reached when there is no change in the state variables under continuing deformation. The present micro-polar hypoplastic model is characterized by the following limit condition on the stress, the couple stress, and the void ratio at stationary states [18]:

$$\frac{\|\hat{\boldsymbol{\sigma}}^{d}\|^{2}}{\hat{a}^{2}} + \frac{\|\hat{\boldsymbol{\mu}}\|^{2}}{a_{\mu}^{2}} = 1 \quad \text{and} \quad f_{\mathrm{d}} = 1. \tag{8}$$

This relation shows a coupling between the limit stress and the limit couple stress. It also provides a physical interpretation for the parameters $\hat{a}$ and a_{μ}.

The evolution of the void ratio is governed by the equation:

$$\dot{e} = (1 + e)\mathrm{tr}\ \dot{\boldsymbol{\varepsilon}}, \tag{9}$$

which is a result of mass balance by neglecting the volume change in solid grains. The void ratio of a granular material is bounded by the maximum void ratio e_i for the loosest states and the minimum void ratio e_d for the densest state, and it tends to a critical void ratio e_c at stationary states. e_i, e_d and e_c are pressure-dependent quantities. In the present model, the variation of e_i, e_d and e_c with the mean pressure p is described by the following relation (refer to Figure 1):

$$\frac{e_c}{e_{co}} = \frac{e_d}{e_{do}} = \frac{e_i}{e_{io}} = \exp\{(3p/h_s)^{n}\}. \tag{10}$$

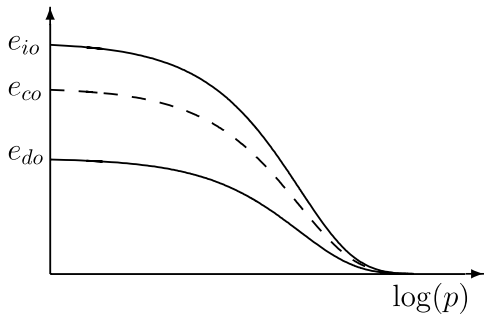


Figure 1. Pressure dependence of the maximum, the critic and the minimum void ratio [17].

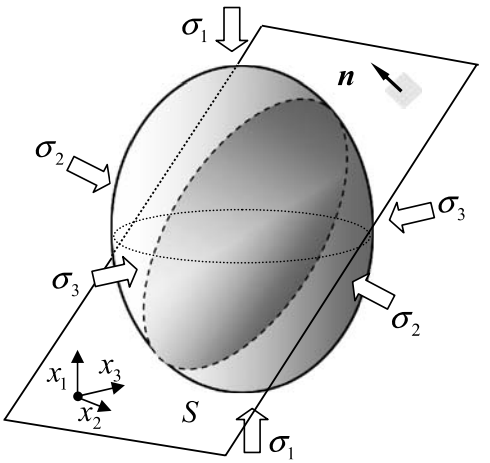


Figure 2. Sketch of weak discontinuity plane in the principal stress space.

Herein h_s and n are two material constants that determine the shape of the void ratio-pressure curves, and e_{io}, e_{do} and e_{co} are three material constants which represent the maximum, minimum and critical void ratio at the stress-free state. The density factor f_d is related to the pressure-dependent relative density according to

$$f_\mathrm{d} = \left(\frac{e - e_d}{e_c - e_d}\right)^{\alpha}, \tag{11}$$

where α is a material constant which scales the peak stress state under loading. Note that dilative deformation corresponds to an increase in the value for f_d. The densest state is characterized by $f_\mathrm{d} = 0$, while $f_\mathrm{d} = 1$ corresponds to $e = e_c$. Therefore, the critical void ratio is reached at a stationary state, which is in accordance with experimental observations for granular materials.

The stiffness factor f_s has the following representation:

$$f_\mathrm{s} = \frac{h_s}{n(\hat{\boldsymbol{\sigma}} : \hat{\boldsymbol{\sigma}})} \left(\frac{e_i}{e}\right)^{\beta} \frac{1 + e_i}{e_i} \left(\frac{p}{h_s}\right)^{(1-n)} [3\hat{a}_i^2 + 1 + \sqrt{3}\hat{a}_i f_{di}]^{-1}. \tag{12}$$

Herein f_{di} and a_i represent the values of f_d and $\hat{a}$ at isotropic states. From (10) and (11), f_{di} has a constant value $f_{di} = \{(e_{io} - e_{do})/(e_{co} - e_{do})\}^{\alpha}$. As described later, a_i has a constant value, which is related to the critical friction angle ϕ_c. The term $(e_i/e)^{\beta}$ reflects the influence of the density on the stiffness factor with $\beta \approx 1$ being another material constant.

For the special case where the couple stress and micro-curvature rate vanish and the micro-spin coincides with the macro-spin, the non-polar continuum model proposed by Gudehus [16] and Bauer [17] is recovered with

$$\mathcal{L}^{\sigma\sigma} = \mathcal{L} = \hat{a}^2 \mathcal{I} + \hat{\boldsymbol{\sigma}} \otimes \hat{\boldsymbol{\sigma}}, \quad \mathbf{N}^{\sigma} = \mathbf{N} = \hat{a}(\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\sigma}}^d). \tag{13}$$

Correspondingly, the limit stress at a stationary state satisfies the following condition:

$$||\hat{\boldsymbol{\sigma}}^d|| = \hat{a}, \tag{14}$$

which represents a conical surface in the principal stress space with its apex at the origin. Bauer [28] showed that it is possible to embed different limit stress conditions into this model

by adopting different expressions for the parameter $\hat{a}$. For instance, a constant value for $\hat{a}$ corresponds to a Drucker–Prage type limit condition. With the following formulation, the Matsuoka–Nakai limit condition is embedded:

$$\hat{a} = \frac{\sin\varphi_c}{3-\sin\varphi_c}\left[\sqrt{\frac{8/3-3||\hat{\boldsymbol{\sigma}}^{sd}||^2+\sqrt{3/2}||\hat{\boldsymbol{\sigma}}^{sd}||^3\cos(3\theta)}{1+\sqrt{3/2}||\hat{\boldsymbol{\sigma}}^{sd}||\cos(3\theta)}}-||\hat{\boldsymbol{\sigma}}^{sd}||\right], \tag{15}$$

where $\hat{\boldsymbol{\sigma}}^{sd}$ denotes the deviator of the normalized symmetric stress, and θ denotes the Lode angle defined in the deviatoric stress plane for the symmetric stress.

The limit couple stress related parameter a_μ is considered to be constant. For a granular material without preferred grain orientation, this is a reasonable assumption. All the material constants in this model, except a_μ, can be determined easily from elementary laboratory tests as discussed in detail by Herle and Gudehus [29]. a_μ may be determined from a back analysis procedure after measuring the shear band thickness, which requires advanced techniques. For the following discussion we note that the stiffness factor f_s has a value of the order of $h_s^n p^{1-n}$. The granular hardness (as termed in Gudehus [16]) h_s, which has a dimension of stress, has a typical value of the order of $1{\cdot}0\times10^6$ kPa or higher for sands. The dimensionless exponent n falls in the range of $(0{\cdot}25, 0{\cdot}5)$ [29].

3. Bifurcation analysis for the non-polar hypoplastic continuum

Consider a homogeneous specimen under uniform deformation and assume that the velocity and stress fields are continuous up to a certain state where a discontinuity in the velocity gradient across a planar surface becomes possible. This weak discontinuity plane S is then characterized by the following kinematic conditions:

$$[[\mathbf{l}]] = [[\nabla\dot{\mathbf{u}}]] = (\nabla\dot{\mathbf{u}})^1 - (\nabla\dot{\mathbf{u}})^0 = \mathbf{g}\otimes\mathbf{n} \quad \text{and} \quad [[\dot{\mathbf{u}}]] = \dot{\mathbf{u}}^1 - \dot{\mathbf{u}}^0 = 0. \tag{16}$$

Here the superscript 0 and 1 are used to denote the values for a quantity on either side of the discontinuity plane. Equation (16) states that across the discontinuous plane S the velocity gradient is experiencing a jump, which is characterized by a vector $\mathbf{g}$ and the unit normal vector $\mathbf{n}$ to the discontinuity plane, while the velocity field itself is continuous. Corresponding to the discontinuous velocity gradient, jumps in the strain rate and the spin tensor will be encountered on crossing S:

$$[[\dot{\boldsymbol{\varepsilon}}]] = \dot{\boldsymbol{\varepsilon}}^1 - \dot{\boldsymbol{\varepsilon}}^0 = \frac{1}{2}(\mathbf{g}\otimes\mathbf{n}+\mathbf{n}\otimes\mathbf{n}), \tag{17a}$$

$$[[\dot{\boldsymbol{\omega}}]] = \dot{\boldsymbol{\omega}}^1 - \dot{\boldsymbol{\omega}}^0 = \frac{1}{2}(\mathbf{g}\otimes\mathbf{n}-\mathbf{n}\otimes\mathbf{g}). \tag{17b}$$

In response to this, the stress rate becomes discontinuous. A bifurcation condition can be formulated by considering rate-form equilibrium long the possible discontinuity plane. Concerning the fact that the direction of the discontinuity plane varies with time due to the motion, we start to consider the equilibrium in a reference configuration (A reference configuration is a fixed configuration in the space which uniquely designates a one-to-one mapping with material points in motion.). On the image discontinuity plane in the reference configuration, the nominal traction rate must be unique, *i.e.*, $[[\dot{\mathbf{P}}]]\cdot\mathbf{N}=\mathbf{0}$, where $\mathbf{P}$ is the first Piola–Kirchhoff stress and $\mathbf{N}$ is the unit normal of the image discontinuity plane, which is independent of time. Using the well-known relation $\mathbf{P} = J\boldsymbol{\sigma}\cdot\mathbf{F}^{-T}$ and the Nanson's formula (see, for instance, [30,

p. 75]) the following equation is obtained:

$$[[\dot{\boldsymbol{\sigma}} + (\operatorname{div}\dot{\mathbf{u}})\boldsymbol{\sigma} - \boldsymbol{\sigma}\cdot\mathbf{l}^T]]\cdot\mathbf{n} = [[\dot{\boldsymbol{\sigma}}]]\cdot\mathbf{n} + ([[\operatorname{div}\dot{\mathbf{u}}]]\boldsymbol{\sigma} - \boldsymbol{\sigma}\cdot[[\mathbf{l}^T]])\cdot\mathbf{n} = 0. \tag{18}$$

Note that the material response is described in terms of the objective time rates of stress and couple stress. There is an infinite number of possibilities which define an objective time rate for the stress tensor. The most widely used objective stress rates include the Zaremba–Jaumann rate, the Oldroyd rate, the Green–Naghdi rate and the Truesdell rate ([30, Section5.3]). Choosing a proper objective stress rate is, in many cases, a problem of a proper formulation of the constitutive equations. It should also be noted that the discrepancies due to the choice of different objective stress rates become significant only when a large shear deformation has taken place (*e.g.* [28]). At the beginning of shear localization, such discrepancies are negligible. In this study, we take the Zaremba–Jaumann rate as the objective time rate for both the stress tensor and the couple stress tensor:

$$\mathring{\boldsymbol{\sigma}} = \dot{\boldsymbol{\sigma}} - (\dot{\boldsymbol{\omega}}\cdot\boldsymbol{\sigma} - \boldsymbol{\sigma}\cdot\dot{\boldsymbol{\omega}}), \tag{19a}$$

$$\mathring{\boldsymbol{\mu}} = \dot{\boldsymbol{\mu}} - (\dot{\boldsymbol{\omega}}\cdot\boldsymbol{\mu} - \boldsymbol{\mu}\cdot\dot{\boldsymbol{\omega}}). \tag{19b}$$

Substitution of (19a) in (18) yields the following bifurcation condition

$$[[\mathring{\boldsymbol{\sigma}}]]\cdot\mathbf{n} + ([[\operatorname{div}\dot{\mathbf{u}}]]\boldsymbol{\sigma}\cdot\mathbf{n} + [[\dot{\boldsymbol{\omega}}]]\cdot\boldsymbol{\sigma}\cdot\mathbf{n} - \boldsymbol{\sigma}\cdot[[\mathbf{d}]])\cdot\mathbf{n} = 0. \tag{20}$$

Invoking the constitutive relations leads to the following equations for the components of the vector $\mathbf{g}$:

$$\mathbf{K}\cdot\mathbf{g} - \lambda f_{\mathrm{d}}\mathbf{r} = 0. \tag{21}$$

The following notations are introduced in this equation: the scalar factor $\lambda = ||\dot{\boldsymbol{\varepsilon}}^1|| - ||\dot{\boldsymbol{\varepsilon}}^0||$, the vector $\mathbf{r} = -\hat{a}(\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\sigma}}^d)\cdot\mathbf{n}$ and the tensor $\mathbf{K} = \mathbf{K}' + \mathbf{K}''$, where

$$\begin{aligned}\mathbf{K}' &= \tfrac{\hat{a}^2}{2}(\mathbf{I} + \mathbf{n}\otimes\mathbf{n}) + (\hat{\boldsymbol{\sigma}}\cdot\mathbf{n})\otimes(\hat{\boldsymbol{\sigma}}\cdot\mathbf{n}),\\ \mathbf{K}'' &= -\tfrac{3p}{f_{\mathrm{s}}}[\hat{\boldsymbol{\sigma}}\cdot\mathbf{n}\otimes\mathbf{n} - \mathbf{n}\otimes\mathbf{n}\cdot\hat{\boldsymbol{\sigma}}^T].\end{aligned} \tag{22}$$

In the representation for $\mathbf{K}''$, the factor $-3p$ comes out from the stress normalization. Since the stiffness factor $f_{\mathrm{s}} \sim h_s^n p^{1-n}$, it follows that the factor p/f_{s} has a value of the order of $(p/h_s)^n$. Therefore, for most practical stress states with $p << h_s$, $\mathbf{K}'$ is the dominant term in $\mathbf{K}$ and $\mathbf{K}''$ can be neglected. Because $\mathbf{K}'$ is positive-definite and therefore invertible, $\mathbf{K}$ is also invertible. Then we can write $\mathbf{g}$ as a linear function of the scalar factor λ:

$$\mathbf{g} = \lambda f_{\mathrm{d}}\mathbf{K}^{-1}\cdot\mathbf{r}. \tag{23}$$

For $\lambda = 0$, we have only a null solution for $\mathbf{g}$, which indicates no bifurcation at all. Thus the bifurcation problem is equivalent to seeking a non-trivial solution for the scalar factor λ. By introducing the notations

$$\hat{\mathbf{g}} = \mathbf{K}^{-1}\cdot\mathbf{r} \quad \text{and} \quad \Delta\hat{\boldsymbol{\varepsilon}} = \frac{1}{2}(\hat{\mathbf{g}}\otimes\mathbf{n} + \mathbf{n}\otimes\hat{\mathbf{g}}), \tag{24}$$

we can write the strain rate as $\dot{\boldsymbol{\varepsilon}}^1 = \dot{\boldsymbol{\varepsilon}}^0 + \lambda f_{\mathrm{d}}\Delta\hat{\boldsymbol{\varepsilon}}$. It follows that the nonlinear equation for the scalar factor λ can be written as

$$\lambda - \left(\sqrt{(\dot{\boldsymbol{\varepsilon}}^0 + \lambda f_{\mathrm{d}}\Delta\boldsymbol{\varepsilon}) : (\dot{\boldsymbol{\varepsilon}}^0 + \lambda f_{\mathrm{d}}\Delta\boldsymbol{\varepsilon})} - \sqrt{\dot{\boldsymbol{\varepsilon}}^0 : \dot{\boldsymbol{\varepsilon}}^0}\right) = 0. \tag{25}$$

Algebraic manipulation leads to the following equation:

$$(f_{\mathrm{d}}^2 \Delta\hat{\boldsymbol{\varepsilon}} : \Delta\hat{\boldsymbol{\varepsilon}} - 1))\lambda^2 + 2(f_{\mathrm{d}}\dot{\boldsymbol{\varepsilon}}^0 : \Delta\hat{\boldsymbol{\varepsilon}} - \|\dot{\boldsymbol{\varepsilon}}^0\|)\lambda = 0.$$

The non-trivial solution for λ then reads

$$\lambda = 2\frac{\|\dot{\boldsymbol{\varepsilon}}^0\| - f_{\mathrm{d}}\dot{\boldsymbol{\varepsilon}}^0 : \Delta\hat{\boldsymbol{\varepsilon}}}{f_{\mathrm{d}}^2 \Delta\hat{\boldsymbol{\varepsilon}} : \Delta\hat{\boldsymbol{\varepsilon}} - 1} = 2\frac{1 - f_{\mathrm{d}}\|\Delta\hat{\boldsymbol{\varepsilon}}\|\cos\gamma}{f_{\mathrm{d}}^2 \Delta\hat{\boldsymbol{\varepsilon}} : \Delta\hat{\boldsymbol{\varepsilon}} - 1}\|\dot{\boldsymbol{\varepsilon}}^0\| \tag{26}$$

with $\cos\gamma := (\dot{\boldsymbol{\varepsilon}}^0 : \Delta\hat{\boldsymbol{\varepsilon}})/(\|\dot{\boldsymbol{\varepsilon}}^0\|\|\Delta\hat{\boldsymbol{\varepsilon}}\|)$. Note that λ is restricted by the condition

$$\lambda \geq -\|\dot{\boldsymbol{\varepsilon}}^0\|, \tag{27}$$

since $\|\dot{\boldsymbol{\varepsilon}}\| = \lambda + \|\dot{\boldsymbol{\varepsilon}}^0\| \geq 0$. It can be shown that the requirement (27) can be fulfilled only when

$$f_{\mathrm{d}}^2 \Delta\hat{\boldsymbol{\varepsilon}} : \Delta\hat{\boldsymbol{\varepsilon}} = \frac{f_{\mathrm{d}}^2}{2}[\hat{\mathbf{g}} \cdot \hat{\mathbf{g}} + (\hat{\mathbf{g}} \cdot \mathbf{n})^2] \geq 1. \tag{28}$$

Proof. Assume that (28) is not fulfilled, then $f_{\mathrm{d}}\|\Delta\hat{\boldsymbol{\varepsilon}}\| < 1$. The maximum value for λ in (26) corresponds to $\cos\gamma = 1$, which reads

$$\lambda_{\max} = -\frac{2}{1 + f_{\mathrm{d}}\|\Delta\hat{\boldsymbol{\varepsilon}}\|}\|\dot{\boldsymbol{\varepsilon}}^0\| < -\|\dot{\boldsymbol{\varepsilon}}^0\|, \tag{29}$$

which violates the requirement (27).

Remark. Inequality (28) can also be obtained in an alternative way. Note that a triangular inequality $\|\dot{\boldsymbol{\varepsilon}}^1 - \dot{\boldsymbol{\varepsilon}}^0\|^2 \geq (\|\dot{\boldsymbol{\varepsilon}}^1\| - \|\dot{\boldsymbol{\varepsilon}}^0\|)^2$ is generally valid. Substitution of $(\|\dot{\boldsymbol{\varepsilon}}^1\| - \|\dot{\boldsymbol{\varepsilon}}^0\|)^2 = \lambda^2$ and $\|\dot{\boldsymbol{\varepsilon}}^1 - \dot{\boldsymbol{\varepsilon}}^0\|^2 = \lambda^2 f_{\mathrm{d}}^2(\hat{\mathbf{g}} \cdot \hat{\mathbf{g}} + (\hat{\mathbf{g}} \cdot \mathbf{n})^2)/2$ leads to inequality (28). This approach is simpler. However, the approach from (25) to (28) provides more detail about the factor λ and can be repeated for derivation of the bifurcation criterion for the micro-polar continuum model.

Remark. The equality in (28) corresponds to either $||\dot{\boldsymbol{\varepsilon}}^1||/||\dot{\boldsymbol{\varepsilon}}^0|| = \infty$ or $\cos\gamma = 1$. The later is true only when $\dot{\boldsymbol{\varepsilon}}^1$ is co-axial with $\dot{\boldsymbol{\varepsilon}}^0$, which corresponds to pure dilation or contraction in the localized zone. For a uniformly deforming process leading to a continuous onset of shear bifurcation, the equality represents a limit state at which the bifurcation will not occur.

To assess whether the bifurcation condition is met in a loading program, the equality can be used as a determinant. That is, incipience of shear localization is indicated by the following condition:

$$f_{\mathrm{d}}^2 \Phi_0(\hat{\boldsymbol{\sigma}}, \mathbf{n}) - 1 = 0, \tag{30}$$

with

$$\Phi_0(\hat{\boldsymbol{\sigma}}, \mathbf{n}) = \frac{1}{2}[\hat{\mathbf{g}} \cdot \hat{\mathbf{g}} + (\hat{\mathbf{g}} \cdot \mathbf{n})^2] = \frac{1}{2}[(\mathbf{K}^{-1} \cdot \mathbf{r}) \cdot (\mathbf{K}^{-1} \cdot \mathbf{r}) + (\mathbf{n} \cdot \mathbf{K}^{-1} \cdot \mathbf{r})^2]. \tag{31}$$

Note that for $p << h_s$, $\mathbf{K}''$ can be neglected in the representation of $\mathbf{K}$. In the case where $p \sim h_s$ or $p > h_s$, we have $e_i/e \sim 1$, which implies that f_{s} and $\mathbf{K}''$ are almost independent of the void ratio. Therefore, in the criterion (30), Φ_0 can be considered independent of the density. The influence of the density on bifurcation is reflected only by the factor f_{d}, which is separated from Φ_0.

4. Bifurcation analysis for the micro-polar hypoplastic continuum

Similar conditions for weak-discontinuity bifurcation can be formulated for a micro-polar hypoplastic continuum. We start again with a homogeneously deforming specimen in which the velocity, rotation rate, stress and couple stress fields are continuous up to a state where a banded planar weak discontinuity may develop. Across the weak discontinuity plane S, the velocity and the rotation rate fields are initially continuous:

$$[[\dot{\mathbf{u}}]] = \dot{\mathbf{u}}^1 - \dot{\mathbf{u}}^0 = 0 \quad \text{and} \quad [[\dot{\mathbf{w}}^c]] = \dot{\mathbf{w}}^{c1} - \dot{\mathbf{w}}^{c0} = 0\,, \tag{32}$$

while the jump in the velocity gradient and the rotation-rate gradient can be written as

$$[[\nabla\dot{\mathbf{u}}]] = (\nabla\dot{\mathbf{u}})^1 - (\nabla\dot{\mathbf{u}})^0 = \mathbf{g}^u \otimes \mathbf{n}, \tag{33a}$$

$$[[\dot{\boldsymbol{\kappa}}]] = (\nabla\dot{\mathbf{w}}^c)^1 - (\nabla\dot{\mathbf{w}}^c)^0 = \mathbf{g}^w \otimes \mathbf{n}. \tag{33b}$$

Weak discontinuity bifurcation is characterized by vectors $\mathbf{g}^u, \mathbf{g}^w$ and the unit normal vector $\mathbf{n}$ for the discontinuity plane S. With respect to the discontinuity in the velocity gradient, the jumps in the strain rate and in the macro-spin read

$$[[\dot{\boldsymbol{\varepsilon}}^c]] = \dot{\boldsymbol{\varepsilon}}^{c1} - \dot{\boldsymbol{\varepsilon}}^{c0} = [[\nabla\dot{\mathbf{u}} + \boldsymbol{\epsilon}\cdot\mathbf{w}^c]] = \mathbf{g}^u \otimes \mathbf{n}, \tag{34a}$$

$$[[\dot{\boldsymbol{\omega}}]] = \dot{\boldsymbol{\omega}}^1 - \dot{\boldsymbol{\omega}}^0 = \frac{1}{2}(\mathbf{g}^u \otimes \mathbf{n} - \mathbf{n} \otimes \mathbf{g}^u). \tag{34b}$$

Similar to the situation for the nonpolar continuum, consideration of the equilibrium in rate form along the discontinuity plane leads to the following conditions to be fulfilled by the total stress rate and the total couple stress rate:

$$[[\dot{\boldsymbol{\sigma}}]]\cdot\mathbf{n} + ([[\operatorname{div}\dot{\mathbf{u}}]]\boldsymbol{\sigma} - \boldsymbol{\sigma}\cdot[[\mathbf{I}^T]])\cdot\mathbf{n} = 0, \tag{35a}$$

$$[[\dot{\boldsymbol{\mu}}]]\cdot\mathbf{n} + ([[\operatorname{div}\dot{\mathbf{u}}]]\boldsymbol{\mu} - \boldsymbol{\mu}\cdot[[\mathbf{I}^T]])\cdot\mathbf{n} = 0. \tag{35b}$$

Using the expressions for stress rate (19a) and couple-stress rate (19b) and the constitutive Equations (4a) and (4b), the following nonlinear equations are obtained for the vectors $\mathbf{g}^u$ and $\mathbf{g}^w$:

$$\mathbf{K}_{uu}\cdot\mathbf{g}^u + \mathbf{K}_{uw}\cdot\mathbf{g}^w - \lambda^c f_{\mathrm{d}}\mathbf{r}^u = 0, \tag{36a}$$

$$\mathbf{K}_{wu}\cdot\mathbf{g}^u + \mathbf{K}_{ww}\cdot\mathbf{g}^w - \lambda^c f_{\mathrm{d}}\mathbf{r}^w = 0. \tag{36b}$$

Herein the following notations are used: the scalar factor $\lambda^c = [[R]]$, the vectors $\mathbf{r}^u = -\hat{a}(\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\sigma}}^d)\cdot\mathbf{n}$ and $\mathbf{r}^w = -2\hat{a}\hat{\boldsymbol{\mu}}\cdot\mathbf{n}$, and the tensors $\mathbf{K}_{uu} = \mathbf{K}'_{uu} + \mathbf{K}''_{uu}$, $\mathbf{K}_{uw} = \mathbf{K}'_{uw}$, $\mathbf{K}_{wu} = \mathbf{K}'_{wu} + \mathbf{K}''_{wu}$ and $\mathbf{K}_{ww} = \mathbf{K}'_{ww}$ with the following representations:

$$\begin{aligned}
\mathbf{K}'_{uu} &= \hat{a}^2\mathbf{I} + (\hat{\boldsymbol{\sigma}}\cdot\mathbf{n})\otimes(\hat{\boldsymbol{\sigma}}\cdot\mathbf{n}), \quad \mathbf{K}'_{uw} = d_{50}(\hat{\boldsymbol{\sigma}}\cdot\mathbf{n})\otimes(\hat{\boldsymbol{\mu}}\cdot\mathbf{n}),\\
\mathbf{K}'_{wu} &= (\hat{\boldsymbol{\mu}}\cdot\mathbf{n})\otimes(\hat{\boldsymbol{\sigma}}\cdot\mathbf{n}), \quad \mathbf{K}'_{ww} = d_{50}[a_\mu^2\mathbf{I} + (\hat{\boldsymbol{\mu}}\cdot\mathbf{n})\otimes(\hat{\boldsymbol{\mu}}\cdot\mathbf{n})],\\
\mathbf{K}''_{uu} &= -\tfrac{3p}{f_s}[\hat{\boldsymbol{\sigma}}\cdot\mathbf{n}\otimes\mathbf{n} - \mathbf{n}\otimes\mathbf{n}\cdot\hat{\boldsymbol{\sigma}}^T],\\
\mathbf{K}''_{wu} &= -\tfrac{3p}{f_s}[\hat{\boldsymbol{\mu}}\cdot\mathbf{n}\otimes\mathbf{n} - \mathbf{n}\otimes\mathbf{n}\cdot\hat{\boldsymbol{\mu}}^T].
\end{aligned} \tag{37}$$

For a concise representation, we define $\boldsymbol{\Pi}$ according to

$$\boldsymbol{\Pi} := \begin{bmatrix} \mathbf{K}_{uu} & \mathbf{K}_{uw} \\ \mathbf{K}_{wu} & \mathbf{K}_{ww} \end{bmatrix}. \tag{38}$$

By neglecting $\mathbf{K}''_{uu}$ and $\mathbf{K}''_{wu}$ in $\mathbf{K}_{uu}$ and $\mathbf{K}_{wu}$ for $p << h_s$, we can show that $\mathbf{\Pi}$ is invertible (see appendix A). We can then write

$$\begin{Bmatrix} \mathbf{g}^u \\ \mathbf{g}^w \end{Bmatrix} = \lambda^c f_\mathrm{d} \mathbf{\Pi}^{-1} \cdot \begin{Bmatrix} \mathbf{r}^u \\ \mathbf{r}^w \end{Bmatrix}. \tag{39}$$

Again the bifurcation vectors $\mathbf{g}^u$ and $\mathbf{g}^w$ are linearly related to the scalar factor λ^c. Thus the problem is equivalent to seeking a non-trivial solution for the scalar factor λ^c. By introducing the following notations

$$\begin{Bmatrix} \hat{\mathbf{g}}^u \\ \hat{\mathbf{g}}^w \end{Bmatrix} = \mathbf{\Pi}^{-1} \cdot \begin{Bmatrix} \mathbf{r}^u \\ \mathbf{r}^w \end{Bmatrix} \tag{40}$$

and

$$\Delta\hat{\boldsymbol{\varepsilon}}^c = \hat{\mathbf{g}}^u \otimes \mathbf{n}, \quad \Delta\hat{\boldsymbol{\kappa}} = \hat{\mathbf{g}}^w \otimes \mathbf{n}, \tag{41}$$

the equation for the factor λ^c can be written as

$$\begin{aligned} &\lambda^c + \sqrt{\dot{\boldsymbol{\varepsilon}}^{c0} : \dot{\boldsymbol{\varepsilon}}^{c0} + \delta^2 \dot{\boldsymbol{\kappa}}^0 : \delta^2 \dot{\boldsymbol{\kappa}}^0} \\ &\quad = \sqrt{(\dot{\boldsymbol{\varepsilon}}^{c0} + \lambda^c f_\mathrm{d} \Delta\hat{\boldsymbol{\varepsilon}}^c) : (\dot{\boldsymbol{\varepsilon}}^{c0} + \lambda^c f_\mathrm{d} \Delta\hat{\boldsymbol{\varepsilon}}^c) + \delta^2 (\dot{\boldsymbol{\kappa}}^0 + \lambda^c f_\mathrm{d} \Delta\hat{\boldsymbol{\kappa}}) : (\dot{\boldsymbol{\kappa}}^0 + \lambda^c f_\mathrm{d} \Delta\hat{\boldsymbol{\kappa}})}. \end{aligned} \tag{42}$$

The non-trivial solution for λ^c reads

$$\lambda^c = 2 \frac{1 - f_\mathrm{d} (\dot{\boldsymbol{\varepsilon}}^{c0} : \Delta\hat{\boldsymbol{\varepsilon}}^c + \delta^2 \dot{\boldsymbol{\kappa}}^0 : \Delta\boldsymbol{\kappa}) / \sqrt{\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2}}{f_\mathrm{d}^2 (\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2 \|\Delta\hat{\boldsymbol{\kappa}}\|^2) - 1} \times \sqrt{\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2}. \tag{43}$$

Here λ^c is restricted by a requirement similar to (27), which reads

$$\lambda^c \geq -\sqrt{\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2}. \tag{44}$$

In a similar way, we can show that this requirement can be fulfilled only when

$$f_\mathrm{d}^2 (\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2 \|\Delta\boldsymbol{\kappa}\|^2) \geq 1. \tag{45}$$

Proof. Since $\dot{\boldsymbol{\varepsilon}}^{c0} : \Delta\hat{\boldsymbol{\varepsilon}}^c \leq \|\dot{\boldsymbol{\varepsilon}}^{c0}\| \|\Delta\hat{\boldsymbol{\varepsilon}}^c\|$ and $\dot{\boldsymbol{\kappa}}^0 : \Delta\hat{\boldsymbol{\kappa}} \leq \|\dot{\boldsymbol{\kappa}}^0\| \|\Delta\hat{\boldsymbol{\kappa}}\|$, we have

$$(\dot{\boldsymbol{\varepsilon}}^{c0} : \Delta\hat{\boldsymbol{\varepsilon}}^c + \delta^2 \dot{\boldsymbol{\kappa}}^0 : \Delta\boldsymbol{\kappa})^2 \leq (\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2)(\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2 \|\Delta\boldsymbol{\kappa}\|^2).$$

If follows that the following inequality is valid.

$$f_\mathrm{d} (\dot{\boldsymbol{\varepsilon}}^{c0} : \Delta\hat{\boldsymbol{\varepsilon}}^c + \delta^2 \dot{\boldsymbol{\kappa}}^0 : \Delta\boldsymbol{\kappa}) / \sqrt{\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2} \leq f_\mathrm{d} \sqrt{\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2 \|\Delta\boldsymbol{\kappa}\|^2}.$$

Assume that (45) is not fulfilled, then $f_\mathrm{d} \sqrt{\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2 \|\Delta\boldsymbol{\kappa}\|^2} < 1$. The maximum value for λ^c in (43) can be obtained as

$$\lambda^c_{\max} = -\frac{2}{1 + f_\mathrm{d} \sqrt{\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2 \|\Delta\boldsymbol{\kappa}\|^2}} \sqrt{\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2} < -\sqrt{\|\dot{\boldsymbol{\varepsilon}}^{c0}\|^2 + \delta^2 \|\dot{\boldsymbol{\kappa}}^0\|^2},$$

which violates the requirement (44). Therefore, the occurrence of shear bifurcation in the micro-polar hypoplastic continuum can be judged by the following criterion:

$$f_\mathrm{d}^2 \Phi_1 (\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\mu}}, \mathbf{n}) - 1 = 0, \tag{46}$$

where

$$\Phi_1(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\mu}}, \mathbf{n}) = (\|\Delta\hat{\boldsymbol{\varepsilon}}^c\|^2 + \delta^2\|\Delta\boldsymbol{\kappa}\|^2) = \hat{\mathbf{g}}^u \cdot \hat{\mathbf{g}}^u + \delta^2 \hat{\mathbf{g}}^w \cdot \hat{\mathbf{g}}^w. \tag{47}$$

It has been widely recognized that polar effects are usually associated with pronounced strain gradients. A homogeneous specimen under uniform deformation is almost always free from polar effects, that is, the couple stress and the micro-curvature will not develop up to a shear bifurcation state. Therefore we have $\boldsymbol{\mu} = \mathbf{0}$ at the incipience of bifurcation, and the bifurcation condition (36b) is reduced to yield $\mathbf{g}^w = \mathbf{0}$. Thus the bifurcation criterion (46) for the micro-polar hypoplastic continuum is simplified to

$$f_d^2 \Phi_1^*(\hat{\boldsymbol{\sigma}}, \mathbf{n}) - 1 = 0 \tag{48}$$

and

$$\Phi_1^*(\hat{\boldsymbol{\sigma}}, \mathbf{n}) = \hat{\mathbf{g}}^u \cdot \hat{\mathbf{g}}^u = (\mathbf{K}_{uu}^{-1} \cdot \mathbf{r}^u) \cdot (\mathbf{K}_{uu}^{-1} \cdot \mathbf{r}^u). \tag{49}$$

5. Examination of bifurcation condition

Granular materials show strong pressure- and density-sensitive behavior. For an initially dense specimen under monotonic loading to a stationary state, the deviatoric stress will show a peak with the specimen experiencing dilation after an initial contraction. The denser the specimen is packed, the higher the peak and the stronger the dilation will be. The peak will be less pronounced at a higher pressure level. If a specimen is initially in a very loose state, no peak state will be displayed and the specimen will experience a consistent contraction without dilation. This pressure and density effect has been taken into account in the present hypoplastic model by including the void ratio as a state variable. For monotonic loading paths, accessible stress states are bounded by the peak stress states, which form a conical surface in the principal stress space with its apex at the origin and its axis aligned with the hydro-static pressure axis.

With regard to criteria (30) and (46), we note that shear bifurcation depends on the stress and density for a non-polar continuum and the stress, couple stress and density for a micro-polar continuum (the couple stress often equals zero in a homogeneously deforming specimen). Shear bifurcation will occur in a homogeneously deforming specimen as the criterion (30) or (46) is met by the varying stress and density. For the non-polar continuum model, it has been shown that shear bifurcation occurs before the peak stress state is reached in biaxial compression tests [23,24]. The bifurcation point and the inclination of the shear band are influenced by the initial density and the mean pressure. Wu and Sikora [25] and Wu [26], in their analysis for bifurcation and failure in a density-independent non-polar hypoplastic continuum, have shown that the peak stress state may be reached before the bifurcation condition is met for some loading paths. This means a homogeneously deforming specimen may experience either a homogeneous failure or a localized failure, depending on the loading path.

In order to investigate the possibility of bifurcation in a general way for the present continuum models, we will compare the bifurcation states with the peak stress states in the deviatoric stress plane, as done by Wu [26] for a density-independent hypoplastic model. For this purpose, a mathematical representation for the peak stress state is formulated in the following.

5.1. Peak stress state

A peak stress state is defined by a vanishing stress rate with $\dot{e} > 0$. Note that a peak state differs from a stationary state in that the void ratio vanishes simultaneously with the latter. We consider loading programs such that the directions of the principal stresses do not change while the deviatoric stress varies. Triaxial compression and extension, as well as biaxial compression and extension tests, are examples of these loading programs. In these cases, macro-spin does not develop and the peak stress state corresponds to $f_s(\mathcal{L}:\dot{\boldsymbol{\varepsilon}} - f_d\mathbf{N}||\dot{\boldsymbol{\varepsilon}}||) = \mathbf{0}$, which yields

$$\dot{\boldsymbol{\varepsilon}} = f_d\mathcal{L}^{-1}:\mathbf{N}, \tag{50}$$

where $\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}/\|\dot{\boldsymbol{\varepsilon}}\|$ is the normalized strain rate. Since $||\dot{\boldsymbol{\varepsilon}}|| = 1$, the following condition for the peak stresses is obtained:

$$\Phi_p(\hat{\boldsymbol{\sigma}}) = (\mathcal{L}^{-1}:\mathbf{N}):(\mathcal{L}^{-1}:\mathbf{N}) = 1/f_d^2. \tag{51}$$

Herein Φ_b is a function of $\hat{\boldsymbol{\sigma}}$ (or $\hat{\boldsymbol{\sigma}}^d$) only. Note that this peak stress state representation is also relevant to the micro-polar hypoplastic model if couple stress and micro-curvature rate are zero. Given the representations for $\mathcal{L}$ and $\mathbf{N}$ in (13), $\mathcal{L}^{-1}$ and $\Phi_b(\hat{\boldsymbol{\sigma}})$ become

$$\mathcal{L}^{-1} = \frac{1}{\hat{a}^2}\mathcal{I} - \frac{\hat{\boldsymbol{\sigma}}\otimes\hat{\boldsymbol{\sigma}}}{\hat{a}^2(\hat{a}^2 + ||\hat{\boldsymbol{\sigma}}||^2)}, \tag{52}$$

$$\Phi_p(\hat{\boldsymbol{\sigma}}) = \frac{1}{\hat{a}^2}[\eta^2||\hat{\boldsymbol{\sigma}}||^2 + (2\eta+1)||\hat{\boldsymbol{\sigma}}^d||^2] \tag{53}$$

with $\eta = (\hat{a}^2 - ||\hat{\boldsymbol{\sigma}}^d||^2)/(\hat{a}^2 + ||\hat{\boldsymbol{\sigma}}||^2)$.

It can be seen in Equation (51) that the peak state is influenced by the density through the factor f_d. In the vicinity of the peak state, $f_d < 1$ holds. For a loading program starting from an isotropic (hydrostatic) state, numerical tests show that f_d normally falls within a range of $(0{\cdot}85, 1{\cdot}0)$ around the peak state. A geometric representation of the peak state can be presented in a deviatoric stress plane by searching the radial and circumferential directions for points at which (51) is satisfied (refer to Figure 3a). The peak stress states in the deviatoric stress plane for some assumed values for f_d are shown in Figure 3b. The closed loop represents an intersection of the conical peak surface with the deviatoric stress plane. It can be seen that a decrease in the value of f_d corresponds to an expansion of the peak state loop. A value $f_d = 1$ corresponds to stationary states with $\dot{e} = 0$, at which (51) is reduced to $\|\hat{\boldsymbol{\sigma}}^d\| = \hat{a}$. Starting from the same initial relative density, f_d can be closer to 1 for the peak state at a higher mean pressure level. Therefore, the generatrix of the peak state cone is slightly curved, rather than linear, in the meridional plane.

5.2. Bifurcation states versus peak stress states

Now we consider whether bifurcation states will be encountered before the peak stress states are reached. We search the stress states in the deviatoric stress plane by starting from the hydro-static pressure axis and increasing the radius in the deviatoric stress plane, while keeping the Lode angle and the mean pressure constant (refer to Figure 3a). At each point, the bifurcation criteria (30) and (46) and the peak state condition (51) are checked.

In the numerical search algorithm, a maximization procedure is employed to obtain the maximum value of the functions $f_d^2\Phi_0 - 1$ and $f_d^2\Phi_1^* - 1$ for all possible bifurcation directions

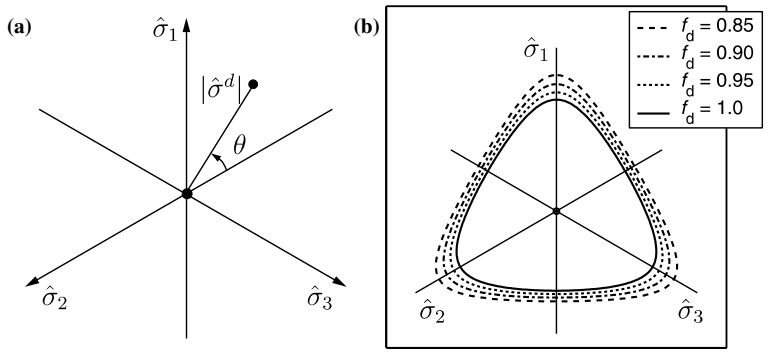


Figure 3. Representation of peak stress on the deviatoric stress plane.

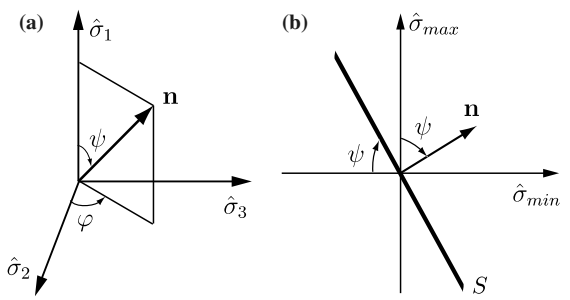


Figure 4. Determination of the direction for the weak discontinuity plane.

defined by the vector $\mathbf{n} = [\cos\psi, \sin\psi\cos\varphi, \sin\psi\sin\varphi]$ (refer to Figure 4). Numerical results, presented in Figure 5, compare the bifurcation points with the peak stress points in the deviatoric stress plane for a constant value $f_d = 0{\cdot}85$ (Figure 5a) and $f_d = 1{\cdot}0$ (Figure 5b).

It can be seen that the peak stress points and the bifurcation points for the non-polar continuum and the micro-polar continuum form three closed loops in the deviatoric stress plane. They are three-fold symmetric or periodic at the Lode angle increment $\Delta\theta = 120^\circ$. The loop of bifurcation points for the non-polar continuum intersects with the peak stress loop, whereas the loop of bifurcation points for the micro-polar continuum lies completely outside the peak stress loop. It should be noted that f_d is not a constant in a real loading program. In the vicinity of the peak states, f_d is increasing as a result of dilation and has a value less than

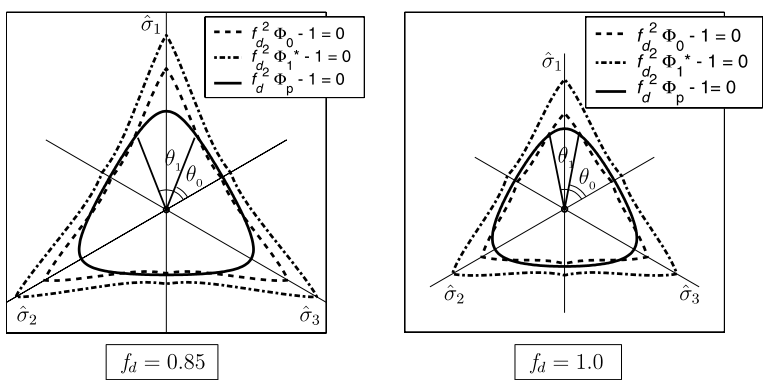


Figure 5. Representation of peak stress points (solid curves) and bifurcation points (dashed and dash-dotted curves) on the deviatoric stress plane.

1·0. Since an increase in f_d corresponds to a shrinking of the peak state and the bifurcation loops, the distance between the bifurcation points and the peak point in an arbitrary direction may be smaller. However, in the specific directions where the bifurcation point coincides with the peak point, f_d must have a unique value at this point. Therefore, the relative positions of these loops are correctly shown in Figure 5.

Bearing in mind that only stress states inside the peak stress loop are accessible, we can interpret these results as follows. In the non-polar continuum, shear bifurcation may occur before the peak state is reached for stress paths with a Lode angle $\theta \in [0, \theta_0)$ and $\theta \in (\theta_1, 120°]$. For stress paths with $\theta \in (\theta_0, \theta_1)$, which includes the stress path for a triaxial compression test, the peak state is reached before bifurcation. This indicates that a homogeneous loading will lead to a homogeneous failure or peak failure rather than a localized failure. A similar result was also obtained in [26] with an amorphous hypoplastic model (*i.e.*, a hypoplastic model with the Cauchy stress being the only state variable).

In contrast to the possible shear bifurcation in the non-polar continuum, Figure 5 shows that no bifurcation point will be reached before the peak state in the micro-polar continuum, even though the bifurcation states lie close to the peak states for a smaller Lode angle. This result means that there is no solution for the discontinuity vector $\mathbf{g}^{\mu}$ within the accessible stress domain (which is an area in the principal stress space surrounded by a cone-shaped peak-stress surface with its apex at the origin), which indicates that only homogeneous failure or peak failure will occur in a homogeneously deforming continuum body. As no restriction has been put on vector $\mathbf{g}^{\mu}$, the result also rules out a co-axial solution of $\mathbf{g}^{\mu}$ with respect to $\mathbf{n}$. In other words, pure compression or tension localization is excluded too. In an earlier analysis of localized failure with a micropolar elastoplastic model, Iordache and Willam [27] found that the micropolar continuum description suppresses localization bifurcation in shear. It may not suppress localization bifurcation in pure tension. This is, however, not in contradiction with our results, since the micropolar hypoplastic model used in this study is defined only in a compressive sub-domain in the principal stress space, as a cohesionless granular material can not sustain tension. While localized failure is widely observed experimentally in biaxial compression tests [31,32], peak failures have also been observed in triaxial compression tests [33] and in true triaxial tests [34]. It should be pointed out that in the bifurcation analysis, ideally homogeneous states are assumed in the continuum. However, in a real granular medium some packing inhomogeneity is inevitable. On the micro-scale, the void ratio varies significantly from point to point even though a macroscopically homogeneous condition is maintained. Shahinpoor [35] showed that even within a granular body composed of equal-sized spheres, the void ratio or porosity will not be uniform. This inhomogeneity of the void ratio can lead to a fluctuation in stress as well on the micro-scale when a granular specimen is loaded uniformly on its boundary. Numerical results have shown that such a state fluctuation is sufficient to initiate shear localization in a granular specimen undergoing a uniform loading process [36,37]. Therefore, in the micro-polar hypoplastic description, shear localization may occur, not in the form of sudden shear bifurcation, but rather in the form of continuous development of deformation inhomogeneity as a result of state fluctuation and strain softening.

For the non-polar continuum, the intersection points of the bifurcation and the peak state loops vary with respect to f_d (Figure 5). This result can be explained by the density-dependence of the inclination of the weak discontinuity plane. The latter, which is defined by the angles φ and ψ (refer to Figure 4), varies not only with the Lode angle θ but also with the

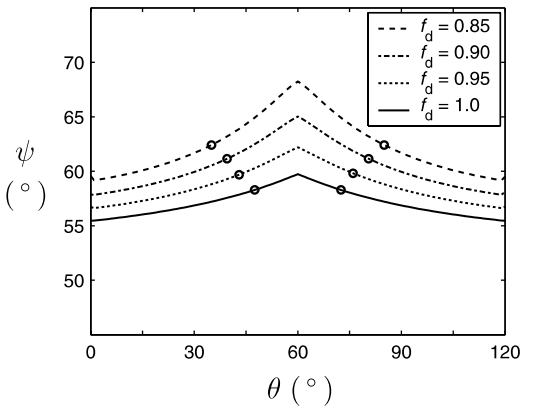


Figure 6. Inclination of the weak discontinuity plane in the non-polar hypoplastic continuum.

density factor f_d. Numerical results for the non-polar continuum show that

$$\varphi = \begin{cases} 180^\circ \text{ for } \theta \in [0, 60^\circ), \\ 90^\circ \quad \text{for } \theta \in (60^\circ, 120^\circ] \end{cases}$$

which means that the weak discontinuity plane has its normal perpendicular to the intermediate principal stress direction. In other words, the weak discontinuity will occur in the plane defined by the maximum and minimum principal stresses. The inclination of the discontinuity plane is, however, influenced by the intermediate principal stress since ψ varies with the variation of the Lode angle θ (Figure 6). An increase in the inclination of the discontinuity plane is obtained for a smaller value for f_d. The intersection points between the bifurcation loop and the peak loop are marked in Figure 6. Between these marks a homogeneous failure is predicted.

6. Conclusion

Shear localization in granular materials has been studied at the constitutive model level as a bifurcation problem. The materials have been modelled as a non-polar continuum and as a micro-polar continuum using a hypoplastic description. The bifurcation conditions have been formulated in a general manner for the two incrementally nonlinear constitutive models. These conditions indicate that shear bifurcation depends on the stress and density state in the non-polar continuum and on the couple stress state as well in the micro-polar continuum. The possibility of shear localization has been examined using a geometric interpretation for the bifurcation states and the peak stress states in the principal stress space. The peak stress states form a conical surface in the principal stress space bounding the accessible stress domain. The stress states for bifurcation have been identified on the deviatoric stress plane and compared with the peak stress states. The results show that, in the non-polar hypoplastic continuum, weak discontinuity bifurcation will occur in certain loading paths, whereas in the micro-polar hypoplastic continuum, weak discontinuity bifurcation will never occur.

In the non-polar continuum description of material behaviour, occurrence of shear localization is often attributed to the weak discontinuity bifurcation, a mathematical property associated with the constitutive models of this type. The bifurcation analyses for the micro-polar hypoplastic continuum in this work and for a micro-poplar elastoplastic continuum by Iordache and Willam suggest that the property of weak discontinuity bifurcation may no longer

be associated with a micro-polar constitutive model. And shear localization at the constitutive model level may generally be suppressed in the micro-polar continuum description of material behaviour. While shear localization has been widely observed in experiments and engineering practice, this physical phenomenon now can be interpreted as only a result of structure response. The inhomogeneity at the micro-structure level, the inevitably existing state fluctuation and the strain softening are the main causes leading to shear localization in a homogeneous-on-the-macro-level granular body.

Acknowledgement

The financial support by Australia Research Council (grant DP0453056) is acknowledged by the first author.

Appendix A

Let $\mathbf{A}$ be a partitioned square matrix in the form of

$$\mathbf{A}=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}.$$

If $\mathbf{A}$ is invertible, its inverse

$$\mathbf{B}=\mathbf{A}^{-1}=\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix}$$

can be obtained by

$$\mathbf{B}_{22}=(\mathbf{A}_{22}-\mathbf{A}_{21}\cdot\mathbf{A}_{11}^{-1}\cdot\mathbf{A}_{12})^{-1},$$

$$\mathbf{B}_{12}=-\mathbf{A}_{11}^{-1}\cdot\mathbf{A}_{12}\cdot\mathbf{B}_{22},$$

$$\mathbf{B}_{21}=-\mathbf{B}_{22}\cdot\mathbf{A}_{21}\cdot\mathbf{A}_{11}^{-1},$$

$$\mathbf{B}_{11}=\mathbf{A}_{11}^{-1}-\mathbf{B}_{12}\cdot\mathbf{A}_{21}\cdot\mathbf{A}_{11}^{-1}.$$

This representation shows that the invertibility of $\mathbf{A}$ depends on $\mathbf{A}_{11}$ and $\mathbf{A}_{22}-\mathbf{A}_{21}\cdot\mathbf{A}_{11}^{-1}\cdot\mathbf{A}_{12}$ being invertible. If $\mathbf{A}$ stands for $\mathbf{\Pi}$ and we neglect the terms $\mathbf{P}''_{uu}$ and $\mathbf{P}''_{wu}$ for $p << f_{\mathrm{s}}$, we have

$$\mathbf{A}_{11}=\mathbf{P}'_{uu},\ \mathbf{A}_{12}=\mathbf{P}_{uw},\ \mathbf{A}_{21}=\mathbf{P}'_{wu},\ \mathbf{A}_{22}=\mathbf{P}_{ww}.$$

Referring to expression (37), $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are basically invertible. In particular,

$$\mathbf{A}_{11}^{-1}=\frac{1}{\hat{a}^2}\left(\mathbf{I}-\frac{\hat{\mathbf{t}}_n\otimes\hat{\mathbf{t}}_n}{\hat{a}^2+\|\hat{\mathbf{t}}_n\|^2}\right),$$

where $\hat{\mathbf{t}}_n=\hat{\boldsymbol{\sigma}}\cdot\mathbf{n}$ denotes the normalized traction on plane S. Furthermore, by inserting all components in this expression, the following representation can be obtained:

$$\mathbf{A}_{22}-\mathbf{A}_{21}\cdot\mathbf{A}_{11}^{-1}\cdot\mathbf{A}_{12}=d_{50}\left[a_{\mu}^2\mathbf{I}+\left(1-\frac{1}{\hat{a}^2+\|\hat{\mathbf{t}}_n\|^2}\right)\hat{\mathbf{m}}_n\otimes\hat{\mathbf{m}}_n\right],$$

where $\hat{\mathbf{m}}_n=\hat{\boldsymbol{\mu}}\cdot\mathbf{n}$ denotes the normalized couple traction on plane S. Obviously $\mathbf{A}_{22}-\mathbf{A}_{21}\cdot\mathbf{A}_{11}^{-1}\cdot\mathbf{A}_{12}$ is invertible, too.

References

1. J.R. Rice, The localization of plastic deformation. In: W.T. Koiter (ed.), *Theoretical and Applied Mechanics*. North-Holland Publishing Company, Amsterdam (1976) pp. 207–220.
2. J.W. Rudnicki and J.R. Rice, Conditions for the localization of deformation in pressure-sensitive dilatant materials. *J. Mech. Phy. Solids* 23 (1975) 371–394.
3. I. Vardoulakis, Bifurcation analysis of the plane rectilinear deformation on dry sand sample. *Int. J. Solids Struct.* 17 (1981) 1085–1101.
4. N.S. Ottosen and K. Runesson, Properties of discontinuous bifurcation solutions in elasto-plasticity. *Int. J. Solids Struct.* 27 (1976) 401–421.
5. J.P. Bardet and I. Vardoulakis, The asymmetry of stress in granular media. *Int. J. Solids Struct.* 38 (2001) 353–367.
6. W. Ehlers, E. Ramm, S. Diebels and G.A. D'Addetta, From particle ensembles to Cosserat contiua: homogenization of contact forces towords stress and couple stresses. *Int. J. Solids Struct.* 40 (2003) 6681–6702.
7. H.-B. Mühlhaus, Scherfugenanalyse bei granularem Material in Rahmen der Cosserat-Theorie. *Ingenieur-Archiv* 56 (1986) 389–399.
8. R. de Borst, Simulation of strain localization: a reappraisal of the Cosserat-continuum. *Engng. Comp.* 8 (1991) 317–332.
9. A. Dietsche, P. Steinmann and K.J. Willam, Micropolar elasto-plasticity and its role in localization analysis. *Int. J. Plast.* 9 (1993) 813–831.
10. W. Ehlers and W. Volk, On theoretical and numerical methods in the theory of porous media based on polar and non-polar solid materials. *Int. J. Solids and Struct.* 35 (1998) 4597–5616.
11. J. Tejchman, Modelling of shear localisation and autogeneous dynamic effects in granular bodies. *Veröffentlichungen des Institutes für Bodenmechanik und Felsmechanik der Universität Fridericiana in Karlsruhe*, Heft 140 (1997).
12. J. Tejchman and E. Bauer, Numerical simulation of shear band formation with a polar hypoplastic model. *Comp. Geotech.* 19 (1996) 221–244.
13. E. Bauer and W. Huang, Numerical study of polar effects in shear zones. In: Pande and Pietruszczak and Schweiger (eds.), *Numerical Models in Geomechanics–NUMOG VII*. Rotterdam: A.A. Balkema (1999) pp. 133–138.
14. J. Tejchman and G. Gudehus, Shearing of a narrow granular layer with polar quantities. *Int. J. Numer. Anal. Meth. Geomech.* 25 (2001) 1–28.
15. W. Huang and E. Bauer, Numerical investigations of shear localization in a micro-polar hypoplastic material. *Int. J. Numer. Anal. Meth. Geomech.* 5 (2003) 124–148.
16. G. Gudehus, A comprehensive constitutive equation for granular materials. *Soils Found.* 36 (1996) 1–12.
17. E. Bauer, Calibration of a comprehensive hypoplastic model for granular materials. *Soils Found.* 36 (1996) 13–26.
18. W. Huang, K. Nübel and E. Bauer, Polar extension of a hypoplastic model for granular materials with shear localization. *Mech. Mat.* 34 (2002) 563–576.
19. D. Kolymbas, Bifurcation analysis for sad samples with a non-linear constitutive equation. *Ingenieur-Archiv* 50 (1981) 131–140.
20. R. Chambon and J. Desrues, Bifurcation par localisation et non linéarité incrémentale: un example heuristique d'analyse complété. In: *Plasticity Instability*. Paris: Press ENPC. (1985) pp. 101–119.
21. J. Desrues and R. Chambon, Shear band analysis for granular materials: The question of incremental non-linearity. *Ingenieur-Archiv* 59 (1989) 187–196.
22. R. Chambon, S. Crochepeyer and J. Desrues, Localization criteria for non-linear constitutive equations of geomaterials. *Mech. Cohesive-Fric. Mater.* 5 (2000) 61–82.
23. E. Bauer and W. Huang, The dependence of shear banding on pressure and density in hypoplasticity. In: Adachi, Oka and Yashima (eds.), *Localization and Bifurcation Theory for Soils and Rocks*. Rotterdam: A.A. Balkema (1998) pp. 81–90.
24. E. Bauer, Analysis of shear band bifurcation with a hypoplastic model for a pressure and density sensitive granular mateiral. *Mech. Mater.* 31 (1999) 597–609.
25. W. Wu and Z. Sikora, Localized bifurcation in hypoplasticity. *Int. J. Engng. Sci.* 29 (1991) 195–201.
26. W. Wu, Non-linear analysis of shear band formation in sand. *Int. J. Numer. Anal. Meth. Geomech.* 24 (2000) 245–263.

27. M.-M. Iordache and K. Willam, Localized failure analysis in elastoplastic Cosserat continua. *Comput. Methods Appl. Mech. Engng.* 151 (1998) 559–586.
28. E. Bauer, Conditions for embedding Casagrande's critical states into hypoplasticity. *Mech. Cohesive-Frict. Mater.* 5 (2000) 124–148.
29. I. Herle and G. Gudehus, Determination of parameters of a hypoplastic constitutive model from properties of grain assemblies.. *Mech. Cohesive-Fric. Mater.* 4 (1999) 461–486.
30. G.A. Holzapfel, *Nonlinear Solid Mechanics–A Continuum Approach for Engineering*. Chichester: John Wiley (2001) 455 pp.
31. I. Vardoulakis and B. Graf, Calibration on constitutive models for granular materials using data from biaxial experiments. *Géotechnique* 35 (1985) 299–317.
32. D. Mokni and J. Desrues, Strain localization measurements in undrained plane-strain biaxial tests on Hostun RF sand. *Mech. Cohesive-Fric. Mater.* 4 (1998) 419–441.
33. P.V. Lade, Localization effects in triaxial tests on sand. In: Proc. IUTAM Symposium on *Deformation and Failure of Granular Materials*, Delft (1982) pp. 461–471.
34. Q. Wang and P.V. Lade, Shear banding in true triaxial tests and its effect on failure in sand. *ASCE J. Engng. Mech.* 127 (2001) 754–761.
35. M. Shahinpoor, Statistical mechanical considerations on storing bulk solids. *Bulk Solids Handling* 1 (1981) 31–36.
36. K. Nübel, Experimental and numerical investigation of shear localization in granular material. *Veröffentlichungen des Institues für Bodenmechanik und Felsmechanik der Universität Fridericiana in Karlsruhe*, Heft 159 (2003) 159 pp.
37. K. Nübel and W. Huang, A study of localized deformation pattern in granular media. *Compt. Meth. Appl. Mech. Engng.* 193 (2004) 2719–2743.

Journal of Engineering Mathematics (2005) 52: 185–198
DOI 10.1007/s10665-004-1771-8

Large-strain dynamic cavity expansion in a granular material

V.A. OSINOV
Institute for Geotechnical Engineering, University of Natural Resources and Applied Life Sciences, Feistmantelstr. 4, A-1180 Vienna, Austria (vladimir.osinov@boku.ac.at)

Received 20 June 2004; accepted in revised form 26 July 2004

Abstract. The dynamic problem of the symmetric expansion of a cylindrical or spherical cavity in a granular medium is considered. The constitutive behaviour of the material is governed by a hypoplasticity relation for granular soils capable of describing both monotonic and cyclic deformation. The problem is solved numerically by a finite-difference technique. A nonreflecting boundary condition used at the outer boundary of the computational domain makes it possible to model a continuous multi-cycle loading on the cavity wall. The solution is illustrated by numerical examples. Possible geomechanical applications to the modelling of the vibratory compaction and penetration in granular soils are discussed.

Key words: dynamic cavity expansion, granular material, hypoplasticity, transparent boundary condition

1. Introduction

The present paper is devoted to the numerical solution of the dynamic problem of the symmetric expansion of a cylindrical or spherical cavity in a granular medium. The study is orientated towards geomechanical applications. The cavity-expansion problem is widely used in geomechanics for the modelling of various processes (*e.g.* penetration) in soils and rocks [1]. As far as granular soils are concerned, most analyses of the cavity-expansion problem are carried out for the quasi-static monotonic expansion in order to determine the limit pressure and/or the pressure-expansion curve [2–7].

For the modelling of processes in which the rate of loading is high, the use of a quasi-static formulation of the cavity problem is inadequate, and a dynamic formulation is necessary. The difference between the quasi-static and the dynamic solutions may be substantial. For instance, for a cylindrical cavity in a linear elastic medium, the stresses in the quasi-static problem vary as $1/r^2$, while in the dynamic problem the stresses decay asymptotically for $r \to \infty$ as $1/\sqrt{r}$.

In the geotechnical context, it is clear intuitively that the modelling of such processes as vibratory pile driving or deep vibratory compaction (also called vibroflotation [8]) requires a dynamic formulation. However, because of the nonlinear behaviour of soils at large deformations, it may be difficult to make a quantitatively justified *a priori* judgement about the necessity of using the dynamic formulation for a particular problem. In the case of the cavity problem with a periodic boundary condition, a quantitative criterion for using the dynamic formulation is that the wavelength for a given frequency is of the same order as or smaller than the cavity diameter. The difficulty in using this criterion consists in the fact that the incremental stiffness of a soil changes considerably during large plastic deformation. The speed of propagation of small-amplitude elastic waves, which is often known for a particular soil, is not applicable to the estimation of the wavelength when the large-strain expansion of a cavity is studied: the wave speed corresponding to the incremental stiffness may be much

lower and, moreover, hardly predictable as it depends on the strain. In such cases the necessity of using a dynamic formulation can only be justified after the solutions to the dynamic problem have been obtained. This issue is discussed in more detail below in Section 5 where numerical examples are considered.

In this paper we solve the dynamic problem allowing for the possibility of prescribing a multi-cycle loading on the cavity wall expressed either through the cavity pressure or the cavity radius. Unlike problems with a monotonic or single-cycle loading, the cavity expansion problem with a multi-cycle boundary condition requires the use of a proper constitutive model which realistically describes the behaviour of a granular material under repeated loading and, in particular, its gradual compaction under cyclic shearing. In this study we employ an extended constitutive relation of the hypoplasticity theory developed in [9] and capable of describing both monotonic and cyclic deformation of granular materials such as sand. This hypoplasticity relation was used earlier in [10–12] for the numerical solution of dynamic plan-wave problems for granular soils. A number of particular solutions to the dynamic-cavity problem with a simple version of the hypoplasticity relation were obtained in [13].

The constitutive theory used in the present study is outlined in Section 2. The mathematical formulation of the initial-boundary-value problem is described in Section 3. The problem is solved numerically by a finite-difference technique described in Section 4. Numerical examples are given in Section 5, and possible geotechnical applications are briefly discussed at the end of the paper.

2. Hypoplasticity relations

The theory of hypoplasticity originated as an alternative to elasto-plasticity theories for the purpose of describing the plastic deformation of granular materials without the introduction of a yield surface, a flow rule and without the decomposition of the deformation into an elastic and a plastic part. Attempts to find a suitable mathematical form including the pressure and density dependence of the stiffness had led to an equation written in rectangular coordinates as [14, 15]

$$\frac{\mathrm{d}T_{ij}}{\mathrm{d}t} = \mathcal{L}_{ijkl}(\mathbf{T}, e) D_{kl} + N_{ij}(\mathbf{T}, e)\|\mathbf{D}\|, \tag{1}$$

where $\mathbf{T}$ is the Cauchy stress tensor, $\mathbf{D}$ is the stretching tensor (the symmetric part of the velocity gradient), e is the void ratio, t is time, and $\mathrm{d}(\,)/\mathrm{d}t$ stands for the material time derivative (for simplicity, we write the material time derivative instead of an objective stress rate; for the velocity fields considered in this paper the objective Jaumann stress rate used in hypoplasticity coincides with the material time derivative). The fourth-order stiffness tensor $\mathcal{L}$ and the second-order tensor $\mathbf{N}$ in (1) are responsible, respectively, for the linear and the nonlinear parts of the constitutive relation. The term $\|\mathbf{D}\| = \sqrt{D_{ij} D_{ij}}$ makes the constitutive equation incrementally nonlinear. Material parameters involved in a specific form of the tensors $\mathcal{L}$ and $\mathbf{N}$ are independent of the stresses and the density. Equation (1) also allows the critical-state concept to be incorporated into the theory. The tensors $\mathcal{L}$ and $\mathbf{N}$ are given in detail in the Appendix.

A limitation of the hypoplasticity concept expressed in the form (1) is the so-called ratcheting which manifests itself in an unrealistically high rate of accumulation of deformation or stress during repeated multi-cycle loading with small strain amplitudes (below 10^{-2}) [16, 17]. This is inevitable as long as the state of a granular material and therefore its stiffness are determined only by stresses and density.

In order to better describe the behaviour of granular materials and, in particular, their response to alternating cyclic loading, the concept of intergranular strain was introduced in [9] as an extension of hypoplasticity. This concept as such is not related to a specific hypoplastic equation and is applicable to any equation of the form (1). The extended constitutive theory contains a so-called intergranular-strain tensor $\boldsymbol{\delta}$ as a new state variable which carries the information about the history of the deformation and determines the state of the material along with the stress tensor and the void ratio.

The hypoplasticity relation in the extended form becomes

$$\frac{\mathrm{d}T_{ij}}{\mathrm{d}t} = \mathcal{M}_{ijkl}(\mathbf{T}, \mathbf{D}, \boldsymbol{\delta}, e) D_{kl}, \tag{2}$$

where the tensor $\mathcal{M}$ includes the tensors $\mathcal{L}$ and $\mathbf{N}$ from (1). The intergranular-strain tensor is determined by its evolution equation

$$\frac{\mathrm{d}\delta_{ij}}{\mathrm{d}t} = F_{ij}(\mathbf{D}, \boldsymbol{\delta}). \tag{3}$$

For detailed forms of $\mathcal{M}$ and $\mathbf{F}$, see the Appendix.

The function $\mathbf{F}$ in (3) is such that, under monotonic loading with a fixed direction of deformation $\mathbf{D}/\|\mathbf{D}\|$, the tensor $\boldsymbol{\delta}$ tends asymptotically to a certain value. In turn, the tensor $\mathcal{M}$ is constructed in such a way that, as $\boldsymbol{\delta}$ approaches its asymptotic value, the constitutive response of (2) approaches the response of the basic equation (1). Thus, during monotonic unidirectional deformation the extended equation gradually turns into the original one. On the other hand, for small-amplitude alternating deformation the response produced by (2), (3) is close to being linearly elastic, if the amplitude is small enough.

An initial value of $\boldsymbol{\delta}$ required for the integration of Equation (3) is often indeterminate when solving a particular problem. Note that the indeterminacy of the initial values of certain state variables is typical of plasticity models in which these variables are determined by the foregoing deformation. In problems with multi-cycle deformation, the influence of the initial value of $\boldsymbol{\delta}$ vanishes after 2–3 cycles. In the numerical calculations in this paper the initial value of $\boldsymbol{\delta}$ is taken to be zero.

Examples of how the extended hypoplasticity relation works for cyclic loading can be found in [12].

3. Boundary-value problem

In the dynamic problem, constitutive equations (2), (3) are supplemented with the evolution equation for the void ratio (with incompressible grains)

$$\frac{\mathrm{d}e}{\mathrm{d}t} = (1+e)\ \mathrm{tr}\,\mathbf{D} \tag{4}$$

and the equation of motion (without mass forces)

$$\mathrm{div}\,\mathbf{T} = \varrho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}, \tag{5}$$

where $\mathbf{v}$ is the velocity field, and ϱ is the density of the material.

We study the problem of the symmetric expansion of a cavity in which all quantities are functions of the radial coordinate and time. In the cylindrical coordinates r, φ, z, a symmetric expansion under plane strain conditions (infinitely long cavity) is described by the velocity component v_r, the stress components $T_{rr}, T_{\varphi\varphi}, T_{zz}$, the intergranular-strain components

$\delta_{rr}, \delta_{\varphi\varphi}$ and the void ratio e. The stretching tensor has two nonzero components $D_{rr} = \partial v_r / \partial r$ and $D_{\varphi\varphi} = v_r / r$. In what follows, for brevity we will write $v, T_r, T_\varphi, T_z, \delta_r, \delta_\varphi$, respectively, for $v_r, T_{rr}, T_{\varphi\varphi}, T_{zz}, \delta_{rr}, \delta_{\varphi\varphi}$.

In cylindrical coordinates, Equations (5), (4) are written, respectively, as

$$\frac{\partial T_r}{\partial r} + \frac{1}{r}(T_r - T_\varphi) = \varrho \frac{\mathrm{d}v_r}{\mathrm{d}t}, \tag{6}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = (1+e)\left(\frac{\partial v}{\partial r} + \frac{v}{r}\right). \tag{7}$$

For brevity, we do not write out the constitutive functions in detail; for the considered case of cylindrical symmetry this can easily be done using the functions given in the Appendix. We write the constitutive equations in general form showing only the dependence of the time derivatives on the other functions involved in the problem:

$$\frac{\mathrm{d}T_r}{\mathrm{d}t} = H_r\left(T_r, T_\varphi, T_z, \frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi, e\right), \tag{8}$$

$$\frac{\mathrm{d}T_\varphi}{\mathrm{d}t} = H_\varphi\left(T_r, T_\varphi, T_z, \frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi, e\right), \tag{9}$$

$$\frac{\mathrm{d}T_z}{\mathrm{d}t} = H_z\left(T_r, T_\varphi, T_z, \frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi, e\right), \tag{10}$$

$$\frac{\mathrm{d}\delta_r}{\mathrm{d}t} = F_r\left(\frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi\right), \tag{11}$$

$$\frac{\mathrm{d}\delta_\varphi}{\mathrm{d}t} = F_\varphi\left(\frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi\right). \tag{12}$$

The initial-boundary-value problem for Equations (6–12) with the unknown functions $v, T_r, T_\varphi, T_z, \delta_r, \delta_\varphi, e$ is formulated for a domain $r \in [r_a(t), r_b(t)], t \geqslant 0$, where $r_a(t)$ is the radius of the cavity, and $r_b(t)$ is the outer radius of the domain where the solution is sought. For a cavity in an infinite body, $r_b = \infty$. The motion of the boundaries $r_a(t), r_b(t)$ is determined by the integration of the velocity starting with initial values r_a^0, r_b^0. Initial conditions for all the functions are prescribed on the interval $[r_a^0, r_b^0]$.

The boundary condition at $r_a(t)$ and $r_b(t)$ can be taken as either the velocity or the stress component T_r as a function of time. For a cavity in an infinite body it is natural to put $v(\infty) = 0$ or $T_r(\infty) = \text{const}$. However, in the numerical solution an infinite outer radius, as required by the mathematical formulation of the problem, is replaced with a finite radius r_b. This leads to a spurious reflection from the outer boundary with either of the conditions $v(\infty) = 0$, $T_r(\infty) = \text{const}$. Increasing the outer radius does not eliminate this problem: although the amplitude of the outgoing and, correspondingly, the reflected waves may be small and even negligible far away from the cavity owing to geometrical attenuation, the amplitude of the reflected waves increases again as they approach the cavity because of the same geometrical effect. Even in the presence of material damping the reflected waves in the vicinity of the cavity may essentially distort the solution as compared to that for an infinite domain. This fact makes it necessary to introduce a nonreflecting boundary condition at r_b which will make the boundary transparent for outgoing waves and thus imitate an unbounded domain.

Since the amplitude of a wave substantially decreases as it propagates from the cavity, the wave far away from the cavity can be well approximated by a linear elastic wave. If, in addition, we neglect the cylindrical or spherical geometry and assume the wave at $r \gg r_a$ to be a plane longitudinal one, this allows us to prescribe a simple nonreflecting boundary condition at $r_b \gg r_a$. For a plane elastic wave propagating in the positive x-direction with a wave speed c and described by a function $f(x-ct)$, the partial derivatives are connected with each other as $\partial f/\partial t = -c\partial f/\partial x$. The required boundary condition in terms of velocity thus takes the form

$$\left.\frac{\partial v}{\partial t}\right|_{(r_b,t)} = -c \left.\frac{\partial v}{\partial r}\right|_{(r_b,t)} \tag{13}$$

with the wave speed $c = \sqrt{C/\varrho}$, where the small-strain uniaxial stiffness C at $r = r_b$ can be found directly from the constitutive equation used. The effectiveness of the boundary condition (13) will be demonstrated later by a numerical example.

For a spherical cavity, the symmetric expansion is described in spherical coordinates r, φ, θ by the velocity component v_r, the stress components $T_{rr}, T_{\varphi\varphi} = T_{\theta\theta}$, the intergranular-strain components $\delta_{rr}, \delta_{\varphi\varphi} = \delta_{\theta\theta}$ and the void ratio e. The stretching tensor has three nonzero components $D_{rr} = \partial v_r/\partial r$ and $D_{\varphi\varphi} = D_{\theta\theta} = v_r/r$. The equation of motion, the constitutive equations and the mass-balance equation for a spherical cavity are written, respectively, as

$$\frac{\partial T_r}{\partial r} + \frac{2}{r}(T_r - T_\varphi) = \varrho \frac{\mathrm{d}v_r}{\mathrm{d}t}, \tag{14}$$

$$\frac{\mathrm{d}T_r}{\mathrm{d}t} = H_r\left(T_r, T_\varphi, \frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi, e\right), \tag{15}$$

$$\frac{\mathrm{d}T_\varphi}{\mathrm{d}t} = H_\varphi\left(T_r, T_\varphi, \frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi, e\right), \tag{16}$$

$$\frac{\mathrm{d}\delta_r}{\mathrm{d}t} = F_r\left(\frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi\right), \tag{17}$$

$$\frac{\mathrm{d}\delta_\varphi}{\mathrm{d}t} = F_\varphi\left(\frac{\partial v}{\partial r}, \frac{v}{r}, \delta_r, \delta_\varphi\right), \tag{18}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = (1+e)\left(\frac{\partial v}{\partial r} + 2\frac{v}{r}\right). \tag{19}$$

4. Numerical algorithm

A numerical approach developed in [10, 11] for dynamic plane-wave problems is modified here for the solution of cylindrical and spherical problems. The approach consists in the replacement of a continuous body by a discrete system with the mass of the medium being concentrated at separate surfaces (lumped masses). The equations of motion for the masses and the constitutive equations for the subdomains constitute a finite-difference approximation of the original partial differential equations.

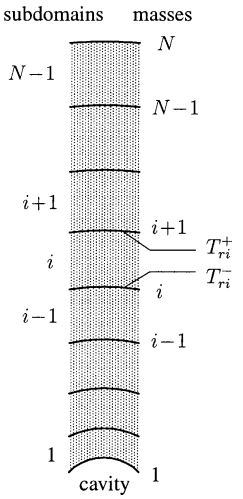


Figure 1. Discretisation of the domain.

4.1. Cylindrical problem

The domain between two concentric cylindrical surfaces of radii r_a^0 and r_b^0 where the solution is sought is divided by N concentric surfaces into $N-1$ subdomains as shown in Figure 1. The mass of a subdomain i is assumed to be concentrated at the surface i. The mass calculated per unit length in the z-direction and per radian in φ is given by

$$m_i = \tfrac{1}{2}\varrho\left(r_{i+1}^2 - r_i^2\right), \tag{20}$$

where r_i, r_{i+1} are the initial radii of the corresponding surfaces, and ϱ stands for the initial density. The mass of the surface N remains unspecified.

The equation of motion for the mass m_i reads

$$\frac{m_i}{r_i}\frac{\mathrm{d}v_i}{\mathrm{d}t} = T_{ri}^{-} - T_{r(i-1)}^{+}, \tag{21}$$

where T_{ri}^{+}, T_{ri}^{-} denote the stresses at the boundaries of the subdomains, see Figure 1, and v_i is the radial velocity of the mass. The mass m_i is independent of the change in the radius of the surface i during the cavity expansion.

The stress components are assumed to vary linearly in the subdomains. If T_{ri} is the value of the radial stress in the middle point of the subdomain i, the stresses T_{ri}^{+}, T_{ri}^{-} at the boundaries are

$$T_{ri}^{+} = T_{ri} + \tfrac{1}{2}(r_{i+1} - r_i)\left.\frac{\partial T_r}{\partial r}\right|_i, \tag{22}$$

$$T_{ri}^{-} = T_{ri} - \frac{1}{2}(r_{i+1} - r_i)\left.\frac{\partial T_r}{\partial r}\right|_i. \tag{23}$$

Once the mass of the subdomains is concentrated at their boundaries, the material in the subdomains is treated as inertia-free. This allows us to find the gradient of the radial stress in (22), (23) from the static equilibrium equation

$$\frac{\partial T_r}{\partial r} + \frac{1}{r}(T_r - T_\varphi) = 0 \tag{24}$$

written for the middle of the subdomain:

$$\left.\frac{\partial T_r}{\partial r}\right|_i = 2\frac{T_{\varphi i} - T_{ri}}{r_i + r_{i+1}}. \tag{25}$$

Substitution of (22), (23), (25) in (21) gives the equation of motion of the mass i in the form:

$$\frac{m_i}{r_i}\frac{\mathrm{d}v_i}{\mathrm{d}t} = T_{ri} - T_{r(i-1)} + \frac{r_{i+1} - r_i}{r_{i+1} + r_i}\left(T_{ri} - T_{\varphi i}\right) + \frac{r_i - r_{i-1}}{r_i + r_{i-1}}\left(T_{r(i-1)} - T_{\varphi(i-1)}\right). \tag{26}$$

Equation (26) is used in the numerical calculations for $i = 2$ to $N - 1$. The motion of the mass m_1 is determined by the equation

$$\frac{m_1}{r_1}\frac{\mathrm{d}v_1}{\mathrm{d}t} = T_{r1} + \frac{r_2 - r_1}{r_2 + r_1}\left(T_{r1} - T_{\varphi 1}\right) - T_{ra}, \tag{27}$$

where T_{ra} is a boundary condition for the radial stress at $r = r_a$. The motion of the mass m_N is determined by the boundary condition (13)

$$\frac{\mathrm{d}v_N}{\mathrm{d}t} = -c\frac{v_N - v_{N-1}}{r_N - r_{N-1}}. \tag{28}$$

Owing to the invariance of the masses m_i with respect to the deformation of the medium, the use of the equations of motion in the form (26), (27) does not require the calculation and the updating of the density ϱ in the case of large strains.

If we substitute (20) for m_i in (26), we obtain

$$\frac{1}{2}\varrho\frac{r_{i+1} + r_i}{r_i}\frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{T_{ri} - T_{r(i-1)}}{r_{i+1} - r_i} + \frac{T_{ri} - T_{\varphi i}}{r_{i+1} + r_i} + \frac{(r_i - r_{i-1})\left(T_{r(i-1)} - T_{\varphi(i-1)}\right)}{(r_{i+1} - r_i)(r_i + r_{i-1})}. \tag{29}$$

It is easily seen that (29) is a finite-difference approximation of the original equation of motion (6). Equation (29) turns asymptotically into (6) as $\max\Delta r \to 0$ if

$$\lim_{\max\Delta r \to 0} \frac{r_i - r_{i-1}}{r_{i+1} - r_i} = 1. \tag{30}$$

The stress components, the intergranular-strain components and the void ratio in each subdomain are calculated by the integration of their evolution equations (8–12), (7), whereas the components of the stretching tensor are found as

$$\left.\frac{\partial v}{\partial r}\right|_i = \frac{v_{i+1} - v_i}{r_{i+1} - r_i}, \tag{31}$$

$$\left.\frac{v}{r}\right|_i = \frac{v_{i+1} + v_i}{r_{i+1} + r_i}. \tag{32}$$

The time integration of the equations is performed implicitly as

$$f(t + \Delta t) = f(t) + [\alpha f'(t) + (1 - \alpha) f'(t + \Delta t)]\Delta t, \tag{33}$$

where f and f' denote the set of the unknown functions and their time derivatives, respectively, and α is the parameter of the scheme, $0 \leqslant \alpha < 1$. Equation (33) is solved by successive approximations.

4.2. Spherical problem

The numerical scheme for the spherical problem can be constructed in a similar way. Omitting intermediate computations, we write out only equations corresponding to (20), (26), (27):

$$m_i = \frac{1}{3}\varrho\left(r_{i+1}^3 - r_i^3\right), \tag{34}$$

$$\frac{m_i}{r_i^2}\frac{\mathrm{d}v_i}{\mathrm{d}t} = T_{ri} - T_{r(i-1)} + 2\frac{r_{i+1}-r_i}{r_{i+1}+r_i}(T_{ri} - T_{\varphi i}) + 2\frac{r_i - r_{i-1}}{r_i + r_{i-1}}\left(T_{r(i-1)} - T_{\varphi(i-1)}\right), \tag{35}$$

$$\frac{m_1}{r_1^2}\frac{\mathrm{d}v_1}{\mathrm{d}t} = T_{r1} + 2\frac{r_2 - r_1}{r_2 + r_1}(T_{r1} - T_{\varphi 1}) - T_{ra}. \tag{36}$$

5. Numerical examples

Numerical solutions presented below are calculated for a cylindrical cavity with an initial radius of 0·2 m, with a homogeneous initial stress state $T_r = T_\varphi = T_z = -100$ kPa and an initial void ratio of 0·9. For dry sand this void ratio corresponds to an initial density of $1{\cdot}4 \times 10^3$ kg/m^3. The outer boundary of the computational domain is taken at $r_b^0 = 30$ m.

To show the numerical performance of the boundary condition (13), let the cavity be loaded by a single pressure impulse on the cavity wall as shown in Figure 2. Two cases of the boundary condition at the outer boundary are considered: the case $T_r = \text{const} = -100$ kPa and the transparent boundary condition (13). The corresponding solutions are compared in Figure 3. It is seen that the inaccuracy related to the boundary condition (13) caused by the cylindrical geometry, the non-elastic effects in the material behaviour and by the numerics itself is negligible: the outgoing wave goes through the boundary without reflection.

In the further examples, the cavity is loaded by a cyclic sinusoidal pressure T_r on the cavity wall varying between -100 and -150 kPa. Figures 4–7 show two cases with different frequencies, 30 and 10 Hz, with all other parameters being the same as before.

The periodic loading results in a steady increase of the cavity radius. Figure 4 shows the change in the radius during the first 30 cycles of the loading. In spite of the fact that the amplitude of the loading is the same in both cases, the rate of expansion (the change in the radius per cycle) in the high-frequency case of 30 Hz is 60 times higher than in the low-frequency case of 10 Hz.

The dynamic pressure-expansion curves at the beginning of the expansion are compared in Figure 5. The curves are essentially different. The hysteresis and, consequently, the work done by the loading in one cycle is much higher in the high-frequency case.

The cyclic loading of the cavity is accompanied by cyclically changing shear stresses in the vicinity of the cavity. According to the general characteristics of granular materials, this

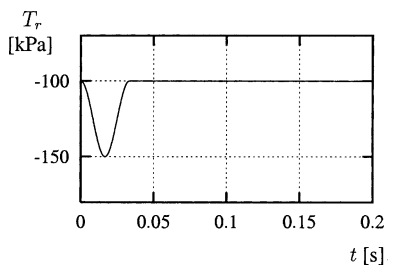


Figure 2. Boundary condition at $r_a = 0{\cdot}2$ m for the solutions shown in Figure 3.

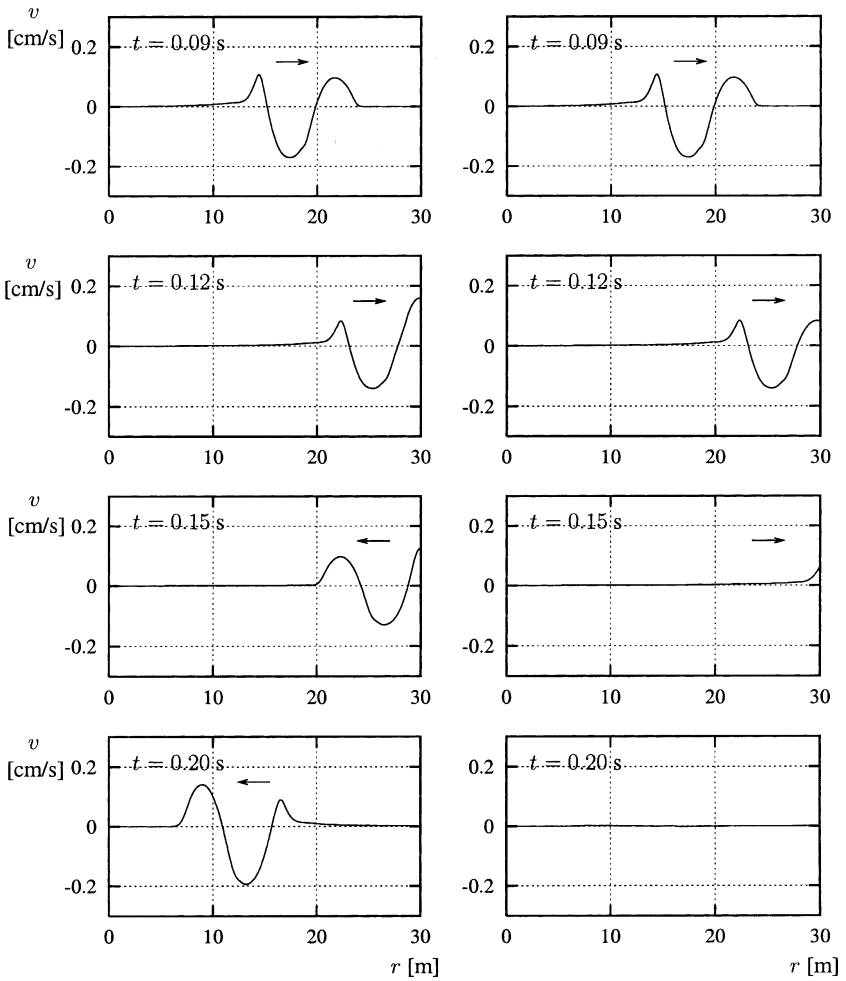


Figure 3. Velocity profiles at different times produced by a pressure impulse (Figure 2) in a cylindrical cavity with $r_a^0 = 0{\cdot}2$ m. Left column: boundary condition $T_r = \text{const}$ at $r_b = 30$ m. Right column: the nonreflecting plane-wave boundary condition (13) at $r_b = 30$ m.

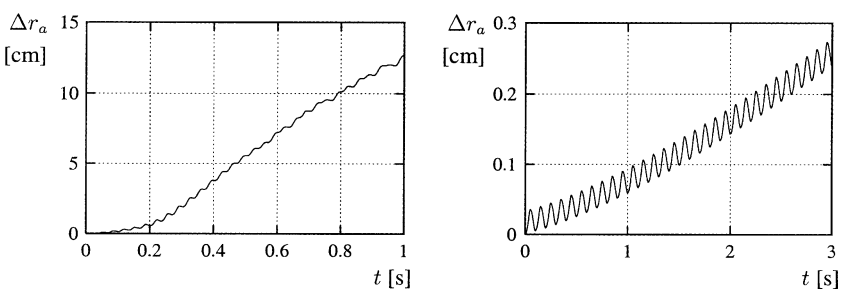


Figure 4. The change in the radius of a cylindrical cavity loaded by a cyclic pressure varying between −100 and −150 kPa with a frequency of 30 Hz (left) and 10 Hz (right). Initial radius: 0·2 m.

results in a gradual compaction of the material. The change in the void ratio in the vicinity of the cavity is shown in Figure 6 which again reveals a pronounced influence of the frequency on the solution. Not only is the rate of compaction much higher in the high-frequency case, but also the compaction spreads much farther from the cavity wall. Obviously,

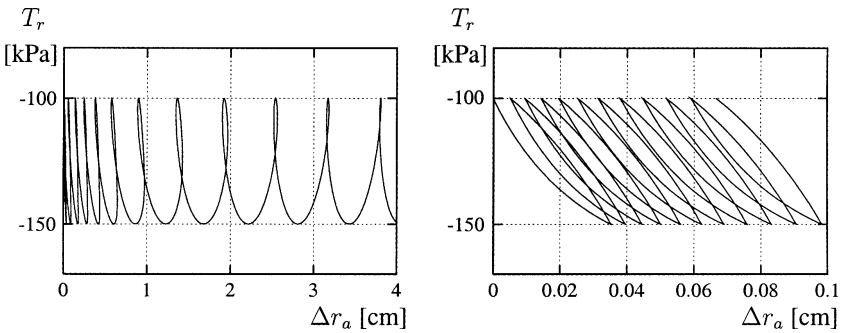


Figure 5. Pressure-expansion curves for the same problems as in Figure 4. Left: 30 Hz, right: 10 Hz.

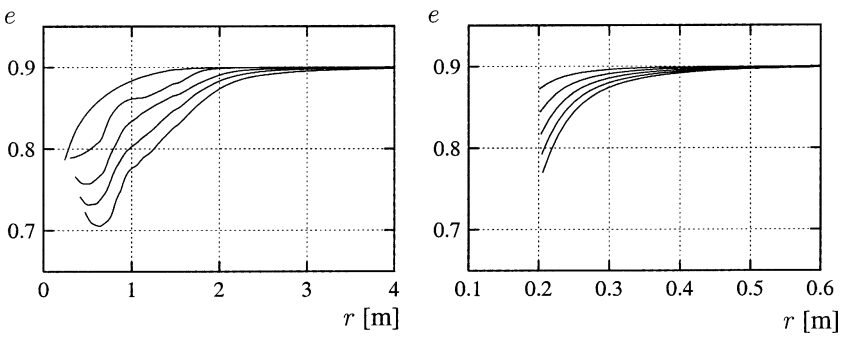


Figure 6. Void ratio in the vicinity of the cavity after 12, 24, 36, 48 and 60 cycles of loading (from the upper to the lower curve, respectively). Left: 30 Hz, right: 10 Hz.

in cavity problems the compaction of the material facilitates the increase of the cavity radius as compared to problems in which the material does not exhibit steady compaction or is incompressible.

The cavity expansion not only leads to the compaction of the material in the vicinity of the cavity but also to the change in the stress state. The time dependence of the stress components in the material element initially situated at $r = 1{\cdot}0$ m is shown in Figure 7. In the high-frequency case, the average absolute values of all stress components are rapidly reduced at the beginning of the loading and then change rather slowly. The ratios between the average values after the reduction do not correspond to a hydrostatic state as was the case initially. In the low-frequency case, apart from the fact that the average values of the stresses change slowly, the amplitudes of the stresses are much smaller.

The significant reduction in the stress level observed in the high-frequency case can be explained if we again recall that, according to the general characteristics of granular materials, cyclic shearing leads to compaction or, if volume change is prohibited by kinematical restrictions, to a reduction in the pressure. Although the material in the vicinity of the cavity is compacted due to the cyclic deformation, this compaction is not enough for the mean pressure to remain at the same level. As a result, the absolute values of the stresses decrease.

As mentioned in the Introduction, the change in the stress amplitudes with radius is different in the quasi-static and the dynamic problems: the stresses decrease slower in the dynamic case and thus spread farther from the cavity. For a cylindrical cavity in an elastic medium, the stress amplitudes are proportional to $1/r^n$ (in the dynamic case – asymptotically for $r \to \infty$) with $n = 2$ and $0{\cdot}5$ for the quasi-static and the dynamic problems, respectively. For the two numerical examples just considered, Table 1 shows the amplitudes of the stress component T_r

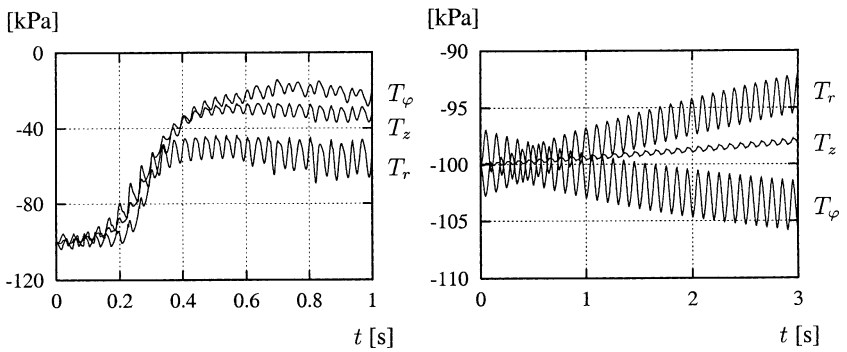


Figure 7. Stress components at $r = 1{\cdot}0$ m versus time. Left: 30 Hz, right: 10 Hz.

Table 1. Comparison of the solutions.

Frequency (Hz)	$\Delta T_r(r_a)$ (kPa)	$\Delta T_r(r)$ (kPa)	n (37)
30	50	12–20	0·75–1·0
10	50	2·9	1·77
30	1	0·056	1·79

at two points with initial coordinates $r_a = 0{\cdot}2$ m (cavity wall) and $r = 1$ m. In the first example (30 Hz) the stress amplitude at $r = 1$ m is taken for $t > 0{\cdot}5$ s. The exponent n given in Table 1 is estimated from the stress amplitudes at the two points by the formula

$$n = \frac{\log[\Delta T_r(r)/\Delta T_r(r_a)]}{\log(r_a/r)}. \tag{37}$$

The exponent n for the low-frequency problem is greater and closer to that for the quasi-static solution. Also shown in Table 1 is a third case for a frequency of 30 Hz and a small loading amplitude of 1 kPa. Surprisingly, despite the high frequency, the attenuation of the stresses in this case is high and results in a high exponent n. The explanation is, however, simple: the small-amplitude loading causes nearly elastic deformations which are characterised by relatively high stiffness and, correspondingly, by a large wavelength. In this case the speed of longitudinal small-amplitude waves, which is estimated from the constitutive equation, is about 260 m/s, which for a frequency of 30 Hz gives a wavelength of 8·7 m – large enough in comparison with the cavity size. Contrary to this, the plastic deformations which the material undergoes during large-amplitude loading are characterised by lower stiffness and shorter wavelengths.

Solutions to the spherical problem are qualitatively similar to those to the cylindrical problem: the expansion is also accompanied by the compaction of the material and by the decrease in the absolute values of stresses in the vicinity of the cavity. There are, however, noticeable quantitative differences: if the spherical problem is solved with the same parameters, the attenuation of the stress amplitude with distance from the cavity is much higher than in the cylindrical case. As a consequence, the rate of the cavity expansion and the rates of the compaction and the stress reduction are much lower.

6. Discussion of applications

The cavity-expansion problem described above allows us to model dynamic processes related to the cyclic deformation of granular soils such as deep vibratory compaction [8] (cylindrical cavity) and vibratory pile driving (spherical cavity).

In the deep vibratory compaction technique, the soil is compacted by a vibrating steel cylinder (vibrator) placed vertically in a soil layer. The vibrator is 30–45 cm in diameter, 3–5 m in length and has inside a motor with a rotating eccentric mass which causes the axis of the vibrator to rotate about the vertical axis with a frequency of 30–60 Hz.

The accurate modelling of deep vibratory compaction is extremely difficult if at all possible (in this connection we refer to [18]). This is not only because of the complicated material behaviour but also due to the fact that the dynamic problem describing the deformation of the soil surrounding the vibrator is spatially three-dimensional. Additional difficulties arise when taking into account the inflow of the new material from above to the immediate vicinity of the vibrator.

The main goal in the modelling of deep vibratory compaction is to reveal the dependence of the compaction on the parameters of the vibrator and the soil in order to optimise the process. The solution of two- or three-dimensional problems is time-consuming and can therefore hardly serve this purpose. The one-dimensional problem considered in this paper involves much less computing time and can thus provide much data required for a detailed modelling.

The two main simplifications entailed in the reduction of the deep vibratory compaction problem to a one-dimensional one consist in the fact that the rotating force produced by the vibrator is replaced with an axially symmetric loading, and that the contribution of the vertical displacement of the soil to the compaction is not taken into account. In order to assess the influence of these simplifications, a number of solutions to the one-dimensional problem should be compared with the corresponding solutions to a two- or three-dimensional problem. This may result in establishing correction factors which could subsequently be used in a one-dimensional analysis.

The introduction of correction factors, either by comparison with experimental data or with more sophisticated numerical calculations, is also necessary when applying the dynamic cavity-expansion theory to the modelling of vibratory pile driving. For the quasi-static penetration this approach was adopted, for instance, in [19].

Appendix. Constitutive functions

The tensors $\mathcal{L}$ and N in (1), (2) are written as [9]

$$\mathcal{L}_{ijkl} = \frac{f_b f_e}{\mathrm{tr}(\hat{\mathrm{T}}^2)}[F^2 \delta^K_{ik}\delta^K_{jl} + a^2 \hat{T}_{ij}\hat{T}_{kl}], \tag{A.1}$$

$$N_{ij} = \frac{f_b f_e f_d a F}{\mathrm{tr}(\hat{\mathrm{T}}^2)}(\hat{T}_{ij} + \hat{T}^*_{ij}), \tag{A.2}$$

where δ^K_{ij} is the Kronecker delta, and

$$\hat{T}_{ij} = \frac{T_{ij}}{\mathrm{trT}}, \qquad \hat{T}^*_{ij} = \hat{T}_{ij} - \tfrac{1}{3}\delta^K_{ij}. \tag{A.3}$$

The factor a in (A1), (A2) is determined by the friction angle φ_c in critical states:

$$a = \sqrt{\frac{3}{8}}\frac{(3 - \sin\varphi_c)}{\sin\varphi_c}. \tag{A.4}$$

The factor F is a function of $\hat{\mathbf{T}}^*$:

$$F=\sqrt{\frac{1}{8}\tan^2\xi+\frac{2-\tan^2\xi}{2+\sqrt{2}\tan\xi\cos 3\theta}}-\frac{1}{2\sqrt{2}}\tan\xi, \tag{A.5}$$

where

$$\tan\xi=\sqrt{3}\|\hat{\mathbf{T}}^*\|,\quad \cos 3\theta=-\sqrt{6}\frac{\mathrm{tr}(\hat{\mathbf{T}}^{*3})}{[\mathrm{tr}(\hat{\mathbf{T}}^{*2})]^{3/2}}. \tag{A.6}$$

The factors a and F determine the critical-state surface in the stress space.

Three characteristic void ratios are specified as functions of the mean pressure: the minimal possible void ratio, e_d, the critical void ratio, e_c, and the void ratio in the loosest state, e_i. The pressure dependence of these void ratios is postulated in the form (compressive stresses are negative)

$$\frac{e_i}{e_{i0}}=\frac{e_c}{e_{c0}}=\frac{e_d}{e_{d0}}=\exp\left[-\left(\frac{-\mathrm{tr}\mathbf{T}}{h_s}\right)^n\right], \tag{A.7}$$

with the corresponding reference values e_{i0}, e_{c0}, e_{d0} for zero pressure ($e_{i0}>e_{c0}>e_{d0}$). The constants e_{i0}, e_{c0}, e_{d0} with the h_s, n are material parameters.

The factor

$$f_d=\left(\frac{e-e_d}{e_c-e_d}\right)^\alpha, \tag{A.8}$$

where α is a material parameter, tends to unity as the state of the material approaches a critical state. The functions f_e and f_b are defined as

$$\begin{aligned} f_e&=\left(\frac{e_c}{e}\right)^\beta,\\ f_b&=\frac{h_s}{n}\left(\frac{1+e_i}{e_i}\right)\left(\frac{e_{i0}}{e_{c0}}\right)^\beta\left(\frac{-\mathrm{tr}\mathbf{T}}{h_s}\right)^{1-n}\left[3+a^2-\sqrt{3}a\left(\frac{e_{i0}-e_{d0}}{e_{c0}-e_{d0}}\right)^\alpha\right]^{-1}, \end{aligned} \tag{A.9}$$

with a parameter β.

The tensor $\mathcal{M}$ in (2) is written as

$$\begin{aligned} \mathcal{M}_{ijkl}=&[\rho^\chi m_T+(1-\rho^\chi)m_R]\mathcal{L}_{ijkl}\\ &+\begin{cases}\rho^\chi(1-m_T)\mathcal{L}_{ijqs}\hat{\delta}_{qs}\hat{\delta}_{kl}+\rho^\chi N_{ij}\hat{\delta}_{kl} & \text{if } \hat{\delta}_{ij}D_{ij}>0,\\ \rho^\chi(m_R-m_T)\mathcal{L}_{ijqs}\hat{\delta}_{qs}\hat{\delta}_{kl} & \text{if } \hat{\delta}_{ij}D_{ij}\leqslant 0,\end{cases} \end{aligned} \tag{A.10}$$

where

$$\hat{\delta}_{ij}=\begin{cases}\delta_{ij}/\|\boldsymbol{\delta}\| & \text{if } \boldsymbol{\delta}\neq\mathbf{0},\\ 0 & \text{if } \boldsymbol{\delta}=\mathbf{0},\end{cases} \tag{A.11}$$

$\rho=\|\boldsymbol{\delta}\|/R$, and R, m_R, m_T, χ are constitutive parameters.

The evolution equation (3) for the intergranular-strain tensor $\boldsymbol{\delta}$ is written as

$$\frac{\mathrm{d}\delta_{ij}}{\mathrm{d}t}=\begin{cases}D_{ij}-\rho^{\beta_r}\hat{\delta}_{ij}\hat{\delta}_{kl}D_{kl} & \text{if } \hat{\delta}_{ij}D_{ij}>0,\\ D_{ij} & \text{if } \hat{\delta}_{ij}D_{ij}\leqslant 0\end{cases} \tag{A.12}$$

with a parameter β_r.

The constitutive parameters used in the numerical examples in this paper are given in Table 2.

Table 2. Constitutive parameters.

$\varphi_c[^\circ]$	h_s[MPa]	e_{c0}	e_{d0}	e_{i0}	α	β	n	R	m_R	m_T	β_r	χ
33	1000	0·95	0·55	1·05	0·25	1·5	0·25	4×10^{-5}	5·0	5·0	0·05	1·5

Acknowledgements

The present study was initiated within the framework of SFB 298 financed by the Deutsche Forschungsgemeinschaft. The study has also been supported by the Stiftung für geotechnische Grundlagenforschung, Universität für Bodenkultur Wien.

References

1. H.S. Yu, *Cavity Expansion Methods in Geomechanics*. Dordrecht: Kluwer (2000) 404 pp.
2. A.S. Vesić, Expansion of cavities in infinite soil mass *ASCE J. Soil Mech. Found. Div.* 98 (1972) 265–290.
3. J.P. Carter, J.R. Booker and S.K. Yeung, Cavity expansion in cohesive frictional soils. *Géotechnique* 36 (1986) 349–358.
4. H.S. Yu and G.T. Houlsby, Finite cavity expansion in dilatant soils: loading analysis. *Géotechnique* 41 (1991) 173–183.
5. I. Collins, M. Pender and W. Yan, Cavity expansion in sands under drained loading conditions. *Int. J. Num. Anal. Meth. Geomech.* 16 (1992) 3–23.
6. D. Shuttle and M. Jefferies, Dimensionless and unbiased CPT interpretation in sand. *Int. J. Num. Anal. Meth. Geomech.* 22 (1998) 351–391.
7. V.A. Osinov and R. Cudmani, Theoretical investigation of the cavity expansion problem based on a hypoplasticity model. *Int. J. Num. Anal. Meth. Geomech.* 25 (2001) 473–495.
8. R.E. Brown, Vibroflotation compaction of cohesionless soils. *ASCE J. Geotech. Eng. Div.* 103 (1977) 1437–1451.
9. A. Niemunis and I. Herle, Hypoplastic model for cohesionless soils with elastic strain range. *Mech. Cohesive-frictional Mater*. 2(4) (1997) 279–299.
10. V.A. Osinov, Wave-induced liquefaction of a saturated sand layer. *Cont. Mech. Thermodyn*. 12 (2000) 325–339.
11. V.A. Osinov, A numerical model for the site response analysis and liquefaction of soil during earthquakes. In: O. Natau, E. Fecker and E. Pimentel (eds.), *Geotechnical Measurements and Modelling*. Lisse: Swets & Zeitlinger (2003) pp. 475–481.
12. V.A. Osinov and G. Gudehus. Dynamics of hypoplastic materials: theory and numerical implementation. In: K. Hutter and N. Kirchner (eds.), *Dynamic Response of Granular and Porous Materials under Large and Catastrophic Deformations*. Berlin: Springer (2003) pp. 265–284.
13. J.M. Hill, Some symmetrical cavity problems for a hypoplastic granular material. *Quart. J. Mech. Appl. Math*. 53 (2000) 111–135.
14. G. Gudehus, A comprehensive constitutive equation for granular materials. *Soils and Foundations* 36 (1996) l–12.
15. P.A. von Wolffersdorff, A hypoplastic relation for granular materials with a predefined limit state surface. *Mech. Cohesive-frictional Mater*. 1 (1996) 251–271.
16. E. Bauer and W. Wu, A hypoplastic model for granular soils under cyclic loading. In: D. Kolymbas (ed.), *Modern Approaches to Plasticity*. Amsterdam: Elsevier (1993) pp. 247–258.
17. A. Niemunis, Hypoplasticity vs. elastoplasticity, selected topics. In: D. Kolymbas (ed.), *Modern Approaches to Plasticity*. Amsterdam: Elsevier (1993) pp. 277–307.
18. W. Fellin, *Rütteldruckverdichtung als plastodynamisches Problem*. Rotterdam: Balkema (2000) 94 pp.
19. R. Cudmani and V.A. Osinov, The cavity expansion problem for the interpretation of cone penetration and pressuremeter tests. *Canadian Geot. J*. 38 (2001) 622–638.

Journal of Engineering Mathematics (2005) 52: 199–229
DOI 10.1007/s10665-004-3950-z

Generalised homogenisation procedures for granular materials

E. PASTERNAK and H.-B. MÜHLHAUS[1]
Department of Civil and Resource Engineering, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia (elena@civil.uwa.edu.au); [1]*Department of Earth Sciences, The University of Queensland, St Lucia, QLD 4072, Australia (muhlhaus@esscc.uq.edu.au)*

Received 11 December 2003; accepted in revised form 27 September 2004

Abstract. Engineering materials are generally non-homogeneous, yet standard continuum descriptions of such materials are admissible, provided that the size of the non-homogeneities is much smaller than the characteristic length of the deformation pattern. If this is not the case, either the individual non-homogeneities have to be described explicitly or the range of applicability of the continuum concept is extended by including additional variables or degrees of freedom. In the paper the discrete nature of granular materials is modelled in the simplest possible way by means of finite-difference equations. The difference equations may be homogenised in two ways: the simplest approach is to replace the finite differences by the corresponding Taylor expansions. This leads to a Cosserat continuum theory. A more sophisticated strategy is to homogenise the equations by means of a discrete Fourier transformation. The result is a Kunin-type non-local theory. In the following these theories are analysed by considering a model consisting of independent periodic 1D chains of solid spheres connected by shear translational and rotational springs. It is found that the Cosserat theory offers a healthy balance between accuracy and simplicity. Kunin's integral homogenisation theory leads to a non-local Cosserat continuum description that yields an exact solution, but does not offer any real simplification in the solution of the model equations as compared to the original discrete system. The rotational degree of freedom affects the phenomenology of wave propagation considerably. When the rotation is suppressed, only one type of wave, *viz*. a shear wave, exists. When the restriction on particle rotation is relaxed, the velocity of this wave decreases and another, high velocity wave arises.

Key words: Cosserat continuum, homogenisation, non-local continuum, rotational degrees of freedom, wave propagation

1. Introduction. Comparative analysis of non-standard continua approach to modelling materials with microstructure

1.1. Non-standard continua

There are cases when the classical continuum mechanics approach is insufficient to model adequately materials with microstructure. In particular, microstructure cannot be ignored when one considers layered material, especially when the layers can slide, blocky structures, granular or fractured media. The consideration of microstructure is necessary when the size of redistribution of an external load is comparable with the microstructure size or if the stress gradients at intergranular contact points (finite contact area, modelled as a contact point) of grains have to be taken into account.

Naturally, there are different approaches (or combinations thereof) to take microstructural effects into account. Each approach results in different types of standard or non-standard continua.

The first step in the adaptation or extension of a standard continuum theory to granular materials is the introduction of rotational degrees of freedom (DOF) in addition to the

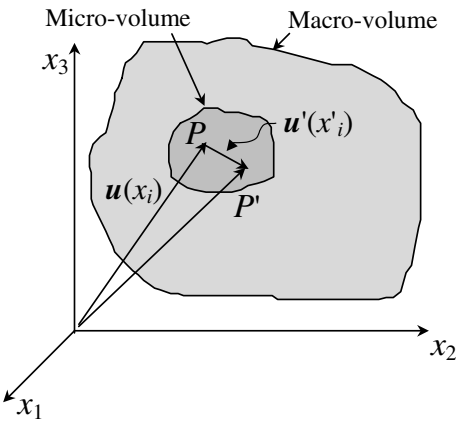


Figure 1. Macro- and micro-volumes and the relationship between the macro- and micro-displacements.

conventional translational ones (if, for instance, the spin of particles is important). This leads to gradient or higher-order gradient theories (when further degrees of freedom have to be included as well), resulting in the introduction of additional strain measures and the corresponding stress tensors.

While the need for independent rotational DOF's (Cosserat type theories, *e.g.* [1]; Nowacki [2]) is quite easy to understand in the context of granular and layered materials, further increase of DOF requires more explanations. Mindlin [3] based his reasoning on the simultaneous consideration of macro- and micro- displacements within a volume element.

In the spirit of Mindlin's discussion we consider a macro-volume, *i.e.*, the domain occupied by a body, and choose a Cartesian coordinate system $x_1x_2x_3$ (see Figure 1). Let P be an arbitrary point of a body, the position of the point in the macro-volume being determined by the macro-coordinates x_i. The macro-motion of this point can be described by the macro-displacement vector $\boldsymbol{u}(x_i)$ and macro-rotation vector $\boldsymbol{\varphi} = \frac{1}{2}\mathrm{rot}\,\boldsymbol{u}$. According to conventional continuum mechanics, the deformation measures at this point are the components of the macro-distortion tensor $u_{i,j}$, the symmetric part of which gives the usual components of a macro-strain tensor $\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$. The antisymmetric part of the macro-distortion tensor gives the macro-rotational vector $\varphi_i = -\frac{1}{2}\varepsilon_{kli}u_{k,l}$, where ε_{ijk} is the alternating tensor. It is seen that the macro-rotation vector and the distortion tensor are fully determined by the components of the displacement vector $\boldsymbol{u}(x_i)$.

Next, assume the material point P as a centroid of a micro-volume of the originally inhomogeneous medium. This volume element defines the scale of resolution of the envisaged continuum theory. Effects below this characteristic scale are ignored. This volume element could not be constricted to the point because of the microstructure of the material. A particular choice of the micro-volume size is not important here. It suffices to mention that, in general, the micro-volume size is supposed to be (a) much larger than the microstructure size to asymptotically satisfy the requirement for the micro-volume to be representative, *i.e.*, contain a sufficient number of elements of the microstructure and; (b) much smaller than the external size such as the dimension of the problem or a characteristic length of the load redistribution (*e.g.*, wavelength), to asymptotically satisfy the requirement for the micro-volume to be infinitesimal. Furthermore, its shape has to reflect the material symmetry of the material to be modelled.

We now introduce another local Cartesian coordinate system with the origin at P. An arbitrary point P' of the micro-volume has the micro-coordinates $x'_1x'_2x'_3$. The vector

connecting P and P' will be called the micro-displacement vector and denoted as $\boldsymbol{u}'(x_i')$. The micro-displacement vector characterises the displacement of the point P' within the micro-particle (micro-volume element).

The displacement of the point P' being a point of the macro-volume is given by the sum of the macro-displacement vector $\boldsymbol{u}(x_i)$ and the micro-displacement vector $\boldsymbol{u}'(x_i')$. Expanding the components of the micro-displacement vector into the Taylor series at the vicinity of the point P, one gets the corresponding coordinates of the displacement vector of the point P':

$$u_i(x) + u_i'(x') = u_i(x) + u_i'(0) + u_{i,j}'(0)x_j' + \frac{1}{2}u_{i,jk}'(0)x_j'x_k' + \cdots, \tag{1.1}$$

where $u_i'(0) = 0$, $x \stackrel{\text{def}}{=} (x_1, x_2, x_3)$, $x' \stackrel{\text{def}}{=} (x_1', x_2', x_3')$.

The underlying assumption behind Equation (1.1) is that the displacements within the representative volume are analytic and can be represented by Taylor expansions around point P. As discussed above (conditions (a) and (b)), the size of the volume element h must be much smaller than the macro-volume characteristic dimension L and much bigger than the micro-structure size a in a sense that the continuum approximation is a double asymptotic as $h/L \to 0$ and $a/h \to 0$ (*e.g.*, [4–6]). As a result of this asymptotic transition, we have a continuum that permits the usual description based on the established rules of differential geometry. The only difference from a conventional continuum is that each point may have additional DOF, namely the higher-order polynomial coefficients in (1.1) enabling the consideration of deviations of the deformation from the mean values within a representative volume element.

The term $u_{i,j}'$ provides us with nine micro-distortion components: three micro-rotations and six micro-strains. If, for simplicity, we take into account only the antisymmetric part, we arrive at a continuum with six DOF (three translational DOF, the macro-displacements u_i, and three rotational ones, microrotations φ_i). This is the Cosserat theory or the theory of micropolar elasticity (*e.g.*, [2, 7–9]). The rotational degrees of freedom are very often referred to as the Cosserat rotations giving tribute to the brothers Cosserat who were the first to propose such a theory.

Further generalisations can be obtained by including the symmetric part of the micrograd-ients into the model as well and/or by keeping the next term of the Taylor expansion $u_{i,jk}'$ bringing the total number of DOFs to 36 [3].

It should be emphasised that the micro-deformations in the expansion (1.1) are independent in general from the macroscopic deformation gradient. The relationship between the macro- and the micro-deformation is established by means of additional constitutive relationships.

The higher-order gradient theories necessitate the introduction of additional stress tensors which are conjugate to the additional deformation measures (*e.g.*, couple or moment stresses in the Cosserat type theories; double-forces tensor in the Mindlin continuum, etc.). In the elasticity theories, these new stress tensors can normally be obtained by differentiating the variation of an elastic potential (the elastic energy density) with respect to the variation of the deformation measures. The equations of equilibrium or motion also have to be obtained for additional stress factors in the higher-order gradient theories. It should also be mentioned that the formulation of boundary-value problems may be in terms of displacements complemented by the additional DOF's (for instance, rotations in the Cosserat theory) or in terms of the stress tensors complemented by the conjugates of the additional DOF's (for instance, couple/moment stresses in the Cosserat theory) or in a mixed form.

It is well known that the governing equations of the continuum have to be translationally and rotationally invariant. This requirement yields exactly 2×3 balance equations. Translational

invariance requires the consideration of linear momentum; rotational invariance requires the consideration of angular momentum. The equations of motion in the standard theories result from the momentum balance, while the moment of momentum balance gives the symmetry of the stress tensor. In the Cosserat theories both momentum balance and moment of momentum balance are used, each contributing three equations. An interesting question for higher-order theories would be where to get the additional equations of motion/equilibrium from. At present, this question is still open[1], as no other fundamental hypotheses similar to the hypotheses of the isotropy and homogeneity of space and time have been formulated yet. That is why gradient-enhanced theories become increasingly popular and used. They do not require additional motion/equilibrium equations but at the same time allow one to include the strain gradients into the formulation. This, in reality, does not add new DOF, but increases the order of the governing partial differential equations.

Mindlin [3] used Hamilton's principle for independent variations, which were his 12 DOF, and obtained 12 equations of motion from the variational equation of motion. However, this approach is applicable only in elastic materials. An application of the method of virtual power for derivation of the balance equations of micropolar and second-gradient continua was discussed in the works of Germain [10,11] and Maugin [12].

The second approach (homogenisation by integral transformation) involves the introduction of a non-local (integral) constitutive law (when the long-range interactions between the particles need to be accounted for, *e.g.,* [13, p.34]). This, in general, means that the stress components depend on the strains at all points of the continuum, albeit with weight decreasing with distance from the point of interest. In essence, this approach shifts the procedure of homogenisation from the definition of deformation measures (by introducing the volume element) to the constitutive relationships. Both approaches can be combined leading to non-local theories with additional degrees of freedom.

Non-local homogenisation strategies of the discrete materials were introduced by Kröner [14], Kröner and Datta [15], Kunin and Waisman [16], Kunin [13, Chapter 2,3], [17, Chapters 2,3] and Eringen [18] for periodical microstructures. The homogenisation was performed by trigonometric interpolation of the discrete field of displacements and rotations of the particles. In those theories the particle centres are assumed to be situated at the nodes of a regular grid. This leads to non-local stress–deformation relationships reflecting the fact that the values of the interpolating polynomial at a point are sensitive to the values at the other points. The kernels of the non-local relationship are expressed through Kunin's analog of the Dirac-delta function which "remembers" the microstructure size.

Specifically, in the case of a three-DOF continuum, this homogenisation procedure leads to a non-local continuum (with the same three DOF), the non-local stress–strain relationship and the non-local stresses satisfying the conventional equations of equilibrium or motion. In the following it will be demonstrated that, in a continuum with six DOF, the Cosserat continuum, this homogenisation strategy leads to a non-local Cosserat continuum. Subsequently, in this continuum one comes up with the non-local constitutive equations in which stresses and moment stresses are expressed through the deformation measures (strains and curvatures). Non-local stresses and moment stresses satisfy exactly the same equations of equilibrium or motion as in the case of a "local" or "conventional" Cosserat theory. We would like

[1]A possible remedy to obtain an extra set of equilibrium/motion equations in situations when the representative volume element is made up of homogeneous but materially distinct sub-volumes is to divide the volume element into smaller ones and write down the corresponding equations of motion/equilibrium for each of them. At least in this way the lever-arm (the subdivision size) can be brought into the formulation in order to get the third-order double-forces tensor with the balance equation for its components.

to mention here that Kunin [17, Section 3.5] also considered a quasi-six-DOF continuum, in which the three rotational DOF's are equal to the three rotations of the standard continuum. In Section 2 we represent a generalisation of Kunin's method to the case of independent rotations.

The major difference between local and non-local continua is that in the latter the stresses and couple or moment stresses become pseudo-stresses, as they no longer refer to an elementary area, but to a finite volume. This obviously contradicts the Cauchy-Euler principle. The question arises why the equations of equilibrium should necessarily hold in their "conventional", local form when the interaction between the parts of the body is not of the surface nature; it is transmitted not only through the surface, but also through the volume.

For the considered periodic case, a justification of the conventional form of the equations of equilibrium/motion will be offered when a non-local orthotropic Cosserat continuum model will be constructed for non-interacting identical chains of granulates. Starting from a discrete medium consisting of, for example, particles having translational and rotational DOF's, it is possible to obtain the Lagrange equation of motion. It turns out that the Lagrange equations are formally identical to the equations of a local Cosserat continuum with non-local constitutive relations. This finding supports the adoption of the "usual" local conservation laws in connection with non-local continua.

In general, the choice of the kernels is based on purely mechanistic or phenomenological considerations (*e.g.,* [19,20]), material symmetry combined with a choice of the size of the domain of influence. Unfortunately, and in particular in 3D, these requirements do not constrain the possible variety of kernel forms significantly. This "inconvenience" is rather difficult to overcome, as the kernel is a function in principle of all variables of the continuum model. This poses considerable difficulties in determining possible kernel forms from experiments.

The phenomenological approach does not address the question of the validity of local conservation laws and the physical significance of Cauchy stresses in the presence of non-local constitutive laws. These issues can only be answered and follow naturally if the model equations are derived from a micro-mechanical model by means of a suitable homogenisation procedure.

Alternatively, non-local operators on strains are used in their own right as substitutes for local strains in the damage loading function [21] leading to non-local damage models (*e.g.,* [22–24]).

Finally we mention a selection of developments (by no means complete) we consider relevant to the topic of this paper: macroscopic modelling of layered materials was conducted by Mühlhaus [25], Zvolinskii and Shkhinek [26], Adhikary and Dyskin [27]. In this 2D modelling the role of the Cosserat rotation (only one rotation in this case) was played by the rotation of the neutral axis of the layer (the deflection gradient), while the moment stresses were the bending moment per unit area in the cross-sections of the layer. Mühlhaus [28] and Mühlhaus and Hornby [29,30] modified the model of layered materials by introducing a different rotation measure, which is the relative deformation gradient. Mühlhaus [25] and Sulem and Mühlhaus [31] developed a model for a blocky material (the material broken into rectangular blocks). Cosserat-type theories for random packing of granulates were also developed, for instance, by Mühlhaus and Vardoulakis [32], Mühlhaus *et al.* [33], Chang and Ma [34]. A combination of a Cosserat continuum and a higher-order gradient continuum for granular materials has been derived by Mühlhaus and Oka [35] and Mühlhaus and Hornby [36].

Additional DOF's also appear in a new numerical fracture-mechanics method [37]. In this method the additional DOF's are associated with the distribution of displacement discontinuity introduced to model crack evolution.

1.2. Homogenisation methods

Homogenisation, as the main method of constructing continuum descriptions of a discrete material, has always been the cornerstone of continuum mechanics. It was the main simplifying factor and for decades provided a powerful means to model solid bodies as a mathematical continuum rather than a collection of elements the solids are composed of. The first and simplest approach was not to consider the scale of the microstructure at all, but to smear it, so to speak. The introduction of the concept of the representative volume element served this purpose perfectly. Many significant problems were and are formulated and solved within the framework of classical elasticity and plasticity theories. However, restrictions of smearing the microstructure were felt mainly where the microstructure was essential to model and thus could not be ignored. As a result, more sophisticated continuum theories became high in demand around the late 1980s and so were adequate homogenisation strategies. The following is a sampling of the most important homogenisation procedures:

1. *Averaging over volume element adopted in the theory of effective characteristics* (*e.g.*, [38–42], [43, Part I], [5], [44, Chapter 3], [45]). Homogenisation produces a set of macroscopic elastic moduli based on the properties of the microstructure. These microstructural constants are called effective characteristics, from which the name of the method derives.
2. *Homogenisation method applied to materials with randomly varying elastic properties* based on averaging over realisations (ensemble averaging), ([46,47]).
3. *A group of methods based on modelling of periodically regular microstructure* such as a periodical system of defects/inclusions, layers, regular granular packing (*e.g.*, [48–50]). This group of methods exploits the fact that the microstructure is positioned in periodical cells and the problem is solved for a representative defect. Then by using the periodicity, one seeks the solution for the whole domain by looking for suitable periodic functions accompanied by corresponding periodic boundary conditions at the cell boundaries (*e.g.*, [51, Chapter 1–4, 6, 7], [43], [45]).
4. *Homogenisation by integral transformations* ([13, Chapter 2, 3], [17, Chapter 2, 3]). This is a special method of homogenisation applied to periodical structures only. The method is based on trigonometrical interpolation. The discrete medium is replaced with a continuous one such that the continuous values of field variables coincide with the discrete ones at the nodes and give some values in between by using the above trigonometrical interpolation.
5. *Homogenisation by differential expansions* (*e.g.*, [35], [52, Chapter 4], [53–55]). The method is based on expanding the field variables into a Taylor series once a strategy to relate the discrete variables to continuum variables is established.

1.3. Continuum modelling of granular materials

Depending on the packing density, granular materials can behave like solids or like fluids. If the packing is dense, granular material behaves solid-like. Large (finite) deformations are not ruled out. For loose packings, granular material may behave fluid-like. Here we concentrate on the first type. For a discussion of the fluid-like regime see, *e.g.*, [36]. We concentrate on elastic models for simplicity, since the emphasis is on the homogenisation procedure and not on constitutive details.

Discrete and continuum models represent two main streams of modelling of mechanical behaviour of granular materials. In the former approach granular materials are modelled as a discrete system using, for example, the discrete-element method. Many models are developed in the framework of this approach (*e.g.*, [56–59]). Equations of motion are solved for each particle in the assembly: the particle is subjected to momentum or even mass transfer from neighbouring particles. The number of equations to solve is naturally quite large requiring powerful computers. On top of that, direct computer simulations would require detailed information (geometrical and mechanical) of all grains in the assembly. This kind of information is usually hard to obtain or not available. In other words, the computational accuracy achievable in principle by this method much exceeds the accuracy of the input data available – hardly an efficient way of modelling.

In the continuum approach the equations of motion are derived for a volume element; governing equations describing constitutive behaviour are formulated by using the continuum stress–strain concepts. Continuum models can be classified as phenomenological and microstructural. Phenomenological modelling (*e.g.*, [60]) is based on postulating the constitutive equations, which necessitates a considerable amount of testing and model calibration.

Microstructural continuum modelling was extensively developed over the past few years as an alternative or a strategy to provide constraints for phenomenological constitutive models. The benefit of the microstructural approach is that it results in rational estimates of the model parameters.

For applications of the microstructural approach see the following papers [48], [49], [61–65], [35]. The principles of microstructural modelling have been revisited recently by Suiker *et al*. [53–55] and Cambou *et al*. [66].

The first simple micro-polar (Cosserat) type theories for random packing granulates were developed by Mühlhaus and Vardoulakis [32], Mühlhaus *et al*. [33], Chang and Ma [34]. Both stresses and moment stresses were introduced, but the contact particle interaction was less sophisticated than in the more recently developed theories [67]. For example, moment stresses were attributed to the tangential component of the contact force and only resistance to the relative particle displacements was introduced into the contact relations (*e.g.*, [33]).

Further development of pure Cosserat-type theories for randomly packed assemblies went in the direction of more sophisticated and refined particle-interaction modelling. This includes both contact force and moment exerted onto a reference particle, as well as the resistance of the particles to both their relative displacements and relative rotations at the contact points. These refinements were implemented by Pasternak and Mühlhaus [68–70]. Orthotropic Cosserat and non-local Cosserat continuum models for non-interacting chains of granulates were developed by Mühlhaus *et al*. [6], Pasternak and Mühlhaus [71, 72]; see details in Section 2.

The wide variety of approaches discussed above suggests that a comparative study of the different methodologies may be useful in order to assess the quality of the approximation of the different homogenisation methods.

For simplicity we consider a model example of microstructure: 1D non-interacting chains. The structural bonds determine the material behaviour only in one direction. This case of non-interacting chains of spheres is artificial and cannot be seriously thought of as an example of a granular material. It is selected to serve a special purpose – to have a structure for which the exact solution would not be very difficult to find and to provide a testing ground for comparison of homogenisation methods. We will then proceed to investigate wave propagation in granular materials with internal rotational degrees of freedom.

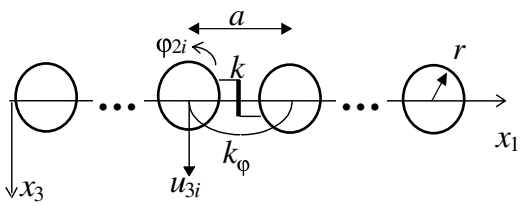


Figure 2. 1D chain of spherical grains connected by translational (shear) and rotational springs.

2. Homogenisation of 1D structures

In the following we demonstrate how additional rotational degrees of freedom arise naturally by mathematical homogenisation of a discrete system. For this purpose, a simple periodic discrete model of spheres connected to each other by both rotational and translational springs will be considered. This model allows the analytical derivation of a general closed-form solution.

The aim of this section is to investigate two homogenisation techniques. One of them produces an anisotropic Cosserat continuum and the other a non-local Cosserat continuum. A boundary-value problem of a vertical duct will be considered in order to establish how accurately the Cosserat and non-local models describe the behaviour of the granular materials.

2.1. Model formulation

In many cases, generalised continuum theories provide a convenient framework for the approximate representation of an originally discrete model. In order to analyse different homogenisation techniques, we consider a simple material consisting of one-dimensional, parallel chains of identical, spherical grains. We suppose that the chains do not interact with each other. The grains in a chain are connected by translational shear springs of stiffness k and rotational springs of stiffness k_φ (Figure 2), r is the sphere radius. The grains in neighbouring chains are not connected and move independently.

The potential (elastic) energy of a single chain in the system reads

$$U_1 = \frac{1}{2}k\sum_i((u_{3i}-u_{3i-1})+(a/2)(\varphi_{2i}+\varphi_{2i-1}))^2+\frac{1}{2}k_\varphi\sum_i(\varphi_{2i}-\varphi_{2i-1})^2 \tag{2.1}$$

with the potential-energy density referred to the ith sphere being:

$$W_i = \left(2\eta a^3\right)^{-1}\left\{k\left((u_{3i}-u_{3i-1})+(a/2)(\varphi_{2i}+\varphi_{2i-1})\right)^2+k_\varphi(\varphi_{2i}-\varphi_{2i-1})^2\right\}. \tag{2.2}$$

Here a designates the spacing of the mass centres of neighbouring spheres, and $a^{-2}\eta^{-1}$ is the number of chains per unit area of cross-section.

We note here that the rotational springs are important in this arrangement, since neglecting the resistance to rotation ($k_\varphi \to 0$) makes the energy lose its positive definiteness.

The equations of motion for the spheres are:

$$m\ddot{u}_{3i}-k(u_{3i+1}-2u_{3i}+u_{3i-1})-k(a/2)(\varphi_{2i+1}-\varphi_{2i-1})=q_{3i}, \tag{2.3$_1$}$$

$$J\ddot{\varphi}_{2i}+k(a/2)(u_{3i+1}-u_{3i-1})+k(a^2/4)(\varphi_{2i+1}+2\varphi_{2i}+\varphi_{2i-1})-k_\varphi(\varphi_{2i+1}-2\varphi_{2i}+\varphi_{2i-1}) = M_{2i}, \tag{2.3$_2$}$$

where u_{3i}, φ_{2i} are the independent Lagrange coordinates, q_{3i} and M_{2i} are applied load and moment at ith sphere respectively, $J=2mr^2/5$ is the moment of inertia of the sphere.

2.2. Homogenisation by differential expansion (Cosserat continuum)

We replace the finite-difference expressions in (2.2) with corresponding differential expressions. Truncation of the Taylor expansions in a after the second-order terms gives the following approximation

$$W(x_1) = \left(2\eta a^3\right)^{-1} \left\{ k \left(\frac{\partial u_3}{\partial x_1}\right)^2 a^2 + 2ka \frac{\partial u_3}{\partial x_1} \varphi_2 a + ka^2 \varphi_2^2 + k_\varphi \left(\frac{\partial \varphi_2}{\partial x_1}\right)^2 a^2 \right\}. \tag{2.4}$$

Differentiation of the energy density with respect to the Cosserat deformation measures, *viz*. strains and curvatures,

$$\gamma_{13} = \frac{\partial u_3}{\partial x_1} + \varphi_2, \qquad \kappa_{12} = \frac{\partial \varphi_2}{\partial x_1}, \tag{2.5}$$

gives

$$\sigma_{13} = k\,(\eta a)^{-1}\,\gamma_{13}, \qquad \mu_{12} = k_\varphi\,(\eta a)^{-1}\,\kappa_{12}. \tag{2.6}$$

Introduction of body force f_3 and moment m_2 and consideration of momentum and angular momentum equilibrium yield

$$\frac{\partial \sigma_{13}}{\partial x_1} + \rho f_3 = \rho \ddot{u}_3, \qquad \frac{\partial \mu_{12}}{\partial x_1} - \sigma_{13} + \rho m_2 = \tilde{J} \ddot{\varphi}_2, \tag{2.7}$$

where $\rho = \frac{m}{a^3 \eta}$ is the density, $\tilde{J} = \frac{J}{a^3 \eta}$ is rotational inertia per unit volume or density of rotational inertia.

Equations (2.5–2.7) formally represent a 1D Cosserat continuum (*e.g.*, [2]). Every point of this continuum has two degrees of freedom: u_3 represents the displacement, while φ_2 is being identified as the independent rotational degree of freedom, the Cosserat rotation. Mechanically speaking, the obtained continuum equations describe a 3D orthotropic Cosserat continuum (all other components of stress and moment stress tensors and corresponding deformation measures are zero). Formally, after suitable reinterpretation of the model parameters[2], one obtains the governing equations of a Timoshenko beam (*e.g.*, [28,73, p. 183]). In this case φ_2 represents the rotation of the beam cross-section and u_3 is the displacement of its neutral fibre.

After substituting the constitutive equations (2.6) in the equations of motion (2.7), we obtain the Cosserat equations of motion (the Cosserat equivalent of Lamé equations):

$$\frac{k}{\eta a} \left(\frac{\partial^2 u_3}{\partial x_1^2} + \frac{\partial \varphi_2}{\partial x_1} \right) + \rho f_3 = \rho \ddot{u}_3, \tag{2.8$_1$}$$

$$\frac{1}{\eta a} \left(k_\varphi \frac{\partial^2 \varphi_2}{\partial x_1^2} - k \frac{\partial u_3}{\partial x_1} - k\varphi_2 \right) + \rho m_2 = \frac{J}{a^3 \eta} \ddot{\varphi}_2. \tag{2.8$_2$}$$

Next we homogenise the discrete equations of motion (2.3$_1$–2.3$_2$) and compare the result with the obtained Cosserat equations of motion (2.8$_1$–2.8$_2$). For the derivation of the continuum version of equations (2.3$_1$–2.3$_2$) we first replace the discrete coordinate by a continuous coordinate, *i.e.*, $ai \to x$ as outlined for example by Mühlhaus and Oka [35] $u_{3i} \mapsto u_3(x_1)$,

$$u_{3i} \mapsto u_3(x_1),\ \ddot{u}_{3i} \mapsto \ddot{u}_3(x_1),\ \varphi_{2i} \mapsto \varphi_2(x_1),\ \ddot{\varphi}_{2i} \mapsto \ddot{\varphi}_2(x_1),\ q_{3i} \mapsto q_3(x_1),\ M_{2i} \mapsto M_2(x_1),$$
$$u_{3i\pm1} \mapsto u_3(x_1 \pm a),\ u_{3i\pm1} \mapsto u_3(x_1 \pm a),\ \varphi_{3i\pm1} \mapsto \varphi_3(x_1 \pm a).$$

[2] k_φ / a is interpreted as the bending stiffness, k/a as the shear modulus.

Equations (2.3_1–2.3_2) can formally be written in a homogenised (continuous) form by introducing continuous functions $u_3(x)$, $\varphi_2(x)$ which coincide with u_{3i} and $\varphi_{2i}(x)$ at discrete points $x=ai$ and assume some values in between:

$$m\ddot{u}_3(x_1)-k[u_3(x_1+a)-2u_3(x_1)+u_3(x_1-a)]-k(a/2)[\varphi_2(x_1+a)-\varphi_2(x_1-a)]=q_3(x_1), \tag{2.9_1}$$

$$J\ddot{\varphi}_2(x_1)+k(a/2)[u_3(x_1+a)-u_3(x_1-a)]+k(a^2/4)[\varphi_2(x_1+a)+2\varphi_2(x_1)+\varphi_2(x_1-a)]$$
$$-k_\varphi[\varphi_2(x_1+a)-2\varphi_2(x_1)+\varphi_2(x_1-a)]=M_2(x_1). \tag{2.9_2}$$

These formal equations will form the starting point for the homogenisation below.

Following Mühlhaus and Oka [35], Equations (2.9_1–2.9_2) can be homogenised by replacing the finite differences with the Taylor-series expansions, keeping terms of the second order in a and normalising the obtained equations by the volume (*i.e.*, dividing by $a^3\eta$). One eventually has

$$\frac{k}{\eta a}[u_3''(x_1)+\varphi_2'(x_1)]+O(1)+\frac{q_3(x_1)}{a^3\eta}=\rho\ddot{u}_3(x_1), \tag{2.10_1}$$

$$\frac{1}{\eta a}\left[k_\varphi\varphi_2''(x_1)-ku_3'(x_1)-k\varphi_2(x_1)\right]+O(1)+\frac{M_2(x_1)}{a^3\eta}=\frac{J}{a^3\eta}\ddot{\varphi}_2(x_1), \tag{2.10_2}$$

where $\rho=m/(a^3\eta)$.

Comparison of (2.8_1–2.8_2) with (2.10_1–2.10_2) leads to the conclusion that the equations of motion in the Cosserat approximation (2.8_1–2.8_2) give the same leading terms as the approximation of the discrete (exact) equations of motion (2.3_1–2.3_2). Hence, the Cosserat theory gives exact terms up to the first order in a inclusive. One could anticipate that the terms of order higher than a would be captured by higher-order theories. The resolution of the theory is a, *i.e.*, all the "events" smaller than a are not seen (recognised) by the Cosserat continuum, which is natural, since the a is a length scale or microstructural parameter of this Cosserat theory.

This Cosserat theory has also another length-scale parameter. The second parameter is given by the square root of the ratio of the stiffness of the rotational spring k_φ to the translational spring stiffness k and has the dimension of length.

Note that the limit $a\to 0$ in both (2.8_1–2.8_2) and (2.10_1–2.10_2) should be understood as $a/L\to 0$, where L is the size of redistribution of the load (*i.e.*, an external size). In other words, in the above calculations the normalisation in which $L=1$ is presumed.

It follows from the above analysis that the Cosserat equations of motion through displacements and rotations (a kind of "Lamé equations" for the Cosserat continuum) can be obtained either by the direct homogenisation of the discrete equations of motion or by homogenisation by differential expansions of the potential-energy density of the discrete system, provided that the same order of approximation is maintained in both cases.

The outlined strategy of homogenisation by differential expansions allows one to formulate the appropriate continuum description of the discrete system. The homogenised potential-energy density of the mechanical system has the meaning of the elastic energy density in the continuum sense with subsequent introduction of the deformation measures. Once the constitutive equations have been recovered, the Cosserat "Lamé equations" are obtained in a usual fashion by substituting the governing equations in the motion equation. Thus, the boundary-value problem can be formulated accompanied by boundary conditions. This approach can be virtually adjusted to any microstructural-particles arrangements.

2.3. Homogenisation by integral transformation (non-local Cosserat continuum)

Here we consider another homogenisation strategy, namely homogenisation by integral transformations. Kunin's [13, pp. 13–14] homogenisation procedure for discrete periodical structures is based on the trigonometrical interpolation of discrete functions. For the material consisting of independent periodical chains of grains we have:

$$\begin{pmatrix} u_3(x_1) \\ \varphi_2(x_1) \end{pmatrix} = a \sum_i \begin{pmatrix} u_{3i} \\ \varphi_{2i} \end{pmatrix} \delta_K(x_1 - ia), \quad \begin{pmatrix} u_{3i} \\ \varphi_{2i} \end{pmatrix} = \int \delta_K(x_1 - ia) \begin{pmatrix} u_3(x_1) \\ \varphi_2(x_1) \end{pmatrix} \mathrm{d}x_1, \quad \delta_K(x) = (\pi x)^{-1} \sin \frac{\pi x}{a}. \tag{2.11}$$

The application of (2.11_1) to the discrete equations of motion (2.3) yields the non-local equations of motion (non-local Lamé equations):

$$\begin{aligned} & k \int_{-\infty}^{+\infty} [2\delta_K(x_1 - y_1) - \delta_K(x_1 - y_1 - a) - \delta_K(x_1 - y_1 + a)] u_3(y_1) \mathrm{d}y_1 + \\ & \quad + k(a/2) \int_{-\infty}^{+\infty} [\delta_K(x_1 - y_1 - a) - \delta_K(x_1 - y_1 + a)] \varphi_2(y_1) \mathrm{d}y_1 = q_3(x_1) - m\ddot{u}_3(x_1), \end{aligned} \tag{2.12$_1$}$$

$$\begin{aligned} & k(a/2) \int_{-\infty}^{+\infty} [\delta_K(x_1 - y_1 + a) - \delta_K(x_1 - y_1 - a)] u_3(y_1) \mathrm{d}y_1 + \\ & \quad + k(a^2/4) \int_{-\infty}^{+\infty} [\delta_K(x_1 - y_1 + a) + 2\delta_K(x_1 - y_1) + \delta_K(x_1 - y_1 - a)] \varphi_2(y_1) \mathrm{d}y_1 + \\ & \quad + k_\varphi \int_{-\infty}^{+\infty} [2\delta_K(x_1 - y_1) - \delta_K(x_1 - y_1 + a) - \delta_K(x_1 - y_1 - a)] \varphi_2(y_1) \mathrm{d}y_1 = M_2(x_1) - J\ddot{\varphi}_2(x_1). \end{aligned} \tag{2.12$_2$}$$

These equations are essentially a representation of (2.9).

In order to obtain the constitutive relationship, we consider the potential energy of a chain (2.1), since the structure we are studying is essentially one-dimensional. Inserting (2.11_1) into (2.1), integrating the result by parts and assuming that the functions u and φ and their derivatives decay strongly enough at infinity to make the non-integral terms zero, we obtain the following representation of the elastic energy of the chain:

$$\begin{aligned} U_1 = {} & \frac{k}{2a} \int\int [2C(x_1 - y_1) - C(x_1 - y_1 - a) - C(x_1 - y_1 + a)] \gamma_{13}(x_1) \gamma_{13}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 + \\ & + \frac{k}{a} \int\int [2K(x_1 - y_1) - K(x_1 - y_1 - a) - K(x_1 - y_1 + a)] \kappa_{12}(x_1) \gamma_{13}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 + \\ & + \frac{k}{2a} \int\int [2K_1(x_1 - y_1) - K_1(x_1 - y_1 - a) - K_1(x_1 - y_1 + a)] \kappa_{12}(x_1) \kappa_{12}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 + \\ & + \frac{k}{2} \int\int [C(x_1 - y_1 - a) - C(x_1 - y_1 + a)] \gamma_{13}(x_1) \kappa_{12}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 + \\ & + \frac{k}{2} \int\int [K(x_1 - y_1 - a) - K(x_1 - y_1 + a)] \kappa_{12}(x_1) \kappa_{12}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 + \\ & + \frac{ka}{8} \int\int [2C(x_1 - y_1) + C(x_1 - y_1 - a) + C(x_1 - y_1 + a)] \kappa_{12}(x_1) \kappa_{12}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 + \\ & + \frac{k_\varphi}{2a} \int\int [2C(x_1 - y_1) - C(x_1 - y_1 - a) - C(x_1 - y_1 + a)] \kappa_{12}(x_1) \kappa_{12}(y_1) \mathrm{d}x_1 \mathrm{d}y_1 \end{aligned} \tag{2.13}$$

where $C''(x) = -\delta_K(x)$, $K'(x) = C(x)$, $K_1''(x) = -C(x)$ and

$$\gamma_{13} = \frac{\partial u_3}{\partial x_1} + \varphi_2, \qquad \kappa_{12} = \frac{\partial \varphi_2}{\partial x_1} \tag{2.14}$$

are components of the Cosserat relative deformation gradient and the curvature tensor. The deformation measures (2.14) are invariant with respect to rigid-body motions: if we consider the rigid-body translation $u_3 = \text{const}$, we have $\varphi_2 = 0$, $\gamma_{13} = 0$, $\kappa_{12} = 0$; similarly for the rigid-body rotation $u_3 = -x_1\varphi_2$, $\varphi_2 = \text{const}$, we find $\gamma_{13} = \partial u_3/\partial x_1 + \varphi_2 = -\varphi_2 + \varphi_2 = 0$, $\kappa_{12} = 0$.

Consider the energy of the whole body $U = N_1 N_2 U_1$, where N_1, N_2 are numbers of chains in the directions x_2 and x_3, respectively. Variation of the energy is

$$\delta U = \int_V \delta W(x_1, x_2, x_3)\mathrm{d}v = \eta a^2 N_1 N_2 \int_{-\infty}^{+\infty} \delta W(x_1)\mathrm{d}x_1 = \eta a^2 N_1 N_2 \int_{-\infty}^{+\infty} (\sigma\,\delta\varepsilon + \mu\,\delta\kappa)\mathrm{d}x_1, \tag{2.15}$$

where $W(x_1, x_2, x_3)$ is the specific potential energy at point $x = (x_1, x_2, x_3)$, and $\eta^{1/2}a$ is the spacing between non-interacting chains of the spatial assembly. Since $\delta U = N_1 N_2 \delta U_1$, the variation of $U_1 = U_1(\gamma_{13}, \kappa_{12})$ in (2.13), with subsequent extraction of δW in (2.15) and the introduction of the stress and moment stress

$$\sigma_{13}(x) = \frac{\delta W}{\delta\gamma_{13}(x)}, \qquad \mu_{12}(x) = \frac{\delta W}{\delta\kappa_{12}(x)}, \tag{2.16}$$

yield the following expressions for the stress and the moment stress:

$$\begin{aligned}\sigma_{13}(x_1) = \left(\sqrt{\eta}a\right)^{-2} \Bigg\{ & E\int_{-\infty}^{+\infty} [2C(x_1 - y_1) - C(x_1 - y_1 - a) - C(x_1 - y_1 + a)]\gamma_{13}(y_1)\mathrm{d}y_1 - \\ & - E\int_{-\infty}^{+\infty} [2K(x_1 - y_1) - K(x_1 - y_1 - a) - K(x_1 - y_1 + a)]\kappa_{12}(y_1)\mathrm{d}y_1 + \\ & + E(a/2)\int_{-\infty}^{+\infty} [C(x_1 - y_1 - a) - C(x_1 - y_1 + a)]\kappa_{12}(y_1)\mathrm{d}y_1 \Bigg\}, \quad E = k/a,\end{aligned} \tag{2.17}$$

$$\begin{aligned}\mu_{12}(x_1) = \left(\sqrt{\eta}a\right)^{-2} \Bigg\{ & E(a/2)\int_{-\infty}^{+\infty} [C(x_1 - y_1 + a) - C(x_1 - y_1 - a)]\gamma_{13}(y_1)\mathrm{d}y_1 + \\ & + E(a^2/4)\int_{-\infty}^{+\infty} [2C(x_1 - y_1) + \\ & + C(x_1 - y_1 - a) + C(x_1 - y_1 + a)]\Omega_{12}(y_1)\mathrm{d}y_1 + \\ & + E_\varphi\int_{-\infty}^{+\infty} [2C(x_1 - y_1) - \\ & - C(x_1 - y_1 - a) - C(x_1 - y_1 + a)]\Omega_{12}(y_1)\mathrm{d}y_1 + \\ & + E\int_{-\infty}^{+\infty} [2K(x_1 - y_1) - K(x_1 - y_1 + a) - \\ & - K(x_1 - y_1 - a)]\gamma_{13}(y_1)\mathrm{d}y_1 + \\ & + E\int_{-\infty}^{+\infty} [2K_1(x_1 - y_1) - K_1(x_1 - y_1 + a) - \\ & - K_1(x_1 - y_1 - a)]\Omega_{12}(y_1)\mathrm{d}y_1 + \\ & + Ea\int_{-\infty}^{+\infty} [K(x_1 - y_1 - a) - \\ & - K(x_1 - y_1 + a)]\Omega_{12}(y_1)\mathrm{d}y_1 \Bigg\}, \quad E_\varphi = k_\varphi/a.\end{aligned} \tag{2.18}$$

It is seen that homogenisation by integral transformations produces non-local constitutive relationships with oscillating kernels. The origin of this particular type of non-locality is in the fact that the interpolation function for a given set of u_{3i}, φ_{2i} is unique, and hence the alteration of any local value leads to the change of the whole function.

Integrating the non-local "Lamé equations" (2.12) by parts, extracting the expressions (2.17) and (2.18) and accounting for volume forces and moments yields the following Euler–Lagrange equations:

$$\frac{\mathrm{d}\sigma_{13}(x_1)}{\mathrm{d}x_1}+\frac{q_3(x_1)}{a^3\eta}=\frac{m}{a^3\eta}\ddot{u}_3(x_1), \qquad \frac{\mathrm{d}\mu_{12}(x_1)}{\mathrm{d}x_1}-\sigma_{13}(x_1)+\frac{M_2(x_1)}{a^3\eta}=\frac{J}{a^3\eta}\ddot{\varphi}_2(x_1). \tag{2.19}$$

The form of the angular momentum balance (2.19_2) is standard and consistent with its Cosserat counterpart (*cf.* (2.7_2)). Stresses σ_{13}, μ_{12} are interpreted conventionally. However, the constitutive relationships are non-local, *i.e.*, determined by the deformations of all parts of a chain. There could be another view on the non-local stresses and moment stresses. Since they are no longer referred to the elementary area, they are supposed to act on in the conventional Cauchy sense; the continuum obtained may be regarded as a pseudo-one. Nevertheless, we prefer the term 'non-local Cosserat continuum'.

In essence, Equations (2.14) and (2.17–2.19) constitute mathematically a 1D non-local Cosserat continuum. Every point of this continuum has two degrees of freedom, the displacement u_3 and the Cosserat rotation φ_2. Mechanically speaking, the obtained continuum equations describe a 3D non-local orthotropic Cosserat continuum (all other components of stress and moment stress tensors and corresponding deformation measures are zero).

Obviously, the non-local "Lamé equations" (2.12_1–2.12_2) can be recovered if one substitutes the non-local constitutive relationships (2.17) and (2.18) in the equations of motion (2.19). This gives a displacement–rotation formulation.

2.4. Invertibility of kernels in non-local constitutive equations

Here we will discuss the issue of invertibility of the kernels in non-local constitutive equations (2.17) and (2.18). One can expect that they have an inverse, since their generating function, the Kunin-delta, has the inverse in contrast to the non-invertible bell-shaped Gaussian kernels often used nowadays (*e.g.* [20]).

The discrete system with $6N$ degrees of freedom, N being the number of the particles in it, is defined by the $6N$ Lagrange coordinates, displacements and rotations, u_i, φ_i. If the forces and moments q_i, M_i are applied to the mechanical system, the one-to-one correspondence between the loads applied to the mechanical system and displacements and rotations is established by the discrete (Lagrange) equations of motion (2.3). Let $\boldsymbol{A}$ be this one-to-one mapping:

$$A\begin{bmatrix}\vec{u}\\ \vec{\varphi}\end{bmatrix}=\begin{bmatrix}\vec{q}\\ \vec{M}\end{bmatrix}. \tag{2.20}$$

If the system of the discrete equations of motion can be solved for displacements and rotations of the particles, then it is given by the inverse mapping $\boldsymbol{A}^{-1}$.

The application of the homogenisation by integral transformations (Kunin's homogenisation) (2.11) gives the homogenised continuum variables $u(x)$, $\varphi(x)$, $q(x)$, $M(x)$, governed by the homogenised equations of motion (2.3) or formally (2.20), for example in the form (2.9) or (2.12). The spaces of the discrete and continuum representations are isomorphic; see Kunin [13, p. 14]. Let $\boldsymbol{H}_1$ be this isomorphism, *i.e.*, $\boldsymbol{H}_1$ establishes the isomorphism between the discrete Lagrange coordinates and their continuum counterparts, between the discrete and non-local equations of motion. Following a procedure similar to that outlined in the previous section, a non-local constitutive relationship can be obtained (*e.g.*, (2.17) and (2.18)). We

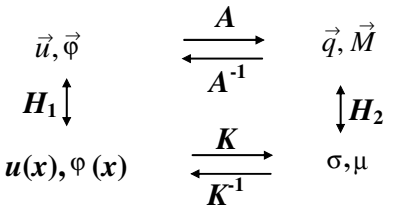


Figure 3. Isomorphism of spaces.

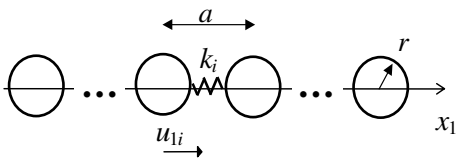


Figure 4. 1D chain of spherical grains connected by translational (normal) springs.

formally write it down as follows

$$\begin{bmatrix} \sigma_{13} \\ \mu_{12} \end{bmatrix} = K \begin{bmatrix} \gamma_{13} \\ \kappa_{12} \end{bmatrix}, \tag{2.21}$$

where $\boldsymbol{K}$ is non-local operator acting on the strains and curvatures and producing non-local stresses and moment stresses.

The space of the discrete loads applied to the system is isomorphic to the space of the non-local stresses and moment stresses, and this isomorphism $\boldsymbol{H}_2$ is established by the equations of equilibrium (2.19) modulo the two free constants determined by the boundary conditions. This is illustrated by the diagram of Figure 3. Thus we conclude that if $\exists \boldsymbol{A}^{-1}$ then $\exists \boldsymbol{K}^{-1}$, *i.e.*, $\boldsymbol{K}$ is invertible. This means that kernels in non-local constitutive equations have their inverses if the initial discrete mechanical system is solvable for displacements and rotations under prescribed loads.

It is worth noting that since $\boldsymbol{H}_1$ establishes the isomorphism between the discrete Lagrange coordinates and their continuum counterparts, *i.e.*, between the discrete and non-local equations of motion, the homogenised equations (2.3) or (2.12) are exact. This means that the solution of a boundary-value problem under this non-local formulation must be exact or, in other words, must coincide with the discrete one.

2.5. Random kernels

As we established above in the non-local relationships (2.17) and (2.18), some of the kernels exhibit an oscillating behaviour. An interesting question is now whether the oscillating behaviour of the kernels in the non-local relationships (2.17) and (2.18) will disappear if some form of randomness is introduced into the mechanical system. In other words, can the randomness help to cure the oscillating nature of the kernels in the non-local constitutive equations, *i.e.*, can we get a nice bell-shape form of the kernels in this case, as presumed *e.g.* by Eringen [20]. We try to find what non-local kernels look like for the irregular arrangements of the balls in chains. Since all the kernels were generated by the Kunin-delta, it would suffice to consider a less sophisticated model in which the nature of the problem is preserved, but the number of kernels would reduce significantly.

In view of this, we consider a simple material consisting of one-dimensional, parallel, non-interacting chains of identical, spherical grains as before, but now the grains in a chain are connected by translational normal springs of stiffness k_i only; r is the sphere radius and a is the inter-ball distance as before (Figure 4). To be able to apply the homogenisation by integral transformation the inter-ball spacing is presumed constant. We assume that the stiffnesses k_i are independent random variables, normally distributed with the mathematical expectations $E(k_i) = k$ and the variance $\mathrm{Var}(k_i) = s^2$. This way the irregularity of the system is achieved. It

will be sufficient to consider only one chain, as our aim is restricted to investigating the form of the kernel, rather than building a proper continuum for this model.

The potential (elastic) energy of a single chain in the system reads

$$U_1 = \frac{1}{2}\sum_i k_i (u_i - u_{i-1})^2. \tag{2.22}$$

Insertion of (2.11_1) into (2.22) yields the following homogenisation for the potential energy:

$$U_1 = \frac{1}{2}\int\int \hat{\Phi}(x, y)[u(x) - u(x-a)][u(y) - u(y-a)]\mathrm{d}x\mathrm{d}y, \tag{2.23}$$

$$\hat{\Phi}(x, y) = \sum_i k_i \delta_K(x - ia)\delta_K(y - ia). \tag{2.24}$$

In the series (2.24) k_i are independent normal variables and $\sum_i [\delta_K(x-ia)\delta_K(y-ia)]^2 < \infty$. Therefore, for fixed (x, y), $\hat{\Phi}(x, y)$ is a normal variable (*e.g.*, [74, p. 170]). Accordingly, the kernel $\hat{\Phi}$ is a normally distributed random function. Since the number of spheres in the chain is large, the finite sums can be replaced with series, the mathematical expectation and the variance are

$$E\hat{\Phi}(x, y) = \sum_i \delta_K(x - ia)\delta_K(y - ia)Ek_i = \frac{k}{a}\delta_K(x - y), \tag{2.25}$$

$$\mathrm{Var}\hat{\Phi}(x, y) = \frac{s^2\pi^{-2}}{(x-y)^2}\left\{-2(a^{-1}\delta_K(x-y) - a^{-2})\sin\frac{\pi x}{a}\sin\frac{\pi y}{a} + a^{-2}\left(\sin\frac{\pi x}{a} - \sin\frac{\pi y}{a}\right)^2\right\}. \tag{2.26}$$

The mathematical expectation is the kernel that one would obtain conducting stochastic experiments and according to (2.25) is given by Kunin-delta function (with a factor) that is oscillating. This means that the imposed randomness modelling irregularity of the particle arrangement does not remove the oscillating nature of the kernel, *i.e.*, after averaging the kernel determined by the random function (2.25) does not have the Gaussian, bell-shaped form.

The variance (2.26) behaves asymptotically as $s^2a^{-4}[(\sin^2(\pi x/a))/3 + 1]$ as $(x - y) \to 0$ and thus does not have singularities when $(x - y) \to 0$, which is sound since the series in (2.24) is convergent.

The normally distributed function $\hat{\Phi}(x, y)$ is fully determined by its mathematical expectation, the variance and the two-point correlation function. The correlation function between any two points $\hat{\Phi}_1 \equiv \hat{\Phi}(x_1, y_1) = \sum_i k_i \delta_K(x_1 - ia)\delta_K(y_1 - ia)$ and $\hat{\Phi}_2 \equiv \hat{\Phi}(x_2, y_2) = \sum_i k_i \delta_K(x_2 - ia)\delta_K(y_2 - ia)$ can be found as follows:

$$\begin{aligned}\mathrm{COV}(\hat{\Phi}_1, \hat{\Phi}_2) = -s^2\pi^{-3}a^{-1}\sin\frac{\pi x_1}{a}\sin\frac{\pi y_1}{a}\sin\frac{\pi x_2}{a}\sin\frac{\pi y_2}{a}\left\{\frac{\cot\frac{\pi x_1}{a}}{(x_1 - x_2)(x_1 - y_1)(x_1 - y_2)} + \right.\\ + \frac{\cot\frac{\pi x_2}{a}}{(x_2 - x_1)(x_2 - y_1)(x_2 - y_2)} + \frac{\cot\frac{\pi y_1}{a}}{(y_1 - x_1)(y_1 - x_2)(y_1 - y_2)} + \\ \left. + \frac{\cot\frac{\pi y_2}{a}}{(y_2 - x_1)(y_2 - x_2)(y_2 - y_1)}\right\}.\end{aligned} \tag{2.27}$$

It can be shown that the formal limiting transition $x_1 \to x_2$, $y_1 \to y_2$ leads to $\mathrm{COV}(\hat{\Phi}, \hat{\Phi}) = \mathrm{Var}\hat{\Phi}$. Indeed, by taking the limit of $(x_1, y_1) \to (x_2, y_2)$ in the last formula, one arrives at the formula for the variance (2.26). This serves as an indirect verification of (2.27).

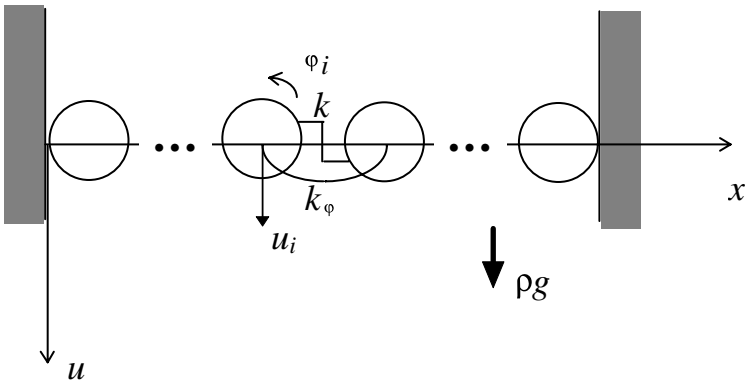


Figure 5. Model of the duct with grains.

2.6. A BOUNDARY-VALUE PROBLEM: VERTICAL DUCT

Let us consider an infinitely long duct occupying the area $0 < x_1 < L, |x_3| < \infty$. The duct is filled with granular material modelled by identical regular chains of balls connected by translational and rotational springs (Figure 5). The boundaries of each chain are subjected to pure clamping which corresponds to the following boundary conditions: $u_3 = 0$, $\varphi_2 = 0$. The volume force of ρg is applied to every ball. Let us assume that all fields depend on x_1 only. The inertia terms $\ddot{u}_3$, $\ddot{\varphi}_2$ are neglected.

For the sake of simplicity, x will be written instead of x_1, u instead of u_3, φ instead of φ_2.

2.6.1. *Exact solution of the discrete equations of equilibrium for the duct*

We now find the exact solution – the solution of the finite-difference equations (2.3) for the static case, $q_i = q$, $(q = -B\eta a^3)$, $M_i = 0$ under the following boundary conditions:

$$u_{j=0} = u_0, \qquad u_{j=N} = u_N, \qquad \varphi_{j=0} = \varphi_0, \qquad \varphi_{j=N} = \varphi_N. \tag{2.28}$$

The general solution of the correspondent homogeneous system:

$$-k(u_{i+1} - 2u_i + u_{i-1}) - k(a/2)(\varphi_{i+1} - \varphi_{i-1}) = 0 \quad (i = 1, \ldots, N-1), \tag{2.29$_1$}$$

$$k(a/2)(u_{i+1} - u_{i-1}) + k(a^2/4)(\varphi_{i+1} + 2\varphi_i + \varphi_{i-1}) - k_\varphi(\varphi_{i+1} - 2\varphi_i + \varphi_{i-1}) = 0 \tag{2.29$_2$}$$

is sought in the form $u_i = C\chi^i$, $\varphi_i = \bar{C}\chi^i$. By substituting it in (2.29), one can find the multiple root of fourth order, $\chi = 1$, of the characteristic equation. Then using the boundary conditions at $j = 0$, one may write the solution of the homogeneous system in the form:

$$u_j^c = u_0 - \left(a\varphi_0 + \left(\frac{a}{6} - \frac{2k_\varphi}{ka}\right)\bar{C}_2\right) j - \frac{a}{2}\bar{C}_1 j^2 - \frac{a}{3}\bar{C}_2 j^3, \qquad \varphi_j^c = \varphi_0 + \bar{C}_1 j + \bar{C}_2 j^2, \tag{2.30}$$

where $\bar{C}_1$ and $\bar{C}_2$ can be obtained from the boundary conditions at the other end, $j = N$.

When u_i is eliminated from the system (2.29), it can be shown that the particular solution of (2.3) for the rotations satisfies the equation

$$\Delta^3\varphi_i = -\frac{aq}{k_\varphi}, \quad \Delta = f(x+1) - f(x). \tag{2.31}$$

A particular solution of Equation (2.31) can be written in the form:

$$\varphi_i^p = -\frac{aq}{k_\varphi}\frac{1}{6}i^3. \tag{2.32}$$

Then, a particular solution in displacements can be found as a solution of the equation

$$\Delta^2 u_i = -\frac{a}{2}\Delta(\Delta+2)\varphi_i - \frac{q}{k}. \tag{2.33}$$

Eventually, the particular solution of Equation (2.33) can be written in the form:

$$u_i^p = \frac{a^2 q}{24k_\varphi} i^4 + \left(\frac{a^2 q}{24k_\varphi} - \frac{q}{2k}\right) i^2. \tag{2.34}$$

The full solution now becomes

$$u_j = u_0 - \left(a\varphi_0 + 2\frac{k(a^2/4) - 3k_\varphi}{3ka}\bar{C}_2\right) j - \frac{a}{2}\bar{C}_1 j^2 - \frac{a}{3}\bar{C}_2 j^3 + \frac{a^2 q}{24k_\varphi} j^4 + \\ + \left(\frac{a^2 q}{24k_\varphi} - \frac{q}{2k}\right) j^2, \tag{2.35$_1$}$$

$$\varphi_j = \varphi_0 + \bar{C}_1 j + \bar{C}_2 j^2 - \frac{aq}{6k_\varphi} j^3. \tag{2.35$_2$}$$

2.6.2. *Non-local Cosserat continuum model of the duct. Solution of the equations of equilibrium in the non-local Cosserat continuum*

According to the established isomorphism (Section 2.4), the non-local Cosserat solution should coincide at sphere centres with the exact (discrete) solution. It is, however, important to see what the continuous non-local solution looks like at the points between the sphere centres. The solution of the equations of equilibrium in the non-local Cosserat continuum (2.12) in the static case with zero volume moment and a constant volume force were obtained by Pasternak and Mühlhaus [72].

Using the conventional Fourier transform, one reduces the non-local "Lamé equations" (2.12) to the following system of equations:

$$4k \sin\frac{\omega a}{2}\delta_{\mathrm{K}}^{\mathrm{F}}\left[\sin\frac{\omega a}{2}\bar{u}(\omega) - \mathrm{i}\frac{a}{2}\cos\frac{\omega a}{2}\bar{\varphi}(\omega)\right] = 2\pi q\delta(\omega), \tag{2.36$_1$}$$

$$\delta_{\mathrm{K}}^{\mathrm{F}}\left[\mathrm{i}k\frac{a}{2}\sin\frac{\omega a}{2}\cos\frac{\omega a}{2}\bar{u}(\omega) + \left(k\frac{a^2}{4}\cos^2\frac{\omega a}{2} + k_\varphi \sin^2\frac{\omega a}{2}\right)\bar{\varphi}(\omega)\right] = 0, \tag{2.36$_2$}$$

$$\delta_{\mathrm{K}}^{\mathrm{F}}(\omega) = \int_{-\infty}^{+\infty}\delta_K(x)\mathrm{e}^{-\mathrm{i}x\omega}\mathrm{d}x, \qquad \bar{u}(\omega) = \int_{-\infty}^{+\infty}u(x)\mathrm{e}^{-\mathrm{i}x\omega}\mathrm{d}x, \qquad \bar{\varphi}(\omega) \\ = \int_{-\infty}^{+\infty}\varphi(x)\mathrm{e}^{-\mathrm{i}x\omega}\mathrm{d}x, \tag{2.37}$$

where $\delta(\omega)$ is the ordinary Dirac-delta function.

The inverse transforms are:

$$\delta_K(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\delta_{\mathrm{K}}^{\mathrm{F}}(\omega)\mathrm{e}^{\mathrm{i}x\omega}\mathrm{d}\omega, \qquad u(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\bar{u}(\omega)\mathrm{e}^{\mathrm{i}x\omega}\mathrm{d}\omega, \\ \varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\bar{\varphi}(\omega)\mathrm{e}^{\mathrm{i}x\omega}\mathrm{d}\omega. \tag{2.38}$$

The Fourier transform of the Kunin-delta function is:

$$\delta_{\mathrm{K}}^{\mathrm{F}}(\omega) = \begin{cases} 1, & |\omega| < \pi/a \\ 0, & |\omega| > \pi/a \end{cases}. \tag{2.39}$$

Correspondingly, we will look for a solution of (2.36) for $|\omega| < \pi/a$. As usual, the full solution of the system (2.36) can be written in the form:

$$\bar{u} = \bar{u}^c + \bar{u}^p, \qquad \bar{\varphi} = \bar{\varphi}^c + \bar{\varphi}^p, \tag{2.40}$$

where the pair $(\bar{u}^c, \bar{\varphi}^c)$ is the homogeneous solution and $(\bar{u}^p, \bar{\varphi}^p)$ is a particular solution of the non-homogeneous system.

The determinant of the homogeneous system is equal to $k_\varphi \sin^4 \frac{\omega a}{2}$ and in the interval $|\omega| < \pi/a$ has a root $\omega = 0$ of the fourth order. Hence, the homogeneous solution has to be sought in the form:

$$X_j(\omega) = 2\pi C_0^j \delta(\omega) + 2\pi C_1^j \frac{\delta(\omega)}{i\omega} + 2\pi C_2^j \frac{2\delta(\omega)}{-\omega^2} + 2\pi C_3^j \frac{6\delta(\omega)}{-i\omega^3} \quad (j = 1, 2), \tag{2.41}$$

where

$$X_1(\omega) = \bar{u}(\omega), \qquad X_2(\omega) = \bar{\varphi}(\omega). \tag{2.42}$$

Insertion of (2.41) into the homogeneous system gives the following relations between the constants

$$C_2^1 = -\frac{1}{2} C_1^2, \qquad C_3^1 = -\frac{1}{3} C_2^2, \qquad C_3^2 = 0, \qquad C_1^1 = -\frac{a^2}{6} C_2^2 - C_0^2 + 2\frac{k_\varphi}{k} C_2^2. \tag{2.43}$$

Assuming u at $x = 0$ to be u_0 and φ at $x = 0$ to be φ_0, one has $C_0^1 = u_0$ and $C_0^2 = \varphi_0$. Finally, the homogeneous solution can be written in the form:

$$u^c(x) = u_0 - \left(a\varphi_0 + 2\frac{k(a^2/4) - 3k_\varphi}{3ka} \bar{C}_2 \right) \frac{x}{a} - \frac{a}{2} \bar{C}_1 \frac{x^2}{a^2} - \frac{a}{3} \bar{C}_2 \frac{x^3}{a^3}, \tag{2.44$_1$}$$

$$\varphi^c(x) = \varphi_0 + \bar{C}_1 \frac{x}{a} + \bar{C}_2 \frac{x^2}{a^2}, \quad \bar{C}_1 = C_1^2 a, \quad \bar{C}_2 = C_2^2 a^2. \tag{2.44$_2$}$$

A particular solution of the system (2.36) reads:

$$\bar{u}^p(\omega) = \frac{a^2 \pi q}{8k_\varphi} \cdot \frac{\cos^2 \frac{\omega a}{2}}{\sin^4 \frac{\omega a}{2}} \delta(\omega) + \frac{\pi q}{2k} \cdot \frac{\delta(\omega)}{\sin^2 \frac{\omega a}{2}}, \qquad \bar{\varphi}^p(\omega) = -\frac{ia\pi q}{4k_\varphi} \cdot \frac{\cot \frac{\omega a}{2}}{\sin^2 \frac{\omega a}{2}} \delta(\omega). \tag{2.45}$$

By performing the inverse Fourier transform and adding the homogeneous solution (2.44), the full solution is obtained as:

$$u(x) = u_0 - \left(a\varphi_0 + \left(\frac{a}{6} - \frac{2k_\varphi}{ka} \right) \bar{C}_2 \right) \frac{x}{a} - \frac{a}{2} \bar{C}_1 \frac{x^2}{a^2} - \\ - \frac{a}{3} \bar{C}_2 \frac{x^3}{a^3} + \left(\frac{a^2 q}{24k_\varphi} - \frac{q}{2k} \right) \frac{x^2}{a^2} + \frac{a^2 q}{24k_\varphi} \frac{x^4}{a^4}, \tag{2.46$_1$}$$

$$\varphi(x) = \varphi_0 + \bar{C}_1 \frac{x}{a} + \bar{C}_2 \frac{x^2}{a^2} - \frac{aq}{6k_\varphi} \frac{x^3}{a^3}. \tag{2.46$_2$}$$

Alternatively, one can solve the above problem by finding the solution of the equations of motion (2.19) with zero inertia terms, in which $q(x)/(a^3\eta) = \rho g = -B$, $M(x) = 0$. Then the strain and curvature can be found from the non-local constitutive relationship (2.17) and (2.18) by solving the system of two integral equations. Subsequently, the displacement and rotation fields can be obtained from (2.14).

Equations (2.19) have the following solutions:

$$\sigma(x) = Bx + \sigma(0), \qquad \mu(x) = B\frac{x^2}{2} + \sigma(0)x + \mu(0), \tag{2.47}$$

where $\sigma(0)$, $\mu(0)$ are yet unknown stress and moment stress at the origin.

The solution can be rewritten in the following form:

$$u(x) = \eta\left[\left(\frac{\sigma(0)}{E} - \frac{\sigma(0)}{12E_\varphi}a^2\right)x - \frac{1}{2E_\varphi}\mu(0)x^2 - \frac{1}{6E_\varphi}\sigma(0)x^3 - \frac{B}{24E_\varphi}x^4 + \frac{B}{2E}x^2 - \frac{a^2B}{24E_\varphi}x^2\right] + u(0) - \varphi(0)x, \tag{2.48$_1$}$$

$$\varphi(x) = \frac{\eta}{E_\varphi}\left[\mu(0)x + \sigma(0)\frac{1}{2}x^2 + \frac{B}{6}x^3\right] + \varphi(0). \tag{2.48$_2$}$$

This form coincides with (2.46) if one sets

$$\mu(0) = \frac{k_\varphi}{\eta a^2}\bar{C}_1, \qquad \sigma(0) = \frac{2k_\varphi}{\eta a^3}\bar{C}_2. \tag{2.49}$$

Using the boundary conditions $u(0) = u(L) = 0$, $\varphi(0) = \varphi(L) = 0$, one obtains the solution:

$$u(x) = -\frac{\eta B}{24E_\varphi}x(x-L)(x^2 + 2px + \hat{q}), \qquad \varphi(x) = \frac{\eta B}{12E_\varphi}x(x-L)(2x-L), \tag{2.50}$$

$$2p = -L, \;\; \hat{q} = -4\left(\frac{3E_\varphi}{E} - \frac{a^2}{4}\right), \tag{2.51}$$

and the constants $\sigma(0)$, $\mu(0)$ are:

$$\sigma_0 = \sigma(0) = -\frac{1}{2}BL, \qquad \mu_0 = \mu(0) = \frac{1}{12}BL^2. \tag{2.52}$$

The normalisation

$$L = 1, \qquad E = 1 \tag{2.53}$$

leads to

$$u(x) = -\frac{\eta B}{24E_\varphi}x(x-1)(x^2 + 2px + \hat{q}), \quad p = -\frac{1}{2}, \;\; \hat{q} = -4(3E_\varphi - a^2/4). \tag{2.54}$$

Note that the coefficient of x^4 in (2.50$_1$) is positive because B is negative. This means that $u(-\infty) = u(+\infty) = +\infty$, *i.e.*, the branches of the fourth-order polynomial $u(x)$ are going downwards at the $\pm\infty$ (the positive direction of u is directed downwards). Because of (2.54) the displacement distribution is symmetrical.

The displacement becomes zero at points:

$$x_{1,2} = \frac{1}{2}\left(1 \pm \sqrt{1 + 16(3E_\varphi - a^2/4)}\right). \tag{2.55}$$

If $3E_\varphi - a^2/4 > 0$, *i.e.*, $\sqrt{3E_\varphi} > a/2$, which is the case when the rotational springs are rather stiff, then $x_1 > 1$, $x_2 < 0$. This means that both roots x_1 and x_2 are outside the duct. If $3E_\varphi - a^2/4 < 0$, *i.e.*, $\sqrt{3E_\varphi} < a/2$, which corresponds to the case of small stiffness of the rotational springs, then $x_1 = 1 - \hat{q} < 1$, $x_2 = \hat{q} > 0$. This means that both roots x_1 and x_2 are inside the

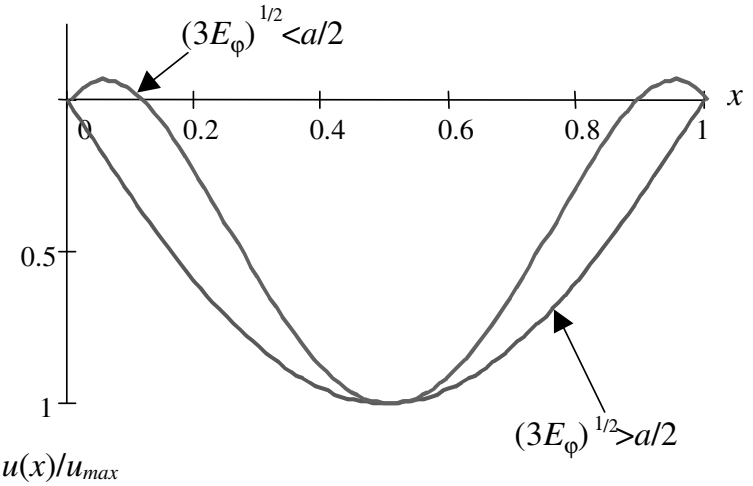


Figure 6. Distribution of normalised displacements for soft ($(3E_\varphi)^{1/2} < a/2$) and stiff ($(3E_\varphi)^{1/2} > a/2$) rotational springs.

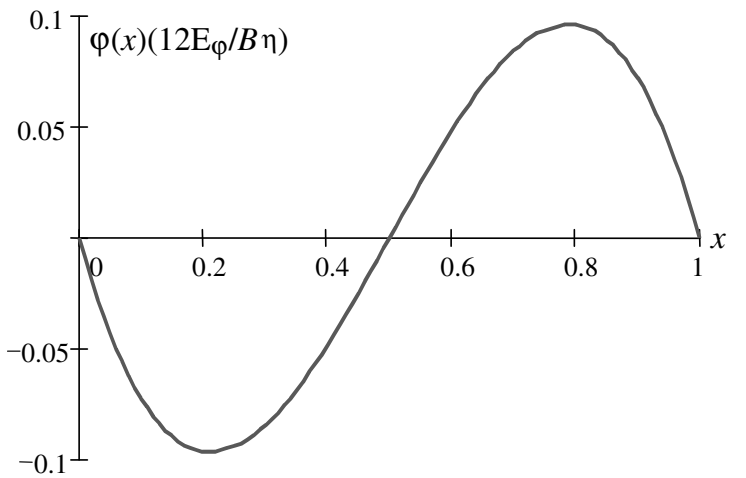


Figure 7. Distribution of normalised rotations.

duct (Figure 6). Thus, for a certain combination of constants the non-local Cosserat continuum solution exhibits a boundary effect consisting of anomalous upward displacements near the boundary.

According to (2.46_2), after the normalisation the rotations become (Figure 7)

$$\varphi(x) = \frac{B\eta}{12E_\varphi} x(x-1)(2x-1). \tag{2.56}$$

2.6.3. *Cosserat continuum model of the duct*

Let us find the solution of the Cosserat equations of equilibrium (2.7) for the case of the constant volume force and zero volume moment ($\rho f_3 = -B$, $\rho m_2 = 0$):

$$\frac{\partial \sigma_{13}}{\partial x_1} + \rho f_3 = 0, \qquad \frac{\partial \mu_{12}}{\partial x_1} - \sigma_{13} = 0. \tag{2.57}$$

Denoting $-\rho f_3 = -\rho g = B$, one gets the solution of Equations (2.57) in the form

$$\sigma_{13} = Bx + \sigma(0), \qquad \mu_{12} = B\frac{x^2}{2} + \sigma(0)x + \mu(0). \tag{2.58}$$

Then taking into account the Cosserat constitutive equations (2.6) one obtains

$$\kappa_{12} = \frac{\eta a}{k_\varphi}\left[B\frac{x^2}{2} + \sigma(0)x + \mu(0)\right], \qquad \gamma_{13} = \frac{\eta a}{k}[Bx + \sigma(0)]. \tag{2.59}$$

Subsequently, upon using the deformation measures (2.5) rotation and displacement fields are found:

$$\varphi_2 = \frac{\eta a}{k_\varphi}\left[B\frac{x^3}{6} + \sigma(0)\frac{x^2}{2} + \mu(0)x\right] + \varphi(0), \tag{2.60}$$

$$u_3 = -\frac{\eta a}{k_\varphi}B\frac{x^4}{24} - \frac{\eta a}{k_\varphi}\sigma(0)\frac{x^3}{6} - \frac{\eta a}{k_\varphi}\mu(0)\frac{x^2}{2} + \frac{\eta a}{k}B\frac{x^2}{2} + \frac{\eta a}{k}\sigma(0)x - \varphi(0)x + u(0). \tag{2.61}$$

After satisfying the boundary conditions

$$u(0) = 0, \qquad u(L) = 0, \qquad \varphi(0) = 0, \qquad \varphi(L) = 0, \tag{2.62}$$

Equations (2.60) and (2.61) become

$$u(x) = \frac{-\eta B}{24E_\varphi}x(x-L)(x^2 + 2px + \hat{q}^c), \qquad \varphi(x) = \frac{\eta B}{12E_\varphi}x(x-L)(2x-L), \tag{2.63}$$

$$2p = -L, \quad \hat{q}^c = -12E_\varphi/E, \qquad \sigma(0) = \sigma_0 = -\frac{1}{2}BL, \qquad \mu(0) = \mu_0 = \frac{1}{12}BL^2. \tag{2.64}$$

Comparing the rotation fields for the Cosserat continuum model (2.63_2) with the non-local Cosserat continuum model (2.50_2), one can see that they coincide, because the constants σ_0 and μ_0 have not changed, while the displacement fields (2.63_1) and (2.50_1) differ in the terms $\hat{q}$ and $\hat{q}^c$. Let us analyse this difference.

After the normalisation (2.53), the zeros of the displacement can be found:

$$x_{1,2} = \frac{1}{2}\left(1 \pm \sqrt{1 + 48E_\varphi}\right). \tag{2.65}$$

It is obvious that $1 + 48E_\varphi > 1$, therefore $x_1 > 1$, $x_2 < 0$. Both roots x_1 and x_2 are always outside the duct. This means that the boundary effects, present in the non-local Cosserat continuum model, disappear in the Cosserat model.

There has to be an explanation for that fact. The length where the boundary effect exists is defined by the value of the parameter q. Let us evaluate

$$|\hat{q}| = |-4(3E_\varphi - a^2/4)| < a^2.$$

Hence, the characteristic size where the boundary effect exists is of the order a^2 . However, the Cosserat theory does not see the lengths smaller than a ($a^2 < a << 1$), the characteristic length parameter which has been used when finite differences were replaced by the partial derivatives. That is why these boundary effects are left invisible in the Cosserat theory. Furthermore, in terms of the original discrete system, no distance smaller than a exists (there are no spheres at such distances). Therefore, the "high resolution" boundary effect is an artefact of the non-local Cosserat continuum resulting from the type of interpolation adopted.

2.6.4. *Comparison of the exact solution with the solution in the non-local Cosserat and Cosserat continua*

Assuming $x = ja$, one has the Cosserat solution (2.60) and (2.61)

$$u_3(ja) = u_0 - \left(a\varphi(0) - \frac{2k_\varphi}{ka}\bar{C}_2\right) j - \frac{a}{2}\bar{C}_1 j^2 - \frac{q}{2k} j^2 - \frac{a}{3}\bar{C}_2 j^3 + \frac{qa^2}{24k_\varphi} j^4, \tag{2.66}$$

$$\varphi_2(ja) = \varphi_0 + \bar{C}_1 j + \bar{C}_2 j^2 - \frac{aq}{6k_\varphi} j^3, \tag{2.67}$$

where

$$\mu(0) = \frac{k_\varphi}{\eta a^2}\bar{C}_1, \qquad \sigma(0) = \frac{2k_\varphi}{\eta a^3}\bar{C}_2, \qquad \eta a^3 B = -q. \tag{2.68}$$

Comparing the Cosserat theory solution (2.66) and (2.67) with the exact solution (2.35) one can conclude that the rotation fields coincide completely, while the displacements differ in the terms $-\frac{a}{6}\bar{C}_2 j$ and $\frac{a^2 q}{24k_\varphi} j^2$.

Putting $ja = x$ in the non-local Cosserat solution (2.46), we immediately see that the non-local Cosserat solution (2.46) coincides completely with the exact one (2.35). This means that the non-local Cosserat solution gives the exact solution at nodes where the centroids are; as we anticipated above, the result is due to the homogenisation by integral transformation. However, being a continuum solution, the non-local Cosserat solution also gives some values in between nodes due to the interpolating nature of the homogenisation by integral transformation.

Figure 8a shows a comparison of the discrete (exact), non-local Cosserat and the Cosserat solutions for a simple case of three balls. Figure 8b shows the configuration before and after deformation.

In the above we developed two continuum models of the discrete model, using two different homogenisation strategies: by differential expansions and integral transformations and compared them against the exact solution of the discrete model. The first approach led to the Cosserat continuum theory. The boundary-value problem is reduced to solving a relatively

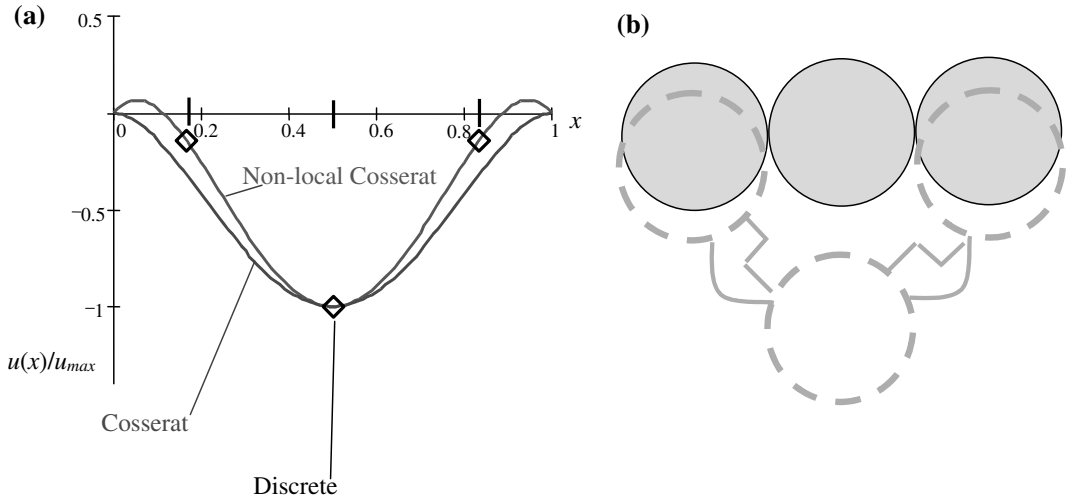


Figure 8. The comparison of the exact, non-local Cosserat and Cosserat solutions: (a) displacement distribution; diamonds indicate the centroids of the spheres; (b) configuration before (solid line) and after (broken line) the deformation in a vertical duct with three spheres.

simple system of two differential equations. The analysis above shows that the Cosserat continuum model of granulates gives both quite good accuracy and relative simplicity of solving the governing equations. The second approach led to the non-local Cosserat continuum theory that gave us the exact solution. The problem is reduced to solving a system of two integral equations which is more complicated than solving the system of differential equations (the first model) and in essence is no simpler than to solve the governing equations of the discrete model, the system of finite-difference equations. Thus, the second homogenisation approach indeed gave us the continuum description of the discrete system, a non-local Cosserat continuum. However, being just an equivalent description of the discrete model, giving the exact solution does not offer any simplification, which generally continuum theories are supposed to do.

3. Wave propagation. Dispersion relationships

For a particular case of $q_3(x_1) = M_2(x_1) = 0$ we consider the propagation of harmonic waves

$$u = A\mathrm{e}^{\mathrm{i}\xi(x-v_p t)}, \qquad \varphi = B\mathrm{e}^{\mathrm{i}\xi(x-v_p t)}, \tag{3.1}$$

where ξ is the wave number and v_p is the phase velocity. For the sake of simplicity x will be written instead of x_1, u instead of u_3 and φ instead of φ_2.

Propagation of these waves will be studied for the original physical model (2.3) and then for the Cosserat (2.8) and non-local Cosserat (2.12) models.

3.1. Wave propagation in the discrete (physical) model

By substituting (3.1) in the governing equations of the original physical model, namely the discrete equations of motion (2.3) or their homogenised (continuous) analogue (2.9), we obtain the following system:

$$-m\xi^2 v_p^2 A + 4k\sin^2\left(\frac{\xi a}{2}\right)A - \mathrm{i}ka\sin(\xi a)B = 0, \tag{3.2$_1$}$$

$$-J\xi^2 v_p^2 B + \mathrm{i}ka\sin(\xi a)A + ka^2\cos^2\left(\frac{\xi a}{2}\right)B + 4k_\varphi\sin^2\left(\frac{\xi a}{2}\right)B = 0. \tag{3.2$_2$}$$

The characteristic equation is biquadratic with respect to the phase velocity

$$mJv_p^4 - 4\left(Jk\sin^2\left(\frac{\xi a}{2}\right) + \frac{1}{4}mka^2\cos^2\left(\frac{\xi a}{2}\right) + mk_\varphi\sin^2\left(\frac{\xi a}{2}\right)\right)\frac{v_p^2}{\xi^2} + \\ + 16kk_\varphi\sin^4\left(\frac{\xi a}{2}\right)\frac{1}{\xi^4} = 0. \tag{3.3}$$

This equation has a positive discriminant. Two real solutions of the equation give the phase velocity. Since $\forall k, k_\varphi, r, a, m, \xi$

$$mJ > 0, \quad 16kk_\varphi\sin^4\left(\frac{\xi a}{2}\right)\frac{1}{\xi^4} > 0,$$
$$-4\left(Jk\sin^2\left(\frac{\xi a}{2}\right) + \frac{1}{4}mka^2\cos^2\left(\frac{\xi a}{2}\right) + mk_\varphi\sin^2\left(\frac{\xi a}{2}\right)\right)\frac{1}{\xi^2} < 0,$$

both solutions for v_p^2 are always positive. They read

$$v_p^2 = \frac{2k}{J\xi^2}\left[\left(\frac{2r^2}{5}+\frac{k_\varphi}{k}\right)\sin^2\left(\frac{\xi a}{2}\right)+\frac{1}{4}a^2\cos^2\left(\frac{\xi a}{2}\right)\pm\right.$$
$$\left.\pm\sqrt{\left(\frac{k_\varphi}{k}-\frac{2r^2}{5}\right)^2\sin^4\left(\frac{\xi a}{2}\right)+\frac{1}{16}a^4\cos^4\left(\frac{\xi a}{2}\right)+\frac{1}{2}a^2\left(\frac{2r^2}{5}+\frac{k_\varphi}{k}\right)\sin^2\left(\frac{\xi a}{2}\right)\cos^2\left(\frac{\xi a}{2}\right)}\right]. \tag{3.4}$$

Let $r = a/2$. One can find the ratio of their amplitudes, for example from Equation (3.2_1):

$$\frac{A}{B} = \mathrm{i}ka\frac{\sin(\xi a)}{4k\sin^2\left(\frac{\xi a}{2}\right)-m\xi^2 v_p^2}, \tag{3.5}$$

where v_p^2 is given by (3.4).

The first type of wave (positive sign before the radical (3.4)) and the second (negative sign) have the following long-wave asymptotics

$$v_p \underset{\xi\to 0}{\sim} \begin{cases} \xi^{-1} \\ \xi \end{cases}. \tag{3.6}$$

These are asymptotics of the same type as obtained by Mühlhaus and Oka [35].

The corresponding asymptotics for the ratio of amplitudes is:

$$\frac{A}{B} \underset{\xi\to 0}{\sim} \begin{cases} 0 \Rightarrow A \sim 0 \text{ (rotational wave)} \\ \infty \Rightarrow B \sim 0 \text{ (shear wave)} \end{cases}. \tag{3.7}$$

Thus, we have two types of waves. The first becomes the rotational wave in the long-wave limit ($\xi \to 0$), while the second is the shear wave. Otherwise, both components are present, but asymptotically one type dominates. For that reason we will call these waves rotational-shear and shear-rotational. One should avoid considering the limiting case of $\xi \to \infty$ (short-length wave), since this case cannot be described properly in terms of the physical model. This is because for large wave numbers, the ball microstructure should be taken into account and the model should be changed accordingly.

The ratio of amplitudes for different ratios of spring stiffnesses is shown in Figure 9. All the plots are given for physically reasonable wavelengths. This is due to the fact that in the considered system the wavelength cannot be shorter than the ball size. Moreover, in the homogeneous models the wavelength should be much greater than the ball size.

The square of the phase velocity for rotational-shear and shear-rotational waves for different ratios of stiffnesses is shown in Figure 10.

3.2. Non-local Cosserat continuum

Substituting (3.1) in the non-local (integral) equations of motion $(2.12_1 - 2.12_2)$ and calculating the corresponding integrals, one can get the same system as obtained for the exact equation of motion (3.2), but with a restriction: $\xi < \pi/a$. (It does not appear mathematically for the exact solution.) This restriction reflects the fact that wavelengths must be larger than the microstructure size. This seems reasonable, since our model was not designed to "see" something less that the microstructure size.

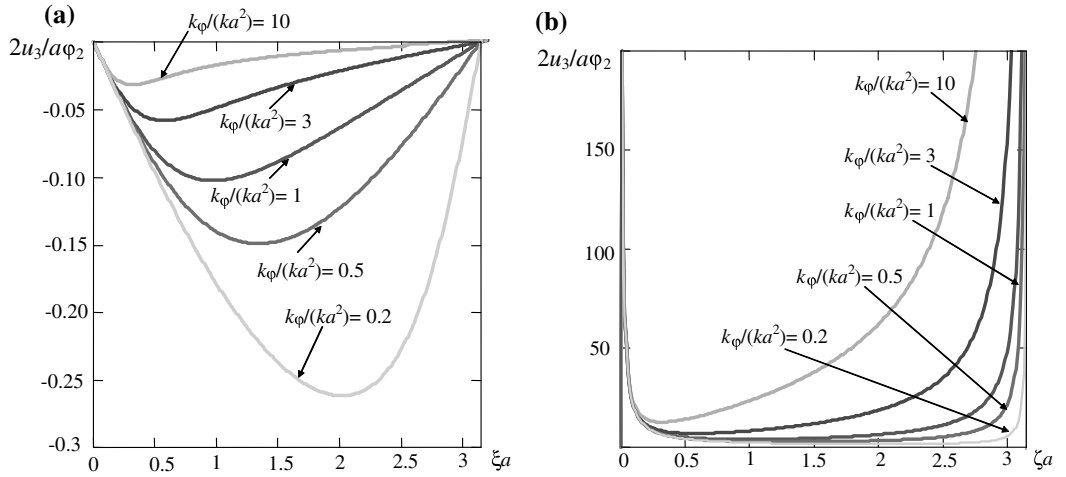


Figure 9. The ratio of amplitudes for rotational–shear (a) and shear–rotational wave (b).

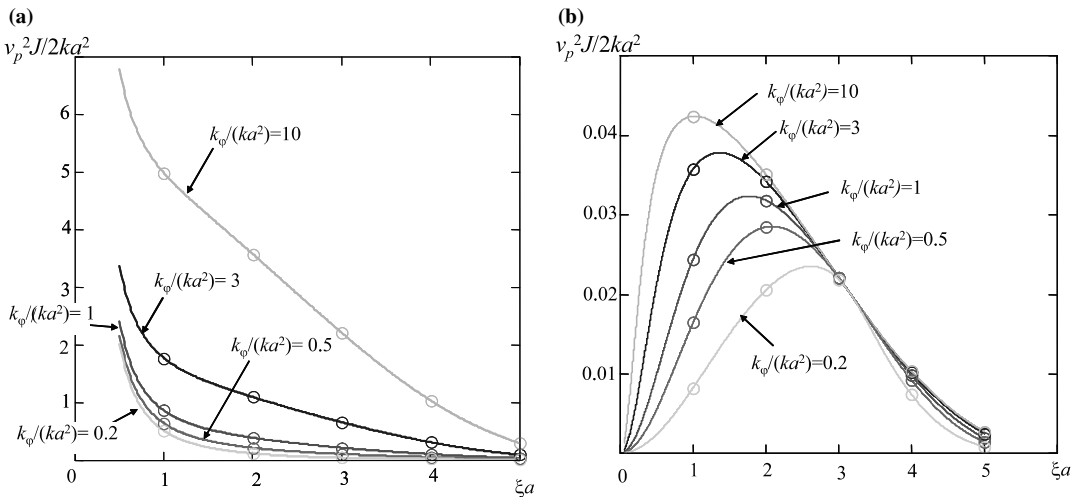


Figure 10. The square of the phase velocity for rotational–shear (a) and shear–rotational wave (b). The circles show velocities corresponding to integer values of the normalised wave number.

3.3. Cosserat continuum approximation

By assuming $f_3 = m_2 = 0$, we can write the equations of motion (2.8_1–2.8_2) in the form:

$$ka^2\left[\frac{\partial^2 u_3}{\partial x_1^2} + \frac{\partial \varphi_2}{\partial x_1}\right] = m\ddot{u}_3, \quad m = \rho a^3 \eta, \tag{3.8$_1$}$$

$$a^2\left[k_\varphi \frac{\partial^2 \varphi_2}{\partial x_1^2} - k\frac{\partial u_3}{\partial x_1} - k\varphi_2\right] = J\ddot{\varphi}_2. \tag{3.8$_2$}$$

Substituting (3.1) in the equations of motion, one can get:

$$\frac{m}{a^2}\xi^2 v_p^2 A - k\xi^2 A + \mathrm{i}\xi k B = 0, \tag{3.9$_1$}$$

$$\frac{J}{a^2}\xi^2 v_p^2 B - k_\varphi \xi^2 B - \mathrm{i}\xi k A - kB = 0. \tag{3.9$_2$}$$

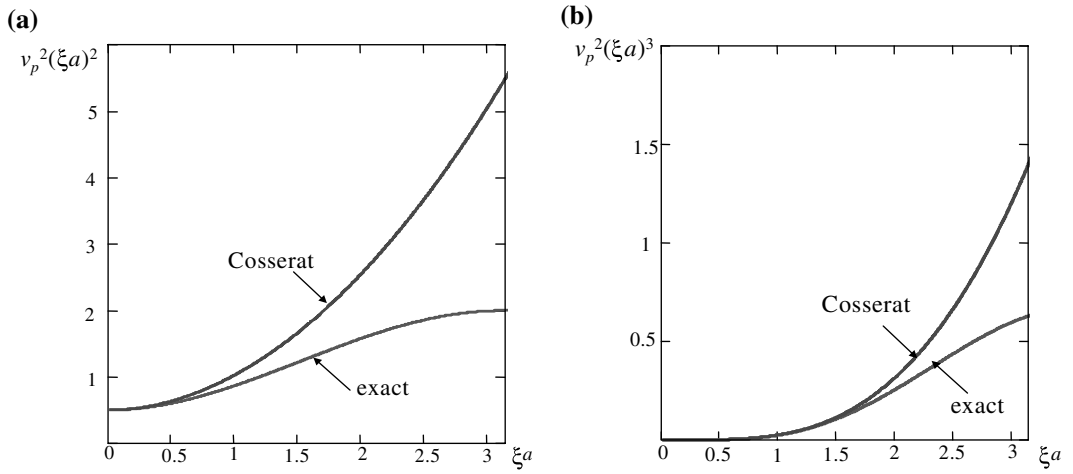


Figure 11. Comparison of the square of the phase velocity for the exact and the Cosserat solutions: (a) rotational-shear wave, (b) shear-rotational wave.

The system has the following biquadratic characteristic equation with respect to the phase velocity

$$\frac{mJ}{a^4}v_p^4 - \frac{1}{a^2}\left(mk_\varphi + Jk + \frac{mk}{\xi^2}\right)v_p^2 + kk_\varphi = 0. \tag{3.10}$$

The discriminant is positive. Two real solutions of the equation give the phase velocity. Since $\forall k, k_\varphi, r, a, m, \xi$

$$\frac{mJ}{a^4} > 0, \qquad -\frac{1}{a^2}\left(mk_\varphi + Jk + \frac{mk}{\xi^2}\right) < 0, \qquad kk_\varphi > 0$$

both solutions for v_p^2 are always positive. They read

$$v_p^2 = \frac{a^2 k}{2J}\left[\left(\frac{2r^2}{5} + \frac{k_\varphi}{k}\right) + \frac{1}{\xi^2} \pm \sqrt{\left(\frac{k_\varphi}{k} - \frac{2r^2}{5}\right)^2 + \frac{1}{\xi^4} + \frac{2}{\xi^2}\left(\frac{2r^2}{5} + \frac{k_\varphi}{k}\right)}\right]. \tag{3.11}$$

This expression is an asymptotic of (3.4) as $\xi \to 0$ with the accuracy $o(\xi)$. This is not surprising, since the Cosserat model is a long-wave (small wave number) approximation.

Figure 11 shows the square of the phase velocity for both rotational-shear and shear-rotational waves comparing the exact and the Cosserat solutions. They are in quite good agreement for the small wave numbers, *i.e.*, in the range where the Cosserat solution approximates properly the exact solution.

The obtained result (3.11) allows us to investigate the effect of the presence of rotational degrees of freedom. Towards this end consider an asymptotic of $k_\varphi/(ka^2) \gg 1$ which is the case when the rotations are almost suppressed. Then assuming $r = a/2$, we have

$$v_{p1}^2 \sim \frac{ka^2}{J}\left[\frac{k_\varphi}{k} + \xi^{-2}\right], \qquad v_{p2}^2 \sim \frac{0{\cdot}1ka^4}{J}\left[1 - \frac{k}{k_\varphi}\xi^{-2}\right]. \tag{3.12}$$

When the rotations are completely suppressed ($k_\varphi/(ka^2) \to \infty$), velocity of the first, rotational–shear, wave tends to infinity, but the amplitude of the displacement oscillations vanishes. The velocity of the second, shear–rotational, wave becomes

$$v_{p2\infty} \sim a^2\sqrt{0{\cdot}1k/J}. \tag{3.13}$$

This corresponds to a conventional shear wave.

In general, when $k_\varphi/(ka^2)$ is finite, the shear-rotational wave is slower than the conventional shear wave, while the rotational-shear wave is faster than the latter. When experimental measurements of wave velocities are conducted by registering the time of first arrival, one can expect that this rotational-shear wave will be registered first. This will lead to the measured wave velocity being higher than predicted by classical elasticity, thus paving the way to experimental observation of Cosserat effects.

4. Conclusions

In many cases it is advantageous to model real materials with internal microstructure as continua based on the well-developed machinery of modern continuum mechanics. This can be accomplished by associating each point of the continuum with a volume element, which on the one hand is large compared to the dimensions of the microstructure but, on the other hand, must be small compared to the characteristic dimensions of the phenomenon to be modelled. The presence of microstructure implies that, at least in principle, relative movements between the microstructure and the average macroscopic deformations are possible. The relative movements may be considered by means of additional degrees of freedom. The introduction of additional degrees of freedom leads to non-classical continua, the simplest being the Cosserat continuum, each point of which possesses both translational and rotational degrees of freedom. Proceeding with further degrees of freedom, one obtains higher-order continua and attains more accuracy in the modelling.

The treatment of a representative volume element as a point of the macroscopic continuum imposes a restriction on the scale of the modelling: details smaller than the representative volume element are beyond the resolution of the model. It is generally believed that this restriction can be overcome by incorporating non-local constitutive laws, where, for instance, the stress at a point depends in an integral sense on the strains within a volume surrounding the point. For that reason, non-local continua do not obey the Cauchy-Euler principle: the stress state of a volume is not completely determined by the stresses at its boundary. This makes it impossible to deduce the continuum equations of motion from first principles, forcing one to either hypothesise on them or to infer them from microstructural considerations.

The introduction of a suitable continuum theory to model a material with a given microstructure requires an appropriate choice of homogenisation procedure. In order to analyse different homogenisation methods, we considered a model system consisting of decoupled periodic 1D chains of solid spheres connected by translational and rotational springs. The model is simple enough to allow complete analytical solutions for both static equilibrium and wave propagation. Two homogenisation techniques were considered: (1) homogenisation by differential expansion and (2) homogenisation by integral transformation (Kunin-type homogenisation). The first technique leads to a local Cosserat continuum, while the second approach gives rise to a non-local Cosserat continuum theory. The former result offered a robust balance between accuracy and simplicity being a long-wave asymptotic approximation to the exact model. The second technique resulted in a non-local continuum description that yielded an exact solution, but at the same time did not really provide any simplification as compared to the exact, discrete model. In fact, there is isomorphism between the discrete model and a non-local Cosserat continuum. Interestingly, the equations of motion derived for this case using the Kunin-type homogenisation of the discrete equations assumed after the introduction of invariant deformation measures, the form expected for a Cosserat continuum. Another feature of this method is that the non-local and the discrete solutions for 1D granulates coincide at the centres of the balls. However, between the discrete points

the Kunin-type homogenisation may lead to unrealistic patterns. In particular, the considered non-local model for a vertical duct under gravity showed near-boundary displacements directed upwards, *i.e.*, against gravity. These boundary effects are, however, limited to distances smaller than the spacing between the particle centres and are simply artefacts of the homogenisation procedure.

Homogenisation by means of integral transformation produces non-local integral relations with oscillating kernels. This oscillation is, however, not a direct consequence of the strict periodicity of the model system. For instance, randomisation of the spring stiffnesses makes the kernels random functions with periodic means.

The analysis of wave propagation in this model system showed that two types of waves exist simultaneously: shear-rotational and rotational-shear waves, the latter being the faster ones. As the wave number tends to zero (long-wavelength limit), the shear component is predominant in the shear-rotational wave, while the rotational component is predominant in the rotational-shear wave. Further analysis of the Cosserat model showed that in the limit of infinite rotational stiffness (when particle rotation is suppressed) the rotational-shear wave disappears, while the velocity of the shear-rotational wave becomes independent of the frequency, indicating the absence of dispersion. The rotational-shear wave was found to be faster than the conventional shear wave. Therefore, when experimental measurements of wave velocities are conducted by registering the time of first arrival, one can expect that this rotational-shear wave will be registered first. This will lead to the measured wave velocity being higher than predicted by the classical elasticity, thus providing a means for the experimental detection of Cosserat effects.

In conclusion, the framework of the Cosserat continuum theory was found to provide accurate descriptions of materials with microstructure. The Cosserat effects are responsible for the increase in measured wave velocities in granular materials as compared to the classical calculations that ignore rotational degrees of freedom.

Acknowledgements

EP acknowledges the financial support from the ARC Postdoctoral Fellowship (2003–2006) and Discovery Grant DP0346148. In addition, both authors acknowledge the support by the Australian Computational Earth Systems Simulator (ACcESS), a major National Research Facility.

References

1. E. Cosserat and F. Cosserat, *Théorie des Corps Déformables*. Paris: A. Herrmann et Fils (1909) 226 pp.
2. W. Nowacki, The linear theory of micropolar elasticity. In: W. Nowacki and W. Olszak (eds.), *Micropolar Elasticity*. Wien, New York: Springer-Verlag (1974) pp.1–43.
3. R.D. Mindlin, Micro-structure in linear elasticity. *Arch. Ration. Mech. Anal.* 16 (1964) 51–78.
4. A.V. Dyskin, R.L. Salganik and K.B. Ustinov, Multi-scale geomechanical modelling. In: T. Szwedziki, G.R. Baird and T.N. Little (eds.), *Proceedings of Western Australian Conference of Mining Geomechanics.* Kalgoorlie, Western Australia: Curtin University, WASM (1992) pp. 235–246.
5. L.N. Germanivich and A.V. Dyskin, Virial expansions in problems of effective characteristics. Part I. General concepts. *J. Mech. Compos. Mater.* 30(2) (1994) 222–237.
6. H.-B. Mühlhaus, A. Dyskin, E. Pasternak and D. Adhikary, Non-standard continuum theories in geomechanics: theory, experiments and analysis. In: R.C. Picu and E. Kremple (eds.), *Proceedings of the Fourth International Conference on Constitutive Laws for Engineering Materials*. Rensselaer Polytechnic Institute: Troy, New York (1999) pp. 321–324.

7. R.D. Mindlin and H.F. Tiersten, Effects of couple-stresses in linear elasticity. *Arch. Ration. Mech. Anal*. 11 (1962) 415–448.
8. A.C. Eringen, Linear theory of micropolar elasticity. *J. Math. Mech.* 15 (1966) 909–923.
9. A.C. Eringen and C.B. Kafadar, Polar field theories. In: A.C. Eringen (ed.), *Continuum Physics,* Volume IV, Part I. New York: Academic Press (1976) pp. 4–73.
10. P. Germain, La méthode des puissances virtuelles en mécanique des milieux continues. Première partie. Théorie du second gradient. *J. Mécanique* 12 (1973) 235–274.
11. P. Germain, The method of virtual power in continuum mechanics. Part 2: Microstructure. *SIAM J. Appl. Math*. 25 (1973) 556–575.
12. G.A. Maugin, The method of virtual power in continuum mechanics: application to coupled fields. *Acta Mech.* 35 (1980) 1–70.
13. I.A. Kunin, *Elastic Media with Microstructure 1. One-dimensional Models*. Berlin, Heidelberg, New York: Springer-Verlag (1982) 291 pp.
14. E. Kröner, The problem of non-locality in the mechanics of solids: review of the present status. In: J.A. Simmons, R. de Wit and R. Bullough (eds.), *Fundamental Aspects of Dislocation Theory*. National Bureau of Standards Special Publication 317, Vol. II. Washington: National Bureau of Standards (1970) pp. 729–736.
15. E. Kröner and B.K. Datta, Non-local theory of elasticity for a finite inhomogeneous medium – A derivation from lattice theory. In: J.A. Simmons, R. de Wit and R. Bullough (eds.), *Fundamental Aspects of Dislocation Theory,* National Bureau of Standards Special Publication 317, Vol. II. Washington: National Bureau of Standards (1970) pp. 737–746.
16. I.A. Kunin and A.M. Waisman, On problems of the non-local theory of elasticity. In: J.A. Simmons, R. de Wit and R. Bullough (eds.), *Fundamental Aspects of Dislocation Theory,* National Bureau of Standards Special Publication 317, Vol. II. Washington: National Bureau of Standards (1970) pp. 747–759.
17. I.A. Kunin, *Elastic Media with Microstructure II . Three-dimensional Models*. Berlin, Heidelberg, New York: Springer-Verlag (1983) 272 pp.
18. A.C. Eringen, Non-local continuum description of lattice dynamics and application. In: J. Chandra and R.P. Srivastav (eds.), *Constitutive Models of Deformation*. Philadelphia: SIAM (1987) pp. 59–80.
19. A.C. Eringen, Non-local polar field theories. In: A.C. Eringen (ed.), *Continuum Physics,* Volume IV, Part III. New York: Academic Press (1976) pp. 205–264.
20. A.C. Eringen, Non-local continuum mechanics and some application. In: A.O. Barut (ed.), *Non-linear Equations in Physics and Mathematics*. Dordrecht: D. Reidel Publishing Company (1978) pp. 271–318.
21. R. de Borst, A. Benallal and R.H.J. Peerlings, On gradient-enhanced damage theories. In: N.A. Fleck and A.C.F. Cocks (eds.), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) pp. 215–226.
22. G. Pijaudier-Cabot and Z.P. Bazant, Non-local damage theory. *J. Engng. Mech*. 113 (1987) 1512–1533.
23. Z.P. Bazant and G. Pijaudier-Cabot, Non-local continuum damage, localization instability and convergence. *J. Appl. Mech*. 55 (1988) 287–293.
24. G. Pijaudier-Cabot, Non-local damage. In: H.-B. Mühlhaus (ed.), *Continuum Models for Materials with Microstructure* (Chapter 4). Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons (1995) pp. 105–143.
25. H.-B. Mühlhaus, Continuum models for layered and blocky rock. In: J.A. Hudson (ed.), *Comprehensive Rock Engineering: Principles, Practice & Projects,* Invited Chapter for Vol. II: Analysis and Design Methods. Oxford, New York: Pergamon Press (1993) pp. 209–230.
26. N.V. Zvolinskii and K.N. Shkhinek, Continual model of laminar elastic medium. *Mech. Solids* 19(1) (1984) 1–9.
27. D.P Adhikary and A.V. Dyskin, A Cosserat continuum model for layered materials. *Comp. Geotechn.* 20 (1997) 15–45.
28. H.-B. Mühlhaus, A relative gradient model for laminated materials. In: H.-B. Mühlhaus (ed.), *Continuum Models for Materials with Microstructure,* (Chapter 13). Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons (1995) pp. 451–482.
29. H.-B. Mühlhaus and P. Hornby, A relative gradient theory for layered materials. *J. Phys. IV France* 8 (1998) 269–276.
30. H.-B. Mühlhaus and P. Hornby, A beam theory gradient continua. In: R. de Borst and E. van der Giessen (eds.), *Material Instabilities in Solids* (Chapter 32). Chichester, New York: John Wiley & Sons (1998) pp. 521–532.

31. J. Sulem and H.-B. Mühlhaus, A continuum model for periodic two-dimensional block structures. *Mech. Cohesive-Frictional Mater.* 2 (1997) 31–46.
32. H.-B. Mühlhaus and I. Vardoulakis, The thickness of shear bands in granular materials. *Géotechnique* 37 (1987) 271–283.
33. H.-B. Mühlhaus, R. de Borst and E.C. Aifantis, Constitutive models and numerical analyses for inelastic materials with microstructure. In: G. Beer, J.R. Booker and J. Carter (eds.), *Computing Methods and Advances in Geomechanics*. Rotterdam: Balkema (1991) pp. 377–385.
34. C.S. Chang and L. Ma, Elastic material constants for isotropic granular solids with particle rotation. *Int. J. Solids Struct.* 29 (1992) 1001–1018.
35. H.-B. Mühlhaus and F. Oka, Dispersion and wave propagation in discrete and continuous models for granular materials. *Int. J. Solids Struct.* 33 (1996) 2841–2858.
36. H.-B. Mühlhaus and P. Hornby, On the reality of antisymmetric stresses in fast granular flows. In: N.A. Fleck and A.C.F. Cocks (eds.), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) pp. 299–311.
37. G.N. Wells and L.J. Sluys, Partition-of-unity for fracture of brittle materials. In: H.-B. Mühlhaus, A.V. Dyskin and E. Pasternak (eds.), *Bifurcation and Localization in Geomechanics*. Lisse: Swets & Zeitlinger (2001) pp. 169–176.
38. R. Hill, Elastic properties of reinforced solids: some theoretical principles. *J. Mech. Phys. Solids* 11 (1963) 357–372.
39. T. Mori and K. Tanaka, Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metal.* 21 (1973) 571–574.
40. R.M. Christensen, *Mechanics of Composite Materials*. New York: John Wiley & Sons (1979) 348 pp.
41. Z. Hashin, The differential scheme and its application to cracked materials. *J. Mech. Phys. Solids* 36 (1988) 719–734.
42. M. Kachanov, Effective elastic properties of cracked solids: critical review of some basic concepts. *Appl. Mech. Rev.* 45(8) (1992) 304–335.
43. S. Nemat-Nasser and H. Horii, *Micromechanics: Overall Properties of Heterogeneous Materials*. Amsterdam, London, New York, Tokyo: North-Holland (1993) 687 pp.
44. D. Krajcinovic, *Damage Mechanics*. Amsterdam, Lausanne, New York, Oxford, Shannon, Tokyo: Elsevier (1996) 761 pp.
45. B. Cambou, Micromechanical approach in granular materials. In: B. Cambou (ed.), *Behaviour of Granular Materials,* CISM Courses and Lectures, No 385. Wien, New York: Springer (1998) pp. 171–216.
46. G.N. Savin and L.P. Khoroshun, The problem of elastic constants of stoichastically reinforced materials. *Mekhanika sploshnoy sredy i rodstvennye problemy analiza* [Mechanics of continuous media and related problems of analysis]. Moscow: Nauka Press (1972) pp. 437–444 (in Russian).
47. L.P. Khoroshun, Methods of theory of random functions in problems of macroscopic properties of microinhomogeneous media. *Soviet Appl. Mech.* 14 (1978) 113–124.
48. J. Duffy and R.D. Mindlin, Stress-strain relation and vibrations of granular medium. *J. Appl. Mech*. 24 (1957) 585–593.
49. H. Deresiewicz, Stress-strain relations for a simple model of a granular medium. *J. Appl. Mech*. 25 (1958) 402–406.
50. S.A. Meguid and A.L. Kalamkarov, Asymptotic homogenization of elastic materials with a regular structure. *Int. J. Solids Struct.* 31 (1994) 303–316.
51. G.A. Vanin, *Micromechanics of Composite Materials*. Kiev: Naukova Dumka (1985) 302 pp (in Russian).
52. G.A. Maugin, *Non-linear Waves in Elastic Crystals*. Oxford: Oxford University Press (1999) 314 pp.
53. A.S.J. Suiker, R. de Borst and C.S. Chang, Micro-mechanically based higher-order continuum models for granular materials. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer (2000) pp. 249–274.
54. A.S.J. Suiker, R. de Borst and C.S. Chang, Micro-mechanical modelling of granular material. Part 1: Derivation of a second-gradient micro-polar constitutive theory. *Acta Mech.* 149 (2001) 161–180.
55. A.S.J. Suiker, R. de Borst and C.S. Chang, Micro-mechanical modelling of granular material. Part 2: Plane wave propagation in finite media. *Acta Mech.* 149 (2001) 181–200.
56. M. Satake, Three-dimensional discrete mechanics of granular materials. In: N.A. Fleck and A.C.F. Cocks (eds.), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) pp. 193–202.

57. U. Tüzün and D.M. Heyes, Distinct element simulations and dynamic microstructural imaging of slow shearing granular flows. In: N.A. Fleck and A.C.F. Cocks (eds.), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) pp. 263–274.
58. C. Thornton, Microscopic approach contributions to constitutive modelling. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer (2000) pp. 193–208.
59. H.-B. Mühlhaus, L. Moresi and H. Sakaguchi, Discrete and continuum modelling of granular materials. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer (2000) pp. 209–224.
60. G. Gudehus, A comprehensive constitutive equation for granular materials. *Soils and Foundations* 36 (1996) 1–12.
61. P.J. Digby, The effective elastic moduli of porous granular rocks. *J. Appl. Mech*. 16 (1981) 803–808.
62. K. Walton, The effective elastic modulus of a random packing of spheres. *J. Mech. Phys. Solids* 35 (1987) 213–226.
63. R.J. Bathurst and L. Rothenberg, Micromechanical aspects of isotropic granular assemblies with linear contact interactions. *J. Appl. Mech*. 55 (1988) 17–23.
64. C.S. Chang, Micromechanical modelling of constitutive relations for granular material. In: M. Satake and J.T. Jenkins (eds.), *Micromechanics of Granular Materials*. Amsterdam: Elsevier Science Publishers B.V. (1988) pp. 271–279.
65. J.T. Jenkins, Volume change in small strain axisymmetric deformations of a granular material. In: M. Satake and J.T. Jenkins (eds.), *Micromechanics of Granular Materials*. Amsterdam: Elsevier Science Publishers B.V. (1988) pp. 245–252.
66. B. Cambou, F. Dedecker and M. Chaze, Relevant local variables for the change of scale in granular materials. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer (2000) pp. 275–290.
67. N.A. Fleck and A.C.F. Cocks (eds.), *IUTAM Symposium on Mechanics of Granular and Porous Materials*. Dordrecht: Kluwer Academic Publishers (1997) 450 pp.
68. E. Pasternak and H.-B. Mühlhaus, Non-classical continua for modelling of granulate materials. In: J.P. Denier and E.O. Tuck (eds.), *The 2001 ANZIAM Applied Mathematics Conference Abstracts*. Barossa Valley, South Australia: University of Adelaide (2001) p. 64.
69. E. Pasternak and H.-B. Mühlhaus, Cosserat continuum modelling of granulate materials. In: S. Valliappan and N. Khalili (eds.), *Computational Mechanics – New Frontiers for New Millennium*. Amsterdam: Elsevier Science (2001) pp. 1189–1194.
70. E. Pasternak and H.-B. Mühlhaus, Large deformation Cosserat continuum modelling of granulate materials. In: L. Zhang, L. Tong and J. Gal (eds.), *Applied Mechanics. Progress and Application. ACAM 2002. The Third Australasian Congress on Applied Mechanics* New Jersey, London, Singapore: World Scientific (2002) pp. 389–396.
71. E. Pasternak and H.-B. Mühlhaus, Cosserat and non-local continuum models for problems of wave propagation in fractured materials. In: X.L. Zhao and R.H. Grzebieta (eds.), *Structural Failure and Plasticity (IMPLAST2000)*. Amsterdam: Pergamon (2000) pp. 741–746.
72. E. Pasternak and H.-B. Mühlhaus, A non-local Cosserat model of heterogeneous materials: 1D structures. In: A.V. Dyskin, X. Hu and E. Sahouryeh (eds.), *Structural Integrity and Fracture*. Lisse: Swets & Zeitlinger (2002) pp. 107–114.
73. K.F. Graff, *Wave Motion in Elastic Solids*. New York: Dover Publications (1975) 649 pp.
74. J.-P. Kahane, *Some Random Series of Functions*, 2nd edition. Cambridge: Cambridge University Press (1985) 305 pp.

Journal of Engineering Mathematics (2005) 52: 231–250
DOI 10.1007/s10665-004-2117-2

Some fundamental aspects of the continuumization problem in granular media

JOHN F. PETERS
U.S. Army Engineer Research and Development Center, Geotechnical and Structures Laboratory, Vicksburg, Mississippi 39180, U.S.A.

Received 18 September 2003; accepted in revised form 6 August 2004

Abstract. The central problem of devising mathematical models of granular materials is how to define a granular medium as a continuum. This paper outlines the elements of a theory that could be incorporated in discrete models such as the Discrete-Element Method, without recourse to a continuum description. It is shown that familiar concepts from continuum mechanics such as stress and strain can be defined for interacting discrete quantities. Established concepts for constitutive equations can likewise be applied to discrete quantities. The key problem is how to define the constitutive response in terms of truncated strain measures that are a practical necessity for analysis of large granular systems.

Key words: continuum mechanics, granular media, homogenization, stress

1. Introduction

Mathematicians and physicists who deal with granular media as a continuum must make a significant departure from traditional models. Attempts to classify granular states in terms of solid–fluid–gas phases have failed to produce a cogent theory of granular-media behavior [1], for while granular media ostensibly display all three states, for each case physical phenomena are observed that seem to lie outside of classical behavior. For most materials, there is a large separation between the micro-scale of atoms and molecules and macro-scale of measurements. This large separation of scales allows effective averaging on microscopic effects in time and space to create a continuum. For granular media, the fundamental particle is itself macroscopic, thus greatly reducing the importance of atomic-scale processes that are typically assumed to control properties. Thus, a clastic, such as sand, has more in common with the clutter flowing out of a closet than with the minerals of the rock from which it is derived. Although granular media might display the distinct states of solid, fluid, and gas, it is difficult to apply traditional continuum descriptions of these phases. The pressure dependence that controls dissipation in granular solids gives rise to profound mathematical difficulties in continuum formulations for stress analysis and wave propagation [2]. Flowing sand exhibits complex dynamic behavior not described by a Naiver–Stokes equation. Shear is not resisted by a simple viscosity, but by a combination of rate-independent processes including dilation, friction, and size separation. At variance to a gas, particle collisions are inelastic. In a gas-like state, granular media displays clustering, with formation of transient structures of particle chains. None of these features can be accommodated easily into a continuum-mechanics framework by simple readjustments of classical relationships of elasticity, plasticity, fluid mechanics, or gas dynamics. To understand particles as a "medium", it is necessary to build a theory from a particle-scale level in much the same manner that

thermodynamics was developed in the late 19th century. This paper draws attention to key aspects to modeling granular media as a continuum solid.

The discrete-element method (DEM) introduced by Cundal and Strack [3] provides a virtual laboratory for the study of granular physics by which it is possible to obtain measurements from simulations that would otherwise be impossible from physical experiments. In the absence of measurements made at the scale of supposed micro-mechanical mechanisms, the theories are essentially phenomenological. With the DEM, particle-scale theories become particularly compelling because they can be built on detailed measurement of deformations and forces at the particle scale. However, most theoretical efforts are couched in terms of continuum mechanics with the expressed goal of devising an equivalent continuum to represent the granular medium. An example is the procedure of Chang and Liao [4,5] where the discrete particle translation and rotation are "driven" by continuum motions. By application of the virtual work principle, continuum stress variables that are conjugate to the deformation variables can be obtained.

The motivation of continuumization granular for medium is reduction of the degrees of freedom that must be dealt with using a DEM approach. A DEM simulation is practically limited to a few million particles, which is equivalent to a handful of sand. Thus, to apply DEM to a prototype-scale problem typically requires use of oversized particles, which introduces scaling errors in the solution. By devising a constitutive equation of the equivalent medium, one can presumably employ a numerical solution to the partial differential equations that result from the continuumization process. However, at the particle scale, granular media do not behave as continua. In many problems, it is found that the kinematical freedom found in the DEM is a major part of the physics and that even a crude particle model can yield realistic results nearly impossible to obtain with a continuum-based approach [6,7]. In addition, the equivalent continuum response is not simple. As a minimum, the constitutive response of the equivalent continuum must contain higher-order terms to avoid mathematical ill-posedness inherent in frictional media. One might ask if the continuum description aids in making analysis of granular media tractable, or does it add a layer of abstraction that hinders progress.

The contribution of this paper to the growing discussion of granular continua is to propose an alternative tack by laying groundwork for analyses based on DEM. The question of continuumization is addressed by asking whether an equivalentcontinuum is in fact required. It is shown that the concepts of stress and strain can be conceived without the assumption of a continuum. Additionally, it is argued that plasticity effects caused by particle slip become more clear in a discrete medium because the oxymoronic notion of non-affine continuum motion is unnecessary. Thus, the principal tools needed to devise a mechanical theory of particulate media are provided. It is proposed that a fruitful avenue of research might be the behavior of coarsened DEM systems, including convergence properties of simulated granular media and averaged inter-particle contact laws that provide the sought-for equivalent behavior.

1.1. Notation

This paper employees two types of notation. When expressions contain subscripts, the standard indices notation of continuum mechanics is used. Tensor components are indicated by subscripts and repeated subscripts indicate summation. The identity tensor is signified as δ_{ij} and e_{ijk} is the permutation symbol. The superscript p signifies a quantity associated with a particle and superscript c signifies a quantity associated with a contact between two particles. Summation is not implied on repeated superscripts, but instead is explicitly indicated by

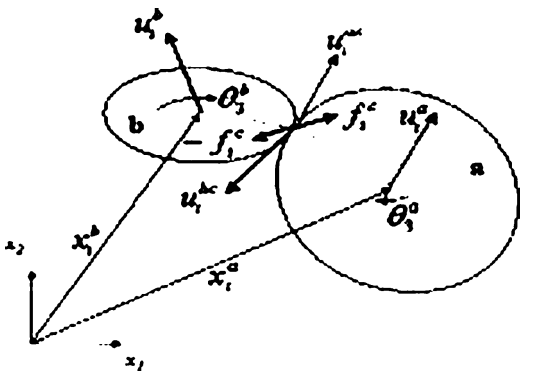


Figure 1. Kinematics of contact by a particle pair.

a summation symbol. The operation $|\cdot|$ is defined as $|\mathbf{f}| \equiv \sqrt{f_i f_i}$. Partial differentiation is denoted by $\nabla_i(\cdot) \equiv \partial(\cdot)/\partial x_i$.

In Section 4 a matrix notation is adopted to simplify the presentation of generalized stress and strain. The matrix operations are presented in the standard notation of linear algebra. The definitions of the matrices are given in Appendix A.

2. The discrete-element method

The discrete-element method provides a computational tool to study particulates without introducing the complications of a continuum theory. In DEM, the particles are treated as distinct interacting bodies. Interactions between particles are described by contact laws that define forces and moments created by relative motions of the particles. The motion of each particle that results from the net forces and moments are obtained by integrating Newton's laws. Thus, the particles are not treated as a medium. Rather, the medium behavior emerges from the interactions of the particles comprising the assemblage. The emergent behavior of the group is governed by relatively simple physical laws that obviate the need for complicated constitutive relationships [8].

The interaction (or contact) forces arise from relative motion between contacting particles. As shown in Figure 1, the motion of each individual particle is described by the velocity of the particle center and the rotation about the center. The branch vector between particle centers, $x_i^A - x_i^B$, is also the difference between the respective radii vectors that link the particle centers to the contact $r_i^A - r_i^B$. With this nomenclature, the relative motion at contact c between particles A and B is given by

$$\dot{\Delta}_i^c = \dot{u}_i^A - \dot{u}_i^B + e_{ijk}\left(r_j^A \dot{\theta}_k^A - r_j^B \dot{\theta}_k^B\right). \tag{1}$$

Under rigid-body motion there is no relative motion at any contacts. Rigid-body translation consists of $u_i^A = u_i^B$, with $\dot{\theta}_k^A = \dot{\theta}_k^B = 0$ for all particle pairs. Rigid-body rotation couples the particle rotation to the rotation of the particle assemblage. In this case, $\dot{\theta}_k^A = \dot{\theta}_k^B = \Omega_k$, where Ω_k is the rotation common to all points in the granular domain. For rigid body rotation therefore,

$$\dot{u}_i^A - \dot{u}_i^B = -e_{ijk}\left(r_j^A - r_j^B\right)\dot{\Omega}_k. \tag{2}$$

In three dimensions, there are three sets of rigid-body translations and three sets of rigid-body rotations. As will be discussed more precisely, the contact slip is equivalent to deformation in the system whereas the contact force is equivalent to stress.

3. Continuumization

The term *continuumization* refers to the process of representing a medium composed of individual discrete particles as a continuum having equivalent mechanical properties. This process is distinguished from more commonly used *homogenization* by recognition that a continuum implies a smoothness of motion with a preservation of connectedness not possessed by the granular media. Generally, the smoothness is imparted on the average motions of the particles, where the average is based on some representative elementary volume (REV). Many schemes have been proposed for finding an equivalent continuum. Chang and Liao [4], and more recently Tordesillas and Walsh [9] have tied particle motions to a smooth velocity using the virtual-work principle. Bardet and Vardoulakis [10] employed a similar method to elucidate the issue of non-symmetric stress tensor in Cosserat media. Bagi [11,12] considered the equilibrium of locally defined volumes obtained from Voronoi tessellation on particle centers. Kruyt [13] inspected graph properties of assemblages to define strain and stress variables.

In fact, there are two hierarchies of the continuum. The most basic concept of the continuum is statistically based in which the REV is chosen such that the fluctuation of the volumetric average is small. In this case, the REV acts as a smoother allowing discrete fields to be modeled as continuous. At the next level, deformation is viewed as affine mapping from some initial state to a final deformed state. Discontinuities can arise in finite number, which are dealt with through appropriate jump conditions. Ubiquitous discontinuities, such as slip, are sub-REV scale features and must be dealt with through abstractions such as internal variables that are effective mathematically but conceptually opaque.

This section deals with the first notion of the continuum. Averaging is performed on the governing equations imposing the momentum balance at the REV scale. The averaged equilibrium of the sampled volume is expressed in terms of stress which is derived from the averaging process. It is not assumed *a priori* that motions are affine. Rather, deformation rate is measured in terms of quantities that are thermodynamic conjugates to the stress. The sampled volume is said to be undergoing rigid-body motion if all relative motions among contacts are zero. The rigid-body motion have a null projection onto the space of deformations, which in effect, define those deformations. The projection of rigid-body motions onto the particle forces give rise to the equilibrium relationships for the sampled volume.

3.1. Properties of the average

For some function or operator f defined in the discrete medium contained within domain Ω, the average $\bar{f}$ of that quantity is defined at a point $x_i \in \Omega$ as

$$\bar{f}(x_i) = \int_{\Omega} \phi(x_i - x_i'; x_i) f(x_i') \mathrm{d}x_i'. \tag{3}$$

The present analysis considers an infinite domain to avoid influence of boundary terms. We specify that $\phi(x_i - x_i'; x_i) = \phi(x_i - x_i')$ where $\phi(x_i - x_i') = 0$ as $|x_i - x_i'| \to \infty$. Specifically, ϕ is assumed to have compact support, thus allowing a finite sampling volume $V \in \Omega$. With these restrictions it is necessary only that the integral of the weighting function over its range satisfy

$$\int_{\Omega} \phi(x_i - x_i') \mathrm{d}x_i' = \int_{V} \phi(x_i - x_i') \mathrm{d}x_i' = 1. \tag{4}$$

3.2. Weak equilibrium statement

The sampled space is composed of N^p particles, each with density ρ_s and volume V^p, accelerating at a_i^p. Interactions among particles occur at contacts through forces, f_i^c and moments m_i^c. For a typical particle, p, having N_c^p contacts, labeled $c \in N_c^p$, conservation of linear momentum requires

$$\sum_c^{N_c^p} f_i^c = V^p \rho^s a_i^p. \tag{5}$$

Each particle likewise rotates with acceleration ω_k^p, with rotational inertia I. For conservation of rotational momentum,

$$\sum_c^{N_c^p} \left(e_{ijk} r_j^c f_i^c + m_k^c \right) = \rho^s I \omega_k^p, \tag{6}$$

where r_j^c is the vector that connects the contact with the rotational center. Equations (5) and (6) must be satisfied for each particle within the domain. We wish to describe the equilibrium over the scale of the sampled volume as a weak form of the particle-scale conservation equations. That is, the weak form of the conservation statements requires that these statements are true when averaged via Equation (3). These averages are approximately

$$\sum_p^{N^p} \phi^p \sum_c^{N_c^p} f_i^c = \sum_p^{N^p} \phi^p V^p \rho^s a_i^p \tag{7}$$

and

$$\sum_p^{N^p} \phi^p \sum_c^{N_c^p} e_{ijk} r_j^c f_i^c = \sum_p^{N^p} \phi^p \rho^s I \omega_k^p. \tag{8}$$

3.3. Linear momentum

The process of averaging removes the spatial dependence on the particle quantities, such that the contact forces become statistical properties of the point for which the average is made. Accordingly, the summations on the left-hand side of in Equation (7) can be written as two summations over contacts within the sampling domain. The first summation is in terms of internal contacts which involve two particles. The second is for contacts at the domain boundary. For the internal contacts the contact force for respective particles A and B must satisfy $f_i^{Ac} = -f_i^{Bc} = f_i^c$, thus

$$\sum_c^{N^I} \left(\phi^A - \phi^B \right) f_i^c + \sum_c^{N^E} \phi^A f_i^{Ac} = \sum_p^{N^p} \phi^p V^p \rho^s a_i^p, \tag{9}$$

where the sum over the external contacts is zero owing to the prescribed properties of ϕ. Using the first term of the Taylor expansion for ϕ, the weighting can be applied at the contacts

$$\sum_c^{N^I} \nabla_j \phi^c \left(r_j^{Ac} - r_j^{Bc} \right) f_i^c = \sum_p^{N^p} \phi^p V^p \rho^s a_i^p, \tag{10}$$

where $\phi^c \equiv \phi(x_i - x_i^c)$, $x_i - x_i^c$ being the coordinate of the contact relative to the sampling point. The weighting function, ϕ^c is the only function of x_i in the summation.

3.4. Continuum stress

Within the domain the averaged linear momentum can be expressed as

$$\nabla_j \overline{\sigma}_{ij} = \overline{\rho a_i},$$

where

$$\overline{\sigma}_{ij} = \sum_c^{N^I} \phi^c \left(r_j^{Ac} - r_j^{Bc} \right) f_i^c. \tag{11}$$

In the event that mineral density ρ^s is the same for all grains and the acceleration is constant

$$\overline{\rho a_i} = \bar{\rho}\, \bar{a}_i,$$

where

$$\bar{\rho} = \rho^s \sum_p^{N^p} \phi^p V^p.$$

The summed quantity is the weighted integral over the solid volume and is the fraction of the sampled volume occupied by the solid phase. The density of that phase is ρ_s. The mean density $\bar{\rho}$ is the mass per *total* volume commonly referred to in soil mechanics as the total density.

3.5. Rotational momentum

Similar to the equation of linear momentum, the balance of rotational momentum can be written in the weak form to produce the counterpart of Equation (9)

$$\sum_c^{N^I} \nabla_l \phi^c \left(e_{ijk} \left(r_l^{Ac} r_j^{Ac} - r_l^{Bc} r_j^{Bc} \right) f_i^c + m_k^c r_l \right) + \sum_c^{N^E} \nabla_l \phi^c \left(e_{ijk} r_l^{Ac} r_j^{Ac} f_i^c + m_k^c r_l \right)$$
$$+ \sum_c^{N^I} \phi^c e_{ijk} \left(r_j^{Ac} - r_j^{Bc} \right) f_i^c + \sum_c^{N^E} \phi^c e_{ijk} r_j^{Ac} f_i^c = \sum_p^{N^p} \phi^p \rho^s I \omega_k^p. \tag{12}$$

Similar to Equation (11), the continuum expression can be written as

$$\nabla_l \overline{\mu}_{kl} + e_{ijk} \overline{\sigma}_{ij} = \overline{\rho^s I \omega_k^p},$$

where the $\overline{\mu}_{kl}$ is a coupled stress defined by

$$\overline{\mu}_{kl} = \sum_c^{N^I} \phi^c \left(e_{ijk} \left(r_l^{Ac} r_j^{Ac} - r_l^{Bc} r_j^{Bc} \right) f_i^c + m_k^c r_l \right). \tag{13}$$

Note that in the absence of contact couples, m_k^c, the coupled stress arises entirely from contact forces. In general the continuum stress $\overline{\sigma}_{ij}$ is not symmetric in the presence of a gradient in $\overline{\mu}_{kl}$ even if the contact moments are zero. The contribution of the contact forces to the asymmetry is a result of the resistance of the *finite* sampling volume to resist flexure-like deformation modes.

3.6. Work on unbalanced rotational forces

The presence of the moment gradient term leaves an unbalanced component in the moments created by the contact forces about a common center given in Equation (12). This moment does work against the rigid body rotation rate Ω_k as given by W_r

$$\dot{W}_r = \sum_c^{N^I} \phi^c e_{ijk} \left(r_j^{Ac} - r_j^{Bc} \right) f_i^c \Omega_k. \tag{14}$$

3.7. Power relationship

The deformation rate of the continuum is described by $\overline{D}_{ij}$, the symmetric part of the deformation tensor, Ω_k, the rotation of the sampling volume, and $\dot{\phi}_{kl}$ the rate of curvature. The deformation and curvature are linked to their respective particle quantities through a power balance whereby the work performed by the stress variables and their conjugate deformation variables is equal to that of contact forces and their conjugate contact motions,

$$\int_V \left(\overline{D}_{ij} \overline{\sigma}_{ij} + \Omega_k e_{ijk} \overline{\sigma}_{ij} + \dot{\phi}_{kl} \overline{\mu}_{kl} \right) \mathrm{d}V = \sum_c^{N_c} f_i^c \left[\Delta u_i^{AB} + e_{ijk} \left(r_j^A \omega_k^A - r_j^B \omega_k^B \right) \right] + W_r. \tag{15}$$

In view of Equations (11), (13), and (14)

$$\begin{aligned} &\sum_c^{N^I} \phi^c f_i^c \left(r_j^A - r_j^B \right) \left(\overline{D}_{ij} + e_{ijk} \Omega_k \right) + \sum_c^{N^I} \phi^c f_i^c e_{ijk} \left(r_l^A r_j^A - r_l^B r_j^B \right) \dot{\phi}_{kl} \\ &\quad = \sum_c^{N_c} f_i^c \left[\Delta u_i^{AB} + e_{ijk} \left(r_j^A \omega_k^A - r_j^B \omega_k^B \right) \right] + W_r. \end{aligned} \tag{16}$$

We immediately conclude that W_r cancels the work term associated with the motion $e_{ijk}\Omega_k$ on the left-hand side. The equality is satisfied for the particular case where

$$\dot{u}_i^{AB} + e_{ijk} \left(r_j^A \omega_k^A - r_j^B \omega_k^B \right) = \phi^c \left(\overline{D}_{ij} \left(r_j^A - r_j^B \right) + e_{ijk} \left(r_l^A r_j^A - r_l^B r_j^B \right) \dot{\phi}_{kl} \right). \tag{17}$$

By this procedure the particle centers are forced to follow the affine motion of the continuum, an obviously crude restriction if the constitutive relationships are to be built from the particle contact laws. The restriction in motion is the result of attempting to describe the many degrees of freedom of the finite sampling volume by the limited degrees of freedom represented by the deformation and curvature. This restriction can only be removed by including higher-order terms in the description of deformation. Another problem with Equation (17) is that the estimation of $\overline{D}_{ij}$ and $\dot{\phi}_{kl}$ from the motions of the particles requires an inversion of Equation (17). Both issues will be addressed from a more generalized conception of stress and strain.

4. Generalized stress and strain

Consider an assemblage of particles for which we wish to have measures of stress, σ, and strain, ϵ. The forces in the assemblage consist of contact forces between particles and external forces acting at the boundary of the sampling volume or as body forces. Whereas each particle has six degrees of freedom, the sampled volume has six N^p degrees of freedom, N_R of which are rigid-body motions that involve no relative movement between particles. The strain must

span a space of $M = N - N_R$ to completely capture all modes possible for the volume. Conjugate to this strain there are M generalized stress quantities. The task at hand is to define the operator that links the particle motion to the generalized strain and the forces to the generalized stress.

The translation and rotation of each particle is assigned to a generalized motion $\mathbf{u}$, which has a conjugate force and moment denoted as the generalized force $\mathbf{f}$ (See Appendix A). The force $\mathbf{f}$ is the vector of forces acting at the particle centers. These forces are the sum for each particle of the contact forces and body forces. As implied by the principle of virtual work, the conjugate pair $\mathbf{f}$ and $\dot{\mathbf{u}}$ should perform the same work as the conjugate σ and $\dot{\epsilon}$. Accordingly, the power balance is given by

$$\mathbf{f}^T \dot{\mathbf{u}} = \sigma^T \dot{\epsilon}. \tag{18}$$

Similarly, the contact forces, $\mathbf{f}_c$ and the conjugate contact motion, $\dot{\mathbf{\Delta}}_c$ should produce the same virtual work as σ and $\dot{\epsilon}$.

$$\mathbf{f}_c^T \dot{\mathbf{\Delta}}_c = \sigma^T \dot{\epsilon}. \tag{19}$$

This equivalence is the result of the fact that all mechanism of storing or dissipating energy are assumed to act at the contacts. (While this is a good approximation for the elastic processes, it is a fiction for the dissipative processes as is discussed in more detail in the section on internal variables.) The relationship between the particle motion and the contact motion of Equation (1) can be constructed in matrix form placing the relationships for each contact in the appropriate columns and rows of $\mathbf{M}$

$$\dot{\mathbf{\Delta}}_c = \mathbf{M}\dot{\mathbf{u}}. \tag{20}$$

By construction of $\mathbf{M}$, the relative motion between all contacts is zero for rigid-body motion. The rigid-body motions are known up to multiplicative constants. The direction vectors for the rigid body motions can be stored in the matrix $\mathbf{m}$, which constitute the null space of the operator $\mathbf{M}$,

$$\mathbf{Mm} = 0. \tag{21}$$

By the virtual work principle, the equilibrium relationship is given by

$$\mathbf{f}^T \mathbf{m} = \mathbf{0}. \tag{22}$$

The strain is a measure of relative movement among the particles, which is collectively perceived as the deformation of the assemblage. The strain rate is a linear transformation of the motion, $\dot{\mathbf{u}}$, given by Equation (23).

$$\dot{\epsilon} = \mathbf{B}\dot{\mathbf{u}}. \tag{23}$$

The null space of the operator $\mathbf{B}$ is given by the rigid-body motion, $\mathbf{m}$, as in Equation (21)[1]

$$\mathbf{Bm} = \mathbf{0}. \tag{24}$$

The relationship between $\mathbf{f}$ and σ immediately follows by substitution of Equation (23) in Equation (18).

$$\mathbf{f} = \mathbf{B}^T \sigma. \tag{25}$$

[1]The operator $\mathbf{B}$ must also satisfy the correct scale dependency. For example, in the particular case where $\mathbf{B}$ is a gradient operator, $\mathbf{Bx} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Similarly, in view of Equations (18) and (19), the total force can be related to the contact force by

$$\mathbf{f} = \mathbf{M}^T \mathbf{f}_c. \tag{26}$$

Therefore, the distribution matrix, **M**, and the strain-displacement matrix, **B**, are equivalent in purpose. Further, the contact conjugate pairs $(\mathbf{f}_c, \dot{\mathbf{\Delta}}_c)$ and stress $(\sigma, \dot{\epsilon})$ should be directly related. To show this, note that the rank of **B** is $N - N_R$ by construction, thus there exists a right inverse

$$\mathbf{BAx} = \mathbf{x}. \tag{27}$$

It follows immediately that the stress is given in terms of the contact forces

$$\sigma = \mathbf{H}^T \mathbf{f}_c, \tag{28}$$

where

$$\mathbf{H} = \mathbf{MA}. \tag{29}$$

Substitute Equation (28) in the power balance, Equation (19), to get

$$\dot{\mathbf{\Delta}}_c = \mathbf{H}\dot{\epsilon}. \tag{30}$$

Thus, the stress and strain are equivalent to the contact force and displacement, respectively. The essential difference between the two conjugate pairs is that the contact pair $(\mathbf{f}_c, \dot{\mathbf{\Delta}}_c)$ depends on specifics of the arrangement of grains in the assemblage whereas the pair $(\sigma, \dot{\epsilon})$ can be viewed as measures for the complete ensemble, which take on standardized forms such as stretch, shear, flexure, and higher-order forms associated with continuum deformation.

4.1. Remark on examples of generalized strain

In Appendix B, examples of generalized strain measures are given for particle assemblages in two dimensions. It is noted that as more particles are added to the assemblage, additional modes appear in the definition of generalized strain. Thus, in a finite volume of particles, motions that describe deformation must satisfy the relationship $M = N_{DOF} N_p - N_R$, where M is the number of deformation modes for the assemblage, N_{DOF} is the number of degrees of freedom for each particle, N_P is the number of particles, and N_R is the number of rigid-body modes for the assemblage. For one-dimensional motion, $N_R = 1$; for two-dimensional motion $N_R = 3$; and for three-dimensional motion $N_R = 6$. The implication to any homogenization procedure is that it is necessary to either define deformation through higher-order terms, or to tie the "internal" higher-order modes to lower-order terms. The later choice requires an approximation that is part of the constitutive response of the assemblage.

4.2. Remark on finite-element stabilization

The construction of the **B** operator in the example computations of Appendix B is somewhat contrived because in finding the vectors orthogonal to **m** to populate **B**, choices were made to produce the traditional strain-displacement operators for finite elements. Clearly, other operators could be constructed from linear combinations of the rows of **B**. In each case, Equation (24) would be satisfied. Each definition of **B** carries with it a definition of stress and its conjugate strain. All definitions are equivalent in the sense that one definition can be derived from the other.

The procedure for obtaining **B** is similar to that for suppressing zero-energy modes in under-integrated finite elements described in [14]. In that case, the zero-energy modes are deformation modes that are not resisted by internal stresses. The finite elements are stabilized by determining a set of orthogonal modes to fully span the space. The forces associated with the added modes are determined from a simple proportionality between the force and deformation. The constant of proportionality is selected as a suitably large number to suppress the spurious modes and thereby stabilize the element.

4.3. Remark on constitutive response

If the generalized stress and strain concepts were to be employed to develop a continuum model, the issue of constitutive response for higher modes would become problematic. In contrast, building the constitutive response for the particle-scale level is straightforward in view of the duality between **B** and **M**. In the DEM approximation, all work is performed by contact forces giving the work balance in Equation (19). Assuming an elastic contact response

$$\dot{\mathbf{f}}_c = \mathbf{K}\dot{\boldsymbol{\Delta}}_c, \tag{31}$$

which immediately leads to (for small deformations),

$$\dot{\sigma} = \mathbf{D}\dot{\epsilon}, \tag{32}$$

where

$$\mathbf{D} = \mathbf{H}^T\mathbf{K}\mathbf{H}. \tag{33}$$

This result is similar to relationships by Tordesillas and Walsh [9], if account is taken of difference in notation.

5. Truncated strain measures

For most micro-mechanical formulations, the measure of strain is restricted to a few terms of a Taylor expansion centered within the REV. Alternatively, for application to coarsened particulate systems alluded to in the introduction, only a few representative particles are present. In either case, the strain measure is truncated leaving an approximation to the contact displacement. By casting Equation (17) in matrix form, the strain is projected to the contact quantities by

$$\dot{\bar{\boldsymbol{\Delta}}}_c = \overline{\mathbf{H}}\dot{\bar{\epsilon}}, \tag{34}$$

where the number of degrees of strain degrees of freedom are less than $N - N_R$. Consider the case where the contact displacements are known, either from a DEM simulation or experiment. The most representative strain is that which minimizes the square of residual

$$R^2 = \frac{1}{2}(\dot{\boldsymbol{\Delta}}_c - \overline{\mathbf{H}}\dot{\bar{\epsilon}})^T(\dot{\boldsymbol{\Delta}}_c - \overline{\mathbf{H}}\dot{\bar{\epsilon}}), \tag{35}$$

which leads to the least-squares approximation

$$\dot{\bar{\epsilon}} = (\overline{\mathbf{H}}^T\overline{\mathbf{H}})^{-1}\overline{\mathbf{H}}^T\dot{\boldsymbol{\Delta}}_c. \tag{36}$$

The strain can be expressed in terms of the particle motions by

$$\dot{\bar{\epsilon}} = (\overline{\mathbf{H}}^T\overline{\mathbf{H}})^{-1}\overline{\mathbf{H}}^T\mathbf{M}\dot{\mathbf{u}}. \tag{37}$$

The stress is given by

$$\bar{\sigma} = \overline{\mathbf{H}}^T \mathbf{f}_c. \tag{38}$$

Note that an implication of Equation (38) is that $\overline{\mathbf{H}}$ can be obtained by converting Equations (11) and (13) into matrix form. Thus, given the averaged stress in terms of particle-scale quantities, the correct definition of the conjugate strain rate, in terms of particle-scale quantities, follows immediately via Equation (36).

5.1. Elastic constitutive response

The form of the relationships for stress and strain are analogous to Equations (23) and (28) and lead to a constitutive relationship that is similar to Equation (32), $\bar{\sigma} = \bar{\mathbf{D}}\bar{\epsilon}$. It is well known that the resulting stress–strain response is too stiff compared to that measured for the assemblage [4]. The stiff response is the result of the constraint imposed on the displacements and it is expected that as more terms are added to the strain, the response will match the assemblage response better. The situation can be understood by considering Equation (33) in a partitioned form.

$$\begin{bmatrix} \mathbf{D}_{uu} & \mathbf{D}_{ul} \\ \mathbf{D}_{ul}^T & \mathbf{D}_{ll} \end{bmatrix} \begin{Bmatrix} \dot{\bar{\epsilon}} \\ \dot{\epsilon}' \end{Bmatrix} = \begin{Bmatrix} \dot{\bar{\sigma}} \\ \dot{\sigma}' \end{Bmatrix}. \tag{39}$$

The barred quantities correspond to those strain terms included in the truncated strain. The primed quantities are the additional terms required to span the space. Note especially that in general $\bar{\mathbf{D}} \neq \mathbf{D}_{uu}$. Possible approaches to account for the truncated terms are considered as the following three cases:

5.1.1. *Case 1:* $\dot{\epsilon}' = 0$

This case corresponds to projecting the truncated strain measure (*e.g.* Equation (17)) such that

$$\mathbf{D}_{uu}\dot{\bar{\epsilon}} = \dot{\bar{\sigma}}, \tag{40}$$

with the internal stress terms given as

$$\dot{\sigma}' = \mathbf{D}_{ul}^T \dot{\bar{\epsilon}}. \tag{41}$$

This case corresponds physically to an interlocked structure that does not allow significant inter-grain motion.

5.1.2. *Case 2:* $\dot{\sigma}' = 0$

In this case, the internal stress terms are absent and the higher-order modes of deformation are not resisted. The constitutive equation for this case would be

$$(\mathbf{D}_{uu} - \mathbf{D}_{ul}\mathbf{D}_{ll}^{-1}\mathbf{D}_{ul}^T)\dot{\bar{\epsilon}} = \dot{\bar{\sigma}}, \tag{42}$$

with

$$\dot{\epsilon}' = -\mathbf{D}_{ll}^{-1}\mathbf{D}_{ul}^T \dot{\bar{\epsilon}}. \tag{43}$$

The higher-order strains become linear functions of the truncated strains. This case can be readily shown to result from minimizing $\dot{\epsilon}^T \mathbf{D} \dot{\epsilon}$ with respect to $\dot{\epsilon}'$, which corresponds physically to a relaxed structure where granular motion is essentially not resisted.

5.1.3. *Case 3:* $\dot{\epsilon}' \neq 0$ *and* $\dot{\sigma}' \neq 0$
The general stiffness relationship is

$$(\mathbf{D}_{uu} - \mathbf{D}_{ul}\mathbf{D}_{ll}^{-1}\mathbf{D}_{ul}^{T})\dot{\bar{\epsilon}} = \dot{\bar{\sigma}} - \mathbf{D}_{ul}\mathbf{D}_{ll}^{-1}\dot{\sigma}', \tag{44}$$

with

$$\dot{\epsilon}' = \mathbf{D}_{ll}^{-1}(\dot{\sigma}' - \mathbf{D}_{ul}^{T}\dot{\bar{\epsilon}}). \tag{45}$$

To close the system of equations a relationship is required to define either $\dot{\sigma}'$ or $\dot{\epsilon}'$. It is seen that in general, the *Case 2*, where $\dot{\epsilon}'$ is a linear function of $\dot{\bar{\epsilon}}$, corresponds to the fully relaxed case, where $\dot{\sigma}' = 0$, whereas *Case 1*, where $\dot{\sigma}'$ is a linear function of $\dot{\bar{\sigma}}$, corresponds to the fully locked where $\dot{\epsilon}' = 0$. The evolution of the structure from one case to another is a dissipative process implying an additional constitutive equation of the type

$$\dot{\sigma}' = \mathbf{A}\dot{\bar{\epsilon}} - \alpha\sigma'.$$

6. Internal variables

In the preceding discussion, a process is proposed in which the granular structure transitions from a completely locked assemblage, which is entirely elastic, to a completely relaxed structure where particle motions are unrestricted. For elastic deformation the topology of the assemblage remains constant such that contacts are not broken nor are new contacts made. Thus, the assemblage can be viewed as an elastic system in which the state of the system is determined entirely by the strain. From the standpoint of solid mechanics, the particulate medium could be homogenized into an elastic solid described by a higher-order strain theory. In the more general case of evolving structure, the state of the system is no longer described completely by the strain because the processes of contact creation and breakage are irreversible, a fact that must be considered for describing any homogenized continuum description of the medium. In continuum mechanics, such non-affine motions are addressed through internal variables, which account for the irreversible mechanisms within the medium. Internal variable theory provides the formal framework for irreversibility such that it is possible to devise models for materials undergoing irreversible changes in internal structure without violating the principles of thermodynamics (*e.g.* [15]). However, the theory does not describe specifically what physical process the internal variables represent, leaving their meaning in phenomenological models somewhat murky. Some authors have related internal variables in plasticity models to slips or diffusion of dislocations in some limited crystalline systems, but models that are more comprehensive generally express their meaning with vague reference to internal processes. More notable is the effort to relate the internal variables to those additional motions that are not captured by the strain. However, as emphasized by Valanis and Lee [16], such deformation must include non-affine motion to qualify as an internal variable. It is implied that such motion is accompanied by changes in the material topology, a process more easily visualized at the particle scale.

A theory that is built up from a particle-scale model has the advantage that the meaning of the internal variable is clear in the sense that internal variables can be computed from particle-scale measurement, whether based on experiments or simulation. The most obvious notion of an internal variable is that of contact slip. Generally, a Coulomb friction law is applied at the contact such that the contact motion includes a slip giving a contact law

$$f_i^c = K_{ij}^c(\Delta_j^c - \delta_j^c), \tag{46}$$

where δ_i^c is the component of contact motion that represents slip. The evolution law for the slip is

$$\dot{\delta}_i^c = \begin{cases} 0 & \text{for } |\tau_i^c| < \mu\sigma^c, \\ \dot{\Delta}_i^c & \text{for } |\tau_i^c| = \mu\sigma^c, \end{cases} \tag{47}$$

where $\sigma^c = f_i^c n_i^c$, $\tau_i^c = f_i^c - f_j^c n_j^c n_i^c$, and μ is the Coulomb friction coefficient. The combined result of Equations (46) and (47) ensure that $|\tau_i^c|$ can never exceed $\mu\sigma^c$.

In reality, slip is not this simple. In a time step, a contact can be created or lost, which at the macro level has the effect of slip. Consider a finite time increment $\Delta t = t_1 - t_o$, during which a particle makes contact and gives rise to a contact force f_i^c, beginning at t_c. The total motion of the particles making the contact is $\dot{\Delta}_i^c \Delta t$. Of this total motion, only the part $\dot{\Delta}_i^c (t_c - t_o)$ actually contributes to the contact force. That is, the average rate is

$$\begin{aligned} \dot{f}_i^c &= K_{ij}^c \left(\dot{\Delta}_j^c \Delta t - \dot{\Delta}_j^c (t_c - t_o) \right) / \Delta t \\ &= K_{ij}^c \left(\dot{\Delta}_j^c - \dot{\delta}_j^c \right) \end{aligned} \tag{48}$$

A simular argument can be made for contacts lost over a time step.

Within the REV many contacts are formed and lost over a time step. These contacts can be binned based on the contact orientation. Each orientation is represented by six rows of matrix **H** (see Appendix A). Thus, $\dot{\Delta}_i^c = H_{ik}^c \dot{\epsilon}_k$. The slip component can similarly be expressed in terms of a macroscopic variable by $\dot{\delta}_i^c = H_{ij}^c q_j$. The portion of the stress, denoted Q_i^c that arises from a particular contact direction is $Q_i^c = H_{ki}^c f_k^c$. The macroscopic law is

$$\dot{Q}_i^c = H_{ki}^c H_{jl}^c K_{kj}^c \left(\dot{\epsilon}_l - \dot{q}_l^c \right), \tag{49}$$

with

$$\sigma_i = \sum_c Q_i^c. \tag{50}$$

A key observation from Equation (49) is that q_j^c is a function of the contact direction implying that no single of q_j^c can describe slips for all contact directions. Thus, the complete constitutive law involves multiple internal variables.

Equations (49) and (50) are equivalent to those produced by an internal variable law based on a Helmholtz free energy $\psi(\epsilon, \mathbf{q}_c) = \sum_c \psi^c(\epsilon, \mathbf{q}^c)$ where

$$\psi^c = D_{ij}^c (\epsilon_i - q_i^c)(\epsilon_j - q_j^c), \tag{51}$$

with

$$\sigma_i = \frac{\partial \psi}{\partial \epsilon_i}, \tag{52}$$

and

$$Q_i^c = -\frac{\partial \psi}{\partial q_i^c}. \tag{53}$$

The evolution of $\mathbf{q}^c$ is an object for research. In any case, the evolution law must satisfy the second law of thermodynamics which implies

$$-\dot{q}_i^c \frac{\partial \psi}{\partial q_i^c} \geq 0. \tag{54}$$

For tangential slip, Equation (47) guarantees that inequality (54) is satisfied. For normal motion at contacts, the situation is more complicated. In DEM simulations, it is necessary to apply some sort of damping to the normal contact law. Physically, a hysteretic law is reasonable, because it maintains rate-independence of the dissipation, although in practice some viscous damping is also needed. These mechanism of energy dissipation obey the inequality (54) as well. The challenge is to make particle-scale measurements that clearly relate these mechanisms of particle interactions to the internal variables. The work of Zhang and Raueszahn [17] is notable in this regard for granular flows.

It should be recognized that $\mathbf{H}^c$ evolves in response to both elastic and inelastic changes to the particulate structure. Its evolution must conform with the second law given by the inequality

$$-\frac{\partial \psi}{\partial \mathbf{H}^c}\dot{\mathbf{H}}^c \geq 0. \tag{55}$$

That is, changes in $\mathbf{H}^c$ due to changes in internal structure cannot increase the free energy.

7. Conclusions

This paper began with a demonstration that continuum equations for the conservation of linear and rotational momentum can obtained from spatial averaging of the particle-scale momentum balance equations to produce the equations of a Cosserat medium. Two points emerged from this effort. First, the stress quantities need not be defined as averages of any local stress but arise naturally as result of spatial averaging of the equilibrium equations. Second, simple projections of conjugate deformation measures cannot be used in micro-scale constitutive relationships because such average motions are overly restricted. It is later shown that either higher-order deformation measures must be introduced, or the average motion as part of the constitutive relationship must drive particle-scale motions. In arriving at that conclusion, it is first shown that the concepts of stress and strain can be generalized to encompass the discrete quantities at the particle scale without recourse to the continuum concept. These generalized stresses and strains are shown to be equivalent to motions and forces at contacts, which allows deriving macro-scale constitutive equations from particle-scale relationship. The deformation conjugates to the averaged stress tensors are observed to be truncated forms of the generalized stress vector. The commonly used least-squares procedure is shown to provide strain measures that are conjugate to the averaged stress measures. Significantly, it is shown that the general concepts associated with constitutive theory for continua can be employed without the specific use of the continuum concept. It is concluded that a coarsened DEM could accordingly be developed.

Appendix A. Matrix nomenclature

The matrix nomenclature is introduced to make the general structure of the theory more clear. The displacement of each particle center is described by six degrees of freedom, $(u_x, u_y, u_z, \theta_x, \theta_y, \theta_z)$, which are arranged in $\mathbf{u}$ matrix as follows:

$$\begin{aligned}\mathbf{u}^T = &[(u_x^1, u_x^2, \ldots, u_x^{N_p}), (u_y^1, u_y^2, \ldots, u_y^{N_p}), (u_z^1, u_z^2, \ldots, u_z^{N_p}),\\ &(\theta_x^1, \theta_x^2, \ldots, \theta_x^{N^P})(\theta_y^1, \theta_y^2, \ldots, \theta_y^{N^P})(\theta_z^1, \theta_z^2, \ldots, \theta_z^{N^P})]\end{aligned} \tag{A.1}$$

The conjugate force vector $\mathbf{f}$ is arranged as $(f_x, f_y, f_z, m_x, m_y, m_z)$, so that

$$\begin{aligned}\mathbf{f}^T = &[(f_x^1, f_x^2, \ldots, f_x^{N_p}), (f_y^1, f_y^2, \ldots, f_y^{N_p}), (f_z^1, f_z^2, \ldots, f_z^{N_p}),\\ &(m_x^1, m_x^2, \ldots, m_x^{N^P})(m_y^1, m_y^2, \ldots, m_y^{N^P})(m_z^1, m_z^2, \ldots, m_z^{N^P})]\end{aligned} \tag{A.2}$$

Therefore, the quantity $\mathbf{f}^T\mathbf{u}$ is the sum of the work for the particle assemblage.

The stress and strain are simply

$$\sigma^T = [\sigma_1, \sigma_2, \ldots, \sigma_N] \tag{A.3}$$

and

$$\epsilon^T = [\epsilon_1, \epsilon_2, \ldots, \epsilon_N]. \tag{A.4}$$

The contact quantities consist of $(\Delta_x, \Delta_y, \Delta_z, \Theta_x, \Theta_y, \Theta_z)$, where Θ_i is the difference in rotation between contacting particles. These are grouped in $\Delta_\mathbf{c}$ in accordance with the contact number. The forces conjugate to the contact degree of freedom are $(f_x^c, f_y^c, f_z^c, m_x^c, m_y^c, m_z^c)$ and are likewise arranged in accordance with the contact number. The matrix $\mathbf{M}$ is constructed using the relationships between particle motion and contact motion,

$$\Delta_i^c = u_i^A - u_i^B + e_{ijk}(r_j^A\dot{\theta}_k^A - r_j^B\dot{\theta}_k^B). \tag{A.5}$$

and

$$\Theta_i = \theta^A - \theta^B. \tag{A.6}$$

The rows of $\mathbf{M}$ correspond to the components of $\mathbf{\Delta}_c$ and $\mathbf{f_c}$, whereas the columns correspond to the ordering of $\mathbf{u}$ and $\mathbf{f}$. For example, for $\mathbf{\Delta}_c = \mathbf{Mu}$,

$$\begin{Bmatrix}\vdots\\ \Delta_x\\ \Delta_y\\ \Delta_z\\ \Theta_x\\ \Theta_y\\ \Theta_z\\ \vdots\end{Bmatrix} = \begin{bmatrix}\ldots & 1 & 0 & 0 & \ldots & -1 & 0 & 0 & \ldots & 0 & -r_z & r_y & \ldots & 0 & r_y & -r_y & \ldots\\ \ldots & 0 & 1 & 0 & \ldots & 0 & -1 & 0 & \ldots & r_z & 0 & -r_x & \ldots & -r_z & 0 & r_x & \ldots\\ \ldots & 0 & 0 & 1 & \ldots & 0 & 0 & -1 & \ldots & -r_y & r_x & 0 & \ldots & r_y & -r_x & 0 & \ldots\\ \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & -1 & 0 & 0 & \ldots\\ \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & -1 & 0 & \ldots\\ \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0 & -1 & \ldots\end{bmatrix}. \tag{A.7}$$

The $\mathbf{H}$ matrix is arranged similarly except that the columns correspond to ordering of the strain components. The rigid body motion is specified up to an arbitrary constant and is given by

$$\begin{aligned}&\underline{\text{Translation}},\ \theta_1 = \theta_2 = \theta_3 = 0\\ &x = 1\\ &y = 1\\ &z = 1\\ &\underline{\text{Rotation}},\\ &u_i = e_{ij1}x_j, \quad \theta_1 = 1, \quad \theta_2 = \theta_3 = 0\\ &u_i = e_{ij2}x_j, \quad \theta_1 = 0, \quad \theta_2 = 1, \quad \theta_3 = 0\\ &u_i = e_{ij3}x_j, \quad \theta_1 = 0, \quad \theta_2 = 0, \quad \theta_3 = 1\end{aligned} \tag{A.8}$$

The **m** matrix is constructed such that its six rows constitute these rigid-body motions and the columns correspond to the ordering used for **u** and **f**.

$$\mathbf{m} = \begin{bmatrix} 1 & 0 & 0\,0\,0\,0 \\ 0 & 1 & 0\,0\,0\,0 \\ 0 & 0 & 1\,0\,0\,0 \\ 0 & -\mathbf{z} & \mathbf{y}\,1\,0\,0 \\ -\mathbf{z} & 0 & \mathbf{x}\,0\,1\,0 \\ \mathbf{y} & -\mathbf{x} & 0\,0\,0\,1 \end{bmatrix} \tag{A.9}$$

where **I** is the identity matrix and

$$\mathbf{1} = [1, 1, 1, \ldots, 1], \tag{A.10}$$

$$\mathbf{0} = [0, 0, 0, \ldots, 0], \tag{A.11}$$

$$\mathbf{x} = [x_1, x_2, x_3, \ldots, x_{NP}], \tag{A.12}$$

$$\mathbf{y} = [y_1, y_2, y_3, \ldots, y_{NP}], \tag{A.13}$$

$$\mathbf{z} = [z_1, z_2, z_3, \ldots, z_{NP}]. \tag{A.14}$$

Appendix B. Example of constructing B

B.1. General case

Construction of the **B** matrix amounts to filling out the **m** matrix given in Equation A.9 with orthogonal rows. Noting the structure of **m**, the process is facilitated by constructing **B** from sub-matrices as shown in Figure B1.

The properties of the sub-matrices follow from the orthogonality requirement, specifically, $\mathbf{L}_i \cdot \mathbf{1} = \mathbf{N}_i \cdot \mathbf{1} = 0$, and $\mathbf{L}_i \cdot \mathbf{x_j} = \mathbf{N}_i \cdot \mathbf{x_j} = 0$ for $i \neq j$. The sub-matrices are scaled such that $\mathbf{L}_i \cdot \mathbf{x_i} = \mathbf{N}_i \cdot \mathbf{x_i} = 1$ The sub-matrices $\mathbf{J}_i$ must satisfy the simpler orthogonality relationship $\mathbf{J}_i \cdot \mathbf{1} = 0$. The sub-matrices $\mathbf{M}_i$ are computed to satisfy the relationship

$$\mathbf{N}_x \cdot \mathbf{x}_x + \mathbf{N}_y \cdot \mathbf{x_y} + \mathbf{M}_z \cdot \mathbf{1} = 0,$$

for all permutations of x, y, and z.

The construction of **B** are given for specific cases in the sections that follow.

B.2. Three-particle assemblage

Consider an example of a three-particle assemblage moving within a plane. The rigid-body motions using the format of Equation (A.9), are

$$\mathbf{m} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0\,0\,0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0\,0\,0 \\ y_1 & y_2 & y_2 & -x_1 & -x_2 & -x_3 & 1\,1\,1 \end{bmatrix}. \tag{B.1}$$

The first row represent uniform translation in the x-direction, the second represents uniform translation in the y-direction, and the third is rigid-body motion in which the particle rotations are equal to the rigid-body rotation of the assemblage about the origin.

It is useful to consider the reduced form shown in Figure B2 that corresponds to two-dimensional motion in the x-y plane. The columns corresponding to translation in the z-direction and rotations about the x and y axes are removed. Accordingly, rows corresponding to gradients of those degrees of freedom are removed.

$$\left\{\begin{array}{c}\mathbf{m}\\ \text{- -}\\ \mathbf{B}\end{array}\right\}=\left[\begin{array}{cccccc}
\mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & -\mathbf{z} & \mathbf{y} & \mathbf{1} & \mathbf{0} & \mathbf{0}\\
-\mathbf{z} & \mathbf{0} & \mathbf{x} & \mathbf{0} & \mathbf{1} & \mathbf{0}\\
\mathbf{y} & -\mathbf{x} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}\\
\text{- -} & \text{- -} & \text{- -} & \text{- -} & \text{- -} & \text{- -}\\
\mathbf{L}_x & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{L}_y & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{L}_z & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_x & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_y & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_z\\
\mathbf{N}_y & \mathbf{N}_x & 0 & 0 & 0 & 0\\
\mathbf{0} & \mathbf{N}_z & \mathbf{N}_y & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{N}_z & \mathbf{0} & \mathbf{N}_x & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_y & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_z & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_x & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_z & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_y\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_x\\
\mathbf{N}_y & -\mathbf{N}_x & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}_z\\
\mathbf{0} & -\mathbf{N}_z & \mathbf{N}_y & \mathbf{M}_x & \mathbf{0} & \mathbf{0}\\
\mathbf{N}_z & \mathbf{0} & -\mathbf{N}_x & \mathbf{0} & \mathbf{M}_y & \mathbf{0}
\end{array}\right]$$

Figure B1. General form of **B**.

$$\left\{\begin{array}{c}\mathbf{m}\\ \text{- - -}\\ \mathbf{B}\end{array}\right\}=\left[\begin{array}{ccc}
\mathbf{1} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{1} & \mathbf{0}\\
\mathbf{y} & -\mathbf{x} & \mathbf{1}\\
\text{- - -} & \text{- - -} & \text{- - -}\\
\mathbf{L}_x & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{L}_y & \mathbf{0}\\
\mathbf{N}_y & \mathbf{N}_x & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{J}_y\\
\mathbf{0} & \mathbf{0} & \mathbf{J}_x\\
\mathbf{N}_y & -\mathbf{N}_x & \mathbf{M}_z
\end{array}\right]$$

Figure B2. Form of **B** for deformation in x–y plane.

The number of rows of the **B** matrix is given by $M = N_{DOF}N_p - N_R$, for two-dimensional motion, $N_{DOF} = 3$ and $N_R = 3$. Thus, $M = 6$ implying that each sub-matrix contains one row. Thus we set $\mathbf{L}_i = \mathbf{N}_i = \mathbf{J}_i$ for all i. The $\mathbf{L}_x$ sub-matrix is orthogonal to $(1, 1, 1)$ and (y_1, y_2, y_3) and can be computed from the cross product between these two vectors. Denoting the cross product as $\mathbf{L}_x^*$, then $\mathbf{L}_x = a\mathbf{L}_x^*$, where $a = 1/(\mathbf{L}_x^* \cdot \mathbf{x})$. The sub-matrix $\mathbf{L}_y$ can be found in a similar manner. It is easily verified by direct substitution that $\mathbf{L}_x^* \cdot \mathbf{x} = \mathbf{L}_y^* \cdot \mathbf{y}$.

$$\mathbf{B} = a\begin{bmatrix}
y_{32} & y_{13} & y_{21} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & x_{23} & x_{31} & x_{12} & 0 & 0 & 0\\
x_{23} & x_{31} & x_{12} & y_{32} & y_{13} & y_{21} & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & y_{31} & y_{13} & y_{21}\\
0 & 0 & 0 & 0 & 0 & 0 & x_{23} & x_{31} & x_{12}\\
x_{23} & x_{31} & x_{12} & y_{23} & y_{31} & y_{12} & M_\theta & M_\theta & M_\theta
\end{bmatrix}, \tag{B.2}$$

where $y_{IJ} = y_I - y_J$, $x_{IJ} = x_I - x_J$, and The sub-matrix $\mathbf{M}_\theta$ are equal and are given by

$$M_\theta = 2(y_{32}x_1 + y_{13}x_2 + y_{21}x_3).$$

The matrix corresponds to the strain-displacement matrix for a constant strain triangular finite element with terms included for the Cosserat rotations. The factor a is equal to twice the area defined by the line segments connecting the three particle centers. The first two rows of **B** correspond to the gradient of velocity in the x- and y-directions, respectively. The third row represents shear. The fourth and fifth rows correspond to the gradients of rotations in the x- and y-directions, respectively. The sixth row corresponds to the difference between the rigid body rotation of the assemblage and the average particle rotation.

B.3. Four-particle assemblage

In the case of a four-particle assemblage with two-dimensional motion, the strain operator would again contain terms for the strains, Cosserat curvatures, and rotations. In this case $N_p = 4$ requiring that $M = 9$. Thus, three rows must be added. However, the matrix shown in Figure B2 remains applicable implying that the sub-matrices contain multiple rows. For the special case where the particle rotation is not included, the matrix shown in Figure B2 is reduced to three rows with five independent vectors required. Thus, of the three vectors required, only one is associated with the rotations.

Consider the case of $\mathbf{L}_x$. The possible vectors that meet the orthogonality conditions can be obtained from a cross product in R^4. For any vector, $\mathbf{l}$ in R^4 that is not parallel with either $\mathbf{1}$ or $\mathbf{y}$, $\mathbf{L}_x$ is given by

$$\mathbf{L}_x = \begin{Bmatrix} l_2(y_4 - y_3) + l_3(y_2 - y_4) + l_4(y_3 - y_2) \\ l_1(y_3 - y_4) + l_3(y_4 - y_1) + l_4(y_1 - y_3) \\ l_1(y_4 - y_2) + l_2(y_1 - y_4) + l_4(y_2 - y_1) \\ l_1(y_2 - y_3) + l_2(y_3 - y_1) + l_3(y_1 - y_2) \end{Bmatrix}. \tag{B.3}$$

For any valid assemblage, $\mathbf{l} = \mathbf{x}$ is can be used to generate $\mathbf{L}_x$. For the particular case of a square array of particles with $\mathbf{x} = (-1, 1, 1, -1)$ and $\mathbf{y} = (-1, -1, 1, 1)$, $\mathbf{L}_x$ is found to be

$$\mathbf{L}_x = \begin{bmatrix} -1 & 1 & 1 & -1 \\ -4 & 4 & -4 & 4 \end{bmatrix}. \tag{B.4}$$

The first row corresponds to a uniform expansion of the particles in the x-direction. The second row represents a flexure mode, commonly referred to as an hour-glass mode in the finite-elements literature. The case for $\mathbf{L}_y$ is similar. The additional mode associated with rotation can be similarly computed by taking the cross product in R^4 for the set $\mathbf{1}$, $\mathbf{L}_x$, and $\mathbf{L}_y$. The result is a pattern corresponding to the flexure mode.

B.4. Nearest-neighbor configuration

The nearest-neighbor configuration consists of a central particle surrounded by three particles such that $\mathbf{x} = (0, -1, 1, 0)$ and $\mathbf{y} = (\sqrt{3}/2, 0, 0, \sqrt{3})$. This arrangement is similar to that used for the micro-mechanical modeling by Tordesillas and Walsh [9]. The relationships for this four particle assemblage is the same as the preceeding example. The sub-matrix $\mathbf{L}_x$ obtained in this case illustrates the difference made by the coordinate location of the fourth particle.

$$\mathbf{L}_x = \begin{bmatrix} 0 & -1 & 1 & 0 \\ -2\sqrt{3} & \dfrac{\sqrt{3}}{2} & \dfrac{\sqrt{3}}{2} & \dfrac{\sqrt{3}}{2} \end{bmatrix}. \tag{B.5}$$

As for the preceding example, the first row corresponds to a uniform stretch in the x-direction. However, the second row corresponds to a motion of the central particle relative to the three surrounding particle, which is quite unlike the flexure modes of the four-corner configuration. It is significant that these two modes are orthogonal; the deformation $\overline{D}_{ij}$ of Equation (17) provides no information on the magnitude of the second mode. Therefore, this mode is absent from the micro-mechanical model. Increasing the coordination number (by adding particles) would bring additional modes that would likewise be unaccounted for in a micro-mechanical model driven by $\overline{D}_{ij}$.

Acknowledgements

This paper resulted from research conducted under the AT22 Research Project Stress Transfer in Granular Media conducted at the U.S. Army Engineer Research and Development Center. Permission was granted by the Chief of Engineers to publish this information.

The author would also like to acknowledge Drs. Antoinette Tordesillas and Bruce Gardiner, and Mr. Stuart Walsh, all from the Department of Mathematics and Physics of the University of Melbourne, for their thoughtful review of this work. Also, the author greatly appreciates Dr. Ernie Heymsfield from the University of Arkansas for the many hours of discussion on fundamental ideas expressed in this paper.

References

1. H.M. Jaeger, S.R. Nager and R.P. Behringer, The physics of granular media. *Physics Today* 49 (1996) 32–38.
2. K.C. Valanis and J.F. Peters, Ill-posedness of the initial and boundary value problems in non-associative plasticity. *ACTA Mech.* 114 (1996) 1–25.
3. P.A. Cundal and O.D.L. Strack, A discrete numerical model for granular assemblies. *Geotechnique* 29 (1979) 47–65.
4. C.S. Chang and C.L. Liao, Constitutive relation for a particulate medium with the effect of particle rotation. *Int. J. Solids Struct.* 26 (1990) 437–453.
5. C.L. Liao, T.C. Chan, A.S.K. Suiker and C.S. Chang, Pressure-dependent elastic moduli of granular assemblies. *Int. J. Num. Anal. Methods Geomech.* 24 (2000) 265–279.
6. R.D. Hryciw, S.A. Raschke, A.M. Ghalib, D.A. Horner and J.F. Peters, Video tracking for experimental validation of discrete element simulations of large discontinuous deformations. *Computers Geotechn.* 21 (1997) 235–253.
7. D.A. Horner, J.F. Peters and A. Carrillo, Large scale discrete element modeling of vehicle-soil interaction. *ASCE J. Eng. Mech.* 127 (1979) 1027–1032.
8. P.A. Cundal, A discrete future for numerical modeling. In: B.K. Cook and R.P. Jensen (eds.), *Discrete Element Methods: Numerical Modeling of Discontinua*. ASCE Geotechnical Special Publication No. 117. Reston (VA): ASCE (2002) pp. 3–4.
9. A. Tordesillas and S. Walsh, Incorporating rolling resistance and contact anisotropy in micromechanical models of granular media. *Powder Technol.* 124 (2002) 106–111.
10. J.P. Bardet and I. Vardoulakis, The asymmetry of stress in granular media. *Int. J. Solids Struct.* 38 (2001) 353–367.
11. B. Bagi, Stress and strain in granular assemblies. *Mech. Materials* 22 (1996) 165–177.
12. B. Bagi, Microstructural stress tensor of granular assemblies with volume forces. *ASME J. Appl. Mech.* 66 (1999) 934–936.
13. N.P. Kruyt, Statics and kinematics of discrete Cosserat-type granular materials. *Int. J. Solids Struct.* 40 (2003) 511–534.
14. J.F. Peters and E. Heymsfield, Application of the 2-D constant strain assumption to FEM elements consisting of an arbitrary number of nodes. *Int. J. Solids Struct.* 40 (2003) 143–159.
15. K.C. Valanis and J.F. Peters, An endochronic plasticity theory with shear-volumetric coupling. *Int. J. Num. and Anal. Methods Geomech.* 15 (1991) 77–102.

16. K.C. Valanis and C.F. Lee, Endochronic plasticity: physical basis and applications. In: C.S. Desai and R.H. Gallagher (eds.), *Mechanics of Engineering Materials*. New York: Wiley (1984) pp. 591–609.
17. D.Z. Zhang and R.M. Raueszahn, Stress relaxation in dense and slow granular flows. *J. Rheology* 45 (1979) 1019–1041.

Journal of Engineering Mathematics (2005) 52: 251–264
DOI 10.1007/s10665-004-5662-9

Compression and shear of a layer of granular material

A.J.M. SPENCER
School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK. E-mail: anthony.spencer @nottingham.ac.uk

Received 20 May 2004; accepted in revised form 27 October 2004

Abstract. A classical problem in metal plasticity is the compression of a block of material between rigid platens. The corresponding problem for a layer of granular material that conforms to the Coulomb-Mohr yield condition and the double-shearing theory for the velocity field has also been solved. A layer of granular material between rough rigid plates that is subjected to both compression and shearing forces is considered. Analytical solutions are obtained for the stress and velocity fields in the layer. The known solutions for steady simple shear and pure compression are recovered as special cases. Yield loads are determined for combined compression and shear in the case of Coulomb friction boundary conditions. Numerical results which describe the stress and velocity fields in terms of the normal and shear forces on the layer at yield are presented for the case in which the surfaces of the platens are perfectly rough. Post-yield behaviour is briefly considered.

Key words: combined loading, compression, geological faults, granular material, shear

1. Introduction

The compression in plane strain of a rectangular block of ideal plastic-rigid material between rough rigid platens is a classical problem in metal plasticity that is described in the standard texts such as [1,2]. For a block that is long in comparison to its thickness there is an analytical solution, also classical, due to Prandtl [3]. The corresponding problem for a layer of granular material, assuming the Coulomb-Mohr failure condition, was studied by Hartmann [4], who obtained the solution for the stress field, and by Marshall [5], who extended Hartmann's stress solution and also determined the velocity field, on the basis of the non-dilatant double-shearing theory proposed by Spencer [6–8]. The problem is a model, for example, for a raft supporting a building resting on a layer of soil covering a rigid substrate, and also has geophysical applications with reference to geological faults.

Another problem in granular material mechanics that has been the subject of extensive study is that of simple shear of a layer of granular material. In principle this is a straightforward problem involving homogeneous stress and velocity fields, and is basic for the definition and measurement of the material parameters that characterize granular materials. Spencer [8,9] considered the simple shear of a layer of granular material using the non-dilatant double-shearing theory, and showed that there is a time-independent steady stress solution in which the material flows uniformly under constant shear stress, but that this solution is unstable (in the remainder of this paper we shall refer to this solution as the 'steady simple shearing solution'). It was also shown in [9] that there is an unsteady solution in which stress is time-dependent and uniform shear flow occurs in conjunction with decreasing shear stress.

However, there seems to be no systematic study of the case in which a granular layer between rough plates is subjected to both compression and shear. This problem is also of interest in modelling geological faults (see, for example, [10,11]). The main purpose of this

paper is to fill this gap by giving an analysis of the problem of compression and shear of a granular layer between rough platens, using the Coulomb condition for cohesionless material and the double-shearing theory for the velocity field.

The basic equations are described in Section 2. In Section 3 we derive an exact solution of these equations for a layer of granular material. In Sections 4 and 5 this solution is applied to the problem of a layer of granular material confined between rough rigid platens that exert both normal and tangential forces on the granular layer; Section 4 develops the solution for the stress field and the velocity field is found in Section 5. In Section 6 we present some numerical results and discussion.

2. General theory

All quantities are referred to a fixed system of rectangular Cartesian coordinates $Oxyz$. The components of the stress tensor σ are denoted as

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}, \tag{2.1}$$

and the components of the velocity $\mathbf{v}$ by (u, v, w). We consider plane strain in the (x, y) planes, so that $w = 0$, u and v are functions of x and y, and the relevant stress components are σ_{xx}, σ_{xy}, and σ_{yy}, all of which depend only on x and y. We write

$$p = -\frac{1}{2}\left(\sigma_{xx} + \sigma_{yy}\right), \quad q = \left\{\frac{1}{4}\left(\sigma_{xx} - \sigma_{yy}\right)^2 + \sigma_{xy}^2\right\}^{\frac{1}{2}}, \quad q \geq 0, \tag{2.2}$$

so that p and q are stress invariants that represent the mean in-plane hydrostatic pressure and the maximum shear stress, respectively. The stress angle ψ is defined by

$$\tan 2\psi = \frac{2\sigma_{xy}}{\sigma_{xx} - \sigma_{yy}}, \tag{2.3}$$

and is the angle that the principal stress axis associated with the algebraically greater principal stress makes with the x-axis (tensile stress is taken to be positive). Then the relevant stress components can be expressed as

$$\sigma_{xx} = -p + q\cos 2\psi, \quad \sigma_{yy} = -p - q\cos 2\psi, \quad \sigma_{xy} = q\sin 2\psi. \tag{2.4}$$

In soil mechanics terminology, the case $\cos 2\psi > 0$ corresponds to passive lateral pressure and $\cos 2\psi < 0$ corresponds to active lateral pressure.

The material is assumed to conform to the Coulomb–Mohr yield condition for cohesionless material

$$q \leq p\sin\phi, \tag{2.5}$$

where ϕ is the angle of internal friction and (2.5) holds as an equality whenever the material is undergoing deformation, in which case

$$\sigma_{xx} = q\left(\cos 2\psi - \operatorname{cosec}\phi\right), \quad \sigma_{yy} = -q\left(\cos 2\psi + \operatorname{cosec}\phi\right), \quad \sigma_{xy} = q\sin 2\psi. \tag{2.6}$$

The equations of equilibrium are, in plane strain and neglecting body forces

$$\frac{\partial \sigma_{xx}}{\partial x}+\frac{\partial \sigma_{xy}}{\partial y}=0,$$
$$\frac{\partial \sigma_{xy}}{\partial x}+\frac{\partial \sigma_{yy}}{\partial y}=0. \tag{2.7}$$

In quasi-static flows, when inertia terms are neglected, the Coulomb–Mohr condition (2.5) (as an equality) and the equilibrium equations can be expressed as a pair of first-order partial differential equations for q and ψ, as follows

$$(\cos 2\psi-\operatorname{co\,sec}\phi)\frac{\partial q}{\partial x}+\sin 2\psi\frac{\partial q}{\partial y}-2q\sin 2\psi\frac{\partial \psi}{\partial x}+2q\cos 2\psi\frac{\partial \psi}{\partial y}=0,$$
$$\sin 2\psi\frac{\partial q}{\partial x}-(\cos 2\psi+\operatorname{co\,sec}\phi)\frac{\partial q}{\partial y}+2q\cos 2\psi\frac{\partial \psi}{\partial x}+2q\sin 2\psi\frac{\partial \psi}{\partial y}=0. \tag{2.8}$$

These equations are hyperbolic, with characteristic curves (termed α-and β-lines) given by

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\tan\left(\psi\pm\frac{\phi}{2}\pm\frac{\pi}{4}\right) \tag{2.9}$$

where the lower and upper signs refer to the α-lines and β-lines respectively.

These equations for the stress field are quite generally accepted. To complete the material description it is necessary to specify a 'flow rule' that relates the stress to the deformation. The choice of an appropriate flow rule is still a matter of debate and has been discussed extensively in the literature. We adopt the 'double-shearing' theory for deformation of granular material. An accessible formulation of this theory, with references to earlier work, is given in Spencer [8]; other recent studies that make comparisons between the double-shearing theory and some alternative theories are by Gremaud [12], Alexandrov [13] and Zhu *et al.* [14]. According to the double-shearing theory, in plane strain the velocity components u and v satisfy

$$\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\cos 2\psi-\left(\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}\right)\sin 2\psi+\sin\phi\left(\frac{\partial u}{\partial y}-\frac{\partial v}{\partial x}+2\Omega\right)=0, \tag{2.10}$$

where

$$\Omega=\dot{\psi}=\frac{\partial \psi}{\partial t}+u\frac{\partial \psi}{\partial x}+v\frac{\partial \psi}{\partial y}, \tag{2.11}$$

is the spin of the principal stress axes through a generic particle. Accordingly, through Ω, the deformation depends on the stress-rate as well as on the stress. Equation (2.11) expresses the assumption that the deformation takes place by shearing on the two families of surfaces on which the Coulomb critical shear stress is mobilized. If in addition the deformation is isochoric, then

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0. \tag{2.12}$$

Although in many applications (2.12) is a good approximation, it is argued in Section 5 that it may not be appropriate in the case considered there. There are two main ways in which the double-shearing theory may be modified to include an element of compressibility. The first was proposed by Spencer and Kingston [15]; in this extended theory (2.10) and (2.11) are unchanged, but (2.12) is replaced by

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=\Delta, \tag{2.13}$$

where the dilatation-rate Δ is a scalar which in general is expected to be complicated function of the deformation and stress history, but for the purposes of this paper does not need to be specified. The second and better-known way to incorporate dilatancy in the theory is due to Mehrabadi and Cowin [16,17] and is based on an assumption that shear of the material is accompanied by an expansion in the direction normal to the shear plane. In practice this expansion-rate decreases as the shear deformation progresses, and so this mechanism is perhaps most important in the initial stages of deformation. The Spencer and Kingston mechanism attempts to incorporate non-directional volumetric effects, such as absorption of fluid, or crushing or uniform compaction of grains under applied pressure. The Mehrabadi and Cowin mechanism describes the expansion normal to the shear plane that arises in shear of initially compacted granular materials as grains overtake their neighbours. This effect is most marked in the initial stages of deformation, and the rate of the normal expansion decays as the flow becomes fully developed or approaches a critical state. The two mechanisms are not exclusive, and there is no reason why they should not both be included in a complete theory that incorporates dilatancy.

3. An exact solution for a layer of granular material

In this section we derive an exact solution of the equations of Section 2, which will be applied to the analysis of the stress and deformation in a granular layer. We choose the coordinate system Oxy to be such that the x-axis is parallel to the layer and y measures distance through the layer thickness, as shown in Figure 1. Then the mid-plane of the layer is $y=0$, its surfaces are $y=\pm h$, and its ends are $x=\pm L$.

The classical solutions of Prandtl [3] and Marshall [5] suggest that we look for stress solutions of the form

$$\psi=\psi(y), \quad q=q(x,y). \tag{3.1}$$

By substituting (3.1) in the equilibrium equations (2.8), we obtain

$$\begin{aligned}(\cos 2\psi-\operatorname{co\,sec}\phi)\frac{\partial q}{\partial x}+\sin 2\psi\frac{\partial q}{\partial y}+2q\cos 2\psi\frac{\mathrm{d}\psi}{\mathrm{d}y}=0,\\ \sin 2\psi\frac{\partial q}{\partial x}-(\cos 2\psi+\operatorname{co\,sec}\phi)\frac{\partial q}{\partial y}+2q\sin 2\psi\frac{\mathrm{d}\psi}{\mathrm{d}y}=0.\end{aligned} \tag{3.2}$$

Hence, by dividing each of (3.2) by q and solving for $\partial(\log q)/\partial x$ and $\partial(\log q)/\partial y$ it follows that

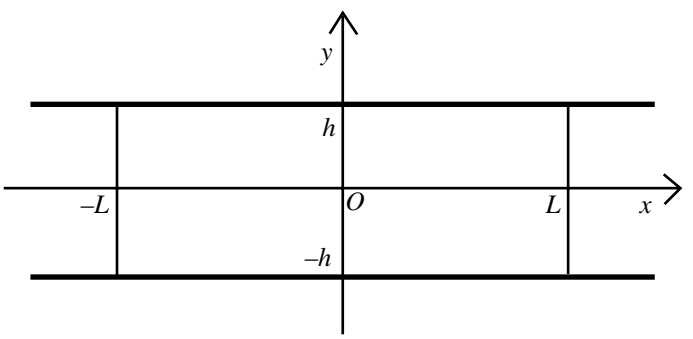


Figure 1. Compression and shear of a layer of granular material between rough rigid plates.

$$\cos\phi\cot\phi\frac{\partial}{\partial x}(\log q)=2(\sin\phi+\cos 2\psi)\frac{\mathrm{d}\psi}{\mathrm{d}y},$$
$$\cos\phi\cot\phi\frac{\partial}{\partial y}(\log q)=2\sin 2\psi\frac{\mathrm{d}\psi}{\mathrm{d}y}. \tag{3.3}$$

Since ψ is independent of x, we have $\partial^2(\log q)/\partial x\partial y=0$, and therefore

$$\frac{\mathrm{d}}{\mathrm{d}y}\left\{2(\sin\phi+\cos 2\psi)\frac{\mathrm{d}\psi}{\mathrm{d}y}\right\}=0 \tag{3.4}$$

which gives, on integration

$$2\psi\sin\phi+\sin 2\psi=a+b\frac{y}{h}, \tag{3.5}$$

where a and b are integration constants. This determines ψ implicitly as a function of y.

It also follows from (3.3_1) and (3.5) that

$$\cos\phi\cot\phi\log q=b\frac{x}{h}+g(y)$$

and by substituting this back in (3.3_2) we find

$$\frac{\mathrm{d}}{\mathrm{d}y}g(y)=2\sin 2\psi\frac{\mathrm{d}\psi}{\mathrm{d}y}$$

and thus

$$g(y)=-\cos 2\psi+c, \tag{3.6}$$

where c is constant. Therefore,

$$\log q=(b\frac{x}{h}-\cos 2\psi+c)\tan\phi\sec\phi$$

or

$$q=C\exp\left\{(b\frac{x}{h}-\cos 2\psi)\tan\phi\sec\phi\right\}, \tag{3.7}$$

where $C=\exp(c\tan\phi\sec\phi)$. Thus the stress solution is determined to within the three constants a, b, and C. For the case $a=0$, (3.5) and (3.7) reduce to the solution for pure compression of a granular layer between rough plates that was derived by Hartmann [4] and also discussed by Marshall [5]. The inclusion of the term a in (3.5) allows consideration of a wider class of solutions.

From (2.9) it follows that α- and β-lines corresponding to (3.5) are given parametrically by the equations

$$\frac{x-x_\alpha}{h}=\frac{1}{b}(\cos 2\psi-2\psi\cos\phi),\quad \frac{y}{h}=-\frac{a}{b}+\frac{1}{b}(\sin 2\psi+2\psi\sin\phi),\quad (\alpha\text{-lines})$$
$$\frac{x-x_\beta}{h}=\frac{1}{b}(\cos 2\psi+2\psi\cos\phi),\quad \frac{y}{h}=-\frac{a}{b}+\frac{1}{b}(\sin 2\psi+2\psi\sin\phi),\quad (\beta\text{-lines}), \tag{3.8}$$

where x_α is constant on a given α-line and x_β is constant on a given β-line. Hence the α- and β-lines are two families of cycloids. Each α-line is generated by a point on the circumference of a circle of radius h/b rolling on a line with slope $-\tan\phi$, and each β-line is generated by a point on the circumference of a similar circle rolling on a line with slope $\tan\phi$.

The steady simple shearing solution corresponds to the case $b=0$. In this event it follows from (3.5) that ψ is constant, and the requirement that $\sigma_{xy}=\sigma_{yy}\tan\phi$ determines $\psi=\psi_0=\frac{1}{2}\phi+\frac{1}{4}\pi$.

For the associated velocity field we consider separately the cases of incompressible and compressible material, as follows:

(*i*) We consider that the material is incompressible and suppose that the layer of granular material is confined between rigid platens at $y=\pm h$ that translate inwards with components of velocity in the y-direction of magnitude U, so that U is the speed of the surface $y=h$, and therefore $U=-\mathrm{d}h/\mathrm{d}t$. We seek solutions of (2.10) and (2.12) of the form

$$u=U\frac{x}{h}+\Phi(y), \quad v=-U\frac{y}{h}. \tag{3.9}$$

Then the isochoric condition (2.12) is identically satisfied, and (2.10) and (2.11) give

$$\frac{\mathrm{d}\Phi}{\mathrm{d}y}(\cos 2\psi+\sin\phi)-2\frac{U}{h}\sin 2\psi+2\sin\phi\left(\frac{\partial\psi}{\partial t}-U\frac{y}{h}\frac{\mathrm{d}\psi}{\mathrm{d}y}\right)=0. \tag{3.10}$$

Since h depends on t, therefore ψ depends on both y and t, and

$$\frac{\partial\psi}{\partial t}=\frac{\mathrm{d}h}{\mathrm{d}t}\frac{\partial\psi}{\partial h}=-U\frac{\partial\psi}{\partial h}=\frac{Uby}{2h^2(\sin\phi+\cos 2\psi)},$$
$$\frac{\mathrm{d}\psi}{\mathrm{d}y}=\frac{b}{h}\frac{1}{2(\sin\phi+\cos 2\psi)},$$

and hence from (2.11) in this case $\Omega=0$ and so (3.10) reduces to

$$\frac{\mathrm{d}\Phi}{\mathrm{d}y}(\cos 2\psi+\sin\phi)-2\frac{U}{h}\sin 2\psi=0. \tag{3.11}$$

Then, provided that $b\neq 0$, it follows from (3.5) that

$$\frac{\mathrm{d}\Phi}{\mathrm{d}y}(\cos 2\psi+\sin\phi)=\frac{\mathrm{d}\Phi}{\mathrm{d}\psi}\frac{\mathrm{d}\psi}{\mathrm{d}y}(\cos 2\psi+\sin\phi)=\frac{b}{2h}\frac{\mathrm{d}\Phi}{\mathrm{d}\psi}.$$

Hence (3.11) becomes

$$b\frac{\mathrm{d}\Phi}{\mathrm{d}\psi}-4U\sin 2\psi=0, \tag{3.12}$$

and it follows that

$$b\Phi=-2U\cos 2\psi+bU_0,$$

where U_0 is a further integration constant that represents a rigid body translation in the x-direction. Hence from (3.9) the associated velocity field is given as

$$u=U\left(\frac{x}{h}-\frac{2}{b}\cos 2\psi\right)+U_0, \quad v=-U\frac{y}{h}. \tag{3.13}$$

In the case $b=0$, then ψ is constant and $U=0$, and the solution reduces to the steady simple shearing motion

$$u=\gamma y, \quad v=0.$$

(*ii*) We suppose that the material is confined between platens as in case (*i*) but that it is capable of undergoing compression. We seek solutions in which the material does not undergo stretching in the x-direction so that there is no slip between the layer and the platens, In order to satisfy this boundary condition it is natural to use the Spencer and Kingston theory [15], with the governing equations (2.10), (2.11) and (2.13). We assume that both of the velocity components u and v depend only on y, of the form

$$u = u(y), \quad v = -U\frac{y}{h}. \tag{3.14}$$

Then $\Omega = 0$ as in case (*i*) and (2.10) and (2.13) become

$$\frac{\mathrm{d}u}{\mathrm{d}y}(\cos 2\psi + \sin\phi) - \frac{U}{h}\sin 2\psi = 0, \quad -\frac{U}{h} = \Delta. \tag{3.15}$$

Hence

$$\frac{\mathrm{d}u}{\mathrm{d}y} = \frac{U}{h}\frac{\sin 2\psi}{(\cos 2\psi + \sin\phi)}, \tag{3.16}$$

and, from (3.5)

$$\frac{\mathrm{d}u}{\mathrm{d}\psi} = 2\frac{U}{b}\sin 2\psi \tag{3.17}$$

and therefore the velocity field is

$$u = -\frac{U}{b}\cos 2\psi + U_0, \qquad v = -U\frac{y}{h}. \tag{3.18}$$

The limiting case of steady simple shearing is an isochoric deformation, and so the response to volume changes is not relevant in this case.

4. Application to compression and shear of a granular layer. Stress field

We now apply the solution developed in Section 3 to a layer of granular material occupying the region $-L \le x \le L$, $-h \le y \le h$, subjected to compression and shear by normal forces N and shear forces S applied through rigid parallel platens at $y = \pm h$. We look for solutions in which the field of α- and β-lines has the form illustrated in Figure 2. For negative values of x, in the region $OA'D'C'B'O$, we assume that the stress is given by (3.5) and (3.7), thus

$$\begin{aligned} &2\psi\sin\phi + \sin 2\psi = a + b\frac{y}{h}, \\ &q = C\exp\left\{(b\frac{x}{h} - \cos 2\psi)\tan\phi\sec\phi\right\}, \end{aligned} \tag{4.1}$$

and the α- and β-lines have the parametric form (3.8), namely

$$\begin{aligned} &\tfrac{x-x_\alpha}{h} = \tfrac{1}{b}(\cos 2\psi - 2\psi\cos\phi),\ \tfrac{y}{h} = -\tfrac{a}{b} + \tfrac{1}{b}(\sin 2\psi + 2\psi\sin\phi),\ (\alpha\text{-lines}) \\ &\tfrac{x-x_\beta}{h} = \tfrac{1}{b}(\cos 2\psi + 2\psi\cos\phi),\ \tfrac{y}{h} = -\tfrac{a}{b} + \tfrac{1}{b}(\sin 2\psi + 2\psi\sin\phi),\ (\beta\text{-lines}). \end{aligned} \tag{4.2}$$

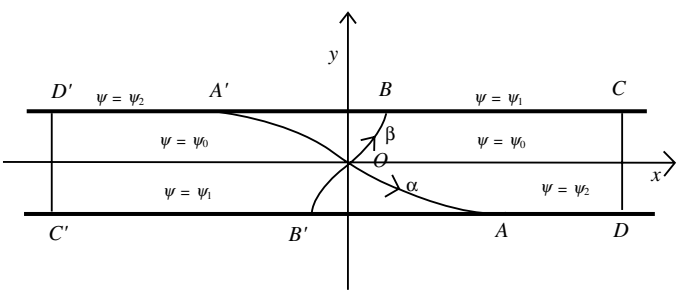


Figure 2. Schematic of slip-line field for compression and shear between rough rigid plates.

For positive values of x, we assume a similar solution, but with b replaced by $-b$, thus in $OADCBO$

$$2\psi \sin\phi + \sin 2\psi = a - b\frac{y}{h},$$
$$q = C \exp\left\{-(b\frac{x}{h} + \cos 2\psi)\tan\phi \sec\phi\right\}, \tag{4.3}$$

and the α- and β-lines have the parametric equations

$$\begin{aligned} &\tfrac{x-x_\alpha}{h} = -\tfrac{1}{b}(\cos 2\psi - 2\psi\cos\phi), \quad \tfrac{y}{h} = \tfrac{a}{b} - \tfrac{1}{b}(\sin 2\psi + 2\psi\sin\phi), \quad (\alpha\text{-lines}) \\ &\tfrac{x-x_\beta}{h} = -\tfrac{1}{b}(\cos 2\psi + 2\psi\cos\phi), \quad \tfrac{y}{h} = \tfrac{a}{b} - \tfrac{1}{b}(\sin 2\psi + 2\psi\sin\phi), \quad (\beta\text{-lines}). \end{aligned} \tag{4.4}$$

We denote by ψ_0 the value of ψ on the mid-plane $y=0$, and hence from (4.1) and (4.3)

$$a = \sin 2\psi_0 + 2\psi_0 \sin\phi. \tag{4.5}$$

Then the curve OA' is a segment of the α-line through O, defined parametrically as

$$\frac{x}{h} = \frac{1}{b}(\cos 2\psi - \cos 2\psi_0 - 2(\psi - \psi_0)\cos\phi), \quad \frac{y}{h} = -\frac{a}{b} + \frac{1}{b}(\sin 2\psi + 2\psi\sin\phi), \tag{4.6}$$

and OB' is the segment of the β-line through O, and is given by

$$\frac{x}{h} = \frac{1}{b}(\cos 2\psi - \cos 2\psi_0 + 2(\psi - \psi_0)\cos\phi), \quad \frac{y}{h} = -\frac{a}{b} + \frac{1}{b}(\sin 2\psi + 2\psi\sin\phi). \tag{4.7}$$

Similarly OA is a segments of the α-line in $x > 0$

$$\frac{x}{h} = -\frac{1}{b}(\cos 2\psi - \cos 2\psi_0 - 2(\psi - \psi_0)\cos\phi), \quad \frac{y}{h} = \frac{a}{b} - \frac{1}{b}(\sin 2\psi + 2\psi\sin\phi), \tag{4.8}$$

and OB is a segment of the β-line

$$\frac{x}{h} = -\frac{1}{b}(\cos 2\psi - \cos 2\psi_0 + 2(\psi - \psi_0)\cos\phi), \quad \frac{y}{h} = \frac{a}{b} - \frac{1}{b}(\sin 2\psi + 2\psi)\sin\phi). \tag{4.9}$$

In the regions OAB' and OBA' the stress is indeterminate, but must be in equilibrium and not violate the yield condition. This form of solution is suggested by the solution for pure compression of a layer by Hartmann [4] and Marshall [5], which in turn are extensions of Prandtl's solution [3]. The present solution reduces to that of Hartmann and Marshall when $a=0$.

It follows from (2.6) that at $y=0$

$$\sigma_{yy} = -q\,(\cos 2\psi_0 + \operatorname{co\,sec}\phi), \quad \sigma_{xy} = q\sin 2\psi_0. \tag{4.10}$$

We look for solutions in which the stress components are large in magnitude near $x=0$ at the centre of the layer, and decay as $x \to \pm L$, and so we require $b > 0$. Also, for definiteness, we assume that $\sin 2\psi_0 > 0$, so that $a > 0$ and $\sigma_{xy} > 0$ at $y=0$, and therefore the shear force on $y=0$ acts in the positive x-direction. There are corresponding solutions with $\sin 2\psi_0 < 0$.

We also denote by ψ_1 the value of ψ on BC and $B'C'$, and by ψ_2 the value of ψ on AD and $A'D'$ so that

$$\begin{aligned} 2\psi_1 \sin\phi + \sin 2\psi_1 &= a - b, \\ 2\psi_2 \sin\phi + \sin 2\psi_2 &= a + b \end{aligned} \tag{4.11}$$

and hence

$$2a = 2(\psi_1 + \psi_2)\sin\phi + \sin 2\psi_1 + \sin 2\psi_2 = 2(2\psi_0 \sin\phi + \sin 2\psi_0),$$
$$2b = 2(\psi_2 - \psi_1)\sin\phi + \sin 2\psi_2 - \sin 2\psi_1. \tag{4.12}$$

Since b is positive we must have $\psi_2 \geq \psi_1$. We also recall that in the case of pure compression considered by Hartmann [4] and Marshall [5], we have $a = 0$, and therefore in this case $\psi_0 = 0$ and $\psi_1 = -\psi_2$. In the case of steady simple shear, then $b = 0$ and $\psi_0 = \psi_1 = \psi_2 = \frac{1}{2}\phi + \frac{1}{4}\pi$. Hence we are concerned with values of ψ_0 in the range $0 \leq \psi_0 \leq \frac{1}{2}\phi + \frac{1}{4}\pi$.

We suppose that equal and opposite tractions act on the end faces CD and $C'D'$. Then, from (4.10), for equilibrium of the upper half of the layer we require

$$N = \int_{-L}^{L} q(\cos 2\psi_0 + \cos\text{ec}\phi)\mathrm{d}x, \quad S = \int_{-L}^{L} q \sin 2\psi_0 \mathrm{d}x, \tag{4.13}$$

and it follows from (4.1) and (4.3) that

$$N = 2Q(\cos 2\psi_0 + \cos\text{ec}\phi)\exp(-\cos 2\psi_0 \tan\phi \sec\phi),$$
$$S = 2Q \sin 2\psi_0 \exp(-\cos 2\psi_0 \tan\phi \sec\phi) \tag{4.14}$$

where

$$Q = C\frac{h}{b}\cos\phi\cot\phi\left\{1 - \exp\left(-\frac{bL}{h}\tan\phi\sec\phi\right)\right\}. \tag{4.15}$$

If the layer is thin, in the sense that $h/L << 1$, then

$$Q \simeq C\frac{h}{b}\cos\phi\cot\phi. \tag{4.16}$$

It follows from (4.14) that the layer has an apparent coefficient of friction $\tan\lambda_0$, where

$$\frac{S}{N} = \tan\lambda_0 = \frac{\sin 2\psi_0 \sin\phi}{\cos 2\psi_0 \sin\phi + 1}, \quad \text{and hence} \quad \sin\phi \sin(2\psi_0 - \lambda_0) = \sin\lambda_0. \tag{4.17}$$

Alternatively, (4.17) can be regarded as determining the value of ψ_0 (and hence of a) that is required to support a force with normal and tangential components N and S applied to the surfaces of the layer. Since $\sin 2\psi_0$ is assumed to be positive, it follows that $\tan\lambda_0$ is positive. It follows from (4.17) that

$$0 \leq \lambda_0 < 2\psi_0 < \pi + \lambda_0.$$

Then in the case $a \geq 0, b \geq 0$ it follows from (4.12) that $\psi_2 \geq \psi_0 \geq \psi_1$.

A further boundary condition is required to determine the constant b. We assume Coulomb friction between the platens and the granular material, with angle of friction μ, so that at the surfaces $y = \pm h$, we have $|\sigma_{xy}| \leq -\sigma_{yy}\tan\mu$ and at yield $|\sigma_{xy}| = -\sigma_{yy}\tan\mu$. Since it has been assumed that the shear force on $y = h$ is in the positive x-direction, it follows that σ_{xy} is positive at points on the surface at which the critical shear stress is realized. We further denote

$$\tan\lambda_1 = \frac{\sin 2\psi_1 \sin\phi}{\cos 2\psi_1 + 1}, \quad \tan\lambda_2 = \frac{\sin 2\psi_2 \sin\phi}{\cos 2\psi_2 + 1}, \quad \sin\delta = \frac{\sin\mu}{\sin\phi}. \tag{4.18}$$

Then if yield occurs on BC and $B'C'$, then $\lambda_1 = \mu$, and if yield occurs on AD and $A'D'$, then $\lambda_2 = \mu$. It follows from (4.18) that when $\lambda_1 = \mu$

$$2\psi_1 = \delta + \mu \tag{4.19}$$

and when $\lambda_2 = \mu$

$$2\psi_2 = \delta + \mu. \tag{4.20}$$

Since $\psi_2 \geq \psi_1$, the critical shear stress is mobilized on AD and $A'D'$, when ψ_2 is given by (4.20). It then follows from (4.12) that

$$2\psi_1 \sin\phi + \sin 2\psi_1 = -(\mu+\delta)\sin\phi - \sin(\mu+\delta) + 2(2\psi_0 \sin\phi + \sin 2\psi_0). \tag{4.21}$$

For a specified value of ψ_0, and with ψ_2 given by the boundary condition (4.20), then (4.21) determines ψ_1.

An important special case is that of a perfectly rough surface, in which $\mu > \phi$ and yield occurs when the Coulomb shear stress $\sigma_{xy} = -\sigma_{yy}\tan\phi$ is mobilized on AD and $A'D'$. In this case the material shears on itself at these surfaces and in the above analysis we take $\mu = \phi$, so that $\sin\delta = 1$ and $\delta = \pi/2$, and

$$2\psi_2 = \phi + \frac{1}{2}\pi. \tag{4.22}$$

This case is adopted for the numerical illustrations given in Section 6.

To complete the analysis of the stress field it remains to determine the constant C, or equivalently Q. For this it is necessary to impose an additional condition on the confining stress in the granular layer, for example by specifying σ_{xx} at $x = \pm L$. Suppose, for example, that a horizontal confining pressure p_0 is applied at $x = \pm L$. Then from (2.6) and (4.3) it follows that

$$\begin{aligned} 2hp_0 &= -\int_{-h}^{h} \sigma_{xx}(L, y)\,\mathrm{d}y = -\int_{-h}^{h} q(L, y)\,(\cos 2\psi - \operatorname{cosec}\phi)\,\mathrm{d}y \\ &= -C\exp\left(-\frac{bL}{h}\tan\phi\sec\phi\right)\int_{-h}^{h}(\cos 2\psi - \operatorname{cosec}\phi)\exp\{-\cos 2\psi\tan\phi\sec\phi\}dy. \end{aligned}$$

Therefore, from (3.5)

$$\begin{aligned} 2hp_0 &= -2\frac{Ch}{b}\exp\left(-\frac{bL}{h}\tan\phi\sec\phi\right)\int_{\psi_2}^{\psi_1}(\cos 2\psi - \operatorname{cosec}\phi)\,(\cos 2\psi + \sin\phi) \\ &\qquad \times \exp\{-\cos 2\psi\tan\phi\sec\phi\}\,\mathrm{d}\psi \\ &= 2\frac{Ch}{b}\exp\left(-\frac{bL}{h}\tan\phi\sec\phi\right)\int_{\psi_2}^{\psi_1}\left(\sin^2 2\psi + \cos 2\psi\cot\phi\cos\phi\right) \\ &\qquad \times \exp\{-\cos 2\psi\tan\phi\sec\phi\}\,\mathrm{d}\psi \\ &= \frac{Ch}{b}\exp\left(-\frac{bL}{h}\tan\phi\sec\phi\right)\cot\phi\cos\phi[\sin 2\psi\exp\{-\cos 2\psi\tan\phi\sec\phi\}]_{\psi_2}^{\psi_1}. \end{aligned} \tag{4.23}$$

Hence, from (4.15)

$$Q = \frac{2hp_0\left\{\exp\left(b\frac{L}{h}\tan\phi\sec\phi\right) - 1\right\}}{[\exp\{-\cos 2\psi_1\tan\phi\sec\phi\}\sin 2\psi_1 - \exp\{-\cos 2\psi_2\tan\phi\sec\phi\}\sin 2\psi_2]}, \tag{4.24}$$

or, if $L >> h$

$$Q \approx \frac{2hp_0\exp\left(b\frac{L}{h}\tan\phi\sec\phi\right)}{[\exp\{-\cos 2\psi_1\tan\phi\sec\phi\}\sin 2\psi_1 - \exp\{-\cos 2\psi_2\tan\phi\sec\phi\}\sin 2\psi_2]}. \tag{4.25}$$

As noted above, OAB' and $OA'B$ are transition zones bounded by the four segments of characteristics through the origin, as illustrated schematically in Figure 2. The stress in this

zone is indeterminate, but must conform to the condition (2.5). There are also regions adjacent to the ends $x=\pm L$ in which the stress depends on the details of the boundary conditions at these surfaces. The analysis by Hill, Lee and Tupper [18] of a related problem in metal plasticity suggests that the required superposed stress field decays rapidly with distance from the ends.

5. Compression and shear of a granular layer. Velocity field

We assume that the origin O is fixed and that the upper and lower platens can move as rigid bodies with velocities that are of equal magnitude but opposite in direction. Let us first consider the case that the deformation is isochoric, so that, from (3.13),

$$\begin{aligned} u &= U\left\{\frac{x}{h}-\frac{2}{b}(\cos 2\psi-\cos 2\psi_0)\right\}, \quad v=-U\frac{y}{h} \quad \text{in} \quad OA'D'C'B', \\ u &= U\left\{\frac{x}{h}+\frac{2}{b}(\cos 2\psi-\cos 2\psi_0)\right\}, \quad v=-U\frac{y}{h} \quad \text{in} \quad OADCB, \end{aligned} \tag{5.1}$$

where the constant U_0 has been chosen so that $u=0$ at the origin. The deformation (5.1) comprises a pure shear deformation, described by

$$u=Ux/h, \quad v=-Uy/h$$

on which is superposed a non-uniform shearing deformation

$$\begin{aligned} u &= -\tfrac{2U}{b}(\cos 2\psi-\cos 2\psi_0), \quad v=0 \quad \text{in} \quad OA'D'C'B', \\ u &= \tfrac{2U}{b}(\cos 2\psi-\cos 2\psi_0), \quad v=0 \quad \text{in} \quad OADCB. \end{aligned} \tag{5.2}$$

Because of the dependence on x in (5.1), this motion implies slip between the material and the platens at $y=\pm h$ such that the magnitude of the slip increases linearly with x. For a thin granular layer this seems unrealistic, because it is clear from the analysis of Section 4 that large compressive stresses and large frictional forces are generated at the surfaces $y=\pm h$, especially in the neighbourhood of $x=0$. We therefore propose that it is more realistic to assume that such slip does not occur, and that the consequent change in volume of the granular layer is accommodated by elastic compression, consolidation, or crushing of the granules. For example the forces applied may result in the crushing strength of the material being exceeded, and then the dilatation-rate Δ depends on the history of the applied pressure, while the surface friction is sufficient to prevent surface slip. In such a case, the theory of Spencer and Kingston [15], with solutions of the form (3.18), is appropriate. Hence we assume

$$\begin{aligned} u &= -\tfrac{U}{b}(\cos 2\psi-\cos 2\psi_0), \quad v=-U\tfrac{y}{h}, \quad \text{in} \quad OA'D'C'B' \\ u &= \tfrac{U}{b}(\cos 2\psi-\cos 2\psi_0), \quad v=-U\tfrac{y}{h}, \quad \text{in} \quad OADCB. \end{aligned} \tag{5.3}$$

Hence if $u=\pm\gamma h$ at $y=\pm h$

$$\gamma h=\frac{U}{b}(\cos 2\psi_1-\cos 2\psi_0), \quad -\gamma h=\frac{U}{b}(\cos 2\psi_2-\cos 2\psi_0) \tag{5.4}$$

and therefore the mean shear strain-rate γ is, from (4.12)

$$\gamma=\frac{U(\cos 2\psi_1-\cos 2\psi_2)}{2hb}=\frac{U(\cos 2\psi_1-\cos 2\psi_2)}{h\left\{2(\psi_2-\psi_1)\sin\phi+\sin 2\psi_2-\sin 2\psi_1\right\}}. \tag{5.5}$$

We note that in the limit $\psi_1 \to \psi_2$, which is the case of uniform steady shear stress

$$\gamma \to \frac{U \sin 2\psi_2}{h(\cos 2\psi_2 + \sin \phi)} \tag{5.6}$$

and in particular that γ is unbounded in the limit $2\psi_2 = 2\psi_1 \to \pi/2 + \phi$, which is the case of steady simple shearing deformations. It is also straightforward to show that if $2\psi_2$ and $2\psi_1$ are close to the limit value $\pi/2 + \phi$, for example

$$2\psi_2 = \frac{\pi}{2} + \phi - \varepsilon_2, \quad 2\psi_1 = \frac{\pi}{2} + \phi - \varepsilon_1, \tag{5.7}$$

with $\varepsilon_1, \varepsilon_2 << 1$, then

$$\gamma = \frac{2U}{h(\varepsilon_1 + \varepsilon_2)} + O(1), \tag{5.8}$$

and so the shear-rate increases rapidly as $2\psi_2$ and $2\psi_1$ tend to $\pi/2 + \phi$.

The transition regions $OA'B$ and OAB' in the vicinity of $x = 0$ move as rigid bodies with speed U, in the positive y-direction for $y < 0$ and in the negative y-direction for $y > 0$.

6. Numerical results and discussion

For the purposes of the calculations presented in this section , we assume that the platen surfaces are perfectly rough, so that $\mu \geq \phi$, and $2\psi_2 = \phi + \frac{1}{2}\pi$, and assign the typical value $\pi/6$ to ϕ. With these values Figure 3 shows the variation of ψ_1 with ψ_0, where from (4.12) and (4.20)

$$2\psi_1 \sin\phi + \sin 2\psi_1 = 2(2\psi_0 \sin\phi + \sin 2\psi_0) - \left(\phi + \frac{1}{2}\pi\right)\sin\phi - \cos\phi. \tag{6.1}$$

We recall that ψ_0 varies from $\psi_0 = 0$, for the case of compression with no resultant shear force, to $\psi_0 = \frac{1}{2}\phi + \frac{1}{4}\pi$ for the case of steady simple shear. Correspondingly, ψ_1 varies from $\psi_1 = -\frac{1}{2}\phi - \frac{1}{4}\pi$ (so that $\psi_1 = -\psi_2$) for pure compression, to $\psi_1 = \frac{1}{2}\phi + \frac{1}{4}\pi$ (or $\psi_1 = \psi_0 = \psi_2$) for steady simple shear.

In Figure 4 we show the variation of $S/N = \tan\lambda_0$ (where λ_0 is given by (4.17)) through this range of values of ψ_0, again for the case $\phi = \pi/6$.

Figure 5 shows the variation of the parameters a and b with ψ_0. These parameters characterize the stress and velocity fields and are defined by (4.12). Then, for given values of ϕ, h, L

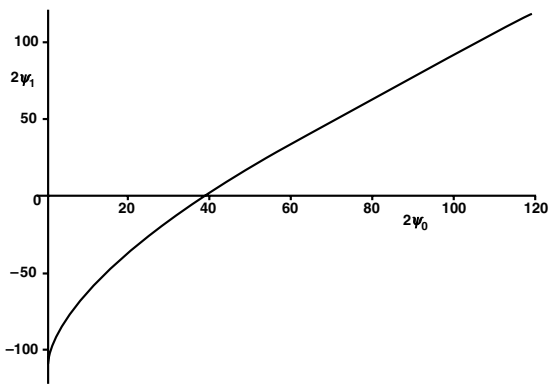


Figure 3. Variation of the angle ψ_1 with the angle ψ_0.

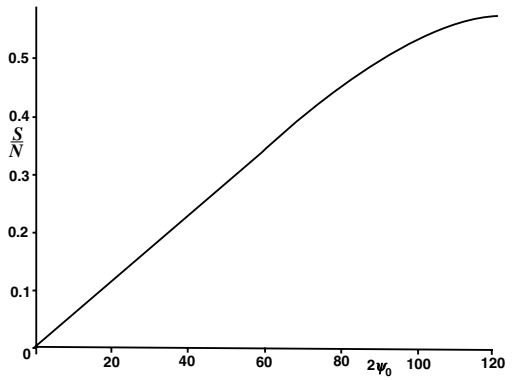


Figure 4. Variation of the ratio S/N of shear force to normal force with the angle ψ_0.

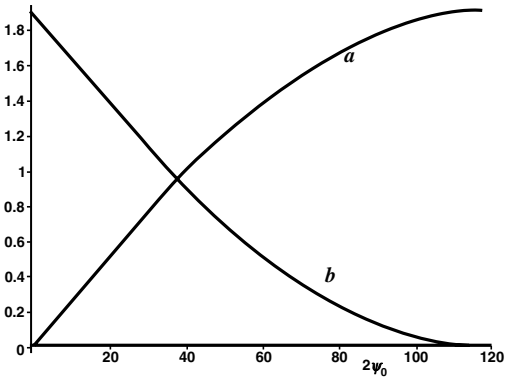


Figure 5. Variation of the parameters a and b with the angle ψ_0.

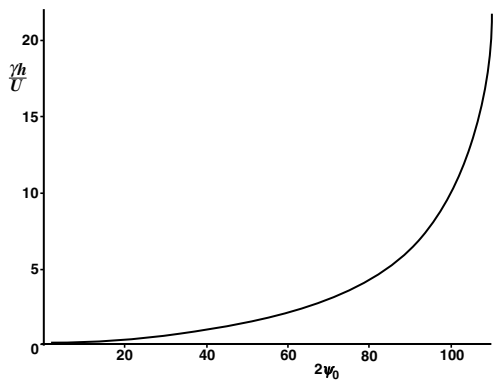


Figure 6. Variation of the shear-rate γ with the angle ψ_0.

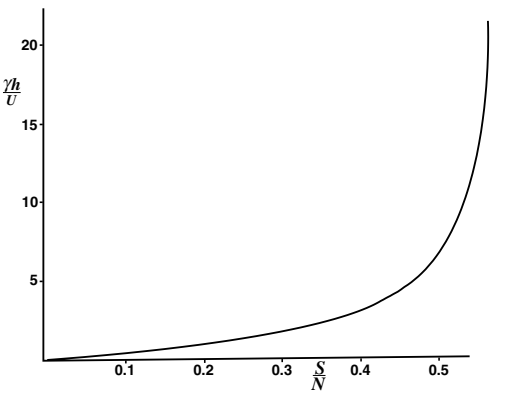


Figure 7. Variation of the shear-rate γ with the ratio S/N of shear force to normal force.

and p_0, Q is determined by (4.22) or (4.23). The values of the resultant normal force N and the resultant shear force S for a given value of ψ_0 then follow from (4.14).

In Figure 6 we show the variation of $h\gamma/U$ with ψ_0, as given by (5.5). Here γ is the average shear strain-rate and U/h is the direct compressive strain-rate. Figure 7 shows how S/N varies with $h\gamma/U$.

Some points of interest arise from these results. It is clear that over most of the range the shear force component has only a minor effect on the deformation and the compressive component predominates. The shearing component of the deformation only becomes significant when S/N is close to $\tan\phi$.

It was pointed out in [8,9] that the steady simple shearing deformation with $\sigma_{xy} = -\sigma_{yy}\tan\phi$ and $\psi_0 = \frac{1}{4}\pi + \frac{1}{2}\phi$ is unstable, in the sense that, if the traction boundary conditions permit it, shear flow of the material is accompanied by a decreasing shear stress. In this paper, we have assumed that S and N are specified, and therefore ψ_0 is fixed. In this case the steady solution is the only possible continuous solution, but the possibility of strain localization or shear band formation cannot be excluded. It appears from the analysis of this paper that if the shear is accompanied by compression, which is our general case, then the compression will have a stabilizing effect. This is because if $U > 0$ then h decreases after yield

occurs, and therefore, from (4.22), if the other parameters are fixed, Q increases after yield takes place, and so the deformation takes place under increasing load. In principle, if the post-yield behaviour of S and N is specified, then the post-yield stress-strain relation may be deduced from (4.14), (4.22), (5.5) and the relation $U = -\mathrm{d}h/\mathrm{d}t$. An unresolved problem is that the equations of the double-shearing theory are known to be mathematically ill-posed, but the present solution does not seem to shed any light on this difficulty.

Acknowledgement

A large part of this paper was written at the Isaac Newton Institute for Mathematical Sciences, Cambridge during the author's participation in the Institute's programme on 'Granular and Particle-laden Flows'.

References

1. R. Hill, *The Mathematical Theory of Plasticity.* Oxford: Oxford University Press (1950) 356 pp.
2. W. Prager and P.G. Hodge, *Theory of Perfectly Plastic Solids*. New York: Wiley (1951) 264 pp.
3. L. Prandtl, Anwendungsbeispiele zu einem Henckyshen Satz über das plastische Gleichgewicht. *Z. angew Math. Mech.* 3 (1923) 401–406.
4. W. Hartmann, *Über die Integration der Differentialgleichungen des ebenen Gleichgewiichtszustandes für den Allgemein-Plastichen Körper*. Thesis: Gottingen (1925).
5. E.A. Marshall, The compression of a slab of ideal soil between rough plates. *Acta Mech.* 3 (1967) 82–92.
6. A.J.M. Spencer, A theory of the kinematics of ideal soils under plane strain conditions. *J. Mech. Phys. Solids* 12 (1964) 337–351.
7. A.J.M. Spencer, Deformation of ideal granular materials. In: H.G. Hopkins and M.J. Sewell (eds.), *Mechanics of Solids; the Rodney Hill Anniversary Volume.* Oxford: Pergamon Press (1982) pp. 607–652.
8. A.J.M. Spencer, Double-shearing theory applied to instability and localization in granular materials. *J. Engng. Math*. 45 (2003) 55–74.
9. A.J.M. Spencer, Instability of shear flows of granular materials. *Acta Mech.* 64 (1986) 77–87.
10. J.C. Savage, D.A. Lockner and J.D. Byerlee, Failure in laboratory fault models in triaxial tests. *J. Geophys. Res.* 101 (1996) 22215–22224.
11. J.R. Rice, Fault stress state, pore pressure distributions, and the weakness of the San Andreas fault. In: B. Evans and T.-F. Wong (eds.), *Fault Mechanics and Transport Properties in Rocks.* San Diego: Academic Press (1992) pp. 475–503.
12. P.A. Gremaud, Numerical issues in plasticity models for granular flows. *J. Volcanology & Geothermal Res.* 137 (2004) 1–9
13. S. Alexandrov, Compression of double-shearing and coaxial models for pressure-dependent plastic flow at frictional boundaries. *J. Appl. Mech.* 70 (2003) 212–219.
14. Huaning Zhu, M.M. Mehrabadi and M. Massoudi, A comparative study of the response of double shearing and hypoplastic models. In: *Proceedings of IMECE: 2002 ASME International Mechanical Engineering Congress and Exposition*, New Orleans Amer. Soc. Mech. Engrs., Materials Div. Publ. MD Vol. 97 (2002) pp. 343–351.
15. A.J.M. Spencer and M.R. Kingston, Plane mechanics and kinematics of compressible ideal granular materials. *Rheol. Acta* 12 (1973) 194–199.
16. M.M. Mehrabadi and S.C. Cowin, Initial planar deformation of dilatant granular materials. *J. Mech. Phys. Solids* 26 (1978) 269–284.
17. M.M. Mehrabadi and S.C. Cowin, Pre-failure and post-failure soil plasticity models. *J. Engng. Mech. Div. Proc ASCE* 106 (1980) 991–1003.
18. R. Hill, E.H. Lee and S.J. Tupper, A method of numerical analysis of plastic flow in plane strain and its application to the compression of a ductile material between rough plates. *J. Appl. Mech.* 8 (1951) 46–52.

Journal of Engineering Mathematics (2005) 52: 265–291
DOI 10.1007/s10665-004-4555-2

An assessment of plasticity theories for modeling the incrementally nonlinear behavior of granular soils

CLAUDIO TAMAGNINI*, FRANCESCO CALVETTI[1] and GIOACCHINO VIGGIANI[2]

*Dipartimento di Ingegneria Civile e Ambientale, Università di Perugia, Italy; [1]Dipartimento di Ingegneria Strutturale, Politecnico di Milano, Italy; [2]Laboratoire 3S, UJF–INPG–CNRS, Grenoble, France; *Author for correspondence; E-mail: tamag@unipg.it*

Received 9 August 2004; accepted in revised form 30 September 2004

Abstract. The objective of this paper is to assess the predictive capability of different classes of extended plasticity theories (bounding surface plasticity, generalized plasticity and generalized tangential plasticity) in the modeling of incremental nonlinearity, which is one of the most striking features of the mechanical behavior of granular soils, occurring as a natural consequence of the particular nature of grain interactions at the microscale. To this end, the predictions of the various constitutive models considered are compared to the results of a series of Distinct Element simulations performed *ad hoc*. In the comparison, extensive use is made of the concept of incremental strain-response envelope in order to assess the directional properties of the material response for a given initial state and stress history.

Key words: Distinct Element Method, granular materials, incremental nonlinearity, plasticity theory

1. Introduction

Nonlinearity and irreversibility are striking features of the mechanical behavior of granular soils which affect the response of virtually all geotechnical structures, such as shallow and deep foundations, excavations, tunnels, etc.

From a mathematical standpoint, the theory of plasticity has been shown to provide a convenient framework to describe these aspects of soil behavior. Early applications of perfect plasticity in geotechnical engineering, already back in the 1950s, dealt with the analysis of limit states, see *e.g.*[1,2] for a thorough account. At that time, nonlinearity and irreversibility were considered to be of importance only when dealing with failure conditions, whereas soil behavior far from failure was typically assumed as linear elastic. After the pioneering work of Roscoe and coworkers in Cambridge [3,4], the application of plasticity theories evolved towards the formulation of fully general constitutive equations for describing soil behavior in all possible initial states and loading conditions.

Over the last 20 years, a number of modifications of the classical theory have been proposed, which attempted to cover further aspects of experimentally observed soil behavior (*e.g.*, non-associativeness, intrinsic or induced anisotropy, hysteresis, soil liquefaction, strain localization into shear bands, etc.) in order to tackle other, more challenging classes of geotechnical engineering problems. These include cyclic or dynamic loading conditions (earthquakes, vibrating machines, wave loading) and soil-structure interaction in deep excavations and tunnelling, where different zones of soil experience widely different stress-paths. The breadth and depth of such scientific production is well portrayed, for example, by the proceedings of the workshops held in Grenoble [5], Cleveland [6], and Horton [7,8].

The objective of this paper is to assess the predictive capabilities of some 'extended' plasticity theories recently proposed for the analysis of cyclic behavior of soils. Rather than considering monotonic loading paths of finite and relatively large size, the main focus of this study is on the qualitative features of the *incremental* response, with particular reference to the effects of loading direction and initial state. The evaluation of the models considered is based on the comparison with a *reference* incremental response as obtained from numerical experiments with the distinct element method (DEM) on an idealized granular material consisting of a random assembly of spheres. This has been shown to be a reliable substitute for a real granular soil which allows to highlight a number of details which are very difficult, if not impossible, to extract from real laboratory experiments [9].

The structure of the paper is as follows. In Section 2, some general principles of constitutive modeling of granular materials are given and the notion of *incremental nonlinearity* is defined. The main features of the DEM model used in this study are described in Section 3, along with the details of the numerical testing program and the criteria adopted in the interpretation of the corresponding results. The details of the particular constitutive models chosen for this investigation are given in Section 4. Their predictions are then compared with the reference behavior in Section 5. Some concluding remarks are finally drawn in Section 6.

As for the notation, boldface lower- and upper-case letters are used to represent vector and tensor quantities. The symbols $\mathbf{1}$ and $\boldsymbol{I}$ are used for the second-order and fourth-order identity tensors, with components:

$$(\mathbf{1})_{ij} = \delta_{ij}, \quad (\boldsymbol{I})_{ijkl} = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{1}$$

The symmetric and skew-symmetric parts of a second-order tensor $\boldsymbol{X}$ are denoted as: $\operatorname{sym}\boldsymbol{X} := (\boldsymbol{X} + \boldsymbol{X}^T)/2$ and $\operatorname{skw}\boldsymbol{X} := (\boldsymbol{X} - \boldsymbol{X}^T)/2$, respectively. The dot product is defined as follows: $\boldsymbol{v}\cdot\boldsymbol{w} := v_i w_i$ for any two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$; $\boldsymbol{X}\cdot\boldsymbol{Y} := X_{ij}Y_{ij}$ for any two second-order tensors $\boldsymbol{X}$ and $\boldsymbol{Y}$. The dyadic product is defined as follows: $[\boldsymbol{v}\otimes\boldsymbol{w}]_{ij} := v_i w_j$ for any two vectors $\boldsymbol{v}$ and$\boldsymbol{w}$; $[\boldsymbol{X}\otimes\boldsymbol{Y}]_{ijkl} := X_{ij}Y_{kl}$ for any two second-order tensors $\boldsymbol{X}$ and $\boldsymbol{Y}$. The quantity $\|\boldsymbol{X}\| := \sqrt{\boldsymbol{X}\cdot\boldsymbol{X}}$ denotes the Euclidean norm of $\boldsymbol{X}$. The usual sign convention of soil mechanics (compression positive) is adopted throughout. In line with Terzaghi's principle of effective stress, all stresses are *effective* stresses, unless otherwise stated. In the representation of stress and strain states, use will sometimes be made of the invariant quantities p (mean stress), q (deviator stress), and θ (Lode angle), defined as

$$p := \frac{1}{3}(\boldsymbol{\sigma}\cdot\mathbf{1}), \quad q := \sqrt{\frac{3}{2}}\|\boldsymbol{s}\|, \quad \sin(3\theta) := \sqrt{6}\frac{(\boldsymbol{s}^3)\cdot\mathbf{1}}{\left[(\boldsymbol{s}^2)\cdot\mathbf{1}\right]^{3/2}} \tag{2}$$

and ϵ_v (volumetric strain), ϵ_s (deviatoric strain), $\dot{\epsilon}_v$ (volumetric strain rate), and $\dot{\epsilon}_s$ (deviatoric strain rate), defined as

$$\epsilon_v := \boldsymbol{\epsilon}\cdot\mathbf{1}, \quad \epsilon_s := \sqrt{\frac{2}{3}}\|\boldsymbol{e}\|, \quad \dot{\epsilon}_v := \dot{\boldsymbol{\epsilon}}\cdot\mathbf{1}, \quad \dot{\epsilon}_s := \sqrt{\frac{2}{3}}\|\dot{\boldsymbol{e}}\|. \tag{3}$$

In Equations (2) and (3), $\boldsymbol{s} := \boldsymbol{\sigma} - p\mathbf{1}$ is the deviatoric part of the stress tensor $\boldsymbol{e} := \boldsymbol{\epsilon} - (1/3)\epsilon_v\mathbf{1}$ and $\dot{\boldsymbol{e}} := \dot{\boldsymbol{\epsilon}} - (1/3)\dot{\epsilon}_v\mathbf{1}$ are the deviatoric parts of the strain and the strain-rate tensors, respectively, while $\boldsymbol{s}^2$ and $\boldsymbol{s}^3$ are the square and the cube of the deviatoric stress tensor, with components $(\boldsymbol{s}^2)_{ij} := s_{ik}s_{kj}$ and $(\boldsymbol{s}^3)_{ij} := s_{ik}s_{kl}s_{lj}$. It is worth noting that in $(3)_4$, with a slight abuse of notation, the symbol $\dot{\epsilon}_s$ has been employed to denote the second (deviatoric) invariant of the strain-rate tensor, which generally does not coincide with the time rate of ϵ_s, as defined in $(3)_2$.

2. Constitutive modeling of granular mterials

According to long-standing experimental evidence, the mechanical response of soils is known to strongly depend on such factors as current state, previous loading history, load increment size and direction and (possibly) time. In principle, this implies that an appropriate description of soil behavior requires the effective stress $\boldsymbol{\sigma}$ to be a function of the deformation history. Mathematically, this is expressed by the following general equation:

$$\boldsymbol{\sigma}(\boldsymbol{x},t)=\overset{\infty}{\underset{\tau=0}{\boldsymbol{\mathcal{G}}}}\left[\boldsymbol{F}^{(t)}(\boldsymbol{X},\tau)\right], \tag{4}$$

where $\boldsymbol{F}^{(t)}(\boldsymbol{X},\tau):=\boldsymbol{F}(\boldsymbol{X},t-\tau)$, with $\tau\geq 0$, is the *history* up to time t of the *deformation gradient* at a material point $\boldsymbol{X}$:

$$\boldsymbol{F}(\boldsymbol{X},t):=\frac{\partial\phi}{\partial\boldsymbol{X}}(\boldsymbol{X},t) \tag{5}$$

associated with the motion $\boldsymbol{x}=\phi(\boldsymbol{X},t)$ carrying point $\boldsymbol{X}$ in the reference configuration to its position $\boldsymbol{x}$ in the current configuration at time t[1]. Note that, since $\boldsymbol{\mathcal{G}}$ is a functional and not a function, the knowledge of the strain at time t is not sufficient to uniquely determine the state of stress.

As shown by [11], an inviscid material whose constitutive functional is differentiable is necessarily *elastic*. Therefore, *non-differentiable*, nonlinear functionals should be employed in the constitutive equation (4) whenever irreversible, inelastic behavior is of concern, which is the norm, rather than the exception, in geotechnical applications. However, working with nonlinear, non-differentiable functionals poses formidable mathematical problems, even in the simplest cases.

An alternative strategy, which is commonly adopted in soil modeling, is to use an incremental (or *rate-type*) formulation, in which the (objective) stress rate is given as a function of the rate of deformation $\boldsymbol{d}:=\mathrm{sym}\nabla\boldsymbol{v}$ ($\boldsymbol{v}:=\mathrm{d}\phi/\mathrm{d}t$ being the spatial velocity) and of the current state of the material:

$$\overset{\triangledown}{\boldsymbol{\sigma}}=\mathbf{G}\left(\boldsymbol{\sigma},\boldsymbol{q},\boldsymbol{d}\right). \tag{6}$$

In (6), $\overset{\triangledown}{\boldsymbol{\sigma}}$ is the Jaumann-Zaremba stress rate, defined as

$$\overset{\triangledown}{\boldsymbol{\sigma}}:=\dot{\boldsymbol{\sigma}}+\boldsymbol{\sigma}\boldsymbol{\omega}-\boldsymbol{\omega}\boldsymbol{\sigma}, \tag{7}$$

where $\boldsymbol{\omega}:=\mathrm{skw}\nabla\boldsymbol{v}$ is the spin tensor, and $\boldsymbol{q}$ represents a set of *internal* state variables, which are introduced to account for the effects of the previous loading history. An additional set of rate equations is then required to define the evolution of the internal variables in time (*hardening laws* in the framework of classical elastoplasticity).

In the following, we will restrict ourselves to rate-independent behavior and linear kinematics. The latter assumption allows to rewrite (6) as

$$\dot{\boldsymbol{\sigma}}=\mathbf{G}\left(\boldsymbol{\sigma},\boldsymbol{q},\dot{\boldsymbol{\epsilon}}\right), \tag{8}$$

where the Jaumann stress rate $\overset{\triangledown}{\boldsymbol{\sigma}}$ has been replaced by the standard time rate $\dot{\boldsymbol{\sigma}}$, and the rate of deformation $\boldsymbol{d}$ with the (linearized) strain rate tensor $\dot{\boldsymbol{\epsilon}}$.

[1]That the state of stress at a point is solely determined by the history of the deformation gradient at that point, and does not depend on the deformation of the neighboring points, is a consequence of the principle of *local action*; see [10].

Rate-independence means that a change in the time scale does not affect the material response, *e.g.*, doubling the strain rate doubles the stress rate. More generally:

$$\boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \lambda\dot{\boldsymbol{\epsilon}}\right) = \lambda \boldsymbol{G}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \dot{\boldsymbol{\epsilon}}\right), \quad \forall \lambda > 0. \tag{9}$$

The above equation states that the function $\boldsymbol{G}$ is *positively homogeneous* of degree one in $\dot{\boldsymbol{\epsilon}}$. This latter property yields the following alternative expression for the constitutive equation (8):

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}\left(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta}\right)\dot{\boldsymbol{\epsilon}}, \tag{10}$$

where $\boldsymbol{D}$ represents the (fourth-order) tangent stiffness tensor at the current state, which depends on the strain rate only through its *direction*, defined by the unit tensor:

$$\boldsymbol{\eta} := \frac{\dot{\boldsymbol{\epsilon}}}{\|\dot{\boldsymbol{\epsilon}}\|}. \tag{11}$$

Equation (10) provides a general representation for rate-independent constitutive equations which includes as particular cases most of the existing theories for geomaterials, including non-linear (hypo)elasticity, classical and extended plasticity, endochronic plasticity and hypoplasticity, see *e.g.* [12] and references therein. In hypoelasticity, the constitutive function $\boldsymbol{G}$ is *linear* in $\dot{\boldsymbol{\epsilon}}$, which implies that the tangent stiffness tensor $\boldsymbol{D}$ does not depend on the strain-rate direction $\boldsymbol{\eta}$. In all the other theories mentioned above, $\boldsymbol{G}$ is a *nonlinear* function of the strain rate, and $\boldsymbol{D}$ explicitly depends on $\boldsymbol{\eta}$. In this case, the material behavior is said to be *incrementally nonlinear*, see [13].

The appropriate description of the observed incrementally nonlinear behavior of granular materials is of paramount importance in constitutive modeling of granular media. This is especially relevant in a number of geotechnical engineering problems, for example in presence of cyclic loading conditions (earthquakes, vibrating machines, wave loading), or whenever different zones of soil experience widely different stress-paths, such as in deep excavations or tunnelling.

3. Incremental behavior of an analogue granular material

3.1. Preliminaries

The most natural and direct way to discriminate among the different classes of constitutive equations which are capable of describing incremental nonlinearity is to compare their predictions with the experimentally observed response of real granular materials. Conventional experiments on monotonic loading paths of finite (and quite large) size do not provide direct information in this respect. Rather, the proper definition of the incrementally nonlinear character of soil response requires the collection of data from experimental programs specifically designed in order to investigate the nature of the material response as a function of loading intensity and direction. To this end, the most natural approach is to perform a program of so-called *stress-probing* experiments, in which incremental strains are typically measured by applying a series of stress increments to a number of 'identical' specimens, with a common initial state. The stress increments (*probes*) should have an identical magnitude and point in different directions of the stress space.

As far as plasticity theory is concerned, such an experimental investigation might help answering such key questions as

(i) is there any domain of finite size in stress space where the response is incrementally linear (*i.e.*, hypoelastic)?
(ii) is the occurrence of irreversible (plastic) strains associated to some particular loading directions?
(iii) is the direction of plastic strains affected by loading direction?
(iv) is the plastic strain rate tensor coaxial to the stress tensor?

Only a few such investigations have been actually performed on sands [14,15], essentially because carrying out stress-probing experiments is, in practice, extremely difficult. First of all, in order to rule out the unavoidable differences between one specimen and another of the same soil, one might be tempted to perform all probes on a single specimen. However, this is not recommended because the application and removal of stress increments may generate irreversible deformations, thus altering the initial state of the material for subsequent probes. Therefore, as many separate specimens are needed as the number of probes, with the associated experimental difficulty of having each time a specimen which is ideally 'identical' to the others – not to speak of the time required to apply to each of them the same loading history up to the initial state. Moreover, the quality of the results crucially depends on the accuracy with which the strains are measured and the stresses are controlled, especially if the size of the probes is relatively small, as it should be to reasonably approximate the infinitesimal (tangential) response of the material. Finally, the distinction between reversible and irreversible components of the total measured strain increment requires either a constitutive assumption (typically, on the elastic behavior), or to perform for each probe a closed loading-unloading cycle.

In an attempt to circumvent these experimental problems, Bardet and Proubet [16] first suggested the use of numerical simulations with the distinct element method as a convenient tool to investigate the incremental behavior of granular materials. The numerical 'experiments' were conducted on a 2D idealized material consisting of a random assembly of circular disks. Further contributions along the same lines were later provided in the early 1990s [17–19]. This approach has been recently revived, also thanks to significant improvements of both computer power and numerical algorithms. Among the key factors, one should mention the dramatic increase of the maximum number of degrees of freedom which can be handled by modern computer architectures, a much better algorithmic description of boundary conditions, and improvements in terms of accuracy, stability and other related numerical issues. In particular, it is possible nowadays to perform fully 3D simulations in which the behavior of the idealized soil can be explored in the three-dimensional principal stress space.

An extensive program of 3D DEM simulations of stress-probing experiments was recently carried out by the authors [9,20,21]. In this paper the results of that study will serve as a reference for evaluating the performance of the elastoplastic constitutive models presented in Section 5.

3.2. Details of the DEM model

All the numerical simulations described in the following were performed with the Distinct Element code PFC-3D developed by ITASCA Consulting Group [22]. This code is the 3D generalization of the DEM code originally introduced for the analysis of rock-mechanics problems [23] and later applied to the micromechanics of analogue 2D granular materials by Cundall and Strack [24].

In PFC-3D, the soil grains are modeled as rigid spherical particles of arbitrary size, which can (slightly) overlap at the interparticle contacts. The particle interaction at the contacts is defined by appropriate contact constitutive equations, characterized by

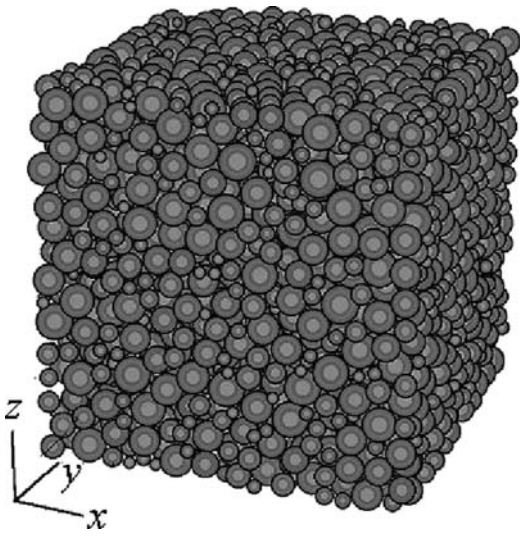


Figure 1. DEM model of a sand specimen (after [9,20,21]).

(i) linear or nonlinear (Hertzian) force-displacement law with finite normal and tangential contact stiffnesses (k_n and k_s);
(ii) intergranular slip controlled by a frictional law of Coulomb type, with friction angle ϕ_μ;
(iii) the possibility of modeling intergranular bonding at the contacts, with finite bond strength, in order to reproduce the behavior of cemented soils or soft rocks.

The equation of motion for each particle and the nonlinear contact constitutive equations at each contact are integrated in time by means of an explicit finite-difference scheme. The interested reader is referred to [25,26] for further details on the application of DEM to the modeling of idealized granular materials.

The DEM model was designed as a small but statistically equivalent sample of a real dense sand (Hostun sand), for which a limited amount of data from stress-probing experiments is available [15]. The sample is composed of about 3500 rigid and weightless spherical particles of different sizes, randomly assembled to form a cubic specimen with a side length of 4·5 mm (Figure 1). As for Hostun sand, the adopted grain-size distribution corresponds to a collection of particles with diameters ranging from 0·15 to 0·45 mm. The initial dense packing of Figure 1 is characterized by the same initial porosity (n = 0·42) of the actual sand specimens tested in the laboratory [15].

As for intergranular contacts, a linear elastic/perfectly-plastic behavior with purely frictional limit condition (no interparticle bonding) was assumed. The boundaries of the cubic specimen are defined by introducing smooth 'wall' elements, to which either stresses (through wall-applied forces) or displacement rates (through prescribed wall speed) can be imposed (walls are not shown in Figure 1).

It is important to emphasize that, contrary to most DEM studies of granular assemblies, in the interpretation of the numerical results the specimen was considered as a 'single element' subject to homogeneous states of stress and strain. No attention was paid to the details of the contact-forces distributions and local grain kinematics, the objective of the experiments being to investigate the material response from a purely phenomenological standpoint. In this respect, it can be noted that, since the walls containing the cubic specimen are frictionless, the principal directions of stresses and strains coincide with the coordinate axes x, y and z shown in Figure 1. Principal strains are

Table 1. Material parameters of the DEM model (after [9,20,21]).

k_n (kN/m)	k_s (kN/m)	$\tan(\phi_\mu)$ (–)
100·0	25·0	0·35

calculated directly from wall displacements, while the corresponding principal stresses are obtained from boundary forces, as in conventional laboratory testing.

Another fundamental issue concerns the possibility of decomposing the total computed strain increments into a reversible ('elastic') part and an irreversible ('plastic') part:

$$\Delta\boldsymbol{\epsilon} = \Delta\boldsymbol{\epsilon}^{\mathrm{e}} + \Delta\boldsymbol{\epsilon}^{\mathrm{p}}. \tag{12}$$

In earlier investigations with 2D models [17,18], a strategy based on the use of closed stress loops was adopted to compute the irreversible part of the strain increment. As for real laboratory tests, this procedure relies on the assumption that the strains occurring during unloading are purely reversible, which is, however, not always the case. A different approach was adopted in [9,20,21], where the reversible component of the strain increment was directly determined by means of an 'elastic' counterpart of the DEM specimen, in which the mechanisms responsible for energy dissipation (interparticle sliding) and structure rearrangement (opening of contacts) are inhibited at the microscale. Plastic strain increments are then computed from the kinematic decomposition (12).

The complete definition of the numerical model required the calibration of normal and shear contact stiffness, k_{n} and k_{s}, and interparticle friction angle, ϕ_μ. These parameters were determined by trial and error in order to provide a reasonable fit with the experimental data reported in [15]. In order to compensate the effect of the spherical shape of the particles in the numerical model – which could otherwise lead to unrealistically low mobilized stress ratios at ultimate failure; see [9] – all the DEM simulations were performed by keeping the particle rotations fixed. Note that the assumed boundary conditions (perfectly rigid and smooth walls) and this additional kinematic constraint tend to inhibit any inhomogeneity in the equivalent macroscopic strain field, *i.e.*, strain localization. The complete set of parameters for the DEM model is given in Table 1.

As has been thoroughly discussed by Calvetti *et al.* [9,20,21], the DEM model proves remarkably effective in describing such a complex behavior as that of a real dense sand, over a large spectrum of loading paths, both from a qualitative and – to a somewhat lesser extent – quantitative point of view. This is even more remarkable if one considers the relative simplicity of the numerical model at the microscopic level (spherical grains, simple elastic-frictional interactions at grain contacts, and fixed rotations) and the extremely low number (only 3) of material parameters required for its complete definition.

3.3. Program of numerical stress probing

As detailed in [20,21], the program of numerical stress-probing tests included the two following types of incremental loading conditions:

(a) axisymmetric probes ($\Delta\sigma_x = \Delta\sigma_y$);
(b) deviatoric probes ($\Delta\sigma_x + \Delta\sigma_y + \Delta\sigma_z = 0$).

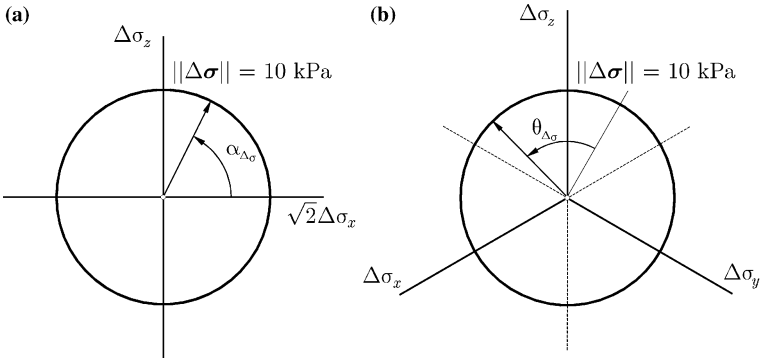


Figure 2. Incremental stress probes: (a) axisymmetric probes; (b) deviatoric probes.

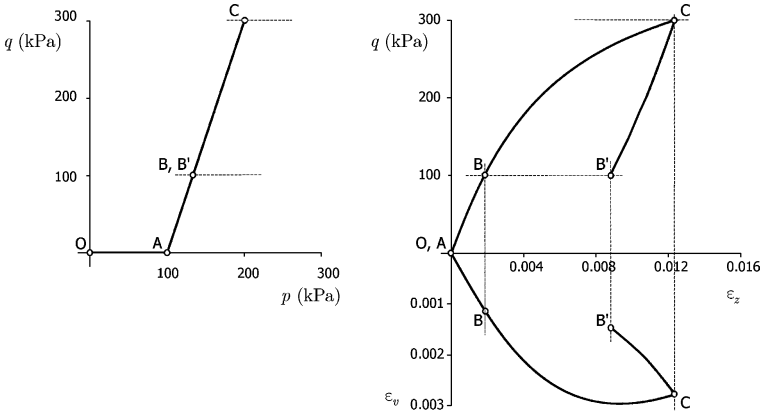


Figure 3. Stress-strain response of the DEM model upon the loading path OABCB′.

The norm of all stress probes is constant and equal to 10 kPa. The stress-probe direction is defined by the angle $\alpha_{\Delta\sigma}$ in the so-called Rendulic plane of stress increments σ_z:$\sqrt{2}\Delta\sigma_x$ for probes of type (a), and by the Lode angle of the stress increment, $\theta_{\Delta\sigma}$, for probes of type (b), see Figure 2.

In [9,20,21] several different initial states were investigated. Herein, only two of them will be considered, both characterized by the following stress state:

$$\sigma_{x0} = \sigma_{y0} = 100{\cdot}0 \text{ kPa}, \quad \sigma_{z0} = 200{\cdot}0 \text{ kPa} \tag{13}$$

but possessing a different loading history, as sketched in Figure 3. Virgin state B was reached by the two-stage stress path OA-AB, corresponding to isotropic compression and standard compression at constant confining stress, respectively. Preloaded state B′ was obtained by first loading along the path OA-AC, and then unloading along CB′.

The overall behavior of the DEM model along the loading-unloading path ABCB′ is also shown in Figure 3, in terms of deviator stress, q, and volume strain, ϵ_v, plotted as functions of axial strain, ϵ_z. Note that the net volume strain increment occurring in the closed loop BCB′ is almost negligible, *i.e.*, the porosity at B′ (n = 0·4178) is practically unchanged as compared to state B (n = 0·4176). Therefore, the two states considered differ from each other only in terms of the previous loading history.

In the following, the response of the granular assembly to the stress-probing program, as well as the predictions of the different models introduced in Section 4, is depicted by using

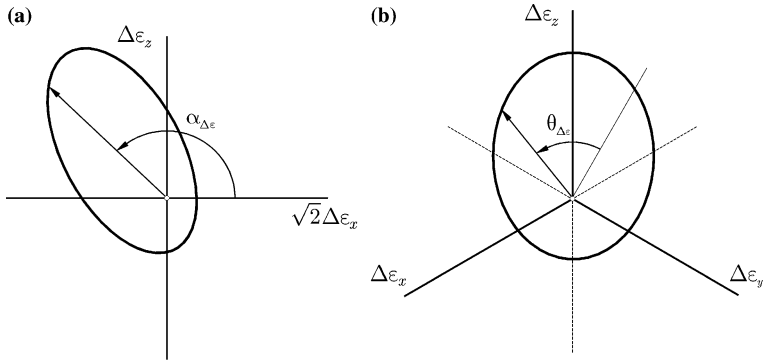


Figure 4. Incremental strain response envelopes: (a) axisymmetric probes; (b) deviatoric probes.

the so-called *incremental strain-response envelope*, as defined in [27]. Such a representation follows directly from the concept of stress-response envelope, first proposed by Gudehus [28] as a convenient tool for visualizing the properties of rate-type constitutive equations. According to Gudehus, a stress(strain)-response envelope is defined as the image in the stress(strain)-rate space of the unit sphere in the strain(stress) rate space, under the map defined by the constitutive equation. By simply replacing rates with finite-size increments, the same definitions apply to the incremental response envelopes. In the general case, an incremental strain-response envelope (RE, hereafter) is a 'surface' in a six-dimensional space. However, in the two particular loading conditions considered, the most natural choice is to represent the section of the REs in the planes of work-conjugated strain increment quantities; see Figure 4.

The size of each strain-increment vector defining the RE can be directly interpreted as a directional secant compliance of the material, for the associated loading direction and stress-increment magnitude. For a vanishing stress-increment norm, the RE asymptotically meets the strain-response envelope defined in terms of rates. Therefore, provided that the stress probes are 'small' enough, the REs obtained from numerical simulations yield useful information on the underlying constitutive behavior of the material. For example, in axisymmetric conditions, it is easy to show that an incrementally linear behavior transforms the unit circle in Figure 2a into an ellipse centered at the origin of the strain-increment plane. Thus, any significant deviation from this kind of response, such as a shift of the center of the RE from the origin or a marked asymmetry of its shape, strongly suggests some form of incremental nonlinearity.

4. Continuum plasticity models

All the models considered here have been developed as extensions of an elastoplastic constitutive model which derives from the isotropic hardening model of Nova [29] and its anisotropic hardening extension proposed by di Prisco [30,31]. The structure of the basic model is detailed in Section 4.1, while three hierarchical extensions of the formulation are presented in Sections 4.2–4.4.

4.1. Anisotropic hardening elastoplastic model

4.1.1. *General formulation*

Following the previous works of Nova and coworkers on granular soils [29–32], a constitutive model has been developed within the framework of the classical theory of anisotropic hardening elastoplasticity, which is capable to reproduce the salient features of the mechanical

response of coarse and fine-grained soils, such as pressure dependence, dilatancy, brittle-ductile transition and critical state.

As usual under the hypothesis of linear kinematics, the strain rate is decomposed additively in an elastic, reversible part, $\dot{\boldsymbol{\epsilon}}^{\mathrm{e}}$, and a plastic, irreversible part, $\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}$:

$$\dot{\boldsymbol{\epsilon}} = \dot{\boldsymbol{\epsilon}}^{\mathrm{e}} + \dot{\boldsymbol{\epsilon}}^{\mathrm{p}}. \tag{14}$$

As in [33], the elastic behavior of the material is defined by postulating the existence of a strain-energy function $\psi(\boldsymbol{\epsilon}^{\mathrm{e}})$ such that

$$\boldsymbol{\sigma}(\boldsymbol{\epsilon}^{\mathrm{e}}) = \frac{\partial \psi}{\partial \boldsymbol{\epsilon}^{\mathrm{e}}}\left(\boldsymbol{\epsilon}^{\mathrm{e}}\right) \tag{15}$$

By differentiation, the hyperelastic constitutive equation in rate form is obtained:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{\mathrm{e}}\left(\boldsymbol{\epsilon}^{\mathrm{e}}\right)\left[\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^{\mathrm{p}}\right] \tag{16}$$

in which:

$$\boldsymbol{D}^{\mathrm{e}} := \frac{\partial^2 \psi}{\partial \epsilon^{\mathrm{e}} \otimes \partial \epsilon^{\mathrm{e}}} \tag{17}$$

is the fourth-order elastic stiffness tensor.

Irreversibility is introduced by requiring the state of the material, defined in terms of the stress tensor $\boldsymbol{\sigma}$ and the internal variables $\boldsymbol{q}$, to lie in the convex set:

$$\mathbb{E}_{\sigma} := \left\{ \left(\boldsymbol{\sigma}, \boldsymbol{q}\right) \,\middle|\, f\left(\boldsymbol{\sigma}, \boldsymbol{q}\right) \leq 0 \right\}, \tag{18}$$

where $f(\boldsymbol{\sigma}, \boldsymbol{q})$ is the yield function. The evolution of plastic strains is defined by prescribing the following classical flow rule:

$$\dot{\boldsymbol{\epsilon}}^{\mathrm{p}} = \dot{\lambda}\, \boldsymbol{n}_g(\boldsymbol{\sigma}, \boldsymbol{q}), \quad \boldsymbol{n}_g := \frac{1}{g^*}\frac{\partial g}{\partial \boldsymbol{\sigma}}, \quad g^* := \left\| \frac{\partial g}{\partial \boldsymbol{\sigma}} \right\| \tag{19}$$

in which $\boldsymbol{n}_g$ defines the plastic flow direction as the normalized gradient of a scalar plastic potential $g(\boldsymbol{\sigma}, \boldsymbol{q})$ and $\dot{\lambda} \geq 0$ is the plastic multiplier.

The evolution of the internal variables is provided by the following hardening law:

$$\dot{\boldsymbol{q}} = \dot{\lambda} \boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{q}), \tag{20}$$

where $\boldsymbol{h}$ is a suitable hardening function.

The plastic multiplier appearing in $(19)_1$ and (20) is subject to the so-called *Kuhn-Tucker complementarity conditions*:

$$\dot{\lambda} \geq 0, \quad f(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0, \quad \dot{\lambda} f(\boldsymbol{\sigma}, \boldsymbol{q}) = 0, \tag{21}$$

stating that plastic deformations may occur only for states on the yield surface. Prager's *consistency condition* requires that, for plastic loading processes ($\dot{\lambda} > 0$):

$$\dot{f} = \frac{\partial f}{\partial \boldsymbol{\sigma}} \cdot \dot{\boldsymbol{\sigma}} + \frac{\partial f}{\partial \boldsymbol{q}} \cdot \dot{\boldsymbol{q}} = 0. \tag{22}$$

When the normalized *loading direction* $\boldsymbol{n}_f$ is defined as

$$\boldsymbol{n}_f := \frac{1}{f^*}\frac{\partial f}{\partial \boldsymbol{\sigma}}, \quad f^* := \left\| \frac{\partial f}{\partial \boldsymbol{\sigma}} \right\| \tag{23}$$

and the constitutive equation (16) and the hardening law (20) are taken into account, whenever plastic loading occurs, then

$$f^* \boldsymbol{n}_f \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}} - \dot{\lambda} \left(f^* \boldsymbol{n}_f \cdot \boldsymbol{D}^{\mathrm{e}} \boldsymbol{n}_g - \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{h} \right) = 0. \tag{24}$$

When, for all admissible states $\left(\boldsymbol{\sigma}, \boldsymbol{q}\right)$, the yield condition, the flow rule and the hardening law are assumed such that the inequality:

$$K_p := \boldsymbol{n}_f \cdot \boldsymbol{D}^{e} \boldsymbol{n}_g + H_p > 0, \quad H_p := -\frac{1}{f^*} \frac{\partial f}{\partial \boldsymbol{q}} \cdot \boldsymbol{h} \tag{25}$$

is *always* satisfied,[2] then Equation (24) provides the following expression for the plastic multiplier:

$$\dot{\lambda} = \frac{1}{K_p} \left\langle \boldsymbol{n}_f \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}} \right\rangle \tag{26}$$

in which $\langle x \rangle := (x + |x|)/2$ is the ramp function. The scalar quantity H_p in (25) is the so-called plastic modulus.

When the above expression for the plastic multiplier is substituted in (19) and (20), the time rates of $\boldsymbol{\sigma}$ and $\boldsymbol{q}$ can be expressed as functions of the corresponding rates of the total strain ϵ:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D} \left(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta} \right) \dot{\boldsymbol{\epsilon}}, \quad \dot{\boldsymbol{q}} = \boldsymbol{H} \left(\boldsymbol{\sigma}, \boldsymbol{q}, \boldsymbol{\eta} \right) \dot{\boldsymbol{\epsilon}} \tag{27}$$

with

$$\boldsymbol{D} := \boldsymbol{D}^{\mathrm{e}} - \frac{\mathcal{H}(\dot{\lambda})}{K_p} \left(\boldsymbol{D}^{\mathrm{e}} \boldsymbol{n}_g \right) \otimes \left(\boldsymbol{n}_f \boldsymbol{D}^{\mathrm{e}} \right), \tag{28}$$

$$\boldsymbol{H} := \frac{\mathcal{H}(\dot{\lambda})}{K_p} \boldsymbol{h} \otimes \left(n_f \boldsymbol{D}^{\mathrm{e}} \right) \tag{29}$$

in which $\mathcal{H}(x)$ is the Heaviside step function, which is equal to 1 if $x > 0$ and 0 otherwise. Note that the dependence of $\boldsymbol{D}$ and $\boldsymbol{H}$ on the strain-rate direction $\boldsymbol{\eta}$ is controlled by the sign of the plastic multiplier; see (26).

4.1.2. *Hyperelastic behavior*

The adopted stored-energy function is the same that used in [33,36]. The function $\psi(\boldsymbol{\epsilon}^{\mathrm{e}})$ is given by the following two-invariant expression:

$$\psi(\boldsymbol{\epsilon}^{\mathrm{e}}) = \bar{\psi}(\epsilon_v^{\mathrm{e}}, \epsilon_s^{\mathrm{e}}) = \tilde{\psi}(\epsilon_v) + \frac{3}{2} \left[G_0 + \frac{\alpha}{\hat{\kappa}} \tilde{\psi}(\epsilon_v) \right] \left(\epsilon_s^{\mathrm{e}} \right)^2, \tag{30}$$

where

$$\tilde{\psi}(\epsilon_{\mathrm{v}}) := \begin{cases} \hat{\kappa}\, p_r \exp\left(\epsilon_v^e / \hat{\kappa} - 1 \right) & (\epsilon_v^e \geq \hat{\kappa}), \\ p_r \epsilon_v^e + p_r \left(\epsilon_v^e - \hat{\kappa} \right)^2 / (2\hat{\kappa}) & (\epsilon_v^e < \hat{\kappa}). \end{cases} \tag{31}$$

[2]In the general case of non-associative hardening plasticity, assumption (25) places a restriction on the amount of allowable softening, as discussed in [34] in the context of a simple one-dimensional case, excluding the possibility of *subcritical softening*, as defined in [35]. In practice, this restriction does not represent a true limitation of the theory, as this phenomenon is not observed in most geomaterials, except perhaps in very stiff, brittle rocks.

In Equations (30), (31), p_r represents a reference mean stress, while $\hat{\kappa}$, G_0 and α are material constants. From (16) and (30) the following expression for the elastic stiffness tensor is obtained:

$$\boldsymbol{D}^{\mathrm{e}} := \left[1 + \frac{3\alpha}{2\hat{\kappa}}\left(\epsilon_s^{\mathrm{e}}\right)^2\right] K_\epsilon \mathbf{1}\otimes\mathbf{1} + 2\left(G_0 + \frac{\alpha}{\hat{\kappa}}\,\tilde{\psi}\right)\left[\boldsymbol{I} - \frac{1}{3}\mathbf{1}\otimes\mathbf{1}\right] + \\ + 2\left(\frac{\alpha}{\hat{\kappa}}\right)\theta_\epsilon\left(\mathbf{1}\otimes\boldsymbol{e}^{\mathrm{e}} + \boldsymbol{e}^{\mathrm{e}}\otimes\mathbf{1}\right), \tag{32}$$

where $\boldsymbol{e}^{\mathrm{e}} := \boldsymbol{\epsilon}^{\mathrm{e}} - (\epsilon_v^{\mathrm{e}}/3)\mathbf{1}$ is the deviatoric elastic strain, and the two functions $\theta_\epsilon(\epsilon_v^{\mathrm{e}})$ and $K_\epsilon(\epsilon_v^{\mathrm{e}})$ are given by

$$\theta_\epsilon := \frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}\epsilon_v^{\mathrm{e}}} = \begin{cases} p_r \exp\left(\epsilon_v^{\mathrm{e}}/\hat{\kappa} - 1\right) & (\epsilon_v^{\mathrm{e}} \geq \hat{\kappa}), \\ p_r\left(\epsilon_v^{\mathrm{e}}/\hat{\kappa}\right) & (\epsilon_v^{\mathrm{e}} < \hat{\kappa}), \end{cases} \tag{33}$$

and

$$K_\epsilon := \frac{\mathrm{d}\theta_\epsilon}{\mathrm{d}\epsilon_v^{\mathrm{e}}} = \begin{cases} (p_r/\hat{\kappa}) \exp\left(\epsilon_v^{\mathrm{e}}/\hat{\kappa} - 1\right) & (\epsilon_v^{\mathrm{e}} \geq \hat{\kappa}), \\ p_r/\hat{\kappa} & (\epsilon_v^{\mathrm{e}} < \hat{\kappa}). \end{cases} \tag{34}$$

In the range $\epsilon_v^{\mathrm{e}} \geq \hat{\kappa}$, the adopted stored-energy function describes a fully nonlinear, pressure-dependent hyperelastic behavior. It is worth noting that the adopted hyperelastic model exhibits volumetric-deviatoric coupling, which is associated with the last term on the RHS of (32). This coupling effect disappears when the state of the material is isotropic ($\mathbf{e}^{\mathrm{e}} = \mathbf{0}$), or if $\alpha = 0$.

4.1.3. *Yield function and plastic potential*

The yield function and plastic potential are given by a suitably modified version of the equations proposed by Lagioia *et al.* [33,37] to account for rotational anisotropy:

$$f(\boldsymbol{\sigma}, p_{\mathrm{s}}, \boldsymbol{\delta}^a) = f(p^a, q^a, \theta^a, p_s) \\ = \left(A_f^a\right)^{K_{1f}/C_f}\left(B_f^a\right)^{-K_{2f}/C_f} p^a - p_s = 0, \tag{35}$$

$$g(\boldsymbol{\sigma}, \boldsymbol{\delta}^a) = g(p^a, q^a, \theta^a) \\ = \left(A_g^a\right)^{K_{1g}/C_g}\left(B_g^a\right)^{-K_{2g}/C_g} p^a - \tilde{p}_s = 0, \tag{36}$$

where

$$A_\alpha^a := 1 + \frac{1}{K_{1\alpha} M_\alpha(\theta^a, \boldsymbol{\delta}^a)}\,\frac{q^a}{p^a}, \tag{37}$$

$$B_\alpha^a := 1 + \frac{1}{K_{2\alpha} M_\alpha(\theta^a, \boldsymbol{\delta}^a)}\,\frac{q^a}{p^a}, \tag{38}$$

$$K_{1\alpha} := \frac{m_\alpha(1-a_\alpha)}{2(1-m_\alpha)}\left\{1 + \sqrt{1 - \frac{4a_\alpha(1-m_\alpha)}{m_\alpha(1-a_\alpha)^2}}\right\}, \tag{39}$$

$$K_{2\alpha} := \frac{m_\alpha(1-a_\alpha)}{2(1-m_\alpha)}\left\{1 - \sqrt{1 - \frac{4a_\alpha(1-m_\alpha)}{m_\alpha(1-a_\alpha)^2}}\right\}, \tag{40}$$

$$C_\alpha := (1-m_\alpha)(K_{1\alpha} - K_{2\alpha}) \tag{41}$$

with $\alpha = f$ or g, and

$$p^a := \frac{1}{3}\boldsymbol{\sigma}\cdot\boldsymbol{\delta}^a, \quad q^a := \sqrt{\frac{3}{2}}\,\|\boldsymbol{s}^a\|, \quad \sin(3\theta^a) = \frac{27}{2}\left(\frac{t^d}{q^d}\right)^3 \tag{42}$$

in which

$$\boldsymbol{s}^a := \boldsymbol{\sigma} - p^a \boldsymbol{\delta}^a, \quad \boldsymbol{d}^a := \mathrm{dev}(\boldsymbol{s}^a), \tag{43}$$

$$\left(q^d\right)^2 := \frac{3}{2}\left\|\boldsymbol{d}^a\right\|, \quad \left(t^d\right)^3 := \frac{1}{3}\,\mathrm{tr}\left(\boldsymbol{d}^{a3}\right). \tag{44}$$

The two functions $M_\alpha(\theta^a, \boldsymbol{\delta}^a)$ appearing in Equations (37) and (38) are provided by the expressions proposed by van Eekelen [38]:

$$M_\alpha(\theta^a, \boldsymbol{\delta}^a) = M_c^\alpha(\theta^a, \boldsymbol{\delta}^a)\, r_M^\alpha(\theta^a) \tag{45}$$

with

$$r_M^\alpha(\theta^a) = k_{1M}^\alpha \left\{1 + k_{2M}^\alpha \sin(3\theta^a)\right\}^{n_\alpha}, \tag{46}$$

$$k_{1M}^\alpha := \frac{1}{2^{n_\alpha}}[1 + (c_M^\alpha)^{1/n_\alpha}]^{n_\alpha}, \quad k_{2M}^\alpha := \frac{1 - (c_M^\alpha)^{1/n_\alpha}}{1 + (c_M^\alpha)^{1/n_\alpha}}, \tag{47}$$

again with $\alpha = f$ or g. The eight quantities a_α, m_α, c_M^α and n_α appearing in (39)–(41), (45) and (46) are material constants, while $\tilde{p}_s$ is a dummy parameter. The nature of the two functions $M_c^\alpha(\theta^a, \boldsymbol{\delta}^a)$ will become clear when the hardening law for $\boldsymbol{\delta}^a$ is discussed in Section 4.1.4.

The typical shape of the yield function described by (35) is shown in Figure 5, in the q-p plane (at $\theta = \pi/6$). The set of internal variables includes the scalar quantity p_s, defining the size of the elastic domain, and the second-order tensor $\boldsymbol{\delta}^a$, which controls its orientation in stress space and accounts for plastic strain-induced anisotropy. Since $\boldsymbol{\delta}^a$ is a director, its norm $\|\boldsymbol{\delta}^a\| = \text{const.} = \sqrt{3}$. The scalars p^a, q^a and θ^a are joint invariants of the second-order tensors $\boldsymbol{\sigma}$ and $\boldsymbol{\delta}^a$, and are similar to those adopted in [39]. The plastic potential and the yield function share the same shape but in general are not coincident. The special case of associative flow ($f \equiv g$) can be obtained by an appropriate choice of the relevant parameters.

4.1.4. *Hardening laws*

The evolution of the scalar internal variable p_s, which plays the same role as the pre-consolidation pressure of classical isotropic-hardening critical-state models (see, *e.g.* [40]), is associated with the first two invariants of the plastic strain rate, as suggested in [41]:

$$\dot{p}_s = \rho_s p_s \left(\dot{\epsilon}_v^{\mathrm{p}} + \xi_s \dot{\epsilon}_s^{\mathrm{p}}\right) = \dot{\lambda}\, \rho_s p_s \left(\hat{T} + \xi_s \hat{N}\right) \tag{48}$$

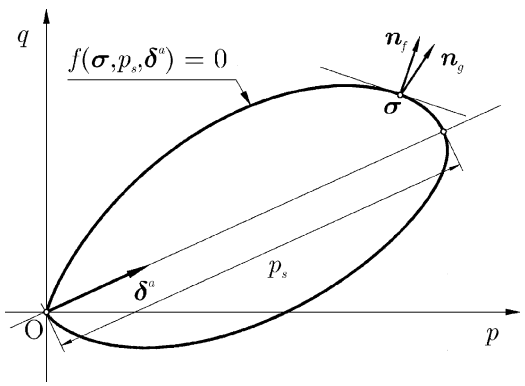


Figure 5. Yield surface of the anisotropic hardening model in the q-p plane.

in which

$$\hat{T} := \mathrm{tr}\left(\boldsymbol{n}_g\right), \quad \hat{N} := \sqrt{\frac{2}{3}}\left\|\mathrm{dev}\left(\boldsymbol{n}_g\right)\right\| \tag{49}$$

and ρ_s and ξ_s are material constants, controlling the rate of expansion/contraction of the yield surface and the balance between deviatoric and volumetric hardening, respectively.

As $\|\boldsymbol{\delta}^a\| = \text{const.}$, the two tensors $\dot{\boldsymbol{\delta}}^a$ and $\boldsymbol{\delta}^a$ must be orthogonal:

$$\dot{\boldsymbol{\delta}}^a \cdot \boldsymbol{\delta}^a = 0.$$

Therefore, the evolution equation for $\boldsymbol{\delta}^a$ is necessarily of the form:

$$\dot{\boldsymbol{\delta}}^a = \dot{\lambda}\, k(\boldsymbol{\sigma}, p_s, \boldsymbol{\delta}^a)\left\{\hat{\boldsymbol{\chi}} - \frac{1}{3}\left(\hat{\boldsymbol{\chi}} \cdot \boldsymbol{\delta}^a\right)\boldsymbol{\delta}^a\right\}, \tag{50}$$

where $\hat{\boldsymbol{\chi}}$ is an arbitrary symmetric second-order tensor. The evolution law (48) inspires the adoption of the following expression for the scalar coefficient k:

$$k(\boldsymbol{\sigma}, \boldsymbol{\delta}^a) = \rho_\delta\left(\hat{T} + \xi_\delta \hat{N}\right), \tag{51}$$

where ρ_δ and ξ_δ are material constants controlling the rate of rotation of the yield surface. Note that, since $\boldsymbol{n}_g$ is independent of p_s, k is a function of $\boldsymbol{\sigma}$ and $\boldsymbol{\delta}^a$ only. Following di Prisco [30,31], the tensor $\hat{\boldsymbol{\chi}}$ is assumed to depend only on the current stress state, according to

$$\hat{\boldsymbol{\chi}} = \cos\hat{\varphi}\,\mathbf{1} + \sqrt{3}\sin\hat{\varphi}\,\frac{\boldsymbol{d}^a}{\left\|\boldsymbol{d}^a\right\|}, \tag{52}$$

where

$$\hat{\varphi}(\theta^a) = \hat{\varphi}_c\, r_\varphi(\theta^a) \tag{53}$$

and

$$r_\varphi(\theta^a) = k_{1\varphi}\left\{1 + k_{2\varphi}\sin(3\theta^a)\right\}^{n_\varphi}, \tag{54}$$

$$k_{1\varphi} := \frac{1}{2^{n_\varphi}}\left[1 + (c_\varphi)^{1/n_\varphi}\right]^{n_\varphi}, \quad k_{2\varphi} := \frac{1 - (c_\varphi)^{1/n_\varphi}}{1 + (c_\varphi)^{1/n_\varphi}}. \tag{55}$$

It is easy to recognize from (50) that the tensor $\hat{\boldsymbol{\chi}}$, with $\|\hat{\boldsymbol{\chi}}\| = \sqrt{3}$, is the asymptotic value of $\boldsymbol{\delta}^a$ in the limit condition

$$\left(\dot{\boldsymbol{\sigma}}, \dot{p}_s, \dot{\boldsymbol{\delta}}^a\right) \to (\mathbf{0}, 0, \mathbf{0}).$$

The scalar quantity $\hat{\varphi}$ represents the maximum inclination of the tensor $\hat{\boldsymbol{\chi}}$ with respect to the isotropic axis, and is assumed to depend on the Lode angle via the van Eekelen expression (53), characterized by the three scalar parameters $\hat{\varphi}_c$, c_φ and n_φ. The scalar constant $\hat{\varphi}_c$ is the value of the inclination $\hat{\varphi}$ in axisymmetric compression ($\theta^a = \pi/6$), while the constant c_φ defines the ratio between the values of $\hat{\varphi}$ in axisymmetric extension ($\theta^a = -\pi/6$) and compression.

Anandarajah and Dafalias [39] and di Prisco [30,31] observed that, to better reproduce the available experimental data, the shape of the yield function must change as it rotates. Therefore they included an additional hardening variable to account for shape-hardening effects. In

this work, the simpler strategy is adopted of linking the values of the functions M_c^f and M_c^g appearing in (45) to a scalar measure of the yield surface rotation. In particular, we set

$$M_c^\alpha(\theta^a, \boldsymbol{\delta}^a) = M_{c0}^\alpha + \left(M_{c1}^\alpha - M_{c0}^\alpha\right) \frac{\varphi}{\hat{\varphi}(\theta^a)} \qquad (\alpha = f, g), \tag{56}$$

where the angle φ is defined as

$$\varphi := \cos^{-1}\left(\frac{1}{3}\,\mathrm{tr}\,\boldsymbol{\delta}^a\right). \tag{57}$$

Equation (56) defines a linear interpolation for M_c^α between the two values M_{c0}^α (at $\varphi = 0$) and M_{c1}^α (at $\varphi = \hat{\varphi}$).

4.2. Bounding-Surface model with radial mapping

An important limitation of classical elastoplasticity as applied to geomaterials is represented by the assumption of a large elastic domain, inside which the response of the material is purely reversible. In view of the concepts introduced in Section 2, classical – perfect or hardening – elastoplasticity is characterized by an incrementally bilinear constitutive equation *only* for states on the yield surface. All elastic states are, by definition, endowed with an incrementally linear response. However, a large body of experimental evidence suggests that soil behavior can be irreversible and path-dependent, even for strongly preloaded states, and that plastic yielding is a rather gradual process. Although such effects can be considered of secondary importance in the simulation of monotonic loading paths, it must be noted that a strong dependence of the small-strain stiffness on the loading-path direction has been observed, *e.g.*, by [42,43] in heavily overconsolidated soils, and that such a feature of soil behavior is of great importance in all practical applications in which strong variations of the stress-path direction are expected in different zones of the soil mass, *e.g.*, in the analysis of excavations. Moreover, irreversible (plastic) strains occurring well inside the locus of admissible stress states are obviously of great importance in cyclic loading processes, and the accurate description of such phenomena as cyclic mobility or liquefaction under repeated loading (see, *e.g.* [44]) requires to take these into proper account.

A number of alternative strategies have been proposed to overcome these limitations of the classical theory of plasticity; see for example the review article by [12]. In this work, we consider the two approaches of *Bounding Surface Plasticity* with radial mapping, after Dafalias and coworkers [39,45–47], and *Generalized Plasticity* after Pastor *et al.* [48]. In this section, the former approach is adopted to develop a Bounding Surface version of the basic model described in Section 4.1, while its extension to Generalized Plasticity is detailed in Section 4.3.

The key concept in the formulation of a Bounding Surface model is the fact that the domain $\mathbb{E}_\sigma$ defined by (18) represents the locus of admissible states, rather than the elastic domain of the material. The Bounding Surface (BS), which separates admissible from impossible states, is given by the equation $f(\boldsymbol{\sigma}, \boldsymbol{q}) = 0$, exactly as a standard yield surface in classical plasticity. However, such a surface is *not* a yield surface, as plastic strains can occur for stress states located in its interior.

For stress states lying on the BS, the evolution equations for stress and internal variables are identical to those of the elastoplastic reference model, *viz*. (27)–(29). For states inside the BS, the flow rule $(19)_1$ is modified as follows:

$$\dot{\boldsymbol{\epsilon}}^{\mathrm{p}} = \dot{\lambda}\,\bar{\boldsymbol{n}}_g, \tag{58}$$

in which $\bar{\boldsymbol{n}}_g$ is the plastic flow direction and $\dot{\lambda} \geq 0$ is the plastic multiplier, given by

$$\dot{\lambda} = \frac{1}{\widetilde{K}_p} \left\langle \bar{\boldsymbol{n}}_f \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}} \right\rangle, \tag{59}$$

where

$$\widetilde{K}_p := \bar{\boldsymbol{n}}_f \cdot \boldsymbol{D}^{\mathrm{e}} \bar{\boldsymbol{n}}_g + \widetilde{H}_p, \tag{60}$$

in which $\bar{\boldsymbol{n}}_f$ is the loading direction, and $\widetilde{H}_p$, by analogy with the standard formulation, plays the role of the plastic modulus.

The definition of $\bar{\boldsymbol{n}}_g$, $\bar{\boldsymbol{n}}_f$ and $\widetilde{H}_p$ requires that, for each stress state $\boldsymbol{\sigma}$ inside the BS a corresponding *image state* $\overline{\boldsymbol{\sigma}}$ is defined on the BS, through a non-invertible *mapping rule*. The mapping rule adopted in this study is a radial mapping with projection center located at the origin of the stress space, as in [45]; see Figure 6. The image state is then given by:

$$\overline{\boldsymbol{\sigma}} = b\,\boldsymbol{\sigma}, \quad b \in [1, \infty). \tag{61}$$

The loading direction, $\bar{\boldsymbol{n}}_f$, and the plastic flow direction, $\bar{\boldsymbol{n}}_g$, are taken as the normalized gradients of the functions f and g at the image state $\overline{\boldsymbol{\sigma}}$:

$$\bar{\boldsymbol{n}}_f := \frac{1}{\overline{f}^*} \frac{\partial f}{\partial \overline{\boldsymbol{\sigma}}}, \quad \bar{\boldsymbol{n}}_g := \frac{1}{\overline{g}^*} \frac{\partial g}{\partial \overline{\boldsymbol{\sigma}}}, \tag{62}$$

while the plastic modulus $\widetilde{H}_p$ is assumed to be a monotonically decreasing function of the distance

$$d := \|\overline{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\| = (b-1)\,\|\boldsymbol{\sigma}\| \tag{63}$$

between the current state and the image state and of the plastic modulus H_p evaluated at $\overline{\boldsymbol{\sigma}}$:

$$\widetilde{H}_p = \widetilde{H}\left(H_p, d\right), \quad \frac{\partial \widetilde{H}}{\partial d} > 0, \quad \widetilde{H}\left(H_p, 0\right) = H_p. \tag{64}$$

After [45] the following expression for the plastic modulus is adopted,

$$\widetilde{H}_p = H_p + \frac{h_{0L}\rho_s p_s}{\bar{f}^* \bar{g}^*} \frac{1}{\langle b/(b-1) - s_L \rangle}, \tag{65}$$

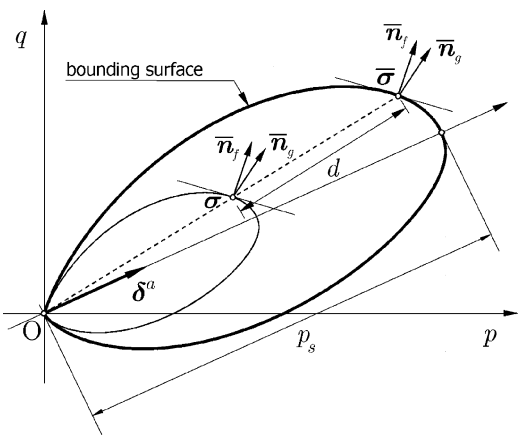


Figure 6. Bounding Surface and radial mapping rule.

in which h_{0L} and $s_L \geq 1$ are two additional material constants. The former controls the stiffness of the material for states far from the BS, while the latter is used to define a region, close to the projection center, known as elastic nucleus. In fact, when b is large enough for the term $[b/(b-1)-s_L]$ to be negative or zero, the plastic modulus in (65) goes to infinity, and the plastic strain rate vanishes. When s_L is close to 1, the elastic nucleus is small (vanishing for $s_L = 1$), whereas it tends to coincide with the BS as $s_L \to \infty$.

4.3. Generalized Plasticity model

A further step towards the extension of classical plasticity to deal with complex loading conditions, taking into account such phenomena as progressive yielding at small strain levels, history-dependence and hysteretic behavior, liquefaction, cyclic mobility or densification, is provided by the framework of *Generalized Plasticity*, first suggested by Zienkiewicz and Mroz [49] and subsequently developed by Pastor *et al.* [48] for applications to cyclic behavior of fine- and coarse-grained soils.

The main feature of Generalized Plasticity is that no plastic potential, yield surface or bounding surface need to be explicitly defined, and no consistency condition is enforced. In Generalized Plasticity the elastic constitutive equation – (15) or (16) – remains the same, whereas the flow rule $(19)_1$ is now replaced by the following evolution laws:

$$\dot{\boldsymbol{\epsilon}}^{\mathrm{p}} = \dot{\lambda}_L \boldsymbol{n}_{gL} \quad \text{if} \quad \boldsymbol{n}_L \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}} \geq 0, \tag{66}$$

$$\dot{\boldsymbol{\epsilon}}^{\mathrm{p}} = \dot{\lambda}_U \boldsymbol{n}_{gU} \quad \text{if} \quad \boldsymbol{n}_L \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}} < 0, \tag{67}$$

in which the plastic multipliers for loading ($\dot{\lambda}_L$) and unloading ($\dot{\lambda}_U$) are given by

$$\dot{\lambda}_L = \frac{1}{\widehat{K}_{pL}} \boldsymbol{n}_L \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}}, \quad \dot{\lambda}_U = \frac{1}{\widehat{K}_{pU}} \boldsymbol{n}_L \cdot \boldsymbol{D}^{\mathrm{e}} \dot{\boldsymbol{\epsilon}} \tag{68}$$

with

$$\widehat{K}_{pL} := \boldsymbol{n}_L \cdot \boldsymbol{D}^{\mathrm{e}} \boldsymbol{n}_{gL} + \widehat{H}_{pL}, \tag{69}$$

$$\widehat{K}_{pU} := \boldsymbol{n}_L \cdot \boldsymbol{D}^{\mathrm{e}} \boldsymbol{n}_{gU} + \widehat{H}_{pU}. \tag{70}$$

In Equation (66)–(70), $\boldsymbol{n}_L$, $\boldsymbol{n}_{gL}$ and $\boldsymbol{n}_{gU}$ define the loading direction, the plastic-flow direction for plastic loading and the plastic-flow direction for plastic unloading (*reverse loading*), respectively, while the scalars $\widehat{H}_{pL}$ and $\widehat{H}_{pU}$ are the corresponding plastic moduli for (plastic) loading and unloading. All these quantities are considered as prescribed functions of the state variables $(\boldsymbol{\sigma}, \boldsymbol{q})$, to be determined from the main features of the observed behavior of the material, but not necessarily deriving from any yield function, plastic potential, or consistency condition.

In order to develop a Generalized Plasticity model which could be thought of as a hierarchical extension of the elastoplastic model presented in Section 4.1 and of the corresponding Bounding Surface version of Section 4.2, the choice is made here to assume

$$\boldsymbol{n}_L \equiv \bar{\boldsymbol{n}}_f, \quad \boldsymbol{n}_{gL} \equiv \bar{\boldsymbol{n}}_g, \tag{71}$$

with $\bar{\boldsymbol{n}}_f$ and $\bar{\boldsymbol{n}}_g$ given by (62). This amounts to identifying the loading tensor with the unit normal to the bounding surface $f(\boldsymbol{\sigma}, p_{\mathrm{s}}, \boldsymbol{\delta}^a) = 0$ at the image state $\bar{\boldsymbol{\sigma}}$ – given by (61) – and the plastic-flow direction in loading with the normalized gradient of the plastic potential at $\bar{\boldsymbol{\sigma}}$ (see Figure 7). As for the plastic-flow direction in unloading, guided by the results

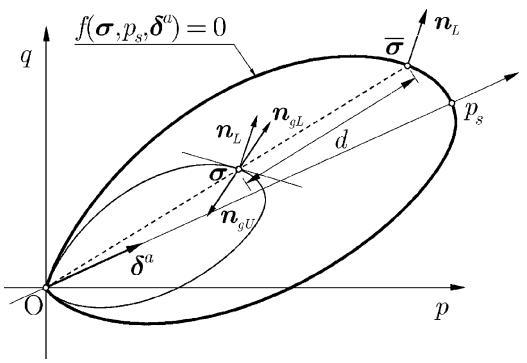


Figure 7. Definition of loading tensor $\boldsymbol{n}_L$ and plastic flow directions for loading, $\boldsymbol{n}_{gL}$, and unloading, $\boldsymbol{n}_{gU}$.

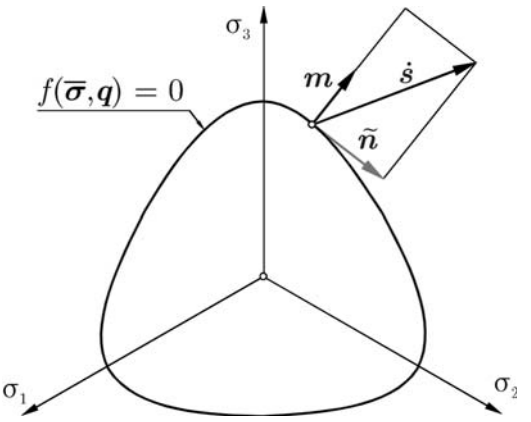


Figure 8. Geometrical interpretation of tensors $\boldsymbol{m}$ and $\tilde{\boldsymbol{n}}$ in the deviatoric plane.

obtained in DEM simulations of deviatoric probing tests, see [20], the simple choice of setting $\boldsymbol{n}_{gU}=-\boldsymbol{n}_{gL}$ has been adopted.

For states on the bounding surface, the plastic modulus $\widehat{H}_{pL}$ is given by $(25)_2$, as in the classical elastoplastic model, while $\widehat{H}_{pU}\to\infty$. For states inside the bounding surface, both plastic moduli are continuous functions of $H_p(\overline{\boldsymbol{\sigma}})$ and of the distance d between the current state and the image state:

$$\widehat{H}_{pL}=H_p+\frac{h_{0L}\rho_s p_s}{\overline{f}^*\overline{g}^*}\,\frac{1}{\langle b/(b-1)-s_L\rangle}, \tag{72}$$

$$\widehat{H}_{pU}=H_p+\frac{h_{0U}\rho_s p_s}{\overline{f}^*\overline{g}^*}\,\frac{1}{\langle b-s_U\rangle^{\beta}}, \tag{73}$$

where h_{0L}, $s_L\geq 1$, h_{0U}, $s_U\geq 1$ and β are material constants. Equation (72) is identical to (65). The simple expression adopted for the plastic modulus in unloading, $\widehat{H}_{pU}$, provides very high values close to the bounding surface ($b=\infty$ for $b\leq s_U$), which decrease with increasing distance between the current state and the image state.

4.4. Generalized tangential plasticity

The third and last extension of the basic elastoplastic model presented in Section 4.1 has been obtained by equipping the Generalized Plasticity model of Section 4.3 with the concept of a *tangential* flow rule, as proposed by Papamichos and Vardoulakis [50]. This choice is motivated by the DEM results presented in [20], which indicate that the direction of plastic-strain increments in the deviatoric plane may not be unique, and depends, in general, on the stress-increment direction.

According to [50], the additive strain decomposition in Equation (14) must be modified as follows:

$$\dot{\boldsymbol{\epsilon}}=\dot{\boldsymbol{\epsilon}}^{\mathrm{e}}+\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\parallel}+\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\perp}. \tag{74}$$

In the above equation, $\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\parallel}$ represents the conventional plastic-strain rate, given by the flow rule (66) – loading – or (67) – unloading – while $\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\perp}$ is an additional contribution to the plastic strain rate, known as *tangential* plastic strain rate.

The term tangential used for $\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\perp}$ refers to its direction with respect to the unit normal $\boldsymbol{n}_{\mathrm{L}}$ to the bounding surface. In fact, the amount of tangential plastic strain rate is governed by the following flow rule:

$$\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\perp} = \frac{1}{h_1}\widetilde{\boldsymbol{n}}, \tag{75}$$

where

$$\widetilde{\boldsymbol{n}} := \boldsymbol{I}_{\perp}\dot{\boldsymbol{s}}, \quad \boldsymbol{I}_{\perp} := \boldsymbol{I} - \boldsymbol{m}\otimes\boldsymbol{m}, \quad \boldsymbol{m} := \frac{\operatorname{dev}\left(\boldsymbol{D}^{\mathrm{e}}\boldsymbol{n}_L\right)}{\left\|\operatorname{dev}\left(\boldsymbol{D}^{\mathrm{e}}\boldsymbol{n}_L\right)\right\|} \tag{76}$$

and h_1 is a material constant, known as tangential plastic modulus. In the particular case of isotropic elasticity, $\boldsymbol{m}$ is normal to the bounding surface in the deviatoric plane, and $\widetilde{\boldsymbol{n}}$ represents the projection of the deviatoric stress rate, $\dot{\boldsymbol{s}}$, onto the tangent plane to the bounding surface at the image point (see Figure 8).

The particular choice of $\widetilde{\boldsymbol{n}}$ allows to retain the expressions given by (68) for the plastic multipliers $\dot{\lambda}_L$ and $\dot{\lambda}_U$. In fact, for states on the bounding surface, the consistency condition upon plastic loading reads

$$\dot{f} = f^{*}\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\left[\dot{\boldsymbol{\epsilon}} - \dot{\lambda}_L\boldsymbol{n}_{gL} - \frac{1}{h_1}\widetilde{\boldsymbol{n}}\right] + \dot{\lambda}_L\frac{\partial f}{\partial \boldsymbol{q}}\cdot\boldsymbol{h} = 0$$

which yields

$$\begin{aligned}\dot{\lambda}_L &= \frac{1}{\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\boldsymbol{n}_{gL} + H_p}\left(\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\dot{\boldsymbol{\epsilon}} - \frac{1}{h_1}\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\widetilde{\boldsymbol{n}}\right)\\ &= \frac{1}{\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\boldsymbol{n}_{gL} + H_{\mathrm{p}}}\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\dot{\boldsymbol{\epsilon}},\end{aligned} \tag{77}$$

since, according to (76), $\boldsymbol{n}_L\cdot\boldsymbol{D}^{\mathrm{e}}\widetilde{\boldsymbol{n}} = \widetilde{\boldsymbol{n}}\cdot\boldsymbol{D}^{\mathrm{e}}\boldsymbol{n}_L = 0$. The above result motivates the adoption of the same expressions (68), even for states inside the bounding surface, with the plastic moduli now given by (72) and (73).

Combining the elastic constitutive equation in rate form with the flow rules for the two components of the plastic-strain rate, we have:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}^{ep}\dot{\boldsymbol{\epsilon}} - \boldsymbol{D}^{\mathrm{e}}\dot{\boldsymbol{\epsilon}}^{\mathrm{p}}_{\perp}, \tag{78}$$

where

$$\boldsymbol{D}^{ep} := \boldsymbol{D}^{\mathrm{e}} - \frac{1}{\widehat{K}_{pL/\mathrm{U}}}\left(\mathbf{D}^{\mathrm{e}}\mathbf{n}_{gL/U}\right)\otimes\left(\boldsymbol{n}_L\boldsymbol{D}^{\mathrm{e}}\right) \tag{79}$$

is the elastoplastic tangent stiffness tensor, with plastic moduli and plastic-flow directions properly selected for loading or unloading according to the strain-rate direction.

The tangential plastic-strain rate appearing in the last term on the RHS of (78) can be expressed as a function of the total strain rate as follows. Let

$$\boldsymbol{I}_{\mathrm{dev}} := \boldsymbol{I} - \frac{1}{3}\mathbf{1}\otimes\mathbf{1}, \quad \boldsymbol{N} := \boldsymbol{I}_{\perp}\boldsymbol{I}_{\mathrm{dev}} = \boldsymbol{I} - \boldsymbol{m}\otimes\boldsymbol{m} - \frac{1}{3}\mathbf{1}\otimes\mathbf{1}. \tag{80}$$

Then, according to Equations (75), (72), (73) and (79), we have:

$$\dot{\boldsymbol{\epsilon}}_{\perp}^{\mathrm{p}} = \frac{1}{h_1}\boldsymbol{N}\dot{\boldsymbol{\sigma}} = \frac{1}{h_1}\boldsymbol{N}\boldsymbol{D}^{\mathrm{e}}\left(\dot{\boldsymbol{\epsilon}} - \dot{\lambda}_{L/U}\boldsymbol{n}_{gL/U} - \dot{\boldsymbol{\epsilon}}_{\perp}^{\mathrm{p}}\right)$$
$$= \frac{1}{h_1}\boldsymbol{N}\boldsymbol{D}^{ep}\dot{\boldsymbol{\epsilon}} - \frac{1}{h_1}\boldsymbol{N}\boldsymbol{D}^{\mathrm{e}}\dot{\boldsymbol{\epsilon}}_{\perp}^{\mathrm{p}}. \tag{81}$$

Solving for $\dot{\boldsymbol{\epsilon}}_{\perp}^{\mathrm{p}}$, we obtain

$$\dot{\boldsymbol{\epsilon}}_{\perp}^{\mathrm{p}} = \boldsymbol{K}^{-1}\boldsymbol{N}\boldsymbol{D}^{ep}\dot{\boldsymbol{\epsilon}} \tag{82}$$

in which

$$\boldsymbol{K} := h_1\boldsymbol{I} + \boldsymbol{N}\boldsymbol{D}^{\mathrm{e}}. \tag{83}$$

Finally, substituting (83) in (78), we obtain the constitutive equation in rate form as:

$$\dot{\boldsymbol{\sigma}} = \left\{\boldsymbol{D}^{\mathrm{ep}} - \boldsymbol{D}^{\mathrm{e}}(\boldsymbol{K}^{-1}\boldsymbol{N})\boldsymbol{D}^{\mathrm{ep}}\right\}\dot{\boldsymbol{\epsilon}}. \tag{84}$$

It is worth noting that the constitutive equation of the Generalized Plasticity model is recovered from (84) as $h_1 \to \infty$.

5. Predicted *vs.* experimental responses

In the following, the DEM response obtained from the stress-probing programs discussed in Section 3.3 is compared to the predictions of the three extended plasticity models discussed in Section 4. Herein, the comparison is intended mainly as a tool to evaluate the *qualitative* features of the directional response of the different constitutive equations. Accordingly, no detailed calibration of the relevant material parameters has been performed for the three models. Rather, the different sets of material parameters adopted in the simulations (reported in Table 2) have been assigned taking as a starting point the results of a single DEM simulation of an axisymmetric, drained compression test, as reported in [9]. Some quantitative differences between model predictions and DEM responses are therefore to be expected.

5.1. Axisymmetric probes

5.1.1. *Virgin initial state B*

The discrete DEM response to axisymmetric stress probes (Figure 2a) with $\|\Delta\boldsymbol{\sigma}\| = 10\,\mathrm{kPa}$, starting from the (virgin) initial state B, is shown in Figure 9. The corresponding predictions of the bounding surface model (BS), generalized plasticity model (GP) and generalized tangential plasticity model (GTP) are shown in Figures 10 and 11, respectively. Note that for this particular loading condition, the response of the GP and GTP models are identical, as $\widetilde{\boldsymbol{n}} = \boldsymbol{0}$.

As has already been observed in our previous publications, see *e.g.* [20], the overall behavior of the DEM model at the virgin state B conforms to that of a classical elastoplastic model with a single plastic mechanism. In particular, Figure 9 shows that the total and elastic response envelopes are nearly coincident within a significant portion of the Rendulic plane of strain increments, and that the plastic RE nearly collapses into a single line, which indicates that the plastic flow direction does not depend on the imposed loading direction.

The envelopes predicted by the BS model (Figure 10) reproduce quite well this pattern of response. A good agreement is also observed between DEM simulations and the predictions of the two Generalized Plasticity models (Figure 11), with the only possible exception of a slight overestimation of plastic-strain increments in unloading conditions, which are not

Table 2. Material parameters adopted for BS, GP and GTP models.

Model	α (–)	$\hat{\kappa}$ (–)	G_0 (MPa)	p_r (kPa)	a_f (–)	m_f (–)	c_M^f (–)	M_{c0}^f (–)	M_{c1}^f (–)
BS	0·0	0·0037	24·1	100·0	0·7	0·9999	0·9	0·7	0·6
GP	0·0	0·0037	24·1	100·0	0·7	0·9999	0·9	0·7	0·6
GTP	0·0	0·0037	24·1	100·0	0·7	0·9999	0·9	0·7	0·6
Model	a_g (–)	m_g (–)	c_M^g (–)	M_{c0}^g (–)	M_{c1}^g (–)	ρ_s (–)	ξ_s (–)	ρ_δ (–)	ξ_δ (–)
BS	0·8	0·9999	0·9	1·2	1·08	250·0	0·47	100·0	0·2
GP	0·8	0·9999	0·9	1·2	1·08	250·0	0·47	100·0	0·2
GTP	0·8	0·9999	0·9	1·2	1·08	250·0	0·47	100·0	0·2
Model		$\hat{\phi}_c$ (–)	h_{0L} (–)	s_L (–)	h_{0U} (–)	s_U (–)	β (–)	h_1 (MPa)	
BS		0·2	1·0	1·005	—	—	—	—	
GP		0·2	1·0	1·005	1·5	1·005	1·5	—	
GTP		0·2	1·0	1·005	1·5	1·005	1·5	200·0	

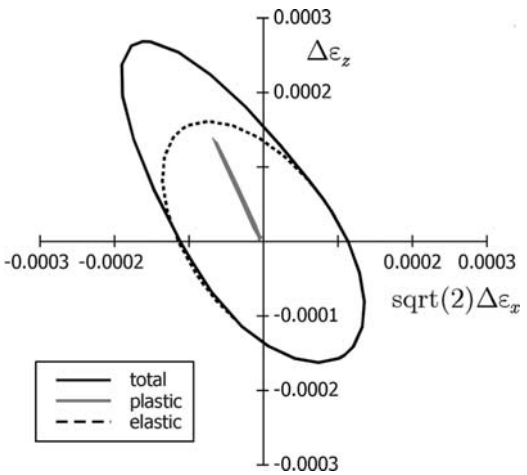


Figure 9. Axisymmetric response envelopes for virgin initial state B: DEM results.

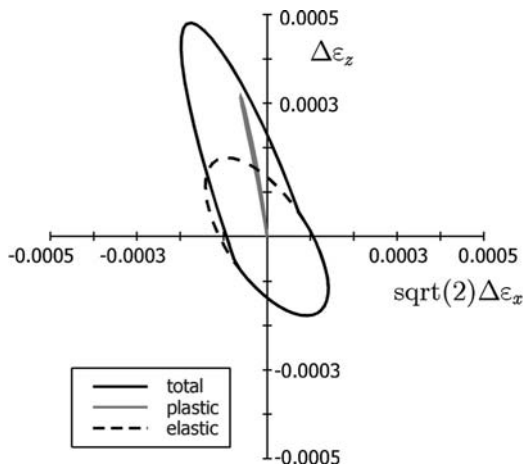


Figure 10. Axisymmetric response envelopes for virgin initial state B: BS model predictions.

detectable in the DEM results. Actually, the DEM behavior at this particular initial state could be equally well reproduced by the classical elastoplastic model outlined in Section 4.1.

5.1.2. *Preloaded initial state B′*

The most striking feature of the incremental response exhibited by the DEM model upon axisymmetric stress probing starting from the preloaded state B′, see Figure 12, is that plastic strains do occur for almost all the prescribed loading directions. These plastic-strain increments are definitely lower in magnitude as compared to those observed at the corresponding virgin state B. In a first approximation, they could be neglected if the main interest is in the

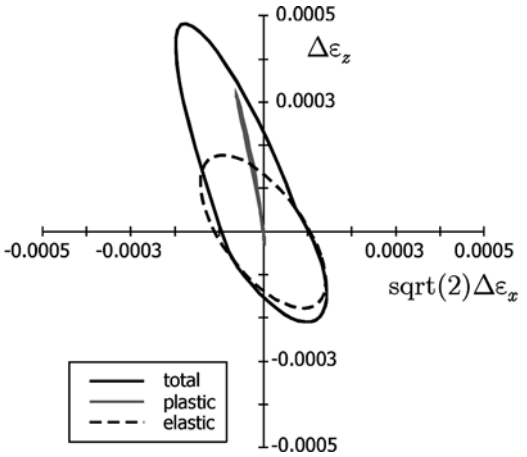


Figure 11. Axisymmetric response envelopes for virgin initial state B: GP and GTP model predictions.

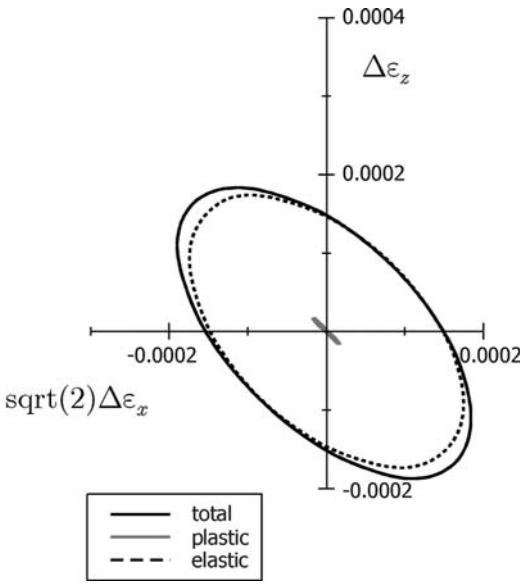


Figure 12. Axisymmetric response envelopes for preloaded initial state B′: DEM results.

modeling of gross yield phenomena at medium to large strain levels. In this case, the use of classical plasticity would be appropriate. If, however, the focus is on the constitutive modeling of material behavior at small strains, then the data clearly show the need for a more general constitutive framework. More specifically, the DEM response at the preloaded state indicates:

(i) the lack of a discernible elastic range, as plastic loading occurs for all the stress probes except for two 'neutral loading' directions, symmetric about the origin;
(ii) the existence of two distinct plastic-flow directions which, for all practical purposes, can be assumed to be coincident, but with opposite orientation.

While the BS model can predict reasonably well the pattern of plastic strains for those stress probes which induce plastic loading conditions, its response to stress probes in the opposite directions remains elastic; see Figure 13. On the contrary, the response of the GP and GTP models, shown in Figure 14, reproduces quite well these features. Although some discrepancies exist and are noticeable in the plastic-flow direction, these are solely due to details of the model formulations and/or inaccurate characterization of the relevant material parameters.

Overall, the GP and GTP models – due to their ability of developing plastic deformations in unloading – are capable of correctly reproducing the observed pattern of irreversible behavior. In particular, the fact that in generalized plasticity the plastic loading/unloading conditions are controlled by a single loading direction is consistent with feature (i), whereas it is the observation (ii) which motivated our choice about the plastic-flow direction in unloading.

5.2. Deviatoric probes

Figure 15 shows the discrete DEM responses for deviatoric stress probes with $\|\Delta\boldsymbol{\sigma}\| = 10\,\text{kPa}$ (see Figure 2b). Only the virgin state B is considered hereafter. As observed in [20], a clear dependence of the direction of plastic-strain increments on the stress-probe direction is apparent. In the deviatoric plane, the plastic RE does not collapse into a single line, but rather is a closed loop, symmetric about the $\Delta\epsilon_z$ axis. Note that the direction of the plastic-strain increments in the Rendulic plane, not shown here, is almost constant and coincident

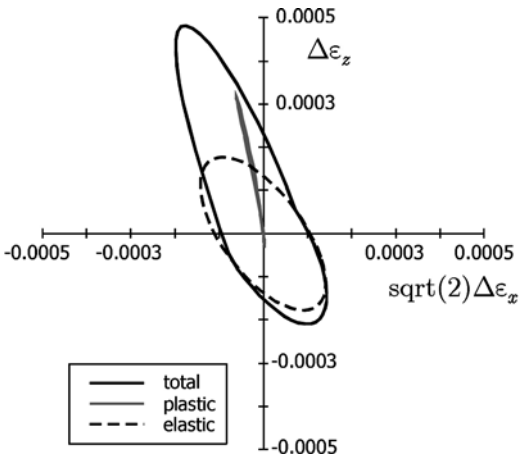


Figure 13. Axisymmetric response envelopes for pre-loaded initial state B′: BS model predictions.

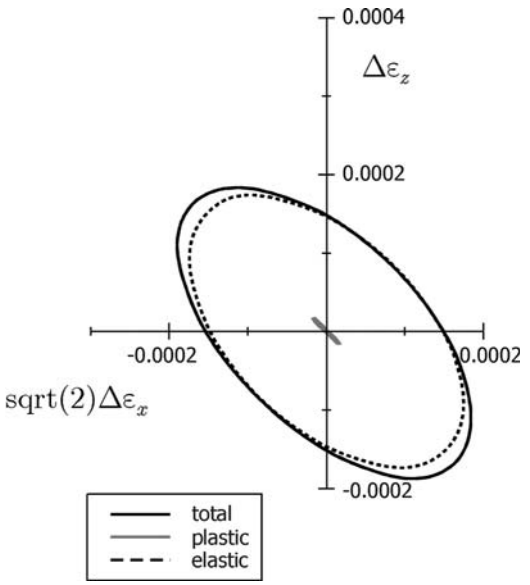


Figure 14. Axisymmetric response envelopes for pre-loaded initial state B′: GP and GTP model predictions.

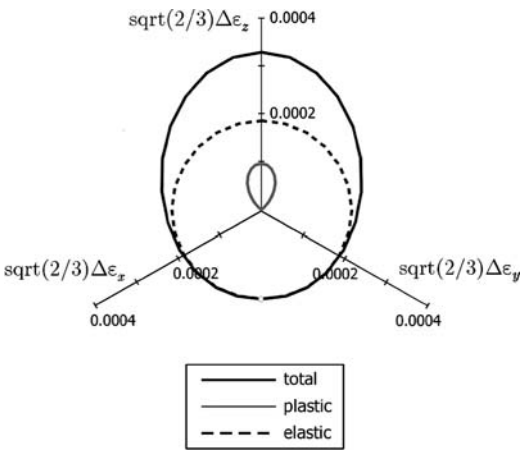


Figure 15. Deviatoric response envelopes for virgin initial state B: DEM results.

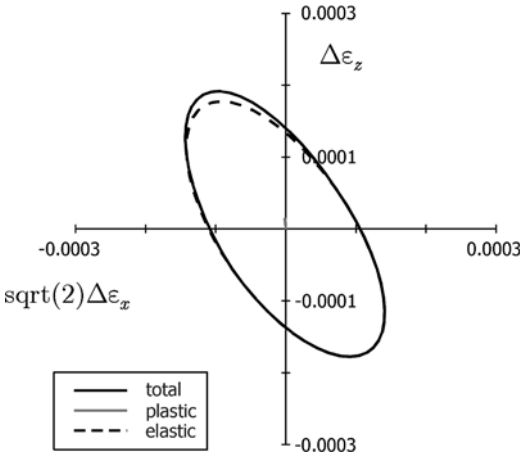


Figure 16. Deviatoric response envelopes for virgin initial state B: BS model predictions.

with the direction of incremental plastic-strains as computed in the axisymmetric case (Figure 9). Results also indicate that the plastic-strain increments corresponding to stress probes orthogonal to the $\Delta\sigma_z$ axis are not negligible.

The response of the BS model is shown in Figure 16. For infinitesimal stress probes, the plastic RE should be a single line parallel to the $\Delta\epsilon_z$ axis. However, this is not the case here, as the envelope does show some dependence of plastic-strain increments on the applied stress increment direction, although this effect is small, *i.e.*, the plastic RE is rather flat about the $\Delta\epsilon_z$ axis. It is worth noting that this particular feature of the plastic-response envelope is not to be attributed to any incremental nonlinearity of the constitutive equation, but is rather a consequence of the finite size of the stress increment and of the high curvature of the plastic potential at the initial stress state. For probes which deviate from the $\Delta\sigma_z$ axis, the current state is no longer axisymmetric as soon as the stress

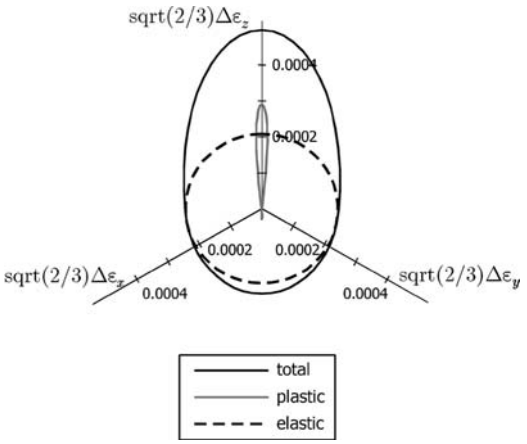


Figure 17. Deviatoric response envelopes for virgin initial state B: GP model predictions.

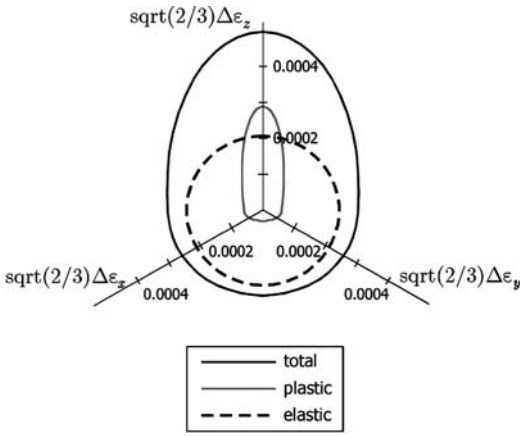


Figure 18. Deviatoric response envelopes for virgin initial state B: GTP model predictions.

increment brings the state of the material out of the so-called triaxial plane. In such conditions, the plastic strain rate changes continuously direction in the deviatoric plane, following the curvature of the plastic-potential function.

The response of the GP model, shown in Figure 17, is quite close to that of the BS model, except for the appearance of some small plastic-strain increments in unloading, which are not detected in the DEM simulations. The assumption of a single loading mechanism, controlled by tensor $\boldsymbol{n}$, is reflected in the existence of two 'neutral loading' directions, orthogonal to the $\Delta\sigma_z$ axis, according to the symmetry of the initial stress state and specimen microstructural features (induced anisotropy). This is however in sharp contrast with the observed DEM behavior (see Figure 15), as in the DEM simulations total and elastic-strain increments appear to be nearly coincident in a region of the strain-increment space which is approximately bounded by the traces of the $\Delta\epsilon_x$ and $\Delta\epsilon_y$ axes, forming an angle of about $2\pi/3$.

Such a deficiency of the GP model, which is shared by all plasticity theories with a single loading mechanism, was the main motivation for the introduction of tangential plasticity in the GP formulation. This has the effect of allowing plastic deformation for all the possible loading directions, as shown in Figure 18. The plastic RE, resulting from the sum of the normal and tangential contributions (see Figure 19), has now a lower aspect ratio, as compared to the plastic RE of the GP model. In particular, as the tangential component of the plastic strain attains its maximum for the two loading directions which are orthogonal to the axisymmetric stress plane (*i.e.*, for the two neutral loading directions of the GP model), the plastic envelope is rather large along the horizontal direction in the deviatoric plane of strain increments. However, such a feature of the plastic RE is not apparent in the plastic RE obtained in the DEM simulations, and therefore, the introduction of tangential plasticity in conventional or extended elastoplastic formulations as a mean to provide the constitutive equations with a certain degree of incremental non linearity does not appear completely satisfactory – from both a quantitative and a qualitative point of view.

6. Concluding remarks

The emphasis of this paper has been on the evaluation of extended plasticity theories for modeling the qualitative features of material response, as observed for an analogue granular material, as a function of the previous stress history and loading conditions. A key assump-

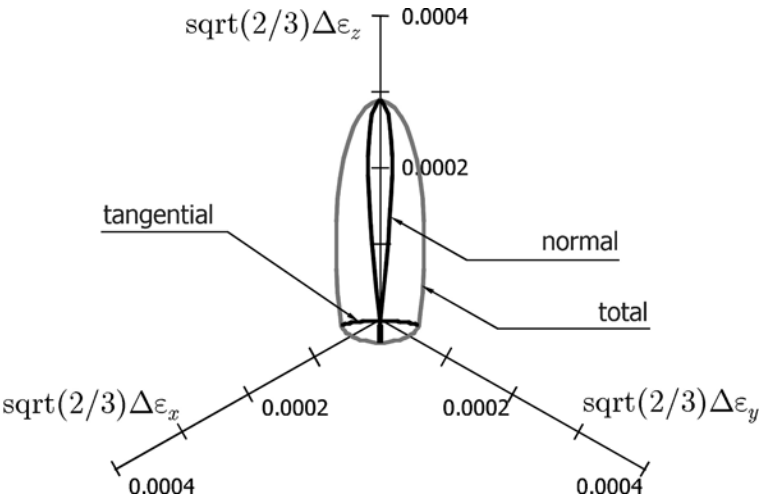


Figure 19. Deviatoric probing at virgin initial state B: detail of plastic response envelope of GTP model.

tion of the present work is that the adopted DEM model sand is capable of displaying all the relevant features of the incremental behavior of real granular soils. In fact, this has been confirmed in a number of previous studies; see *e.g.* [9,21].

The main results obtained in this investigation can be summarized as follows:

1. The observed behavior of the model sand upon stress-probing in the deviatoric plane, starting from a virgin, axisymmetric stress state, departs from the basic assumptions of classical plasticity, as the direction of the plastic strain vector in the strain increment space does depend on the stress-probe direction.
2. The observed behavior of the model sand upon axisymmetric probing from a preloaded state also departs from the basic assumptions of classical plasticity, as non-negligible plastic-strain increments are observed, even inside the so-called 'state boundary surface', separating admissible from impossible stress states in stress space.
3. Upon axisymmetric probing from a preloaded state, plastic strains do occur for (almost) all probing directions. This is clear evidence of the occurrence of reverse plastic loading, which cannot be properly described by such generalizations of classical plasticity as radial mapping Bounding Surface models.
4. Generalized Plasticity appears more suitable than Bounding Surface plasticity for capturing reverse-loading effects observed in axisymmetric probing tests.
5. Tangential Plasticity is effective in introducing a certain degree of dependence of plastic strain increment direction on loading direction (*i.e.*, incremental nonlinearity). However, the qualitative response obtained does not appear to be completely satisfactory. Whether more general approaches to incremental nonlinearity, such as, for example, the theory of hypoplasticity [51,52], can perform better in this respect, is still an open question.

References

1. V.V. Sokolowski, *Statics of Granular Media*. Oxford: Pergamon (1965) 270 pp.
2. W.F. Chen, *Limit Analysis and Soil Plasticity*. Amsterdam: Elsevier (1976) 637 pp.
3. K.H. Roscoe and J.B. Burland, On the generalised stress–strain behaviour of 'wet' clay. In: J. Heyman and F.A. Leckie (eds.), *Engineering Plasticity*. Cambridge: Cambridge Univ. Press (1968) pp. 535–609.
4. A.N. Schofield and C.P. Wroth, *Critical State Soil Mechanics*. London: McGraw-Hill (1968) 310 pp.

5. G. Gudehus, F. Darve and I. Vardoulakis (eds.), *Constitutive Relations for Soils*. Rotterdam: Balkema (1984) 497 pp.
6. A.S. Saada and G.F. Bianchini (eds.), *Constitutive Equations for Granular Non-Cohesive Soils*. Rotterdam: Balkema (1989) 733 pp.
7. D. Kolymbas (ed.), *Modern Approaches to Plasticity*. Amsterdam: Elsevier (1993) 780 pp.
8. D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer (1999) 554 pp.
9. F. Calvetti, G. Viggiani and C. Tamagnini, A numerical investigation of the incremental behavior of granular soils. *Rivista Italiana di Geotecnica*, 37 (2003) 11–29.
10. C.A. Truesdell and W. Noll, The non-linear field theories of mechanics. In: S. Flügge (ed.), *Encyclopedia of Physics*, vol. III/3. Berlin: Springer (1965) pp. 1–602
11. D.R. Owen and W.O. Williams, On the time derivatives of equilibrated response functions. *Arch. Rat'l. Mech. Anal.* 33 (1969) 288–306.
12. C. Tamagnini and G. Viggiani, Constitutive modelling for rate-independent soils: a review. *Revue Française de Génie Civil.* 6 (2002) 933–974.
13. F. Darve, The expression of rheological laws in incremental form and the main classes of constitutive equations. In: F. Darve (ed.), *Geomaterials: Constitutive Equations and Modelling*. Amsterdam: Elsevier (1990) pp. 123–148.
14. A. Anandarajah, K. Sobhan and N. Kuganenthira, Incremental stress–strain behavior of a granular soil. *J. Geotech. Engng. ASCE* 121 (1995) 57–68.
15. P. Royis and T. Doanh, Theoretical analysis of strain response envelopes using incrementally non-linear constitutive equations. *Int. J. Num. Anal. Meth. Geomech.* 22 (1998) 97–132.
16. J.-P. Bardet and J. Proubet, Application of micromechanics to incrementally nonlinear constitutive equations for granular media. In: J. Biarez and R. Gourves (eds.), *Proc. Powders and Grains 1989*. Rotterdam: Balkema (1989) pp. 265–273.
17. J.-P. Bardet, Numerical tests with discrete element method. In: D. Kolymbas (ed.), *Proc. Modern Approaches to Plasticity*. Amsterdam: Elsevier (1993) pp. 179–197.
18. J.-P. Bardet, Numerical simulations of the incremental responses of idealized granular materials. *Int. J. Plasticity* 10 (1994) 879–908.
19. F. Calvetti and C. di Prisco, Fabric evolution of granular materials: a numerical approach. In: *Proc. First Forum Young European Researchers*. Liege, Belgium (1993) pp. 115–120.
20. F. Calvetti, C. Tamagnini and G. Viggiani, On the incremental behaviour of granular soils. In G.N. Pande and S. Pietruszczak (eds.), *Proc. NUMOG VIII*. Rotterdam: Balkema (2002) pp. 3–10.
21. F. Calvetti, G. Viggiani and C. Tamagnini, Micromechanical inspection of constitutive modelling. In: C. Viggiani (ed.), *Constitutive Modelling and Analysis of Boundary Value Problems in Geotechnical Engineering*. Benevento: Hevelius (2003) pp. 187–216.
22. ITASCA. *PFC–3D User Manual*, Itasca Consulting Group, Minneapolis (1995) 110 pp.
23. P.A. Cundall, A computer model for simulating progressive large scale movements in blocky rock systems. In: *Proc. Symp. ISRM*, Nancy, France. Rotterdam: Balkema (1971) Paper no. II–8.
24. P.A. Cundall and O.D.L. Strack, A discrete numerical model for granular assemblies. *Géotechnique* 29 (1979) 47–65.
25. P.A. Cundall and R.D. Hart, Numerical modeling of discontinua. *Engg. Comp*. 9 (1992) 101–113.
26. Y. Kishino and C. Thornton, Discrete element approaches. In: F. Oka and T. Tamura (eds.), *Mechanics of Granular Materials. An Introduction*. Rotterdam: Balkema (1999) pp. 147–223.
27. C. Tamagnini, G. Viggiani, R. Chambon and J. Desrues, Evaluation of different strategies for the integration of hypoplastic constitutive equations: Application to the cloe model. *Mech. Cohesive–Frictional Mater* 5 (2000) 263–289.
28. G. Gudehus, A comparison of some constitutive laws for soils under radially symmetric loading and unloading. In: Wittke (ed.), *3rd Int. Conf. Num. Meth. Geomech.* Rotterdam: Balkema (1979) pp. 1309–1324.
29. R. Nova, Sinfonietta classica: an exercise on classical soil modelling. In: Saada and Bianchini (eds.), *Constitutive Equations for Granular Non-Cohesive Soils*. Rotterdam: Balkema (1988) pp. 501–519.
30. C. di Prisco, *Sand Anisotropy: Experimental Analysis and Mathematical Modelling.* PhD thesis, Politecnico di Milano (1993) 222 pp. (in Italian).
31. C. di Prisco, R. Nova and J.Lanier, A mixed isotropic–kinematic hardening constitutive law for sand. In: D. Kolymbas (ed.), *Modern Approaches to Plasticity*. Amsterdam: Elsevier (1993) pp. 83–124.
32. R. Nova and D.M. Wood, A constitutive model for sand in triaxial compression. *Int. J. Num. Anal. Meth. Geomech.* 3 (1979) 255–278.

33. C. Tamagnini, R. Castellanza and R. Nova, A generalized backward euler algorithm for the numerical integration of an isotropic hardening elastoplastic model for mechanical and chemical degradation of bonded geomaterials. *Int. J. Num. Anal. Meth. Geomech.* 26 (2002) 963–1004.
34. J.C. Simo and T.J.R. Hughes, *Computational Inelasticity*. Berlin: Springer (1997) 392 pp.
35. G. Maier, On associative incremental elastic-plastic constitutive models. *Rend. Ist. Lombardo di Scienze e Lettere* 100 (1966) 809–838. (in Italian).
36. R.I. Borja and C. Tamagnini, Cam–clay plasticity, part III: Extension of the infinitesimal model to include finite strains. *Comp. Meth. Appl. Mech. Engng*. 155 (1998) 73–95.
37. R. Lagioia, A.M. Puzrin and D.M. Potts, A new versatile expression for yield and plastic potential surfaces. *Comput. Geotech.* 19 (1996) 171–191.
38. H.A.M. van Eekelen, Isotropic yield surfaces in three dimensions for use in soil mechanics. *Int. J. Numer. Anal. Meth. Geomech.* 4 (1980) 89–101.
39. A. Anandarajah and Y.F. Dafalias, Bounding surface plasticity. III: Application to anisotropic cohesive soils. *J. Engng. Mech., ASCE* 112 (1986) 1292–1318.
40. D.M. Wood, *Soil Behaviour and Critical State Soil Mechanics*. Cambridge: Cambridge Univ. Press (1990) 462 pp.
41. R. Nova, On the hardening of soils. *Arch. Mech. Stosowanej* 29 (1977) 445–458.
42. J.H. Atkinson, D. Richardson and S.E. Stallebrass, Effect of recent stress history on the stiffness of overconsolidated clay. *Géotechnique* 40 (1986) 531–540.
43. S.E. Stallebrass, *Modelling the Effect of Recent Stress History on the Behaviour of Overconsolidated Soils*. PhD thesis. The City University, London (1990) 164 pp.
44. D.M. Wood, Laboratory investigations of the behaviour of soils under cyclic loading: a review. In: G.N. Pande and O.C. Zienkiewicz (eds.), *Soil Mechanics – Cyclic and Transient Loads*. Chichester: Wiley (1982) pp. 513–582.
45. Y.F. Dafalias and L.R. Herrmann, Bounding surface formulation of soil plasticity. In: G.N. Pande and O.C. Zienkiewicz (eds.), *Soil Mechanics – Cyclic and Transient Loads*. Chichester: Wiley (1982) pp. 253–283.
46. Y.F. Dafalias, Bounding surface plasticity. I: Mathematical foundation and hypoplasticity. *J. Engng. Mech., ASCE* 112 (1986) 966–987.
47. Y.F. Dafalias and L. R. Herrmann, Bounding surface plasticity. II: Application to isotropic cohesive soils. *J. Engng. Mech., ASCE* 112 (1986) 1263–1291.
48. M. Pastor, O.C. Zienkiewicz and A.H.C. Chan, Generalized plasticity and the modelling of soil behaviour. *Int. J. Numer. Anal. Meth. Geomech.* 14 (1990) 151–190.
49. O.C. Zienkiewicz and Z. Mroz, Generalized plasticity formulation and applications to geomechanics. In: C.S. Desai and R.H. Gallagher (eds.), *Mechanics of Engineering Materials*. Chichester: Wiley (1984) pp. 655–679.
50. E. Papamichos and I. Vardoulakis, Shear band formation in sand according to non–coaxial plasticity model. *Géotechnique* 4 (1995) 649–661.
51. D. Kolymbas, An outline of hypoplasticity. *Arch. Appl. Mech.* 61 (1991) 143–151.
52. C. Tamagnini, G. Viggiani and R. Chambon, A review of two different approaches to hypoplasticity. In: D. Kolymbas (ed.), *Constitutive Modelling of Granular Materials*. Berlin: Springer (2000) pp. 107–145.

Journal of Engineering Mathematics (2005) 52: 293–305
DOI 10.1007/s10665-004-1768-3

Incompressible granular flow from wedge-shaped hoppers

GRAHAM J. WEIR
Applied Mathematics, Industrial Research, P.O. B. 31-310, Lower Hutt, New Zealand
E-mail: g.weir@irl.cri.nz

Received 26 March 2003; accepted in revised form 30 July 2004

Abstract. The incompressible plastic flow equations for a Drucker-Prager yield law and a J_2 flow rule are shown not to allow a steady single radial velocity component, for flows from a wedge-shaped hopper. The corresponding equations for two components of velocity are considered, using a series expansion of Kaza and Jackson, which connects asymptotically to Jenike's radial solution. This asymptotic solution gives a poor model of mass flows about the orifice, and an improvement is obtained by considering the pressure variation along the axis of the wedge, but using the angular variations determined by the power-series method. Numerical difficulties occurred for certain parameter values, when solving the two-point boundary-value problem resulting from the asymptotic series method. The region of this parametric sensitivity is associated with an internal maximum in the pressure field, whose appearance tends to offer a conservative estimate for the mass-funnel flow transition.

Key words: granular flow, hopper, mass-funnel flow transition, mathematical model, radial flow

1. Introduction

Previous results from applying plasticity theory to granular flows from hoppers have led to apparently contradictory conclusions. Stability analyses [1,2] suggest the plastic flow equations are inherently ill-posed, which implies that the corresponding granular flows should be transient (since infinitesimal disturbances will grow exponentially rapidly) and the flow fields will contain shear banding, which are regions of high shear.

Experimental observations confirm that granular flows are both transient [3] and contain shear bands [4]. In apparent contradiction to these conclusions are the experimental observations that effectively steady predictable discharges of granular material occur from a large class of hoppers [5–7] and that the properties of these flows are approximately predicted by steady analyses of the plastic-flow equations.

Brennen and Pearce [8] analysed the discharge of an incompressible plastic material from a wedge-shaped hopper, and found promising agreement with their experimental observations of granular flows. The corresponding results for a conical hopper were derived by Nguyen *et al.* [9], again supporting the conclusion that the steady plastic-flow equations provide good agreement with the corresponding granular flows. This does not seem to be compatible with the predictions of stability analyses that granular flows should be inherently pathological.

A major difficulty with the plastic-flow equations is that no one has found an exact solution corresponding to discharge from a hopper. Jenike [10] has obtained an exact 'radial' solution, but as this solution ignores the inertial terms and does not allow for a region of zero pressure, its relationship to hopper flows about the orifice is obscure. The other analyses have all been approximate. Essentially all analyses have assumed or expanded about a single component of velocity. The initial aim of this paper is to ask if such one-component radial flows exist in steady plastic flows from a hopper.

After answering this question in the negative, the asymptotic series expansion of Kaza and Jackson [11] is used, which joins smoothly to the Jenike radial solution far above the orifice. The first term in this series expansion corresponds to the Jenike solution. This series expansion is then truncated at the second term, the resulting post-Jenike terms obtained, and the corresponding modification of the Jenike radial solution discussed, with special interest in the mass-flow–funnel-flow transition.

2. Mathematical analysis

This section assumes that a steady radial plastic flow is occurring in a wedge-shaped hopper, and satisfies a J_2 flow rule and the Drucker-Prager yield law. It is convenient to non-dimensionalise the dimensional pressure components p, the dimensional radius R and dimensional velocity U_1 through

$$p = \rho g R_1 P, \qquad R = R_1 r, \qquad U_1 = \sqrt{g R_1} u_1, \tag{1, 2, 3}$$

where R_1 is the radial position of the orifice, ρ is density, and g is gravitation acceleration. Subscripts of one are used to emphasise the one-dimensional nature of this section.

The variables P, r, and u_1 are now non-dimensional. Since we are assuming only a radial velocity component, and incompressible flow,

$$u_1 = -\frac{T(\theta)}{r}, \tag{4}$$

where T is a function of only θ, the cylindrical angle, and the minus sign has been chosen because the flow direction out of the hopper is opposite to the radial direction.

The non-dimensional Newton-force equations for this steady radial flow are

$$u_1 u_{1,r} + P_{11,r} + \frac{1}{r} P_{12,\theta} + \frac{(P_{11} - P_{22})}{r} = -\cos\theta, \tag{5}$$

$$P_{12,r} + \frac{1}{r} P_{22,\theta} + \frac{2}{r} P_{12} = \sin\theta, \tag{6}$$

where P_{ij} are the components of the stress tensor.

From the J_2 flow rule ($P_{ij} - P\delta_{ij} = -\lambda u_{(i;j)}$), $P_{33} = P$, and so the components of the pressure tensor (the negative of the stress tensor) can effectively be restricted to the 1–2 plane, allowing the Drucker-Prager yield condition to be represented by Sokolovski variables (P, γ)

$$P_{11} = P(1 - \sin\phi\cos 2\gamma), \qquad P_{22} = P(1 + \sin\phi\cos 2\gamma), \tag{7, 8}$$

$$P_{12} = -P\sin\phi\sin 2\gamma, \tag{9}$$

where ϕ is the internal angle of friction, and for $\theta \geq 0$, γ is the clockwise angle from the radial to the maximum stress direction. The choice of signs in (7–9) follows because the largest principal pressure is horizontal (because $\gamma(0) = 0$). The ratio of principal pressures equals $(1 + \sin\phi)/(1 - \sin\phi)$.

The flow rule connects the velocity and the Sokolovski angle through

$$\tan 2\gamma = -\frac{T_{,\theta}}{2T} \tag{10}$$

and so

$$T = T_0 \mathrm{e}^{-2\int_0^\theta \tan 2\gamma \, \mathrm{d}\theta} \tag{11}$$

where T_0 is a constant.

At the walls, where $\theta = \pm\theta_w$, the shear stress $P_{\theta r} = P_{12}$ is assumed equal to the wall coefficient of friction ($\tan\phi_w$) times the normal stress $P_{\theta\theta} = P_{22}$,

$$P_{12} = \mp \tan\phi_w P_{22} \quad \text{at} \quad \theta = \pm\theta_{\mathrm{w}} \tag{12}$$

and so from (7) and (8)

$$2\gamma_w = \phi_w + \arcsin\left(\frac{\sin\phi_w}{\sin\phi}\right). \tag{13}$$

We shall later require that, at the outlet of the hopper, P is zero at $r = 1$ and $\theta = \theta_w$.

The equations above are somewhat unusual, because the two equations in (5) and (6) are partial differential equations involving both r and θ, with the unknown P involvs both r and θ, while γ only involving θ. Consequently, there is the possibility that the equations above may overdetermine the unknowns; showing this is the main aim for the remainder of this section.

Substituting (7–9) in (5) and (6) yields

$$(1 - \sin\phi\cos 2\gamma)P_{,r} - \frac{\sin\phi\sin 2\gamma}{r}P_{,\theta} - \frac{2\sin\phi\cos 2\gamma(1+\gamma_{,\theta})}{r}P = \frac{T^2}{r^3} - \cos\theta, \tag{14}$$

$$\sin\phi\sin 2\gamma P_{,r} - \frac{(1+\sin\phi\cos 2\gamma)}{r}P_{,\theta} + \frac{2\sin\phi\sin 2\gamma(1+\gamma_{,\theta})}{r}P = -\sin\theta, \tag{15}$$

multiplying (14) by $\sin 2\gamma$, and adding to $\cos 2\gamma$ times (15) leads to a linear equation in P

$$\sin 2\gamma P_{,r} - \frac{(\cos 2\gamma + \sin\phi)}{r}P_{,\theta} = \frac{T^2\sin 2\gamma}{r^3} - \sin(2\gamma+\theta) \tag{16}$$

which has as its general solution

$$P = f(r\alpha) + ar - \frac{b}{r^2}, \tag{17}$$

where f is an arbitrary function,

$$\alpha = \exp\left(\int_0^\theta \frac{\sin 2\gamma\,\mathrm{d}\theta}{\cos 2\gamma + \sin\phi}\right), \qquad \beta = \exp\left(-\int_0^\theta \frac{2\sin 2\gamma\,\mathrm{d}\theta}{\cos 2\gamma + \sin\phi}\right) = \alpha^{-2}, \tag{18, 19}$$

$$a = \left(\int_0^\theta \frac{\sin(2\gamma+\theta)\mathrm{d}\theta}{(\cos 2\gamma + \sin\phi)\alpha}\right)\alpha, \qquad b = \left(\int_0^\theta \frac{T^2\sin 2\gamma\,\mathrm{d}\theta}{(\cos 2\gamma + \sin\phi)\beta}\right)\beta, \tag{20, 21}$$

where we have set $a(0) = 0 = b(0)$.

The function f can be found by substituting (17) in (14) and setting θ to zero,

$$(1 - \sin\phi)f_{,r} - \frac{2\sin\phi(\gamma_{,\theta}(0)+1)}{r}f = \frac{T^2(0)}{r^3} - 1. \tag{22}$$

The general solution of (22) is

$$f = Kr^{\omega} + \frac{r}{(\omega-1)(1-\sin\phi)} - \frac{T^2(0)}{(\omega+2)(1-\sin\phi)r^2}, \tag{23}$$

where

$$\omega = \frac{2\sin\phi(1+\gamma_{,\theta}(0))}{1-\sin\phi} \tag{24}$$

and K is a constant. The constant K is found by substituting (23) in (17) and setting P to zero at $r=1, \theta=\theta_w$. This fixes P, given γ and the constant $T(0)$. Consequently the functional form of P is the sum of terms of the form r^{ω}, r, r^{-2}, and so when this result is substituted in (14), three equations are obtained for $\gamma_{,\theta}$, making this system overdetermined. Hence, no solutions to the freely discharging hoppper problem exist with only one radial velocity component.

3. Recap

The conclusion that no wholly radial velocity solution exists for a freely discharging hopper is a surprising and disappointing result. Four alternatives suggest themselves. Firstly, no steady-flow solutions may exist to the plastic flow equations. (Here I am disregarding the radial solution of Jenike [10], which is only valid for zero velocity, and so cannot describe exactly a freely discharging hopper.) The non-existence of steady plastic flows from a freely discharging hopper is a possibility, but has not been proven. However, the exact (non-radial) solution of Weir [12] provides some hope that exact steady solutions may exist.

The second alternative is that steady, though discontinuous velocity solutions exist. Such discontinuous flows have been observed recently in numerical analyses by Gremaud *et al.* [13] for non-inertial flows. Discontinuous flows are perhaps implied in the work of Kaza and Jackson [14], and are compatible with the hyperbolic nature of the plastic-flow equations.

The third alternative is that steady radial solutions may exist, but the flow rule assumed above is incorrect, and should be disregarded. When this is done, exact solutions can be obtained for $P=a_1(\theta)r-b_1(\theta)r^{-2}$, by imposing (4) and (7–9), since then four equations result from (7) and (8), one of which fixes the angular distribution of u_1. Calculations (not shown here) reveal that, for steep hoppers, the typical radial velocity increases with θ, whereas intuition and experiment suggest the opposite. Hence, the alternative of disregarding the flow rule must be dismissed.

The fourth alternative is that steady continuous solutions exist, but the flow necessarily has two components. Within this fourth alternative, two separate limiting possibilities suggest themselves. Firstly, as the radius increases without limit, the solution could tend to the radial solution of Jenike [10]. This is implied in essentially all previous analyses of hopper flow. In this possibility, a plastic flow solution would exist over an infinite domain. In the second possibility, as the radius tends to infinity, the flow could tend to an inner non-plastic vertical funnel flow and an outer stagnant region, and the region of plastic flow is finite about the orifice. Which of these two possibilities arises depends on the rate at which the perturbation to radial flow increases with distance, relative to the radial flow approximation.

Numerical solutions [15] of the non-inertial plastic-flow equations suggest that for large radii the flow does indeed tend towards the (zero velocity) radial solution, supporting the idea of an infinite region of plasticity. However, experimental work by Baxter *et al.* [3] suggests that hopper flows tend from mass to funnel flow as the radius increases, supporting the idea of a finite region of plasticity. In the next section we assume plasticity occurs everywhere.

4. Non-radial flow

Incompressible flow in a wedge-shaped hopper implies the existence of a function χ, with

$$u_1 = \frac{1}{r}\chi_{,\theta}, \qquad u_2 = -\chi_{,r}, \tag{25, 26}$$

where u_1, u_2 are the radial and axial components of velocity.

The J_2 flow rule and the Drucker-Prager yield condition imply that the components of stress still satisfy (7–9), but now γ is a function of both r and θ.

The non-dimensional Newton-force equations are

$$u_1 u_{1,r} + \frac{u_2}{r}u_{1,\theta} - \frac{u_2^2}{r} + P_{11,r} + \frac{1}{r}P_{12,\theta} + \frac{(P_{11} - P_{22})}{r} = -\cos\theta, \tag{27}$$

$$u_1 u_{2,r} + \frac{u_2}{r}u_{2,\theta} + \frac{u_1 u_2}{r} + P_{12,r} + \frac{1}{r}P_{22,\theta} + \frac{2}{r}P_{12} = \sin\theta. \tag{28}$$

From the ratio of $P_{12}/(P_{22} - P)$ a relationship is found connecting γ and χ,

$$\tan 2\gamma = \frac{1/r^2\chi_{,\theta,\theta} - \chi_{,r,r} + \frac{1}{r}\chi_{,r}}{2\left(\frac{\chi_{,\theta}}{r}\right)_{,r}}. \tag{29}$$

Since γ equals γ_w in (13) for $r \geq 1$ and $\theta = \theta_w$, (13) and (29) impose one condition on χ at the walls. Another condition is that u_2 and $\chi_{,r}$ are both zero (χ constant) at the walls, which are assumed to be containing. A third condition on χ for symmetric flows is that $\chi = 0$ along the axis $\theta = 0$. Equation (29) can be used to replace γ in the equations above, and this results in two third-order equations from (27) and (28) for χ. The remaining variable P is contained in two first-order relationships in (27) and (28).

In this section we shall follow the work of Kaza and Jackson [11], and seek a series solution of the equations above using

$$\chi = \Sigma_{n=0}^{\infty}\chi_n(\theta)r^{-3n}, \qquad P = \Sigma_{n=0}^{\infty}P_n(\theta)r^{1-3n}, \qquad \gamma = \Sigma_{n=0}^{\infty}\gamma_n(\theta)r^{-3n}. \tag{30, 31, 32}$$

We seek a solution with χ and P even in θ and γ odd in θ, and satisfying the boundary conditions

$$\chi_n(0) = 0, \qquad \chi_n(\theta_w) = \chi_0(\theta_w)\delta_{n,0}, \tag{33, 34}$$

$$\Sigma_{n=0}^{\infty}P_n(\theta_w) = 0, \qquad \gamma_n(0) = 0, \qquad \gamma_n(\theta_w) = \gamma_w\delta_{n,0}, \tag{35, 36, 37}$$

where $\delta_{n,0}$ is the Kronecker delta function (equalling unity when n is zero, zero otherwise).

The only inertial terms contributing to zero and first order arise from the $u_1 u_{1,r}$ term. The zero; and first-order equations from (27), (28) and (30–32) are

$$(P_0 \sin\phi \sin 2\gamma_0)_{,\theta} = \cos\theta + (1 - 3\sin\phi\cos 2\gamma_0)P_0, \tag{38}$$

$$(P_0(1 + \sin\phi\cos 2\gamma_0))_{,\theta} = \sin\theta + 3P_0\sin\phi\sin 2\gamma_0, \tag{39}$$

$$\sin\phi(2P_0\cos 2\gamma_0\gamma_1 + P_1\sin 2\gamma_0)_{,\theta} = -(\chi_{0,\theta})^2 - 2P_1, \tag{40}$$

$$((1 + \sin\phi\cos 2\gamma_0)P_1)_{,\theta} = (2P_0\sin\phi\sin 2\gamma_0\gamma_1)_{,\theta}, \tag{41}$$

where to second order

$$\cos 2\gamma = \cos 2\gamma_0 - \frac{2\sin 2\gamma_0\gamma_1}{r^3}, \qquad \sin 2\gamma = \sin 2\gamma_0 + \frac{2\cos 2\gamma_0\gamma_1}{r^3}. \tag{42, 43}$$

From (29) the zero- and first-order terms give

$$\chi_{0,\theta,\theta} = -2\tan 2\gamma_0 \chi_{0,\theta}, \qquad \chi_{1,\theta,\theta} + 8\tan 2\gamma_0 \chi_{1,\theta} - 15\chi_1 = -\frac{4\gamma_1 \chi_{0,\theta}}{\cos^2 2\gamma_0} \tag{44, 45}$$

and so

$$\chi_{0,\theta} = -T_0 \exp\left(-2\int_0^\theta \tan 2\gamma_0 \mathrm{d}\theta\right) = -T_0 F, \tag{46}$$

where $T_0 = -\chi_{0,\theta}(0)$ is a positive constant. The boundary conditions in (33) and (34) and (44–45) fix χ_0 and χ_1, and so (25) and (26) fixes the zero- and first-order velocity components.

The equations above can be rewritten in essentially standard form for numerical solution as

$$(\cos 2\gamma_0 + \sin\phi) P_{0,\theta} = \sin(2\gamma_0 + \theta) + \sin 2\gamma_0 P_0, \tag{47}$$

$$\begin{aligned} 2\sin\phi(\cos 2\gamma_0 + \sin\phi) P_0 \gamma_{0,\theta} = {} & \cos\theta + \sin\phi\cos(2\gamma_0 + \theta) \\ & + P_0(1 - 2\sin\phi\cos 2\gamma_0 - 3\sin^2\phi), \end{aligned} \tag{48}$$

$$(\cos 2\gamma_0 + \sin\phi) P_{1,\theta} = 4\gamma_1 \sin\phi P_0 \gamma_{0,\theta} - \sin 2\gamma_0 F^2 T_0^2 - 2P_1 \sin 2\gamma_0, \tag{49}$$

$$\begin{aligned} 2\sin\phi(\cos 2\gamma_0 + \sin\phi) P_0 \gamma_{1,\theta} = {} & \sin^2\phi \sin 2\gamma_0 (P_1(\cos 2\gamma_0)_{,\theta} - 2\gamma_1 (P_0 \sin 2\gamma_0)_{,\theta}) \\ & - (1 + \sin\phi\cos 2\gamma_0)(F^2 T_0^2 + 2P_1) \\ & + \sin\phi\gamma_1 (2P_0 \cos 2\gamma_0)_{,\theta} + \sin\phi P_1 (\sin 2\gamma_0)_{,\theta}, \end{aligned} \tag{50}$$

$$F_{,\theta} = -2\tan 2\gamma_0 F. \tag{51}$$

We shall truncate the series expansions in (30–32) after two terms, which implies the following conditions from (33–37)

$$P_0(0) = \frac{1}{3\sin\phi + 2\sin\phi\gamma_{0,\theta}(0) - 1}, \qquad P_1(0) = \frac{(1 + \sin\phi\cos 2\gamma_w) P_0(\theta_w)}{1 + \sin\phi}, \tag{52, 53}$$

$$P_1(\theta_w) = \frac{(1 + \sin\phi) P_1(0)}{1 + \sin\phi\cos 2\theta_w}, \qquad F(0) = 1, \tag{54, 55}$$

which are useful for initial estimates of the P_0, P_1.

Equations (47–51) are five ordinary equations for the five unknown functions $P_0, \gamma_0, P_1, \gamma_1, F$. There are six boundary conditions above in (33–37) and (55), which allow these equations to be solved, and the constant T_0 to be found. These equations are essentially those in Kaza and Jackson [11], except that here the total flow is unknown, and must be derived from the known geometry. In Kaza and Jackson's work, the flow is given, and the geometry is adjusted to agree with the given total flow.

5. Numerical results

In this section we shall set $\phi_w = 0{\cdot}5\phi$, which is approximately true for aluminium walls. All of the parameters above are then functions of the two variables ϕ, θ_w.

Figure 1 plots the variation of P_0 as a function of θ and ϕ. The equation for P_0 is that for pressure in Jenike's radial solution, except that here the boundary condition on P_0 depends on

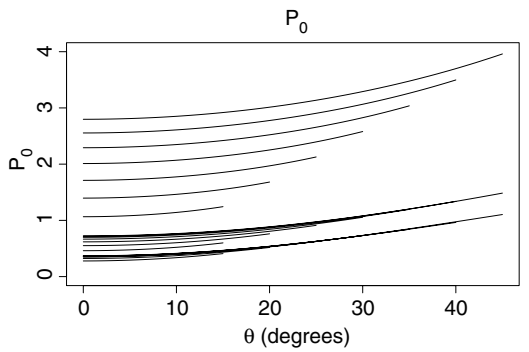


Figure 1. Variation of P_0 as a function of θ for θ_w = 15°, 20°, 25°, 30°, 35°, 40° and 45°; and for ϕ = 20°, 30° and 40°. Each of the 21 curves terminate at its θ_w value. For a given value of θ_w, the three curves for the different values of ϕ are ordered up the page, starting from ϕ = 20° at the bottom, to ϕ = 40° at the top of the figure.

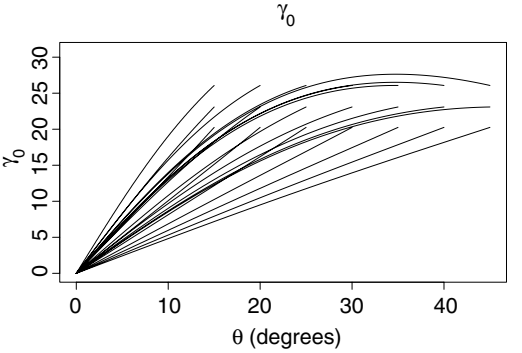


Figure 2. Variation of γ_1 as a function of θ. As for Figure 1.

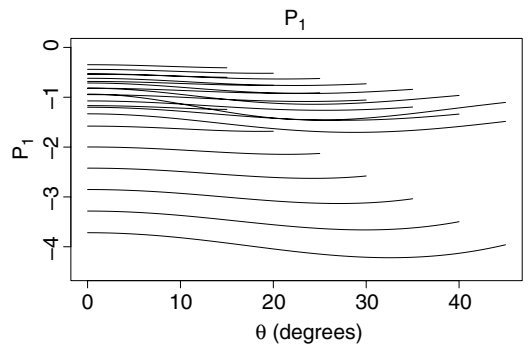


Figure 3. Variation of P_1 as a function of θ. As for Figure 1, but with ϕ values increasing down the page.

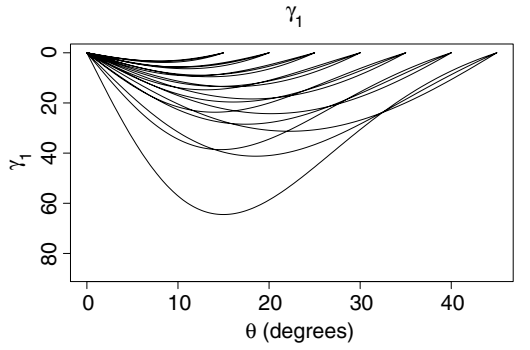


Figure 4. Variation of γ_1 as a function of θ. As for Figure 1, but with ϕ values decreasing down the page to the left of the Figure.

that for P_1 (via (35)). Similarly, the angular dependence of γ_0 is given in degrees in Figure 2, and again is just that of the Sokolovski coordinate in Jenike's radial solution.

The corresponding plot for P_1 is given in Figure 3. In contrast to P_0, P_1 is always negative, in order that P is zero at the orifice edge. Similarly, γ_1 is given in Figure 4, and is also always negative, but as θ increases, the curves for γ_1 show a change of sign in their curvature. The corresponding curves for F are given in Figure 5 for completeness, as these are just the corresponding curves for Jenike's radial solution.

A plot of the total pressure $P_0 r + P_1/r^2$ is given in Figure 6 for $\phi = 30^\circ$ and $\theta_w = 30^\circ$. The concave-downwards contours are typical of the pressure contours when θ_w is not too large for a given ϕ. Only positive pressures are valid, so the region below the zero contour needs to be omitted from consideration.

An extreme example of the variation of P with position is shown in Figure 7, for very high values of friction angle. An internal maximum of P clearly occurs about the orifice along the P = 0 contour, which suggests that a transition to funnel flow should occur in this example, as the pressure field separates the flow internally.

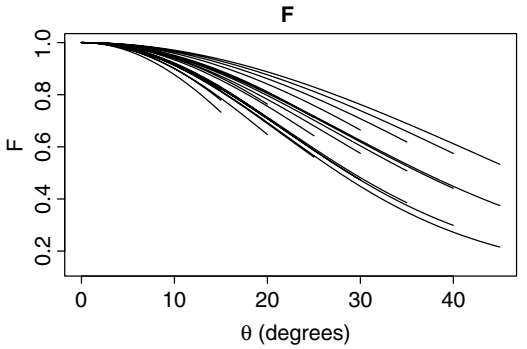


Figure 5. Variation of F as a function of θ. As for Figure 1, but with ϕ increasing down the page.

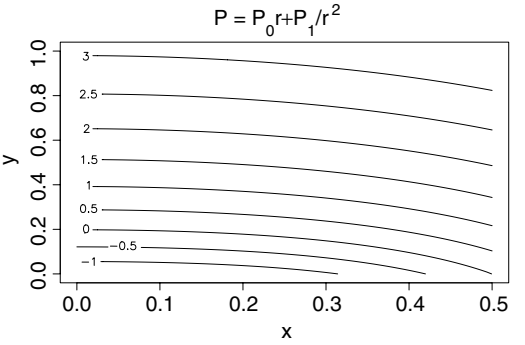


Figure 6. Contour plot of $P_0 r + P_1 r^{-2}$ about the orifice. Lengths are scaled so that the orifice width is 2 $\sin\theta_w$. Ordinate is scaled height above the orifice opening. θ_w = 30° and ϕ = 20°.

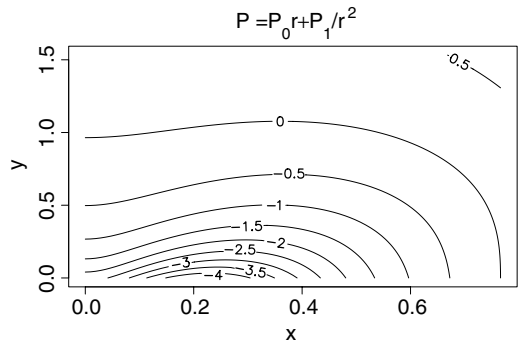


Figure 7. Contour plot of $P_0 r + P_1 r^{-2}$ about the orifice. Lengths are scaled so that the orifice width is 2 $\sin\theta_w$. Ordinate is scaled height above the orifice opening. θ_w = 50° and ϕ = 45°.

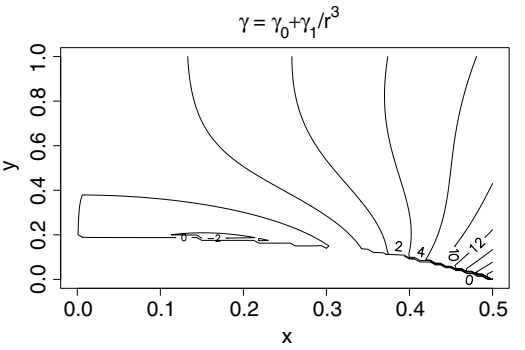


Figure 8. Contour plot of $\gamma_0 + \gamma_1 r^{-3}$ about the orifice, for $\theta_w = 30°$, and $\phi = 20°$. Lengths are scaled so that the orifice width is 2 sin θ_w. Ordinate is scaled height above the orifice opening. Contours are truncated at the $P = 0$ contour.

A plot of γ to second order, $\gamma_0 + \gamma_1 r^{-3}$, is given in Figure 8 about the orifice for $\phi = 20°$ and $\theta_w = 30°$. Far above the orifice, γ tends quickly to the radial function γ_0, but near the middle of the orifice, γ becomes negative, and these negative values extend above the zero pressure contour (where the γ contours are truncated). The surface γ = 0 corresponds to a surface of zero shear stress, from (9).

As θ_w is increased much beyond ϕ, where mass flow will have ceased, the negative values of γ above the zero pressure contour become more extreme, and can drop below $-45°$, which means the analysis above has failed, since then (29) is undefined.This failure can be associated with the failure of mass flow, because a very large γ in (11) implies a very small velocity in (4), *i.e.*, a tendency towards funnel flow. For such extreme and invalid examples, the zero-pressure contour has a maximum away from the centre of the orifice. An example of such extreme behaviour is shown in Figure 9, which shows the zero pressure contour and the invalid values of γ. Clearly, the results in this paper require the values of θ_w to be not too large, for a given value of ϕ. This is also implied in the expansion in (42) and (43).

Parameter values for which mass flow is expected, have axial velocity components u_2 which are much smaller than the corresponding radial velocity component u_1. Examples of these

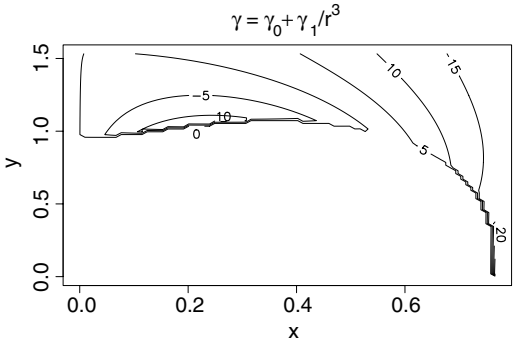


Figure 9. Contour plot of $\gamma_0+\gamma_1 r^{-3}$ about the orifice, for $\theta_w = 50^\circ$ and $\phi = 45^\circ$. Lengths are scaled so that the orifice width is 2 sin θ_w. Ordinate is scaled height above the orifice opening. Contours are truncated at the $P = 0$ contour.

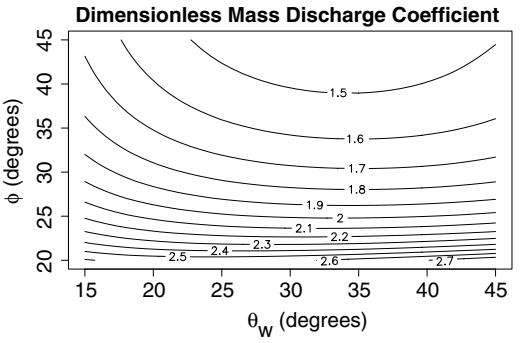


Figure 10. Contour plot of the discharge coefficient C in (57) as a function of θ_w and ϕ.

are not shown here, but the radial velocity component is always negative, and the axial component is always positive, which will tend to slightly divert the velocity flow direction from radial towards the vertical. This was also found by Kaza and Jackson [11].

From (2), (3), (25), (33–34) and (46), the mass discharge from the hopper can be written as

$$\dot{M} = C\rho L W^{\frac{3}{2}}\sqrt{g}, \tag{56}$$

where L is the length of the hopper, W its width, and C the non-dimensional mass-discharge coefficient

$$C = -\frac{\chi_0(\theta_w)}{\sqrt{2}\sin^{\frac{3}{2}}(\theta_w)} = \frac{T_0}{\sqrt{2}\sin^{\frac{3}{2}}(\theta_w)}\int_0^{\theta_w} F\,\mathrm{d}\theta \tag{57}$$

Figure 10 plots contours of C as a function of θ_w and ϕ. In contrast to experimental data, which strongly suggests that the mass discharge is a monotonically decreasing function of θ_w, Figure 10 shows a minimum mass discharge as θ_w varies, for fixed ϕ. This is suggestive of the mass-funnel flow transition, because for θ_w sufficiently large, it is easier for the flow to be confined within a more steeply angled boundary, as occurs in funnel flow.

If only small values of θ are considered, well before the minimum in the contour of C is reached, then C can be written as

$$C = \frac{h(\phi)}{(\tan(\theta_w))^m} \tag{58}$$

for some exponent m, and some function h. For example, from Figure 10, the exponent m for $\phi = 25^\circ, 30^\circ, 40^\circ$ is about 0·17, 0·27 and 0·32, respectively, for θ_w about 15°. Clearly, for these small values of θ_w, the exponent m increases with ϕ. This is opposite to the report in the reference of Laird and Roberts [16] that the exponent m is greater for ballotini than for sand, although ϕ for ballotini should be smaller than for sand. To agree with the empirical discharge law of Nedderman *et al.* [5] requires $C = 1{\cdot}03(\tan\Theta/\tan\theta_w)^{0{\cdot}2}$ where Θ is the (unspecified) angle for the mass-funnel transition.

Finally, contours of the value of r where P is zero on the axis $\theta = 0$ is plotted in Figure 11.

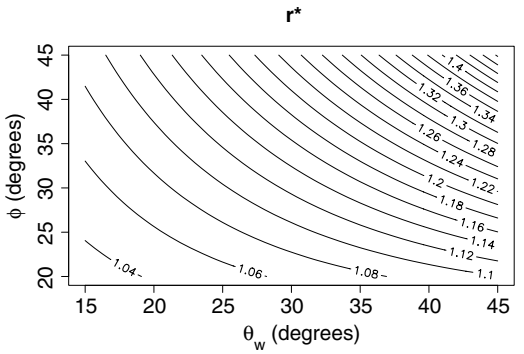


Figure 11. Contour plot of the r value where $P(r, 0)$ is zero, as functions of ϕ and θ_w.

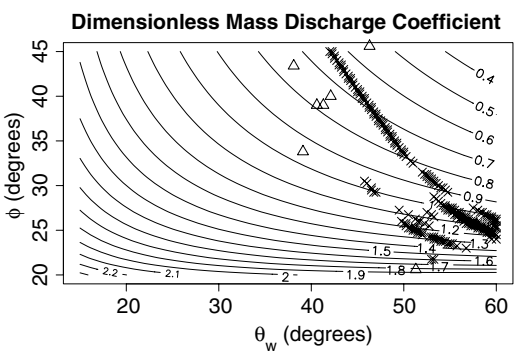


Figure 12. Contour plot of the discharge coefficient C in (57) and (60) as a function of θ_w and ϕ.

6. Axial approximation

Section 4 coupled the radial solution of Jenike with higher-order terms, using the series expansion of Kaza and Jackson [11], thereby allowing the zero-pressure condition to be imposed at the edge of the orifice. However, this analysis was asymptotic, valid for large values of r, and may be poor about the orifice (as found by Kaza and Jackson). Specifically, at the orifice, the momentum terms coupled only weakly to gravity. We would expect the zero-order terms in velocity to couple directly to gravity at the orifice, because these terms are determining the discharge, whereas only the zero-order terms in pressure coupled directly to gravity.

However, when an expansion is sought in which velocity couples directly to gravity, we find the leading terms are $\chi = \chi_0 r^{1\cdot5}$, $P = P_0 r$, $\gamma = \gamma_0$, which are unacceptable because this predicts unbounded velocities for large r. To overcome this problem, the angular variation of variables is chosen from those in Section 4, but the radial variation of pressure is found by setting $\theta = 0$ in (27), giving the axial pressure equation

$$(1-\sin\phi)P_{,r} - \frac{2P(1+\gamma_{0,\theta})(0)\sin\phi}{r} = -1 + \frac{T_0^2}{r^3}, \tag{59}$$

where, from (25) and (30) to first order, along the axis $u_1 = -T_0/r$, and T_0 is a constant. The solution of (59) is of the form $P = Ar - Br^{-2}$ for unlimited hoppers, and so requiring $P = 0$ at $r = r_0$ yields

$$T_0^2 = \frac{2(1+\gamma_{0,\theta}(0)\sin\phi)r_0^3}{3\sin\phi + 2\gamma_{0,\theta}(0)\sin\phi - 1}, \tag{60}$$

where r_0 is the value of r (close to unity) where $P = 0$ on the axis for the functions in Section 4. The dependence of r_0 is given in Figure 11.

This value of T_0 is now used in (57) and the corresponding contours are plotted in Figure 12. Also plotted in Figure 12 as triangles are measurements of the parameter values corresponding to some failure surfaces in wedge-shaped hoppers [17], indicating the transition between mass and funnel flow.

Also plotted in Figure 12 as crosses are the points where the numerical code failed to find solutions to the equations in Section 4 to a high degree of accuracy. The location of these crosses depends on the tolerances requested. Figure 12 has been constructed with only

Table 1. Nomenclature.

a	function of θ, (20)	R	dimensional radial coordinate, (2)
A	constant	R_1	dimensional radial coordinate of orifice boundary, (1)
b	function of θ, (21)		
B	constant	T	angular behaviour of u_1, (4)
C	non-dimensional mass flow coefficient	T_0	T on the axis $\theta = 0$, (11)
E	constant, (A2)	u_1	non-dimensionless radial velocity component, (4)
f	function of θ, (22)		
F	function of θ	u_2	non-dimensionless axial velocity component, (26)
g	gravitational accelertion		
G	function of θ, (A6)	U_1	dimensional radial velocity component, (3)
h	function of ϕ, (58)		
K	constant	W	width of wedge
L	length of wedge	Z	function of θ, (A7)
$\dot{M}$	mass discharge rate	α	function of θ, (18)
m	exponent	β	function of θ, (19)
M	function of θ, (A5)	χ	potential flow function, (25)
n	summation index, (33)	χ_n	function of θ, (33)
p	dimensional pressure	γ	Sokolovski angle, (7)
P	dimensionless pressure	γ_n	function of θ
P_n	dimensionless pressure component, function of θ	γ_w	Sokolovski angle at wall, (13)
		$\delta_{n,0}$	Kronecker delta function
P_{ij}	dimensionless pressure tensor, (5)	θ	angular variable
Q	function of θ, (A4)	θ_w	half hopper angle
r	non-dimensional radial coordinate, (2)	ρ	solid bulk density
r_0	non-dimensional radius where $P(r_0, 0)=0$	ω	constant, (24)
r_1	non-dimensional radial coordinate of orifice boundary	ϕ	internal angle of friction, (7)
		ϕ_w	angle of wall friction

3-significant-figure accuracy, in order to obtain a sharp boundary for the region separating convergence and non-convergence of the numerical code. It is seen that this numerical boundary of parametric sensitivity tends to be slightly greater than the boundary for the mass-funnel flow transition, which is difficult to specify accurately, because of what appears to be significant scatter in the experimental record.

7. Conclusions

This paper contains three new results. Firstly, it was shown in Section 2 that radial-flow solutions do not exist for plastic flows satisfying a Drucker-Prager yield law and the J_2 flow-rule, in a wedge-shaped hopper. This new result is interesting because of the significance of radial flows in previous analytical work.

The mathematical framework developed by Kaza and Jackson [11] was used to obtain approximations for velocity and stresses in a wedge-shaped hopper operating in mass flow. Extensive results were obtained by allowing the internal friction angle and wedge opening angle to vary significantly. Plots were obtained of the zero-pressure contour about the orifice, and the effect of the non-radial flows is to divert the radial flow to being more vertical near the orifice.

Secondly, an improved estimate for mass discharge was obtained in Section 6 by solving the pressure equation along the axis $\theta = 0$, and using the angular variation of variables

found in Section 4. The dimensionless mass-discharge coefficient was plotted in Figure 12, and showed that mass discharge decreases monotonically with internal friction angle ϕ, and (usually) with opening angle θ_w.

The predicted dependence of mass discharge on θ_w and ϕ was more complicated than shown in previous correlations with θ_w. Specifically, the exponent of $\tan\theta_w$ is predicted to vary with ϕ. Exact comparison with experiment was not attempted here, because these previous correlations scale discharge with failure angle, but do not provide a rule for calculating the failure angle. For convenience, all calculations in this paper assumed that the angle of wall friction was one half of the internal friction angle. However, all of the plots in this paper can in principle be varied by allowing the ratio between ϕ_w to ϕ to vary.

Thirdly, the new equations for P_1, γ_1, χ_1 in (49), (50) and (45) respectively, can not always be solved numerically to arbitrary precision as ϕ and θ_w increase. A definite line of singular values exist, shown in Figure 12 as crosses, where the equations did not allow a numerical solution to the requested tolerance. The failure of these equations to yield a numerical solution is closely related to the development of an internal maximum in the pressure field about the orifice. These two effects suggest a conservative estimate for the mass-funnel transition in hopper flows, although both boundaries are subject to significant scatter. Note that the first-order equations do have exact analytical solutions, as shown in Appendix A, suggesting the mass-funnel transition may result from an instability, rather than from a singularity, because the coefficient of T_0^2 in (A8) is always non-zero.

Appendix. Analytical solution of first order terms

From (41),

$$(1+\sin\phi\cos 2\gamma_0)P_1 = (2P_0\sin\phi\sin 2\gamma_0\gamma_1) - E, \tag{A1}$$

where E is a constant of integration.

From (35, 36, 37), $P_0(\theta_w) + P_1(\theta_w) = 0$, and so from (61),

$$E = (1+\sin\phi\cos 2\gamma_w)P_0(\theta_w) \tag{A2}$$

which fixes P_1.

Substituting (A1) in (40) gives a linear equation for γ_1, which can be written in the form

$$Z = (MQ)_{,\theta}, \qquad Q = \frac{2\sin\phi P_0\gamma_1}{(1+\sin\phi\cos 2\gamma_0)}, \qquad M = (\sin\phi+\cos 2\gamma_0)G, \tag{A3, A4, A5}$$

$$G = \exp\left(\int_0^\theta \frac{2\sin 2\gamma_0 \mathrm{d}\theta}{(\sin\phi+\cos 2\gamma_0)}\right), \tag{A6}$$

$$Z = G\left[\frac{2E}{1+\sin\phi\cos 2\gamma_0} + \left(\frac{\sin\phi E\sin 2\gamma_0}{(1+\sin\phi\cos 2\gamma_0)}\right)_{,\theta} - T_0^2F^2\right] \tag{A7}$$

and so from (36) and (37),

$$\int_0^{\theta_w} Z\mathrm{d}\theta = 0, \tag{A8}$$

because Q in (A4) is proportional to γ_1, and so (A8) fixes the the zero-order mean discharge speed T_0, because of the linear appearance of T_0^2 in (A7).

Acknowledgements

The figures have been calculated by Peter McGavin. The author is indebted to referee Dr Scott McCue for pointing out the paper by Kaza and Jackson.

References

1. D.G. Schaeffer, Instability in the evolution equations describing incompresible granular flow. *J. Diff. Equ.* 66 (1987) 19–50.
2. K.C. Valanis and J.F. Peters, Ill-posedness of the initial and boundary value problems in non-associative plasticity. *Acta Mech.* 114 (1996) 1–25.
3. G.W. Baxter, R.P. Behringer, T. Fagert and G.A. Johnson, Pattern formation and time-dependence in flowing sand. In: D.D. Joseph and D.G. Schaeffer (eds.), *Two-Phase Flows and Waves*. Berlin: Springer-Verlag (1990) pp.1–28.
4. H.B. Mulhaus and I. Vardoulaki, The thickness of shear bands in granular materials. *Geotechnique* 37 (1987) 271–283.
5. R.M. Nedderman, U. Tuzun, S.B. Savage and G.T. Houlsby, Review Article Number 10: The flow of granular materials: Discharge rates from hopper. *Chem. Eng. Sci.* 37 (1982) 1597–1609.
6. R.L. Brown and J.C. Richards, Profile of flow of granules through apertures. *Trans. Inst. Chem. Engng.* 38 (1960) 243–250.
7. W.A. Beverloo, H.A. Leniger and J. van de Velde, The flow of granular solids through orifices. *Chem. Eng. Sci.* 15 (1961) 260–269.
8. C. Brennen and J.C. Pearce, Granular material flow in two-dimensional hoppers. *J. Appl. Mech.* 45 (1978) 43–50.
9. T.V. Nguyen, C. Brennen and R.H. Sabersky, Gravity flow of granular materials in conical hoppers. *J. Appl. Mech.* 46 (1979) 529–535.
10. A.W. Jenike, Steady gravity flow of frictional-cohesive solids in converging channels. *J. Appl. Mech.* 31 (1964) 5–11.
11. K.R. Kaza and R. Jackson, The rate of discharge of coarse granular material from a wedge-shaped mass flow hopper. *Powder Technol*. 33 (1982) 223–237.
12. G.J. Weir, The intrinsic cohesion of granular materials. *Powder Technol*. 104 (1999) 29–36.
13. P.A. Gremaud, J.V. Matthews and M. Shearer, Similarity solutions for hopper flows. In: J. Bona, K. Saxton and R. Saxton (eds.), *Nonlinear PDEs, Dynamics and Continuum Physics*. Providence, Rhode Island: AMS Contempory Mathematics, Series 255 (2000) pp. 79–95.
14. K.R. Kaza and R. Jackson, Boundary conditions for a granular material flowing out of a hopper or bin. *Chem. Eng. Sci.* 39 (1984) 915–916.
15. J.J. Johanson and A.W. Jenike, *Stress and Velocity Fields in Gravity Flow of Bulk Solids.* Salt Lake City, Utah: Bulletin No. 116, Bulletin of the University of Utah 53 (21) (1962) 138 pp.
16. B.W. Laird and P.M. Roberts, *Chemical Engineering*. Tripos, Part 2. Research Project Report, University of Cambridge (1979).
17. R.L. Brown and J.C. Richards, Kinematics of the flow of dry powders and bulk solids. *Rheology Acta* 4 (1965) 153–165.

Journal of Engineering Mathematics (2005) 52: 307–320
DOI 10.1007/s10665-004-6011-8

Micromechanic modeling and analysis of unsteady-state granular flow in a cylindrical hopper

H.P. ZHU and A.B. YU
Centre for Simulation and Modelling of Particulate Systems, School of Materials Science and Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

Received 2 June 2004; accepted in revised form 22 July 2004

Abstract. This paper presents a numerical study of the micro- and macro-dynamic behavior of the unsteady-state granular flow in a cylindrical hopper with flat bottom by means of a modified discrete-element method (DEM) and an averaging method. The results show that the trends of the distributions of the microscopic properties such as the velocity and forces, and the macroscopic properties such as the velocity, mass density, stress and couple stress of the unsteady-state hopper flow are similar to those of steady-state hopper flow, and do not change much with the discharge of particles. However, the magnitudes of the macroscopic properties in different regions have different rates of variation. In particular, the magnitudes of the two normal stresses vary little with time in the orifice region, but decrease in other regions. The magnitude of the shear stress decreases with time when far from the bottom wall and central axis of the hopper. The results also indicate that DEM can capture the key features of the granular flow, and facilitated with a proper averaging method, can also generate information helpful to the test and development of an appropriate continuum model for granular flow.

Key words: average method, discrete-element method, granular flow, hopper flow

1. Introduction

Granular materials are quite common in nature and in industry, appearing as, for example, sands, soil, cement, grains, chemical powders, coal, mineral rocks and so on. The study of the fundamentals of these kinds of materials is essential in order to generate a method for solving a wide range of scientific and technological problems in various fields [1,2]. Computer simulation provides a cost-effective alternative to physical experimentation to achieve this goal.

In general, granular material is a discrete system whose physical properties are discontinuous with respect to position and time. Therefore, it can be described by a discrete model on an individual particle scale, which can generate detailed information about the micro-dynamic behavior of the granular material. A major type of discrete model is based on the so-called Discrete Element Method (DEM) originally proposed by Cundall and Strack [3]. In the method, the motion of every particle is traced and the gradients of translation and rotation of a particle are determined in terms of the forces and the torques exerted on it. DEM-based simulation has been recognized as an effective method to study the fundamentals of granular materials [4–9].

Granular material can also be described by a continuum model in an average sense. In the continuum description, the macroscopic behavior of granular flow is described by the balance equations facilitated with constitutive relations and boundary conditions. In the past, two continuum models developed within the framework of plasticity theory and kinetic theory of molecular dynamics have extensively been used to study the dynamic behavior of granular materials [10,11], [12, pp. 1–46], [13, Chapters 6, 7, 9]. They have been shown to be applicable

to quasi-static and rapid flow regimes, respectively. However, they do not satisfactorily apply to a system in which different flow regimes coexist such as hopper flow. An alternative approach is the combined approach of a discrete method and an averaging method, which takes into account the discrete nature of granular materials without any global assumption, thus allowing a better understanding of the fundamental mechanisms of granular flow [14–17]. In the approach, supported by a proper averaging method, the macroscopic quantities, such as density, velocity and stress can be obtained in terms of the microscopic quantities, such as velocities of particles and interaction forces and torques between particles.

In this paper, the three-dimensional granular flow in a cylindrical hopper with flat bottom is investigated with a modified DEM and an averaging method. The modified DEM and the averaging method used in this work are first briefly described. The microscopic properties directly related to the velocity and forces of particles in the hopper flow are then examined. Finally, the macroscopic properties, such as velocity, mass density, stress and couple stress are investigated. The proper use of the weighting function in the averaging method is also discussed.

2. Mathematical models

2.1. Discrete-element method

In the DEM simulation, a granular material is modeled based on a finite number of discrete, semi-rigid spherical or polygon-shaped particles interacting by means of contact or non-contact forces, and the translational and rotational motions of every single particle in a considered system are described by Newton's laws of motion. For simplicity, our present study is limited to granular systems only composed of spherical particles, in which the effect of interstitial fluid and non-contact forces, such as the van der Waals and electrostatic forces, can be ignored. Therefore, the governing equations for translational and rotational motion of particle i can be given by

$$m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = \sum_j \mathbf{f}_{ij} + m_i \mathbf{g}, \tag{1}$$

$$I_i \frac{\mathrm{d}\omega_i}{\mathrm{d}t} = \sum_j \mathbf{m}_{ij}, \tag{2}$$

where $\mathbf{v}_i$, ω_i are, respectively, the translational and angular velocities of particle i with mass m_i and moment of inertia I_i. The forces involved are the gravitational force, $m_i\mathbf{g}$, and interaction force $\mathbf{f}_{ij}$ between particles due to the plastic and elastic deformation resulting from particle collision, sliding and rolling. Further, $\mathbf{m}_{ij}$ is the torque acting on particle i by particle j arising from their interaction.

In general, the total interaction force between particles i and j, $\mathbf{f}_{ij}$, can be expressed as a normal component, $\mathbf{f}_{ij}^{\mathrm{n}}$, and a tangential one, $\mathbf{f}_{ij}^{\mathrm{t}}$. Various approaches have been proposed to model $\mathbf{f}_{ij}^{\mathrm{n}}$ and $\mathbf{f}_{ij}^{\mathrm{t}}$. One of the most popular force models was developed based on the consideration of contact elastic force and viscous contact damping force [4,7–9]. The normal force includes elastic and damping components, whilst the tangential force includes frictional and damping components. The normal elastic force, $\mathbf{f}_{ij}^{\mathrm{ne}}$, described by the Hertz theory [18, pp. 84–106], can be given by

$$\mathbf{f}_{ij}^{\mathrm{ne}} = \frac{4}{3} E_i^* \sqrt{R_{ij}^*} \left(\delta_{ij}^n\right)^{3/2} \mathbf{n}_{ij}, \tag{3}$$

where E_i^* and R_{ij}^* are the reduced modulus of elasticity and particle radius, δ_{ij}^{n} is the relative normal displacement at contact, and $\mathbf{n}_{ij}$ is the unit contact vector. E_i^*, R_{ij}^*, δ_{ij}^{n} and $\mathbf{n}_{ij}$, are, respectively, given by $E_i^* = E_i/\left[2(1-\nu_i^2)\right]$, $R_{ij}^* = \left(R_i R_j\right)/\left(R_i + R_j\right)$, $\delta_{ij}^n = R_i + R_j - \left|\mathbf{r}_j - \mathbf{r}_i\right|$, and $\mathbf{n}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/\left|\mathbf{r}_i - \mathbf{r}_j\right|$, where E_i and ν_i are the modulus of elasticity and Poisson's ratio of particle i; R_i and R_j are the radii of particles i and j, and $\mathbf{r}_i$ and $\mathbf{r}_j$ are the position vectors of particles i and j. The normal damping force, $\mathbf{f}_{ij}^{\mathrm{nd}}$, which is modeled as a dashpot that dissipates a proportion of the relative kinetic energy, is given by

$$\mathbf{f}_{ij}^{\mathrm{nd}} = -c_{ij}^n \left(8 m_{ij} E_i^* \sqrt{R_{ij}^* \delta_{ij}^n}\right)^{1/2} \left(\mathbf{v}_{ij} \cdot \mathbf{n}_{ij}\right) \mathbf{n}_{ij}, \tag{4}$$

where c_{ij}^n is the normal damping coefficient, $\mathbf{v}_{ij}$ is the relative velocity of particles i and j at contact, given by $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j + \omega_i \times \mathbf{C}_{ij} - \omega_j \times \mathbf{C}_{ji}$, $\mathbf{v}_i$ and $\mathbf{v}_j$ are the velocities of particle i and j, and $\mathbf{C}_{ij}$ is a vector from the mass center of the particle to the contact point.

The frictional component of the tangential force is given by

$$\mathbf{f}_{ij}^{\mathrm{te}} = -\mu_{ij} \left|\mathbf{f}_{ij}^{\mathrm{ne}}\right| \left(1 - \left(1 - \min\left(\left|\mathbf{v}_{ij}^{\mathrm{t}}\right|, \delta_{ij}^{\max}\right)/\delta_{ij}^{\max}\right)^{3/2}\right) \hat{\mathbf{v}}_{ij}^{\mathrm{t}}, \tag{5}$$

where μ_{ij} is the sliding friction coefficient between particles i and j, $\hat{\mathbf{v}}_{ij}^{\mathrm{t}} = \mathbf{v}_{ij}^{\mathrm{t}}/|\mathbf{v}_{ij}^{\mathrm{t}}|$, $\mathbf{v}_{ij}^{\mathrm{t}}$ is the vector of the relative tangential displacement of particles i and j, and $\delta_{ij}^{\max}$ is the relative tangential displacement when the sliding starts, given by $\delta_{ij}^{\max} = \mu_{ij}\delta_{ij}^n(2-\nu_i)/[2(1-\nu_i)]$. Equation (5) suggests that the friction is described by the theory proposed by Mindlin and Deresiewicz [19] prior to the sliding (*i.e.*, when $|\mathbf{v}_{ij}^{\mathrm{t}}| < \delta_{ij}^{\max}$) and the Coulomb friction model when the particles at contact start to slide relatively. The tangential damping force is given by

$$\mathbf{f}_{ij}^{\mathrm{td}} = 2c_{ij}^t \left(1{\cdot}5\mu_{ij} m_{ij} \left|\mathbf{f}_{ij}^{\mathrm{ne}}\right| \sqrt{1 - |\mathbf{v}_{ij}^{\mathrm{t}}|/\delta_{ij}^{\max}}/\delta_{ij}^{\max}\right)^{1/2} (\mathbf{v}_{ij} \times \mathbf{n}_{ij}) \times \mathbf{n}_{ij}, \tag{6}$$

where c_{ij}^t is the tangential damping coefficient. Note that (6) implies that, when sliding occurs, there is no contribution of the tangential damping to the tangential force.

In general, the contact between two spheres is not at a single point but is a finite area due to the deformation of both spheres. The inter-particle forces act over the contact region between particles i and j, rather than the mass center of the particle, and they will generate a torque, $\mathbf{m}_{ij}$. The torque $\mathbf{m}_{ij}$ causing particle i to rotate is contributed by the tangential and normal components of the traction distribution, $\mathbf{m}_{ij}^{\mathrm{t}}$ and $\mathbf{m}_{ij}^{\mathrm{n}}$, respectively. Implemented in a discrete simulation, $\mathbf{m}_{ij}^{\mathrm{t}}$ is usually expressed as

$$\mathbf{m}_{ij}^{\mathrm{t}} = \mathbf{C}_{ij} \times \mathbf{f}_{ij}^{\mathrm{t}}. \tag{7}$$

For many inelastic bodies such as viscoelastic particles, the normal traction distribution on the contact area between two contacting particles is asymmetrical due to the relative rotation of particles. The asymmetry leads to a non-zero torque, $\mathbf{m}_{ij}^{\mathrm{n}}$, given by

$$\mathbf{m}_{ij}^{\mathrm{n}} = -\min\left\{\mu_{r,ij}\left|\mathbf{f}_{ij}^{\mathrm{n}}\right|, \mu'_{r,ij}\left|\omega_{ij}^{\mathrm{n}}\right|\right\} \hat{\omega}_{ij}^{\mathrm{n}}, \tag{8}$$

where $\mu_{r,ij}$ is the rolling friction coefficient, $\mu'_{r,ij}$ is the rotational stiffness and $\hat{\omega}_{ij}^{\mathrm{n}} = \omega_{ij}^{\mathrm{n}}/|\omega_{ij}^{\mathrm{n}}|$, ω_{ij}^{n} is the component of the vector of the relative angular velocity of particles i and j in their contact plane. The vector $\mathbf{m}_{ij}^{\mathrm{n}}$ is often referred to as the rolling friction torque which provides a resistance to relative rolling motion between particles [20–22].

2.2. Averaging method

By use of a proper averaging procedure, the discrete system considered above can be transferred into a corresponding continuum system. Assume that the mass density ρ, the velocity $\mathbf{u}$, the angular velocity ω are respectively

$$\rho = \int_{T_t} \sum_i h_i m_i \mathrm{d}s, \tag{9}$$

$$\mathbf{u} = \frac{1}{\rho} \int_{T_t} \sum_i h_i m_i \mathbf{v}_i \mathrm{d}s, \tag{10}$$

$$\omega = \frac{1}{\lambda} \int_{T_t} \sum_i h_i I_i \omega_i \mathrm{d}s, \tag{11}$$

where $h_i = h(\mathbf{r}_i - \mathbf{r}, s - t)$, $h = h(\mathbf{r}, t)$ is the weighting function, which is positive in the limited domain $\Omega = \{(\mathbf{r},t) \,|\, \mathbf{r} \in \Omega_p \subset R^3, t \in T_t \subset R\}$ and zero otherwise. The variables $\mathbf{r}$ and t are, respectively, the position vector and time to which the average value of the considered quantities is assigned, and $\lambda = \int_{T_t} \sum_i h_i I_i \mathrm{d}s$. Differentiating ρ, $\rho\mathbf{u}$ and $\lambda\omega$ by use of Equations (1), (2), (9–11), we can obtain the balance equations of mass and linear momentum and angular momentum of the continuum system, respectively given by [16,17]

$$\mathrm{D}(\rho) + \rho \nabla \cdot \mathbf{u} = 0, \tag{12}$$

$$\mathrm{D}(\rho\mathbf{u}) + \rho\mathbf{u}\nabla \cdot \mathbf{u} = \nabla \cdot \mathbf{T} + \rho\mathbf{g}, \tag{13}$$

$$\mathrm{D}(\lambda\omega) + \lambda\omega\nabla \cdot \mathbf{u} = \nabla \cdot \mathbf{M} + \mathbf{M}', \tag{14}$$

where $\mathrm{D}(\cdot)$ represents the material derivative of a tensor. Further, $\mathbf{T}$ is the stress tensor, $\mathbf{M}$ is the couple-stress tensor, and $\mathbf{M}'$ is the rate of supply of internal spin to particles, respectively given by

$$\mathbf{T} = -\int_{T_t} \sum_i h_i m_i \mathbf{v}_i' \otimes \mathbf{v}_i' \mathrm{d}s + \int_{T_t} \sum_i \sum_{j>i} g_{ij} \mathbf{d}_{ij} \otimes \mathbf{f}_{ij} \mathrm{d}s, \tag{15}$$

$$\mathbf{M} = -\int_{T_t} \sum_i h_i I_i \mathbf{v}_i' \otimes \omega_i \mathrm{d}s + \frac{1}{2} \int_{T_t} \sum_i \sum_{j>i} g_{ij} \mathbf{d}_{ij} \otimes (\mathbf{m}_{ij} - \mathbf{m}_{ji}) \mathrm{d}s, \tag{16}$$

$$\mathbf{M}' = \frac{1}{2} \int_{T_t} \sum_i \sum_{j>i} (\mathbf{m}_{ij} + \mathbf{m}_{ji})(h_i + h_j) \mathrm{d}s \tag{17}$$

where $\mathbf{v}_i' = \mathbf{v}_i - \mathbf{u}$ is the fluctuant velocity of particle i with respect to the average velocity, $\mathbf{d}_{ij}$ is the part of the branch vector connecting the mass centers of particles i and j within domain Ω_p; g_{ij} is the weighting coefficient determined by the weighting function $h(\mathbf{r}, t)$, given by $g_{ij} = \int_0^1 h(\bar{\mathbf{r}}_i + r\mathbf{d}_{ij} - \mathbf{r}, s - t)\mathrm{d}r$, $\bar{\mathbf{r}}_i = \mathbf{r}_i$ if $\mathbf{r}_i - \mathbf{r} \in \Omega_p$; otherwise, $\bar{\mathbf{r}}_i$ is the position vector of the point of intersection of vector $\mathbf{d}_{ij}$ and boundary $\partial\Omega_p$ of domain Ω_p.

A particle may collide with a wall. Therefore, the interaction between particles and walls should be included in the continuum description. The contribution of the interaction between particles and walls to stress and couple stress can be given by

$$\mathbf{T}^b = \int_{T_t} \sum_i g_i^b \mathbf{d}_i^b \otimes \mathbf{f}_i^b \mathrm{d}s, \tag{18}$$

and

$$\mathbf{M}^b = \int_{T_t} \sum_i g_i^b \mathbf{d}_i^b \otimes \mathbf{m}_i^b \mathrm{d}s, \tag{19}$$

where $\mathbf{d}_i^b$ is the ray from the center of mass of particle i to boundary $\partial\Omega_p$ of domain Ω_p, via a point on the wall and perpendicular to the tangential plane of the point; g_i^b is the weighting coefficient corresponding to walls, given by $g_i^b = \int_0^1 g(\mathbf{r}_i + r\mathbf{d}_i^b - \mathbf{r}, s-t)\mathrm{d}r$. Equations (9–11), (15–19) give the relationship between microscopic and macroscopic quantities of granular materials.

The microscopic dynamic behavior of granular flows is often described in Cartesian coordinates in DEM simulation. However, for granular flows with a geometry such as a cylindrical hopper, it is more convenient to describe their macroscopic dynamic behavior in orthogonal curvilinear coordinates. Therefore, it is necessary to determine the transformation between the variables under the two coordinates. For this reason, let $\{x^i\}$ be the Cartesian coordinates of the three-dimensional Euclidian space R^3, $\{y^i\}$ ($y^i = y^i(x^1, x^2, x^3)$, $i = 1, 2, 3$) be the orthogonal curvilinear coordinates in this space, u, ω, T and M be the coordinates of the macroscopic quantities $\mathbf{u}$, ω, $\mathbf{T}$ and $\mathbf{M}$ of granular matter under $\{x^i\}$, and $\bar{u}$, $\bar{\omega}$, $\bar{T}$ and $\bar{M}$ be the coordinates under $\{y^i\}$. It can be readily obtained that

$$\bar{u} = uA, \bar{\omega} = \omega A, \bar{T} = A^{\mathrm{T}} T A, \bar{M} = A^{\mathrm{T}} M A. \tag{20}$$

where $A = \left(\partial y^i / \partial x^j\right) \bar{G}$, $\bar{G}$ is 3×3 matrix whose components are $\sqrt{g_{ij}}\,(i, j = 1, 2, 3)$, $g_{ii} = \mathbf{e}_i' \cdot \mathbf{e}_i' = (\partial x^1/\partial y^i)^2 + (\partial x^2/\partial y^i)^2 + (\partial x^3/\partial y^i)^2 (i = 1, 2, 3)$, and $g_{ij} = 0$ if $i \neq j$. Equation (20) gives the expressions of $\mathbf{u}$, ω, $\mathbf{T}$ and $\mathbf{M}$ under the orthogonal curvilinear coordinates $\{y^i\}$.

Cylindrical coordinates are special orthogonal curvilinear coordinates. Assuming that $\{r, \theta, z\}$ is the cylindrical coordinates, we observe that there exists a relationship between cylindrical and Cartesian coordinates, namely

$$x^1 = r\cos\theta, x^2 = r\sin\theta, x^3 = z. \tag{21}$$

It can be obtained that

$$A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{22}$$

Therefore, by use of (20) and (22), the expressions of $\mathbf{u}$, ω, $\mathbf{T}$ and $\mathbf{M}$ under the cylindrical coordinates can be readily obtained. Note that the mass density is independent of the coordinate transformation, as it is a scalar.

3. Results and discussion

3.1. Simulation conditions and measurement strategy

To generate results that can be directly compared with those under steady-state flow conditions, the geometry of the hopper used in this work is similar to that in our previous work [9]. It is cylindrical in shape, of diameter of $20d$ (d is the maximum particle diameter) and with a circular orifice of diameter of $8d$ at the center of its flat bottom. Twenty-four thousand multisized spherical particles of uniform size distribution in a range of $0{\cdot}8 - 1{\cdot}0d$ are considered. Other physical parameters for particles and wall are as follows. Young's modulus and Poisson ratio of particles are set to $50000\pi\rho_p dg/6$ (ρ_p is the mass density of particle, and g is

the gravitational acceleration) and 0·3, respectively. The sliding and rolling friction coefficients between particles are set to 0·6 and $0{\cdot}001d$, respectively. The normal and tangential damping coefficients are set to 0·3. The wall is assumed to have the same properties as the particles, except the sliding friction coefficient, which is 0·5. The time step used is $0{\cdot}001\sqrt{d/g}$.

The microscopic properties of granular flows can be generated by means of the DEM simulation, whereas their macroscopic properties can be obtained by use of the micro-dynamic information generated and the average method described above. In this work, the microscopic quantities of the hopper flow, including velocity field and force structure, are analyzed in Cartesian coordinates $\{x, y, z\}$ fixed at the center of the orifice of the hopper with z-axis at the central axis of the hopper. The values of the macroscopic quantities, such as density, velocity, stress and couple stress at the probe point far from the bottom and vertical walls (larger than $4d$) in the coordinates, are calculated by use of Equations (9), (10), (15) and (16), whereas those at the probe point adjacent to the walls by use of Equations (18) and (19) in addition to these equations. Since the resulting macroscopic variables should be reasonably axially symmetric, consistent with the practical treatment [23, pp. 298–372], all computed macroscopic quantities are averaged from the values of two symmetrical points and shown as a function of radius r and height z in cylindrical coordinates. The cylindrical coordinates $\{r, \theta, z\}$ starts at the center of the orifice of the hopper. The velocity, v, mass density, ρ, three independent components of planer stress, T_{zz}, T_{rr} and T_{rz}, and two dominant components of couple stress, $M_{r\theta}$ and $M_{z\theta}$, have been investigated in this coordinate system. Their magnitudes are obtained by means of Equations (20) and (22).

3.2. Microscopic analysis

In the DEM simulation, the particles with random initial velocities are first allowed to gradually settle onto the hopper under gravity, giving a packing with height of about $47d$. These particles are then discharged when the hopper outlet is removed, as shown in Figure 1. Three distinct flow patterns are observed. In the upper part, particles flow downward layer by layer except those very near the vertical walls due to the sliding resistance by the walls. In the lower part, particles flow in a V-shape and are discharged except for a small number of particles staying at the bottom corner of the hopper.

The present microscopic analysis is focused on two key aspects: the velocity field and force structure. Since the trajectories of and contact forces acting on individual particles are traced in a DEM simulation, information for such analysis can be readily established. Figure 2

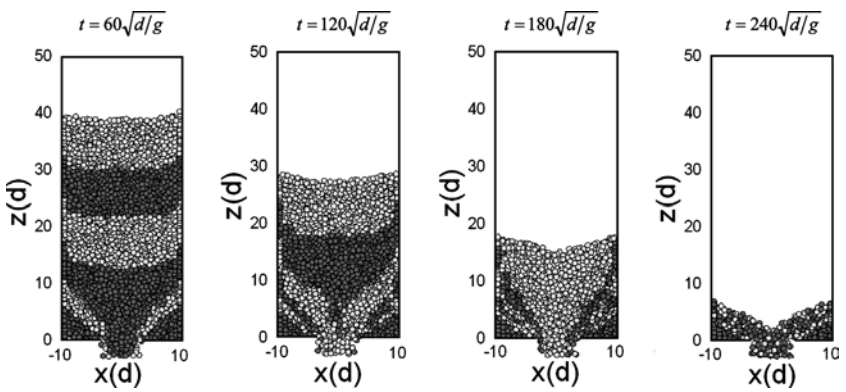


Figure 1. Discharging process of the hopper flow, represented by particles in the central vertical section at different times.

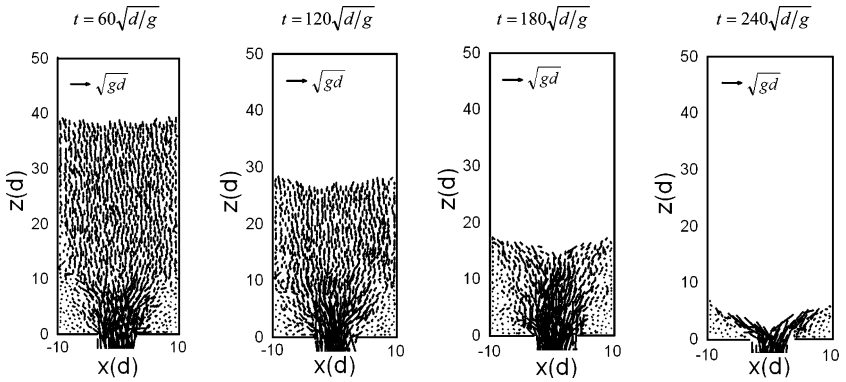


Figure 2. Velocity profiles at different times for particles in the central vertical section.

shows the velocity fields in a vertical mid-section of the hopper at $t = 60\sqrt{d/g}$ (time of discharge), $t = 120\sqrt{d/g}$, $t = 180\sqrt{d/g}$ and $t = 240\sqrt{d/g}$, where each vector represents the components of the velocity of a particle in a section as long as its center is located between $\pm 0{\cdot}5d$ of the position of the section. It can be seen that, in the upper part, the velocity in the vertical direction dominates and is almost uniform across the hopper, except for a narrow shear zone adjacent to the vertical wall. In the lower part, the particles have much larger velocities at the orifice and very low velocities around the bottom corner of the hopper. Large velocities in the radial direction concentrate at the two sides of the orifice. These results are qualitatively consistent with the previous experimental and numerical observations for such a hopper flow [23–25], [26, pp. 298–372]. That is, there are four different zones: a stagnant zone at the bottom corner of the hopper, a plug flow zone in the upper part, a converging-flow zone in the lower part, and a transition zone from plug flow to converging flow. Moreover, the plug-flow zone is observed to reduce with the discharge of particles, whereas the converging flow zone varies little before $t = 180\sqrt{d/g}$ and then reduces with the discharge of the particles in this zone. The flow regimes can be more clearly shown in terms of the distribution of the averaged velocity in the section on macroscopic analysis.

The force structure is another important local characteristic of granular matter. It has been used to depict the flow structure, force transmission and other dynamic behavior of particles [4, 5, 7, 9]. Figure 3 shows the force network in the vertical mid-section at different times, where the thickness of each stick connecting two particle centers is proportional to

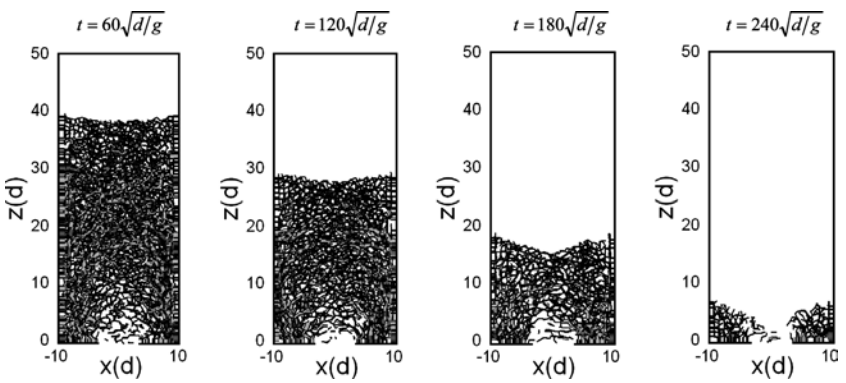


Figure 3. Force networks at different times for particles in the central vertical section.

the magnitude of the force. The normal contact force is probably the most important, as the contact tangential force is related to it and is relatively small. Therefore, only the normal force is included in this figure. Note that the present analysis is focused on the relative values of the contact forces for a given time; different scales have been used for different times for better illustration of the flow and force structures. It can be observed that, similar to the steady-state hopper flow [9], large forces are exerted on the particles at the bottom and vertical walls to support the particles in the top part. The forces gradually propagate from the walls into the particle bed. Relatively large forces are also observed in the middle height of the flow, which corresponds to the transition from plug flow to converging flow in the hopper flow. In the orifice region, the forces are very small and the force chains are broken. A force chain means a connection between particles which confines the relative movement between particles. The breakage of force chains may imply the relative movements between the particles. Therefore, the flow and force structures are related. These characteristics do not change with time.

3.3. Macroscopic analysis

The DEM results can be used to calculate the macroscopic variables such as mass density, velocity, stress and couple stress according to (9), (10), (15) and (16). However, strictly speaking, these macroscopic quantities depend on the weighting function $h(\mathbf{r}, s)$. Therefore, the selection of a proper weighting function is important in applying the equations formulated above. The need to find a suitable weighting function has been noticed; a weighting function has been recommended, and given by [17]

$$h(\mathbf{r}, s) = f(s)g(\bar{s}), \tag{23}$$

where $\bar{s} = |\mathbf{r}|$, and $f(s)$ is the simplified S_B distribution function [27], whilst $g(\bar{s})$ is an extension of the univariate distribution to three variants. They can be explicitly written as

$$f(s) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{2L_t}{(L_t^2 - s^2)} \exp\left(-\frac{1}{2}\log^2 \frac{L_t + s}{L_t - s}\right), & |s| < L_t \\ 0, & |s| \geq L_t \end{cases}, \tag{24}$$

$$g(\bar{s}) = \begin{cases} \frac{c}{4\pi L_p (L_p^2 - \bar{s}^2)} \exp\left(-\frac{1}{2}\log^2 \frac{L_p + \bar{s}}{L_p - \bar{s}}\right), & \bar{s} < L_p \\ 0, & \bar{s} \geq L_p \end{cases}, \tag{25}$$

where L_t and L_p are the distribution parameters, c is the normalized constant of the distribution function $g(\bar{s})$. The functions $f(s)$ and $g(\bar{s})$ are smooth in the entire space, and decrease monotonically with increasing s and $\bar{s}$, respectively.

Parameter L_p or L_t in the weighting function (23) determines the amount of the contributing particles to a probe point or time, and the magnitude of contribution of every particle as well, as illustrated in Figure 4. The specification of appropriate values of L_p and L_t is a significant part of the application of the averaging method described above. In fact, as shown in Figures 5 and 6, L_p and L_t do not affect the trend of the distribution of velocity and stress (only the vertical velocity and the trace of the stress tensor are shown, since the other components of velocity and stress lead to the same conclusion). However, if L_p and L_t are too small, the resulting macroscopic quantities fluctuate, giving unreasonable average results. On the other hand, if they are too large, the local properties of these quantities may be eliminated. The results in Figures 5 and 6 suggest that, to balance the friction, we can choose $L_p = 4{\cdot}0d$ and $L_t = 4{\cdot}0\sqrt{d/g}$. Such a selection can ensure that the averaged quantities smoothly vary in a considered domain and retain the local properties of these quantities as much as possible. The results below were all obtained with these two values.

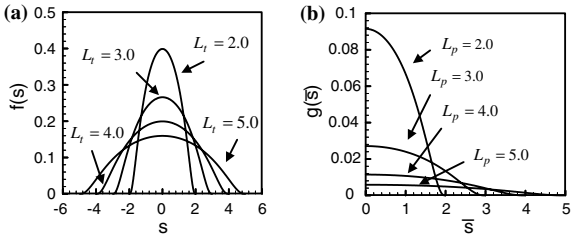


Figure 4. Variation of the weighting function with parameter L_p and L_t: (a) function $f(s)$ and (b) function $g(\bar{s})$.

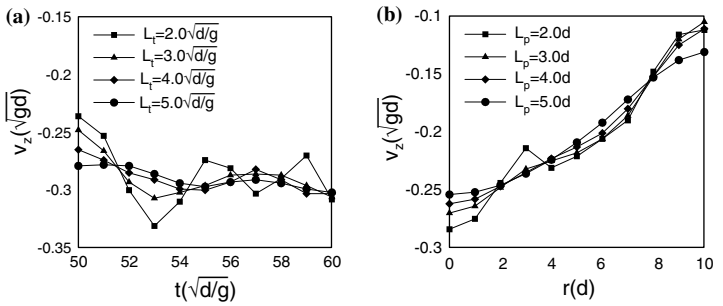


Figure 5. Effect of the parameters in the weighting function on vertical velocity v_z: (a) the variation of v_z with time t for different L_t when $L_p = 4{\cdot}0d$ for a probing point located at r=0 and $z = 12{\cdot}0d$ and (b) the variation of v_z with radius r for different L_p when $L_t = 4{\cdot}0\sqrt{d/g}$ and $t = 50\sqrt{d/g}$ at height $z = 12{\cdot}0d$.

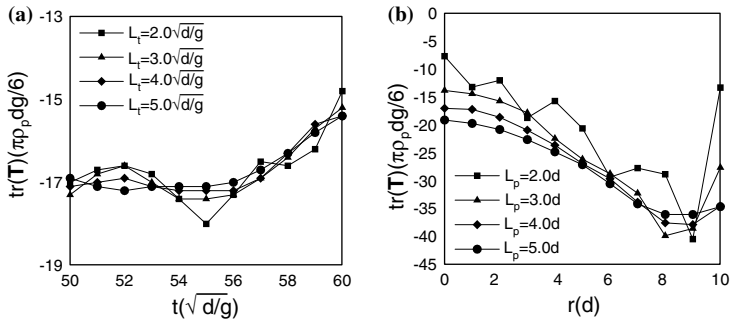


Figure 6. Effect of the parameters in the weighting function on the trace of stress tensor tr(**T**): (a) the variation of tr(**T**) with time t for different L_t when $L_p = 4{\cdot}0d$ for a probing point located at $r = 0$ and $z = 12{\cdot}0d$; and (b) the variation of tr(**T**) with radius r for different L_p when $L_t = 4{\cdot}0\sqrt{d/g}$ and $t = 50\sqrt{d/g}$ at height $z = 12{\cdot}0d$.

Figure 7 shows the mass density or bulk density profiles at $t = 60\sqrt{d/g}$, $t = 120\sqrt{d/g}$ and $t = 180\sqrt{d/g}$. It can be seen that the mass density is lowest at the orifice and increases away from the orifice in the lower part. The magnitude of mass density in this part does not change much with time. In the upper part, the mass density is almost constant in the central region and, because of the dilatancy that takes place during shearing, decreases with the distance to the axis of the hopper in the region close to the wall. The magnitude of mass density in the central upper part decreases with the discharge of particles. Figure 8 shows the velocity profiles of the hopper flow in the vertical mid-section of the hopper at different times. As expected, the distributions are in agreement with the micro-dynamic analysis as above.

Theoretically, the stress contains two parts: a kinetic contribution related to the transport of particles and a collisional contribution related to the interaction force between particles and between particles and walls. For hopper flow, the stress is attributed mainly to collisions

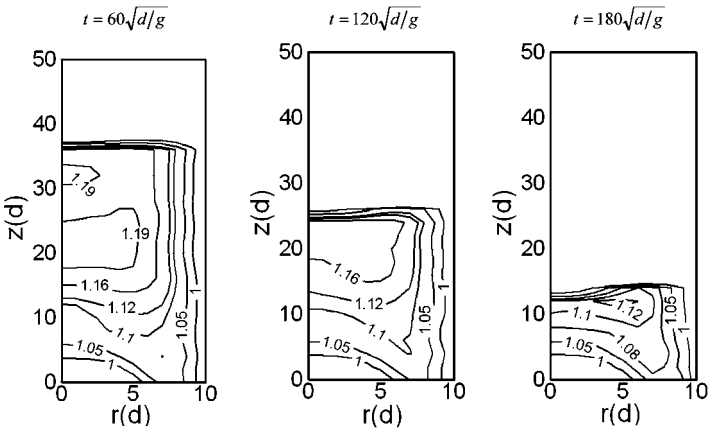


Figure 7. Contour plots of mass density of the unsteady-state hopper flow at different times (the units for mass density are $\pi\rho_p/6$).

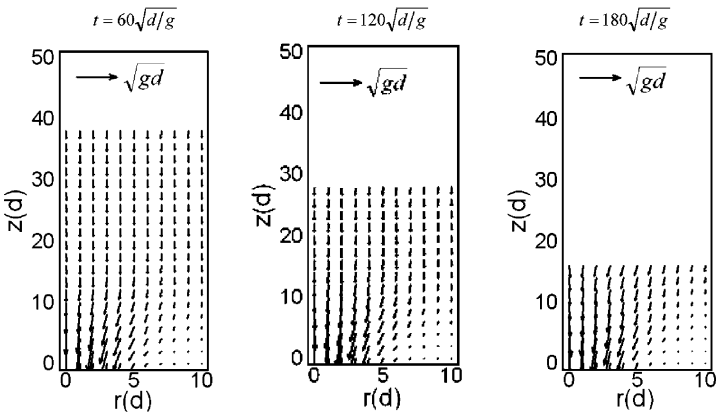


Figure 8. Velocity fields of the unsteady-state hopper flow at different times.

between the particles and between particles and walls. It is related to the magnitude and direction of the interaction forces. However, close to the orifice, the transport of particles may also play a role. To ensure the accuracy of the stress distribution, the two contributions are included in the present calculation. Figure 9 shows the radial distribution of the three components of the stress tensor, that is, two normal stress, T_{zz} and T_{rr}, and a shear stress, T_{rz}. Another shear stress, T_{zr}, has a similar distribution to T_{rz} in the whole domain except in a small region adjacent to the walls, which is not shown for brevity. The results indicate that the distributions of these stresses are similar to those of steady-state hopper flow [28], and have similar trends for different times. It can be observed from Figures 9 (a) and (b) that the two normal stresses are low in a region close to the orifice, and large in a region near the bottom corner. In the upper part, the two stresses vary little in the central section but vary somewhat near the wall. Furthermore, the two normal stresses in this upper part are almost the same in the central region, but not so in a region adjacent to the wall. The magnitude of the radial normal stress T_{rr} is large as well in the middle height in the vertical direction, which is, as expected, in accordance with the force network in Figure 3. As shown in Figure 9(c), the shear stress has its largest magnitude in a region close to the vertical wall and decreases away from this region. In the regions adjacent to the bottom wall and central axis of the hopper, the shear stress is very low. The magnitudes of the two normal stresses vary little with time in

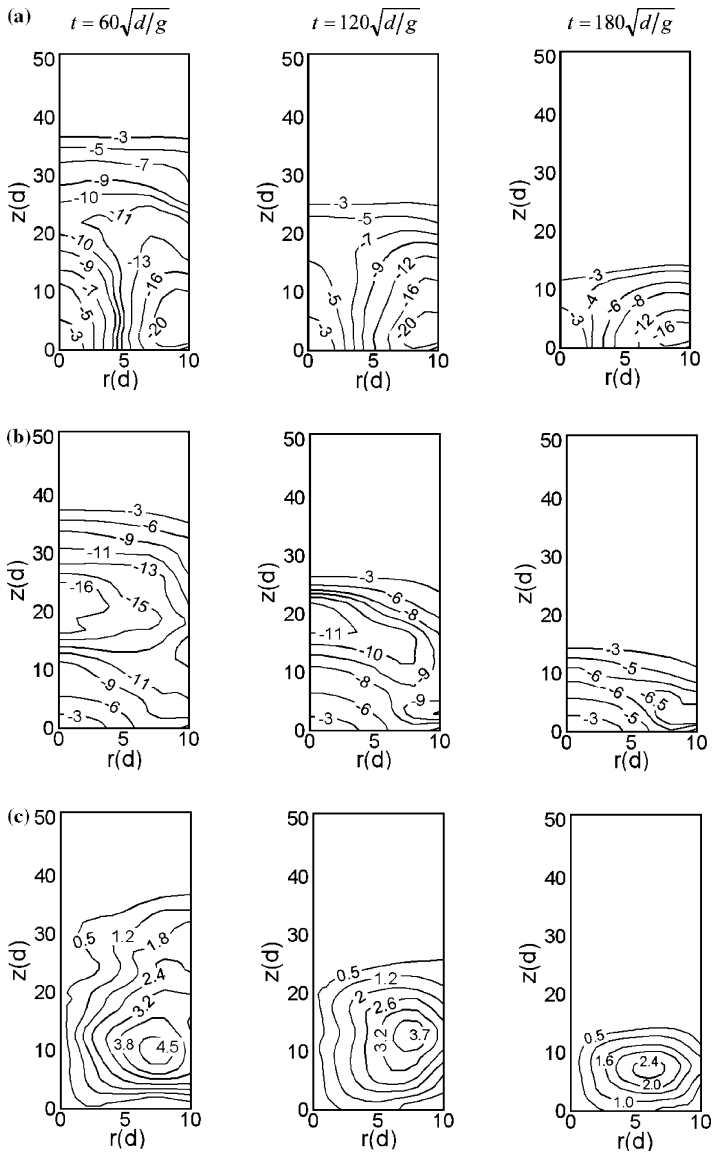


Figure 9. Contour plots of the distribution of stress of the unsteady state hopper flow at different times (the units for stress are $\pi\rho_p dg/6$): (a) T_{zz}; (b) T_{rr} and (c) T_{rz}.

the orifice region, but decrease in other regions. The magnitude of the shear stress decreases with time at locations far from the bottom wall and central axis of the hopper.

For steady-state hopper flow, the most dominant components in the couple stress are $M_{r\theta}$ adjacent to the vertical wall of the hopper and $M_{z\theta}$ close to the bottom wall [28]. This phenomenon has also been observed in the present simulation of unsteady-state hopper flow, as shown in Figure 10, where the distributions of the two dominant components at different times are given. In general, the couple stresses at points far from a wall result from the transport of particles and the collisions between particles. Close enough to the wall they will also be affected by the collisions between particles and wall. The observed sharp increase of the magnitudes of $M_{r\theta}$ and $M_{z\theta}$ in Figure 10 can be attributed to the effect of the vertical and bottom walls. The fluctuation of the couple stress far from the walls results mainly from

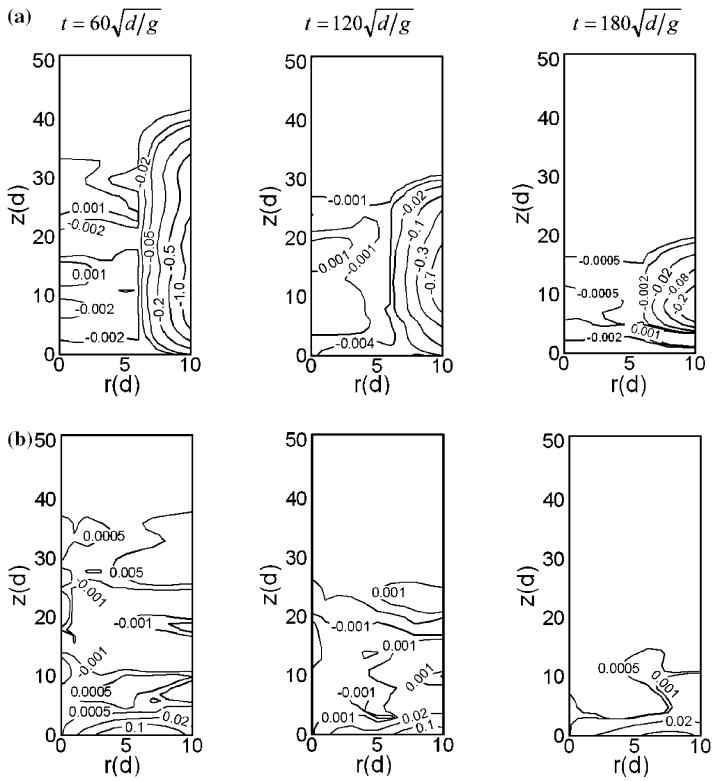


Figure 10. Contour plots of the distribution of couple stress of the unsteady-state hopper flow at different times (the units for couple stress are $\pi\rho_p d^2 g/6$): (a) $M_{r\theta}$ and (b) $M_{z\theta}$.

collisions between particles. The width or height of the region with a large magnitude of $M_{r\theta}$ or $M_{z\theta}$ is almost the same for different times, that is $4d$ for $M_{r\theta}$ and $2d$ for $M_{z\theta}$. These facts indicate that the couple stress related to the gradient of particle rotation should be considered when the effect of walls is included, which highlights the complexity of the macro-dynamic analysis of granular flow adjacent to physical boundaries.

4. Conclusions

A numerical investigation of the micro- and macro-dynamic behavior of the unsteady-state granular flow in a cylindrical hopper with flat bottom has been carried out by means of a modified DEM and an averaging method. It has been found that there are four distinct flow zones in the hopper: stagnant, plug flow, converging flow and transition zone from plug flow to converging flow. The plug flow zone reduces with time, while the converging flow zone varies little and then reduces with the discharge of the particles from this zone. Force arching in the flow is strong and related to the flow structure of particles. The trends of the distributions of the macroscopic properties, such as the velocity, mass density, stress and couple stress of the hopper flow, are similar to those of steady-state hopper flow. They do not vary much with time. However, the magnitudes of these properties in different regions vary at different rates. In particular, the magnitudes of the two normal stresses vary little with time in the orifice region, but decrease in other regions; the magnitude of the shear stress decreases with discharge of particles when far from the bottom wall and central axis of the hopper, and the width or height of the region with a large magnitude of $M_{r\theta}$ or $M_{z\theta}$ is almost the same for

different times. These results about macroscopic properties should be useful in formulating and testing continuum models of hopper flow.

Acknowledgements

The authors are grateful to the Australian Research Council for the financial support of this work.

References

1. P.G. de Gennes, Granular matter: a tentative view. *Rev. Mod. Phys.* 71 (1999) 374–382.
2. S.F. Edwards and D.V. Grinev, Transmission of stress in granular materials as a problem of statistical mechanics. *Physica A* 302 (2001) 162–186.
3. P.A. Cundall and O.D.L. Strack, A discrete numerical model for granular assemblies. *Geotechnique* 29 (1979) 47–65.
4. P.A. Langston, U. Tüzün and D.M. Heyes, Discrete element simulation of internal stress and flow fields in funnel flow hoppers. *Powder Technol*. 85 (1995) 153–169.
5. R. Gutfraind and O. Pouliquen, Study of the origin of shear zones in quasi-state vertical chute flows by using discrete particle simulations. *Mech. Mat.* 24 (1996) 273–285.
6. P.W. Cleary and M.L. Sawley, DEM modeling of industrial granular flows: 3D case studies and the effect of particle shape on hopper discharge. *Appl. Math. Model.* 26 (2002) 89–111.
7. K. Tanaka, M. Nishida, T. Kunimochi and T. Takagi, Discrete element simulation and experiment for dynamic response of two-dismensional granular matter to the impact of a spherical projectile. *Powder Technol.* 124 (2002) 160–173.
8. A.B. Yu, DEM-an effective method for particalate matter. In: *Discrete Element Methods: Numerical Modeling of Discontinua*. Virginia: Ameriacan Society of Civil Engineers (2002) pp. 17–22.
9. H.P. Zhu and A.B. Yu, Steady-state granular flow in a 3D cylindrical hopper with flat bottom using DEM simulation: microscopic analysis. *J. Phys. D, Appl. Phys.* 37 (2004) 1497–1508.
10. C.K.K. Lun, S.B. Savage, D.J. Jeffrey and N. Chepurniy, Kinetic theories for granular flow: inelastic particles in Couette flow and slightly inelastic particles in a general flowfield. *J. Fluid Mech.* 140 (1984) 223–256.
11. C.S. Campbell, Rapid granular flows. *Annu. Rev. Fluid Mech.* 22 (1990) 57–92.
12. R.M. Nedderman, *Statics and Kinematics of Granular Materials*. New York: Cambridge University Press (1992) 352 pp.
13. W. Powrie, *Soil Mechanics: Concepts and Applications*. London: E & FN Spon (1997) 420 pp.
14. O.R. Walton and R.L. Braun, Viscosity, granular-temperature, and stress calculations for shearing assemblies of inelastic, frictional disks. *J. Rheology* 30 (1986) 949–980.
15. Y. Zhang and C.S. Campbell, The interface between fluid-like and solid-like behaviour in two-dimensional granular flows. *J. Fluid Mech.* 237 (1992) 541–568.
16. M. Babic, Average balance equations for granular materials. *Int. J. Eng. Sci.* 35 (1997) 523–548.
17. H.P. Zhu and A.B. Yu, Averaging method of granular materials. *Phys. Rev. E* 66 (2002) 021302.
18. K.L. Johnson, *Contact Mechanics*. Cambridge, UK: Cambridge University Press (1985) 452 pp.
19. R.D. Mindlin and H. Deresiewicz, Elastic spheres in contact under varying oblique forces. *J. Appl. Mech.* 20 (1953) 327–344.
20. K. Iwashita and M. Oda, Rolling resistance at contacts in the simulation of shear band development by DEM. *J. Engrg. Mech., ASCE* 124 (1998) 285–292.
21. A. Tordesillas and S. Walsh, Incorporating rolling resistance and contact anisotropy in micromechanical models of granular media. *Powder Technol.* 124 (2002) 106–111.
22. H.P. Zhu and A.B. Yu, The effects of wall and rolling resistance on the couple stress of granular materials in vertical flow. *Physica A* 325 (2003) 347–360.
23. J.P.K. Seville, U. Tüzün and R. Clift, *Processing of Particulate Solids,* Landon: Blackie Academic and Professional (1997) 372 pp.
24. G.P. Deutsch and L.C. Schmidt, Pressures on silo walls. *J. Engrg. Industry, ASME, Series B* 91 (1969) 450–459.
25. U. Tüzün, G.T. Houlsby, R.M. Nedderman and S.B. Savage, The flow of granular-materials - 2. velocity distributions in slow flow. *Chem. Eng. Sci.* 37 (1982) 1691–1709.

26. H. Sakaguchi and F. Ozaki, Analysis of the formation of arches plugging the flow of granular Materials. In: C. Thornton (ed.), *Powders and Grains*. Rotterdam: A.A. Balkema (1993) 351–355.
27. N.L. Johnson, Systems of frequency curves generated by methods of translation, *Biometrika* 36 (1949) 149–176.
28. H.P. Zhu and A.B. Yu, Steady-state granular flow in a 3D cylindrical hopper with flat bottom using DEM simulation: macroscopic analysis. (2004) (submitted).