Accelerated granular matter simulation

Da Wang

PhD thesis, 2015
Department of Physics
Umeå University
SE-901 87 Umeå
Sweden

Department of Physics
Umeå University
SE-901 87 Umeå
Sweden

This work has been generously supported by Algoryx Simulation, LKAB (dnr 223-2442-09), Umeå University and VINNOVA (2014-01901).

ISBN: 978-91-7601-366-3

Cover picture: Da Wang

Electronic version available at http://umu.diva-portal.org/
Printed by Print & Media, Umeå University
Umeå, Sweden 2015

Abstract

Modeling and simulation of granular matter has important applications in both natural science and industry. One widely used method is the discrete element method (DEM). It can be used for simulating granular matter in the gaseous, liquid as well as solid regime whereas alternative methods are in general applicable to only one. Discrete element analysis of large systems is, however, limited by long computational time. A number of solutions to radically improve the computational efficiency of DEM simulations are developed and analysed. These include treating the material as a nonsmooth dynamical system and methods for reducing the computational effort for solving the complementarity problem that arise from implicit treatment of the contact laws. This allow for large time-step integration and ultimately more and faster simulation studies or analysis of more complex systems. Acceleration methods that can reduce the computational complexity and degrees of freedom have been invented. These solutions are investigated in numerical experiments, validated using experimental data and applied for design exploration of iron ore pelletising systems.

Sammanfattning

Modellering och simulering av granulära material har viktiga tillämpningar inom såväl naturvetenskap som i industrin. En vanlig metod är den så kallade diskreta elementmetoden (DEM). Den kan användas för simulering av granulära material i alla dess olika faser - gas, flytande och fast form - medan alternativa metoder i allmänt endast stöder en av faserna. Diskret elementanalys av stora system är dessvärre begränsat av långa beräkningstider. Ett antal lösningar för att dramatiskt öka beräkningseffektiviteten i DEM-simuleringar har utvecklas och analyserats. Dessa inkluderar att beskriva materialet som ett system med icke-slät dynamik och metoder för att öka effektiviteten i beräkningarna av de komplementaritetsproblem som en implicit hantering kontaktlagarna medför. Detta möjliggör tidsintegration med stora tidssteg vilket kan omsättas i fler och snabbare simuleringsstudier eller analys av mer komplexa system. Lösningarna undersöks i numeriska experiment, valideras mot experimentella mätningar och tillämpas för design-utforskning av system inom järnmalmspelletisering.

Acknowledgements

Firstly, I would like to express my gratitude and admiration to my supervisor *Martin Servin* for his continuous guidiance and support during my PhD studies. I have learned a lot from him about the process of research, writing articles, and more important, the way of balance work and life. I express my gratitude also to people at Algoryx Simulation. Their great work give me possibility to reach fast and reliable simulations. I especially want to thank: *Nils Hjelte* - I don't know how many questions I sent to him; *Tomas Berglund* - realising and optimising many algorithms I have contributed to and turning them into a useful tool; *Michael Brandl*, *Niklas Melin*, *Anders Backman* and many others that helped me in programming and debugging my prototypes. Thanks to *LKAB* for giving funding, challenge and inspiration to the project and for providing measurements for parametrisation and validation. I am happy for the collaboration with my amiable and knowledgeable industry mentor, *Kjell-Ove Mickelsson*, and for all his good advice and many ideas. I have spent five years working in *UMIT Research Lab.* In the time I have received help and advice from *Claude Lacoursière*, *John Nordberg*, *Mattias Linde*, *Eddie Wadbro*, and *Juan Carlos Araujo-Cabarcas* in many ways. Finally, I would like to apologize to my parents and my wife *Feifei Ding* for my absence in these years.

Preface

This thesis consists of an introduction and the following papers:

I Martin Servin, Da Wang, Claude Lacoursiére, Kenneth Bodin, *Examining the smooth and nonsmooth discrete element approaches to granular matter*, Journal for Numerical Methods in Engineering, vol. 97, no. 12, 878-902, 2014.

II Da Wang, Martin Servin, and Kjell-Ove Mickelsson, *Outlet design optimization based on large-scale nonsmooth DEM simulation*, Powder Technology, vol. 253, pp. 438 - 443, 2014.

III Da Wang, Martin Servin, Tomas Berglund, Kjell-Ove Mickelsson, and Stefan Rönnbäck, *Parametrization and validation of a nonsmooth discrete element method for simulating flows of iron ore green pellets*, Powder Technology, Powder Technology vol. 283, 475-487, 2015.

IV Da Wang, Martin Servin, and Tomas Berglund, *Warm starting the projected Gauss-Seidel algorithm for granular matter simulation*, submitted 2015.

V Martin Servin, and Da Wang, *Adaptive model reduction for nonsmooth discrete element simulation*, submitted 2015.

1 Background

Granular matter is collection of discrete macroscopic particles that interact locally by interfacial contact forces that are highly dissipative [1]. Examples are natural grains of corn and minerals and manufactured pharmaceuticals pills and pellets. It has been estimated that more than 50 % of the sales in the world deal with products that involve handling of granular materials at some stage. In most large-scale systems, experiments and measurements are impossible for practical and economical reasons. Modeling and computer simulation is indispensable for deeper understanding of the nature of granular materials and for making radical improvements and innovating entirely new solutions in processing and transportation systems. The nature of granular materials is very complex and rich in phenomena despite the simplicity of the models at grain level. The materials are strongly dissipative and meta-stable with critical phenomena like jamming transitions and avalanches. There is a strong relation between the macroscopic bulk properties and the microscopic properties, such as grain size, shape, friction and elasticity. Granular materials can switch quickly between solid, liquid and gas state [2]. Especially characteristic is the occurrence of strong force chains - that reach through the material and cause arching phenomena that give structural strength. This is fundamental for the design of storage silos, and make grains in an hourglass either jam or flow at steady rate. Another fascinating phenomena is size segregation. When granular materials are excited by periodic perturbation, such as vibration, larger grains move upwards and smaller ones downwards despite having the same mass density. Popularly this is known as the *Brazil nut effect* and explains why the largest nuts are found on the top in a container of mixed nuts. This can be understood as a convection phenomena and make it extremely difficult to produce homogeneous mixtures of granular materials. For many phenomena, it is necessary to explicitly model the individual particles and their interactions. This is most challenging task considering that one cubic meter of sand consists of roughly one billion grains. Each sample can have big variations in the microscopic properties, it can be loosely or densely packed and be influenced in numerous different ways: shaken, rotated, compressed, sheared, suspended in fluid etc. This renders a tremendous large space of possible behaviour to investigate. The goal of this thesis is to provide new knowledge and methods to accelerate the modeling and exploration of complex granular systems using computer simulation.

2 Computational Granular Dynamics

There is a multitude of methods for simulating granular materials [4]. One widely used method is the discrete element method (DEM). It can be used for simulating granular matter in the gaseous, liquid as well as solid regime whereas alternative methods are in general applicable to only one regime. With DEM, and other particle based methods, the characteristic dynamics of granular materials emerges naturally

through the collective behaviour of the interacting particles and with direct relation to the microscopic properties. Resolving the systems at grain level come with considerable computational cost, however. At large enough scales it becomes intractable to use DEM. In the time of writing this thesis, the largest reported number of particles in a scaling experiment with a nonsmooth discrete element method is 2.8×10^{10} non-spherical particles and 1.1×10^{10} contacts using peta-scale supercomputers with almost 500 thousands processor cores [3]. To increase the size further, reduce computational time or the need for supercomputers, a more coarse model must be used. Essentially there are two alternatives to DEM: cellular automata models and continuum mechanics. Cellular automata models make use of a regular lattice where each grid cell attain a certain state that represent the local state of the granular media. The system evolves in time by a set of rules that determine the new state of each cell in terms of the current state of the cell and the states of the neighbouring cells. Cellular automata models have been applied mostly to granular gases, described by kinetic theory and with the automata rules derived from statistical collision models. This has for instance helped explaining transport and pattern formation of sand dunes [5]. On large enough length scales it is reasonable to treat the granular material using continuum mechanics, either as a solid, liquid or a multiphase model [1]. Simulations based on these models involve solving partial differential equations discretised using finite elements, finite differences or meshfree methods. Each discrete unit typically represent many thousand or millions of particles. For natural reasons the continuum models fail at phenomena that occur on the length scales approaching the particle diameter. The main challenge is to find constitutive laws that relate the stress and strain fields that are of general validity. Predicting and handling the transitions between elastic, plastic and viscous behaviour is particularly difficult. Discrete element methods play a central role in this development as detailed analysis can be made of the relationships between stress, strain and microscopic particle properties.

2.1 Discrete Element Method (DEM)

The discrete element method was developed in the late 1960s by Cundall and Stracks [6] as an extension of molecular dynamics to model macroscopic slightly deformable solid grains. Each particle is modeled as a rigid body. The bodies interact via viscoelastic contact forces obeying the Coulomb friction law. For a rigid body, a, the notations $\vec{\boldsymbol{x}}_{[a]}, \vec{\boldsymbol{v}}_{[a]}, \vec{\boldsymbol{f}}_{[a]}$ and $m_{[a]}$ are used for position, velocity, force and mass, and $\vec{\boldsymbol{e}}_{[a]}, \vec{\boldsymbol{\omega}}_{[a]}, \vec{\boldsymbol{\tau}}_{[a]}$ and $\boldsymbol{I}_{[a]}$, for orientation, angular velocity, torque and inertia tensor. These are combined into generalised position, velocity, force and mass, denoted $\mathbf{x}_{[a]}, \mathbf{v}_{[a]}, \mathbf{f}_{[a]}$ and $\mathbf{M}_{[a]}$, with $\mathbf{v}_{[a]} = (\vec{\boldsymbol{v}}_{[a]}^{\mathrm{T}}, \vec{\boldsymbol{\omega}}_{[a]}^{\mathrm{T}})^{\mathrm{T}}$ etc. and $\mathbf{M}_{[a]} = \mathrm{diag}(m_{[a]}\mathbf{1}_{3\times 3}, \boldsymbol{I}_{[a]})$. These are in turn collected in full system state vectors $\mathbf{x} = (\mathbf{x}_{[1]}, \mathbf{x}_{[2]}, \ldots)$, $\mathbf{v} = (\mathbf{v}_{[1]}, \mathbf{v}_{[2]}, \ldots)$ etc. Contact forces and velocities may be decomposed in the directions of contact normals, $\vec{\mathbf{n}}$ and tangents, $\vec{\mathbf{t}}$, see Fig.1. The gap function $g(\mathbf{x})$ measures the magnitude of overlap between the contacting bodies. The equations of motion follows

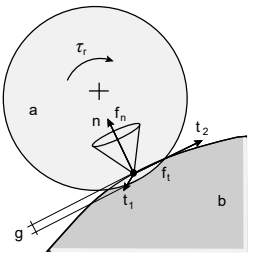

Figure 1: Illustration of the contact between two particles a and b.

Newton's law

$$\begin{aligned} \dot{\mathbf{x}}_{[a]} &= \mathbf{v}_{[a]} \\ \dot{\mathbf{v}}_{[a]} &= \mathbf{M}_{[a]}^{-1}\mathbf{f}_{[a]}(\mathbf{x},\mathbf{v}) \end{aligned} \tag{1}$$

where the force on particle a is the sum of all contact forces acting on it plus external forces, $\mathbf{f}_{[a]} = \mathbf{f}_{\text{ext}} + \sum_{b\in\mathcal{N}_c(a)} \mathbf{f}_{[ab]}$. A common force model is the nonlinear Hertz model that follows from the theory of linear viscoelastic materials. Friction is usually modeled as a spring in the tangential direction. The tangential spring extension is computed by integrating the slide velocity, and the force is projected onto the friction cone to obey the Coulomb law. Similarly, rolling resistance is modeled as torque that counteract relative rolling motion, also limited by a Coulomb-like law. For spherical particles, the contact forces are:

$$\begin{aligned} \vec{\mathbf{f}}_{\text{n}} &= k_{\text{n}}\left(g^{3/2} + cg^{1/2}\dot{g}\right)\vec{\mathbf{n}} \\ \vec{\mathbf{f}}_{\text{t}} &= \text{proj}_{\mu|\vec{\mathbf{f}}_{\text{n}}|}\left(-\int k_{\text{t}}\vec{\mathbf{u}}_{\text{t}}\text{d}t\right) \\ \vec{\tau}_{\text{r}} &= \text{proj}_{\mu_r|\vec{\mathbf{f}}_{\text{n}}|}\left(-\int k_{\text{r}}\vec{\mathbf{w}}_{\text{t}}\text{d}t\right) \end{aligned} \tag{2}$$

where $\vec{\mathbf{u}}_{\text{t}}$ and $\vec{\mathbf{w}}_{\text{t}}$ are the tangential relative velocity and relative angular velocity at the contact point. The normal stiffness and damping coefficients are $k_{\text{n}} = \text{E}\sqrt{2d}/3(1-\nu^2)$, where E is the Young's modulus, ν is the Poisson ratio, $d = (d_{[a]}^{-1} + d_{[b]}^{-1})^{-1}$ is the effective diameter, the dissipation coefficient is $c = 4(1-\nu^2)(1-2\nu)\eta/15\text{E}\nu^2$, and the material viscosity constant is η. For the friction spring coefficient, k_{t}, and rolling resistance coefficient, k_{r}, there are no such relations to fundamental material parameters and they must be determined from experiments. A DEM simulation consists of numerical integration of Eq. (1), which are ordinary differential equations (ODE). However, due to the occurrence of contact events and of the friction law in (2), the force is in general not differentiable in time. High-order integration algorithms are therefore not applicable [7]. Instead, the semi-implicit Euler or Verlet algorithm is a common choice, although the ODEs are very stiff and the time-step Δt become limited by the time-scale given by the elastic interaction time $\sqrt{m/k_{\text{n}}}$. The simulation algorithm of DEM involve contact detection, force computation and velocity and position update, $(\mathbf{x}_i, \mathbf{v}_i) \to (\mathbf{x}_{i+1}, \mathbf{v}_{i+1})$ from time t_i to $t_{i+1} = t_i + \Delta t$.

2.2 Nonsmooth Discrete Element Method (NDEM)

The nonsmooth contact dynamics method was introduced in the late 1980s by J. J. Moreau [8,9] and further developed by Jean [10] and others. In the case of rigid bodies this is also known as the nonsmooth discrete element method (NDEM) [7] and can be understood as an implicit version of DEM using convex analysis to remedy the fact that the contact laws cannot be written as simple mappings of position and velocity to force. In the nonsmooth DEM, impacts and frictional stick-slip transitions are considered as instantaneous events making the velocities discontinuous in time. The contact forces and impulses are modeled in terms of kinematic constraints and complementarity conditions between force and velocity, e.g., by the Signorini-Coulomb law for unilateral non-penetration and dry friction. The contact network becomes strongly coupled and any dynamic event may propagate through the entire system instantaneously. The benefit of nonsmooth DEM is that it allows integration with much larger simulation step-size than for smooth DEM and thus potentially faster. There are several equivalent formulations of NDEM. The formulation in terms of multibody dynamics in descriptor form will be used here [12] and with constraint stabilization [13]. The explicit contact force in Eq. (1) is replaced by a constraint forces $\mathbf{G}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{c}}$ with Lagrange multiplier $\boldsymbol{\lambda}_{\mathrm{c}}$ and Jacobian matrix $\mathbf{G}_{\mathrm{c}}$. The constraint force is implicitly given by the contact laws expressed as kinematic constraints in form of algebraic equations, inequalities or complementarity conditions. One advantage of this particular formulation is that it is automatically unified with the framework of multibody system dynamics for articulated mechanisms and power transmission systems. The constrained equations of motion can be written

$$\mathbf{M}\dot{\mathbf{v}} = \mathbf{f}_{\mathrm{ext}} + \mathbf{G}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{c}} + \mathbf{G}_{\mathrm{j}}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{j}} \tag{3}$$

$$\varepsilon_{\mathrm{j}}\boldsymbol{\lambda}_{\mathrm{j}} + \eta_{\mathrm{j}}\mathbf{g}_{\mathrm{j}} + \tau_{\mathrm{j}}\mathbf{G}_{\mathrm{j}}\mathbf{v} = 0 \tag{4}$$

$$\mathrm{law}[\mathbf{v}, \boldsymbol{\lambda}_{\mathrm{c}}] = \mathrm{true} \tag{5}$$

where (4) is a generic constraint equation that may model kinematic joints and motors. With $\varepsilon_{\mathrm{j}}, \tau_{\mathrm{j}} = 0$ and $\eta_{\mathrm{j}} = 1$, it becomes an ideal holonomic constraint $\mathbf{g}(\mathbf{x}) = 0$. For $\varepsilon_{\mathrm{j}}, \eta_{\mathrm{j}} = 0$ and $\tau_{\mathrm{j}} = 1$, the constraint become an ideal Pfaffian constraint $\mathbf{G}\mathbf{v} = 0$. With $\varepsilon_{\mathrm{j}}, \eta_{\mathrm{j}}, \tau_{\mathrm{j}} \neq 0$ it represent a generic constraint with compliance and damping. Eq. (5) represent the contact laws that are imposed on the system, for instance the Signorini-Coulomb and rolling resistance law

$$0 \leq \varepsilon_{\mathrm{n}}\boldsymbol{\lambda}_{\mathrm{n}} + \mathbf{g}_{\mathrm{n}} + \tau_{\mathrm{n}}\mathbf{G}_{\mathrm{n}}\mathbf{v} \perp \boldsymbol{\lambda}_{\mathrm{n}} \geq 0 \tag{6}$$

$$\gamma_{\mathrm{t}}\boldsymbol{\lambda}_{\mathrm{t}} + \mathbf{G}_{\mathrm{t}}\mathbf{v} = 0\ , \quad |\boldsymbol{\lambda}_{\mathrm{t}}| \leq \mu_{\mathrm{t}}|\mathbf{G}_{\mathrm{n}}^{\mathrm{T}}\lambda_{\mathrm{n}}| \tag{7}$$

$$\gamma_{\mathrm{r}}\boldsymbol{\lambda}_{\mathrm{r}} + \mathbf{G}_{\mathrm{r}}\mathbf{v} = 0\ , \quad |\boldsymbol{\lambda}_{\mathrm{r}}| \leq r\mu_{\mathrm{r}}|\mathbf{G}_{\mathrm{n}}^{\mathrm{T}}\lambda_{\mathrm{n}}| \tag{8}$$

where the contact constraint (c) have been decomposed into normal (n), tangential friction (t) and rolling constraint (r). Eq. (6)-(7) are the Signorini-Coulomb conditions with constraint regularization and stabilization terms ε_{n}, τ_{n} and γ_{t}. With $\varepsilon_{\mathrm{n}} = 0$ and $\tau_{\mathrm{n}} = 0$, Eq. (6) state that bodies should be separated or have zero overlap, $\mathbf{g}_{\mathrm{n}}(\mathbf{x}) \geq 0$,

and if so the normal force should be non-cohesive, $\boldsymbol{\lambda}_\mathrm{n} \geq 0$. With $\gamma_\mathrm{t} = 0$, Eq. (7) state that contacts should have zero relative slide velocity, $\mathbf{G}_\mathrm{t}\mathbf{v} = 0$, provided that the friction force remain bounded by the Coulomb friction law with friction coefficient μ_t. Eq. (8) similarly constrains relative rotational motion of contacting bodies, provided the constraint torque do not exceed the rolling resistance law with coefficient μ_r and contact radius r. The Lagrange multiplier become an auxiliary variable to solve for, in addition to position and velocity. The regularization and stabilization terms, ε and γ, introduce compliance and dissipation when constraints are violated. Regularized constraints may be viewed as Legendre transforms of a potential and Rayleigh dissipation function of the form $U_\varepsilon(\mathbf{x}) = \frac{1}{2\varepsilon}\mathbf{g}^\mathrm{T}\mathbf{g}$ and $\mathcal{R}_\gamma(\mathbf{x},\mathbf{v}) = \frac{1}{2\gamma}(\mathbf{G}\mathbf{v})^\mathrm{T}(\mathbf{G}\mathbf{v})$ [15, 16]. This is a key point to map the constraint parameters to conventional force models, such as the Hertz contact law.

2.2.1 Time integration of NDEM

The numerical time integration scheme used in this thesis is based on the SPOOK stepper [16] derived from a discrete variational principle for the augmented system $(\mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}, \dot{\boldsymbol{\lambda}})$ and applying a semi-implicit discretization. The stepper is linearly stable and $\mathcal{O}(\Delta t^2)$ accurate for constraint violations [16] and involve solving a mixed complementarity problem (MCP) or equivalent quadratic programming problem (QP) [17]

$$\begin{aligned} \mathbf{H}\mathbf{z} + \mathbf{b} &= \mathbf{w}_l - \mathbf{w}_u \\ 0 \leq \mathbf{z} - \mathbf{l} &\perp \mathbf{w}_l \geq 0 \\ 0 \leq \mathbf{u} - \mathbf{z} &\perp \mathbf{w}_u \geq 0 \end{aligned} \tag{9}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{M} & -\mathbf{G}_\mathrm{n}^\mathrm{T} & -\mathbf{G}_\mathrm{t}^\mathrm{T} & -\mathbf{G}_\mathrm{r}^\mathrm{T} \\ \mathbf{G}_\mathrm{n} & \boldsymbol{\Sigma}_\mathrm{n} & 0 & 0 \\ \mathbf{G}_\mathrm{t} & 0 & \boldsymbol{\Sigma}_\mathrm{t} & 0 \\ \mathbf{G}_\mathrm{r} & 0 & 0 & \boldsymbol{\Sigma}_\mathrm{r} \end{bmatrix} \tag{10}$$

$$\mathbf{z} = \begin{bmatrix} \mathbf{v}_{i+1} \\ \boldsymbol{\lambda}_{\mathrm{n},i+1} \\ \boldsymbol{\lambda}_{\mathrm{t},i+1} \\ \boldsymbol{\lambda}_{\mathrm{r},i+1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\mathbf{M}\mathbf{v}_i - \Delta t\mathbf{M}^{-1}\mathbf{f}_\mathrm{ext} \\ \frac{4}{\Delta t}\boldsymbol{\Upsilon}_\mathrm{n}\mathbf{g}_\mathrm{n} - \boldsymbol{\Upsilon}_\mathrm{n}\mathbf{G}_\mathrm{n}\mathbf{v}_i \\ 0 \\ 0 \end{bmatrix} \tag{11}$$

The solution vector $\mathbf{z}$ contains the new velocities at time t_{i+1} and the Lagrange multipliers $\boldsymbol{\lambda}_\mathrm{n}$, $\boldsymbol{\lambda}_\mathrm{t}$ and $\boldsymbol{\lambda}_\mathrm{r}$. A factor Δt has been absorbed in the multipliers such that the constraint force reads $\mathbf{G}^\mathrm{T}\boldsymbol{\lambda}/\Delta t$. The upper and lower limits, $\mathbf{u}$ and $\mathbf{l}$ in Eq. (9), follow from the Signorini-Coulomb and rolling resistance law with the friction and rolling resistance coefficients μ_t and μ_r. Since the limits depend on the solution, this is a partially nonlinear complementarity problem. $\mathbf{w}_l$ and $\mathbf{w}_u$ are temporary slack variables. The diagonal matrices $\boldsymbol{\Sigma}_\mathrm{n}$, $\boldsymbol{\Sigma}_\mathrm{t}$, $\boldsymbol{\Sigma}_\mathrm{r}$ and $\boldsymbol{\Upsilon}_\mathrm{n}$ are related to the constraint

parameters are as follows

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathrm{n}} &= \frac{4}{\Delta t^2}\frac{\varepsilon_{\mathrm{n}}}{1+4\frac{\tau_{\mathrm{n}}}{\Delta t}}\mathbf{1} \\
\boldsymbol{\Sigma}_{\mathrm{t}} &= \frac{\gamma_{\mathrm{t}}}{\Delta t}\mathbf{1} \\
\boldsymbol{\Sigma}_{\mathrm{r}} &= \frac{\gamma_{\mathrm{r}}}{\Delta t}\mathbf{1} \\
\boldsymbol{\Upsilon}_{\mathrm{n}} &= \frac{1}{1+4\frac{\tau_{\mathrm{n}}}{\Delta t}}\mathbf{1}
\end{aligned} \tag{12}$$

Since the stepping scheme (9) is an implicit method the time-step is no longer restricted to the viscoelastic time-scale of the normal force. Instead, it is limited by the tolerance in acceptable constraint satisfaction. Typically, a time-step many orders in magnitude larger than for smooth DEM can be used. The precise mapping of the parameters in Eq. (12) to material parameters are addressed in Paper I and IV in this thesis. Increasing the parameters ε, γ and τ make the contacts more compliant. This improve the solvability of the MCP. The stiffness go to infinity in the limit of $\varepsilon, \gamma, \tau \to 0$.

2.2.2 Projected Gauss-Seidel (PGS) Solver

The computational properties of the solution algorithms for NDEM simulations is largely open questions. There are no general proofs of existence and uniqueness of solutions and few theoretical results on convergence and numerical stability [14]. The nonlinear projected Gauss-Seidel (PGS) algorithm is widely popular because of its low computational cost per iteration, small memory footprint and smooth distribution of truncation errors that favour stable simulation. The Schur complement form of Eq. (9) is

$$(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}} + \boldsymbol{\Sigma})\boldsymbol{\lambda}_{i+1} = \mathbf{q} - \mathbf{G}\mathbf{M}^{-1}\mathbf{p} \tag{13}$$

subject to law$[\mathbf{v}_{i+1}, \boldsymbol{\lambda}_{i+1}]$ = true, $\mathbf{p} = \mathbf{M}\mathbf{v}_i + \mathbf{f}_{\mathrm{ext}}$, and $\mathbf{q} = [(\boldsymbol{\Upsilon}\mathbf{G}_{\mathrm{n}}\mathbf{v}_i - \frac{4}{\Delta t}\boldsymbol{\Upsilon}\mathbf{g})^{\mathrm{T}}, 0, 0]^{\mathrm{T}}$ and the velocity update reads

$$\mathbf{v}_{i+1} = \mathbf{v}_i + \Delta t\mathbf{M}^{-1}\mathbf{f}_{\mathrm{ext}} + \mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{\lambda}_{i+1} \tag{14}$$

The PGS algorithm can be considered as a fixed-point linear iteration scheme computing the increments of $\boldsymbol{\lambda}_{i+1}$ and $\mathbf{v}_{i+1}$ using Eq. (13) and (14) repeatedly while projecting the solution to maintain the nonlinear contact law at each iteration step. After integrating the velocities and positions an impact stage follows. It include solving a similar MCP with the Newton impact law $\mathbf{G}_{\mathrm{n}}\mathbf{v}_+ = -e\mathbf{G}_{\mathrm{n}}\mathbf{v}_-$.

3 Accelerated DEM Simulation

The time required for simulating t_{real} seconds of evolution using DEM is the product of the number of time-steps and the computational time for each step

$$t_{\text{comp}} = \frac{t_{\text{real}}}{\Delta t} \cdot [t_{\text{col}} + t_{\text{solve}} + t_{\text{aux}}] \tag{15}$$

with time serially separated in collision detection (t_{col}), solver (t_{solve}) and auxiliary processing time (t_{aux}). In conventional smooth DEM the computations of force, velocity and position update are very cheap and parallelises well. Collision detection is therefore the main computational bottleneck such that $t_{\text{comp}}^{\text{DEM}} \propto N_{\text{p}}/S_{\parallel}^{\text{DEM}}$, where $S_{\parallel}^{\text{DEM}}$ is the parallel speedup. In NDEM, the solver is the main computational bottleneck. When using a PGS solver, the computational time scale as $t_{\text{comp}}^{\text{NDEM}} \propto N_{\text{p}}^{1+\gamma_\epsilon}/\epsilon S_{\parallel}^{\text{NDEM}}$ depending on error tolerance ϵ. The scaling exponent, $\gamma_\epsilon > 0$, reflect that the number of PGS iterations depend on the system size, connectivity and error tolerance as $N_{\text{it}} \propto N_{\text{p}}^{\gamma}/\epsilon$ [25]. The auxiliary process time include overheads for communication, memory management and processing of the DEM data structures that is necessary for post-processing or for optimisation of the other computational steps. It is interesting to study the relative computational time of the smooth and nonsmooth DEM

$$\frac{t_{\text{comp}}^{\text{DEM}}}{t_{\text{comp}}^{\text{NDEM}}} \propto \epsilon N_{\text{p}}^{-\gamma_\epsilon} \frac{\Delta t_{\text{NDEM}}}{\Delta t_{\text{DEM}}} \tag{16}$$

The following general observations can be made from this expression. The conventional smooth DEM is increasingly favourable in the limit of large systems ($N_{\text{p}} \to \infty$) and high accuracy ($\epsilon \to 0$). The relative efficiency of nonsmooth DEM increase with material stiffness and with slow dynamics, since $\Delta t_{\text{DEM}} \lesssim \sqrt{m/k} \to 0$ when $k \to \infty$ and $\Delta t_{\text{NDEM}} \lesssim \epsilon d/u_{\text{n}}$. To accelerate DEM simulations is to diminish the terms in Eq. (15). The following approaches have been identified:

Adaptive time-step and number of iterations

Depending on the requirements of the simulation and on the material parameters, a considerable speedup can be achieved by simply making the best choice between smooth and nonsmooth DEM. When NDEM is used, additional speedup can be gained by adjusting time-step size and number of iterations to the desired error tolerance and dynamic state of the granular material.

Parallelisation

The parallelisation of DEM subsystems that are weakly or not at all connected to each other is easy to implement, and has almost ideal scaling. Parallelisation of collision detection and PGS solver is more difficult but a well developed area [3, 18] and implemented in the software used in the thesis.

Convergence

There are several ways to improve the convergence of the PGS solver, either modify it or replace it entirely by alternative solvers, such as the accelerated projected gradient descent method [19] or the proximal point method [20]. With increased convergence less number of iterations can be used to reach the same level of accuracy and a speedup is gained if the computational cost of the modification is not too large. This is highly experimental work since there are few theoretical results on existence, uniqueness, stability and convergence of MCP and QP solvers for nonsmooth dynamics [14]. Modifications of the PGS algorithm include successive over-relaxation (SOR) and warm-starting based on some expectation of the solution.

Model reduction

Another way to accelerate the computations is to reduce the model complexity by somehow lowering the number of degrees of freedom in the system. This is known as model reduction and is widely used in solid and fluid mechanics, dynamical systems and control theory [21, 22]. Replacing a DEM model with a finite element model based on a continuum approximation of the same system is an example of this but is in general not applied to accelerate DEM simulations.

4 The Iron Ore Pelletising Process

The work presented in this thesis was inspired and motivated by the need for deeper understanding and specialised simulation tools for the processing of iron ore into pellets at the company LKAB. The pelletising process produces spherical pellets 9-$-$ 16 mm in diameter from ore material. The process consists of the following main stages [24], illustrated in Fig. 2. Comminuted fine size ore, *fines*, is first mixed with binder material. Agglomeration into soft ore balls, *green ore pellets*, occur in balling circuits, where fines, water and undersized pellets are circulated through rotating drums. The main agglomeration mechanisms are nucleation, layering, coalescence and breakage. The green pellets leave the drum through an outlet and are size distributed through a roller sieve, see Fig. 3. Under-sized particles are fed back to the drum. Over-sized pellets are crushed and mixed with the fines. The balling process can be controlled by regulating the drum velocity and feed rate of fine material, binding agencies and moisture. On-size pellets are conveyed to the induration furnace where they form hard pellets by oxidation and sintering. After being cooled, the pellets are ready for further transportation by train and ship. A typical iron ore balling circuit have drum diameter ranging between $3-5$ m, $8-10$ m long and circulate about $400-1200$ ton/h producing $100-300$ ton/h of on-size pellets. For an optimal performance, the material flow on the wide belt conveyor beneath the outlet and onto the sieve need to be as homogeneous in space and time as possible. The right image in Fig. 3 is an example of a failed outlet design that would cause substantial loss in productivity or poor pellet quality.

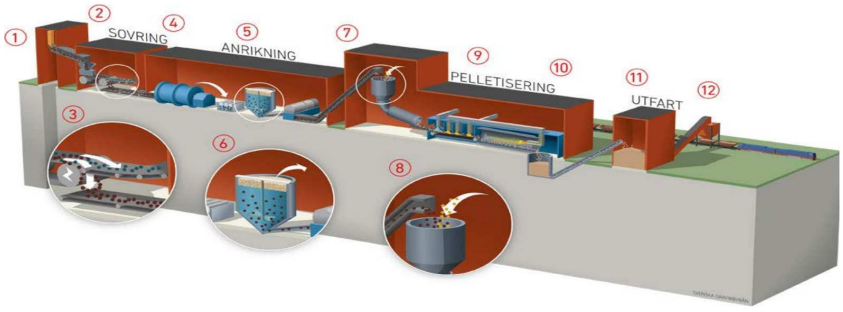

Figure 2: Illustration of the iron ore pelletising process. The balling process occur between step 7 and 9. Image courtesy of LKAB.

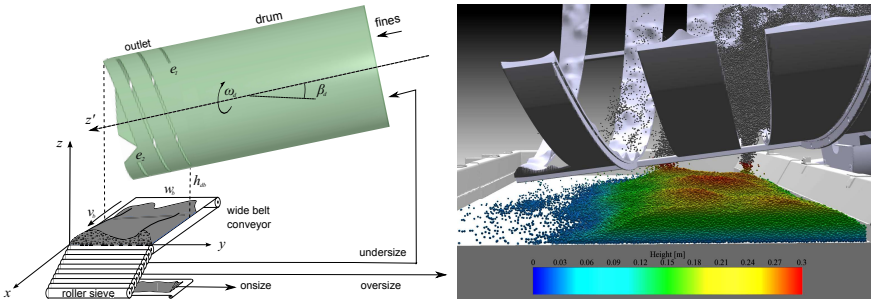

Figure 3: Illustration of the balling circuit and application of NDEM simulation to exploration of outlet designs.

5 Simulation Software

The project has involved a substantial amount of implementation work and numerical experiments. This has been carried out using a combination of the simulation software AgX Dynamics [23] and a series of prototype codes were developed and managed by the author. This code has been a mixture of C++, Python, Lua and Matlab scripts. AgX Dynamics is a software library supporting multidomain dynamics simulation for industrial and scientific applications. It is particularly well suited for realtime interactive simulation involving contacting rigid multibody systems. One large area of applications is heavy machinery simulators. AgX Dynamics also has a number of solutions for computer aided design and engineering. This include a module for modeling and simulation of granular materials and bulk handling systems. The core library include a direct solver for quadratic programming problems using a block pivot method with highly optimised algorithms and data structures for multibody systems with contacts and nonsmooth dynamics. A parallelised version of the PGS

algorithm using spatial partitioning for granular materials is implemented. The direct and iterative solvers can be run as coupled hybrid solvers.

6 Thesis Contribution and Future Work

A short description of the findings of the thesis' papers is given below. All papers have multiple authors and a short description of the main contribution by the author of the thesis is therefore included. All papers have been written in close collaboration and with good opportunity to provide input to and influence the development of the manuscripts.

Paper I

In this paper the computational properties of smooth and nonsmooth DEM are compared and a regularized version of NDEM based on the Hertz-Mindlin contact law is introduced, thus relating the NDEM solver parameters directly to fundamental material parameters. It has the both the conventional DEM and NDEM as limit cases. The required number of iterations needed to obtain a certain error tolerance is investigated in order to understand the true performance of NDEM in relation to accuracy and material properties. A surprising finding is that the required number of iterations do not always grow with the number of bodies in the system. Analogously to the Jansen effect of pressure saturating with depth in a granular column, due to force arching, also the required number of iterations can saturate and become independent of system size.

My contribution: development of a prototype NDEM simulation code with Hertzian contact model and PGS solver; setup and management of simulation system and post-processing pipeline; data analysis; establishing the convergence formula.

Paper II

NDEM simulation is shown to be a feasible tool for exploring the geometric design space of ore pellet balling drums. This is formulated as a design optimization problem. The objective function take the material distribution on the conveyor belt as input. A downscaled drum with a two-parameter outlet design is considered. Over 2000 simulations of different design are run with the identical drum flow to find an optimal shape. The conclusions include that the design principle is fundamentally flawed, as the design can be optimal only for a specific flow rate. It is also observed that the spherical idealization of the particles and absence of rolling resistance lead to significant disagreement between the simulated and observed pile formation.

My contribution: modeling of the balling circuit; setup and management of a system for batch simulations and post-processing; analysis and summary of simulation result.

Paper III

The practical steps of parameter identification, verification and validation of NDEM is described and applied to iron ore green pellets and the balling circuit. Also, the contact model is extended to include constraint based rolling resistance. The validation tests of simulated bulk behaviour show a good agreement with measurements of material flow in the balling plant. The sensitivity to model and solver parameters are investigated. This reveals that the results are very sensitive to rolling resistance and outlet geometry but not particularly sensitive to elasticity, friction coefficient and pellet diameter. It is also found that the application allows for surprisingly large time-step. Although this introduce big errors to a significant fraction (17 %) of the contacts the effect on the bulk flow is negligible.

My contribution: identification of material parameters and material flow from optical measurements; modeling and analysis of verification and validation systems; implementation and testing of rolling resistance model.

Paper IV

The possibility and efficiency of applying warm-starting to the PGS solver in NDEM simulation is investigated. Two methods are proposed - *history based* and *model based* - and tested. It is found that warm-starting based on the history can significantly reduce the required number of iterations. This primarily improve the convergence of friction forces and rolling resistance. A speedup between 2 to 5 can be achieved.

My contribution: joint work in developing the two warm-starting methods; prototype implementation; modeling, simulation and analysis of the column, pile and drum experiments.

Paper V

This paper presents a model order reduction technique for DEM simulations. Algorithms to substitute rigid aggregate bodies adaptively for collections of contacting particles that collectively co-move as rigid bodies are developed, as well as methods to refine the aggregates to smaller parts when necessary. The complexity of the reduced system can be many times smaller than the original. Two methods for predicting refinement are studied: contact event split and background trial solve split. The method has potential to accelerate NDEM simulations by $5-50$ times for reduction levels of $70-95$ %, if the computational overhead can be kept below given thresholds.

My contribution: algorithm development and prototype development of merge-split algorithms; joint work in development of split conditions; simulation and analysis to identify parameters and artefacts; investigate advantage and drawback of each merge-split method.

Future Work

There are many interesting questions and ideas to pursue in future work. When it comes to the application of iron ore pelletising systems, it is unclear how the presence of fine material and moisture affect the bulk behaviour of ore green pellets. It is also not clear how a mixture of fines and ore green pellets are efficiently modeled and simulated in a nonsmooth DEM framework - if at all possible. In reality, the ore green pellets are plastic and it is also an open question how to extend the NDEM contact model to this. On the computational modeling side, there are also many attractive continuations. In most real granular systems the solid, liquid and gaseous phase are simultaneously present and they typically benefit from different simulation modes, e.g., smooth versus nonsmooth DEM, size of time-step, number of iterations or change of solver. Ideally one should apply different iteration count, time-step and DEM-smoothness in different parts of the system and adaptively change this in order to focus the computational power to the part where it is needed the most. This point at asynchronous time-integration of the NDEM equations and the use of model order reduction interleaved in the PGS iterations. It is also clear that full use of model order reduction require extension from rigid to deformable aggregates.

References

[1] B. Androetti, Y. Forterre, O. Pouliquen, *Granular Media*, Cambridge University Press (2013).

[2] H. Jaeger, S. Nagel, R. Behringer, *Granular solids, liquids, and gases*, Rev. Mod. Phys. 68 (4) (1996) 1259–1273.

[3] T. Precklik, U. Rude, *Ultrascale simulations of non-smooth granular dynamics*, Computational Particle Mechanics, DOI 10.1007/s40571-015-0047-6 (2015).

[4] T. Pöschel, T. Schwager, *Computational Granular Dynamics, Models and Algorithms*, Springer-Verlag, 2005.

[5] B. T. Werner, *Eolian dunes: computer simulations and attractor interpretation.* Geology 23, (1995) 1107-1110.

[6] P. A. Cundall, O. D. L. Strack, *A discrete numerical model for granular assemblies*, Géotechnique, 29, (1979), 47-65.

[7] F. Radjai, F. Dubois, *Discrete-element Modeling of Granular Materials*, ISTE Ltd and John Wiley & Sons, Inc. 2011.

[8] J. J. Moreau, *Unilateral Contact and Dry Friction in Finite Freedom Dynamics*, volume 302 of Non-smooth Mechanics and Applications, CISM Courses and Lectures. Springer, Wien, (1988).

[9] J. J. Moreau, *Numerical aspects of the sweeping process*, Computer Methods in Applied Mechanics and Engineering 177 (1999) 329–349.

[10] M. Jean, *The non-smooth contact dynamics method*, Computer Methods in Applied Mechanics and Engineering 177 (1999) 235–257.

[11] F. Radjai, V. Richefeu, *Contact dynamics as a nonsmooth discrete element method*, Mechanics of Materials 41 (6) (2009) 715–728.

[12] E. J. Haug, *Computer-Aided Kinematics and Dynamics of Mechanical Systems, Volume I: Basic Methods*, Allyn and Bacon, Needham Heights, Massachusetts (1989).

[13] U. Ascher, H. Chin, L. Petzold, S. Reich, *Stabilization of constrained mechanical systems with DAEs and invariant manifolds*, J. Mech. Struct. Machines, vol. 23 (1994), 135–157.

[14] B. Brogliato, A. Ten Dam, L. Paoli, F. Génot, M. Abadie. *Numerical simulation of finite dimensional multibody nonsmooth mechanical systems*, Applied Mechanics Reviews, 55(2), (2002), 107–150.

[15] F. Bornemann, C. Schütte, *Homogenization of Hamiltonian systems with a strong constraining potential*, Phys. D, 102(1-2), (1997) 57–77.

[16] C. Lacoursière, *Regularized, stabilized, variational methods for multibodies*, in: D. F. Peter Bunus, C. Führer (Eds.), The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), Linköping University Electronic Press, 2007, pp. 40–48.

[17] K. G. Murty, *Linear Complementarity, Linear and Nonlinear Programming*, Helderman-Verlag, Heidelberg, 1988.

[18] V. Visseq, P. Alart, D. Dureisseix, *High performance computing of discrete non-smooth contact dynamics with domain decomposition*, International Journal for Numerical Methods in Engineering, vol. 96, no. 9, (2013), 584-598.

[19] H. Mazhar, T. Heyn, D. Negrut, A. Tasora, *Using Nesterov's Method to Accelerate Multibody Dynamics with Friction and Contact* ACM Transactions on Graphics, vol. 34, no. 3 (2015).

[20] S. Schoeder, H. Ulbrich, T. Schindler, *Discussion of the GearGupta-Leimkuhler method for impacting mechanical systems*, Multibody System Dynamics, vol. 31 (2014), 477-495.

[21] A. Antoulas, *Approximation of Large-Scale Dynamical Systems*, Society for Industrial and Applied Mathematics (2005).

[22] G. Kerschen, J-C. Golinval, A. Vakakis, L. Bergman, *The Method of Proper Orthogonal Decomposition for Dynamical Characterization and Order Reduction of Mechanical Systems: An Overview*, Nonlinear Dynamics 41(2005) 147–169.

[23] Algoryx Simulations. AGX Dynamics, October 2015.

[24] S. Forsmo, *Influence of green pellet properties on pelletizing of magnetite iron ore*, Ph.D. thesis, Luleå University of Technology, Luleå (2007).

[25] Unger T, Brendel L, Wolf D, Kertsz J, *Elastic behavior in contact dynamics of rigid particles*, Physical Review E. (2002)

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nme.4612

Examining the smooth and nonsmooth discrete element approaches to granular matter

M. Servin*,†, D. Wang, C. Lacoursière and K. Bodin

UMIT Research Lab, Umeå University, Umeå, SE-90187, Sweden

SUMMARY

The smooth and nonsmooth approaches to the discrete element method (DEM) are examined from a computational perspective. The main difference can be understood as using explicit versus implicit time integration. A formula is obtained for estimating the computational effort depending on error tolerance, system geometric shape and size, and on the dynamic state. For the nonsmooth DEM (NDEM), a regularized version mapping to the Hertz contact law is presented. This method has the conventional nonsmooth and smooth DEM as special cases depending on size of time step and value of regularization. The use of the projected Gauss-Seidel solver for NDEM simulation is studied on a range of test systems. The following characteristics are found. First, the smooth DEM is computationally more efficient for soft materials, wide and tall systems, and with increasing flow rate. Secondly, the NDEM is more beneficial for stiff materials, shallow systems, static or slow flow, and with increasing error tolerance. Furthermore, it is found that just as pressure saturates with depth in a granular column, due to force arching, also the required number of iterations saturates and become independent of system size. This effect make the projected Gauss-Seidel solver scale much better than previously thought. Copyright © 2014 John Wiley & Sons, Ltd.

Received 11 June 2013; Revised 31 October 2013; Accepted 01 November 2013

KEY WORDS: discrete element method; multibody dynamics; granular media; contact; explicit time integration; linear solvers

1. INTRODUCTION

The discrete element method (DEM) of simulation is an important tool for studying many natural phenomena and structures in the fields of material science, statistical physics, and geophysics when granular material is involved. It is also an important tool for design and optimization in the industries of processing, manufacturing, storage, and transportation of granular materials, e.g., grains, minerals, pharmaceutical pills, pellets, sand, and rocks. One of the main challenges for DEM simulation is to reduce the computational time of large-scale simulations with $10^3 - 10^9$ elements. There are two main approaches to discrete element simulation, referred to as *smooth* DEM (SDEM) and *nonsmooth* DEM (NDEM). The difference lies in whether the viscoelastic nature of the contacts is resolved in time or not. The nonsmooth approach considers collisions and stick-slip frictional transitions as instantaneous events, where the velocity may change discontinuously in time, according to a given contact law. This allows for large time step integration with the potential of considerably less computational effort. Surprisingly few comparisons between SDEM and NDEM can be found in literature. One comparison of the computational scaling is made by Brendel *et al*, summarized by Equation (1.37) and Figure 6 in Ref. [1]. According to that analysis, the nonsmooth approach

*Correspondence to: M. Servin, Department of Physics, Umeå University, Umeå, SE-90187, Sweden.
†E-mail: martin.servin@physics.umu.se

is most favorable for dense, quasistatic systems with few particles and high ratio of material stiffness over element mass, while the smooth approach becomes increasingly favorable with increasing number of particles, kinetic energy, and decreasing material stiffness. The comparison in Ref. [1] is, however, overly simplified and does not fully account for the dynamic state or geometric shape of the system and does not include any error tolerance threshold.

The main purpose of this paper is to provide a theoretical basis and empirical data for deciding when to use SDEM or NDEM simulation. Also, a regularized version of NDEM is presented, referred to as *semi-smooth* DEM. It has the smooth and nonsmooth methods as limiting cases and thus constitutes a hybrid method sharing properties of both the other methods and forming a starting point for developing DEM simulation with adaptive transitions between smooth and nonsmooth time integration.

The paper is outlined as follows. In Section 2, an overview of SDEM and NDEM is given with key references. A hybrid version of the SDEM and NDEM is presented in Section 2.4. The computational properties of SDEM and NDEM are outlined and discussed in Section 3. The metrics used for comparison are introduced in Section 4.1. In Section 4, the simulation procedure and chosen test systems are described. The results of the simulations are presented and discussed in Section 5. The main conclusions are finally presented in Section 6.

1.1. Notation

For a rigid body, a, we use the notations $\vec{\boldsymbol{x}}_{[a]}, \vec{\boldsymbol{v}}_{[a]}, \vec{\boldsymbol{f}}_{[a]}$, and $m_{[a]}$ for position, velocity, force, and mass, and $\vec{\boldsymbol{e}}_{[a]}, \vec{\boldsymbol{\omega}}_{[a]}, \vec{\boldsymbol{\tau}}_{[a]}$, and $\boldsymbol{I}_{[a]}$, for orientation, angular velocity, torque, and inertia tensor. We agglomerate into generalized position, velocity, force, and mass, denoted $\mathbf{x}_{[a]}, \mathbf{v}_{[a]}, \mathbf{f}_{[a]}$, and $\mathbf{M}_{[a]}$, with $\mathbf{v}_{[a]} = \left(\vec{\boldsymbol{v}}_{[a]}^{\mathrm{T}}, \vec{\boldsymbol{\omega}}_{[a]}^{\mathrm{T}}\right)^{\mathrm{T}}$ etc. and $\mathbf{M}_{[a]} = \mathrm{diag}\left(m_{[a]}\mathbf{1}_{3\times 3}, \boldsymbol{I}_{[a]}\right)$. These are components of the global system quantities that we denote $\mathbf{x}, \mathbf{v}, \mathbf{f}$, and $\mathbf{M}$. We use quaternions for representing orientations. The theory and algorithms in this paper applies to rigid bodies of general shape, although the simulations are made for spherically shaped rigid bodies, where integration of the orientation can be omitted. These bodies are referred to as particles but do posses rotational degree of freedom. Contact forces and velocities are sometimes decomposed in the directions of contact normals, $\vec{\boldsymbol{n}}$ and tangents, $\vec{\boldsymbol{t}}$. The gap function $g(\mathbf{x})$ measures the magnitude of overlap between contacting bodies. N_{p} denotes the number of particles and N_{c} the number of contacts. The matrix dimension of the global quantities are $\dim(\mathbf{x}) = 7N_{\mathrm{p}} \times 1$, $\dim(\mathbf{v}) = \dim(\mathbf{f}) = 6N_{\mathrm{p}} \times 1$, $\dim(\mathbf{M}) = 6N_{\mathrm{p}} \times 6N_{\mathrm{p}}$, and $\dim(\mathbf{g}) = N_{\mathrm{c}} \times 1$. Time is discretized in steps of size h, and we denote the discrete time points by integer index i such that $t_i = t_{i-1} + h$. We use integer k for solver iteration step. Particle indices are emphasized by square brackets $[a]$ and $[b]$ and contact index by round brackets (α). The relative velocity at a contact point α between bodies a and b can thus be written as $\vec{\boldsymbol{u}}^{(\alpha)} = \vec{\boldsymbol{v}}_{[a]} + \vec{\boldsymbol{d}}_{[a]}^{(\alpha)} \times \vec{\boldsymbol{\omega}}_{[a]} - \vec{\boldsymbol{v}}_{[b]} - \vec{\boldsymbol{d}}_{[b]}^{(\alpha)} \times \vec{\boldsymbol{\omega}}_{[a]}$, where $\vec{\boldsymbol{d}}_{[a]}^{(\alpha)}$ is the position of the contact point relative to $\vec{\boldsymbol{x}}_{[a]}$ and $\vec{\boldsymbol{d}}_{[b]}^{(\alpha)}$ relative to $\vec{\boldsymbol{x}}_{[b]}$. Furthermore, the notation $\mathbf{G}(n_1 : n_2, m_1 : m_2)$ is the standard Matlab notation for submatrices, referring to the rectangular submatrix block of $\mathbf{G}$ that ranges from row-column index (n_1, m_1) to (n_2, m_2).

2. SMOOTH AND NONSMOOTH DISCRETE ELEMENT METHODS

An introduction to the theory and computational aspects of SDEM can be found in Ref. [2]. Comprehensive descriptions of NDEM and its relation to SDEM can be found in Ref. [3] and in Ref. [4]. Important original works include those of Cundall and Strack [5] and Moreau [6]. In Section 2.4, we introduce a *semi-smooth* DEM that share features of both methods.

From a computational perspective, the main difference between SDEM and NDEM is related to explicit and implicit integration. In SDEM, the contact forces are modeled as damped springs or more general penalty functions [7, 8]. The equations of motion are the Newton–Euler equations

DOI: 10.1002/nme

of rigid body motion, which form a set of *ordinary differential equations* (ODE) and are usually integrated with an explicit time stepper using small step size depending on spring stiffness and particle mass. For an overview of explicit time stepping schemes, see [9]. Forces are computed *locally* for each contact pair. In NDEM, the Newton–Euler equations are constrained by the Signorini–Coulomb contact law [3, 4], replacing the contact springs. The velocities are no longer assumed time-continuous. The contact forces and resulting velocities are computed *globally* by solving the constrained equations of motion for the entire contact network using an implicit time-integration algorithm. The nonsmooth method is a direct consequence of implicit time-integration in the presence of inequality conditions, e.g., that the contact force should vanish at separations or be limited by the Coulomb friction cone. As shown in Section 2.1, it is convenient to extend the system with auxiliary variables, *Lagrange multipliers* [8], and explicit use of constraints. Mathematically, this transforms the system from an ODE to a *differential variational inequality* (DVI) [10].

2.1. Relations between implicit and explicit integrators and constrained systems

In this section, we elucidate the relations between implicit and explicit integrators, constrained systems, and nonsmooth dynamics. Consider a point particle of unit mass in one dimension subject to the potential $U = \frac{1}{2\varepsilon}x^2$, yielding the force $f = -\partial U/\partial x = -\varepsilon^{-1}x$. Writing $v \equiv \dot{x}$, an explicit integration using the Verlet method with time step h yields

$$\begin{aligned} v_{i+1} &= v_i + hf_i = v_i - h\varepsilon^{-1}x_i \\ x_{i+1} &= x_i + hv_{i+1}. \end{aligned} \tag{1}$$

Here, only the force at time i is needed to advance the system. Using the implicit midpoint method, we instead have

$$\begin{aligned} v_{i+1} &= v_i + hf\left(\frac{x_{i+1}+x_i}{2}\right) = v_i - \frac{h}{2\varepsilon}\left(x_{i+1}+x_i\right) \\ x_{i+1} &= x_i + \frac{h}{2}\left(v_{i+1}+v_i\right). \end{aligned} \tag{2}$$

Simple manipulations yield

$$\left[1+\frac{h^2}{4\varepsilon}\right]v_{i+1} = \left[1-\frac{h^2}{4\varepsilon}\right]v_i - \frac{h}{\varepsilon}x_i. \tag{3}$$

This is a standard implicit integration, and if the system was of higher dimension, the terms inside the brackets would be matrices and we would have to solve a linear system of equations for v_{i+1}. The terms containing ε^{-1} are problematic in the limit $\varepsilon \to 0$. Now introduce the auxiliary variable $\lambda = \varepsilon^{-1}x$ in continuous time so that, in discretized time,

$$\lambda = -\frac{1}{2\varepsilon}\left(x_{i+1}+x_i\right) = -\frac{1}{\varepsilon}x_i - \frac{h}{4\varepsilon}v_i - \frac{h}{4\varepsilon}v_{i+1}. \tag{4}$$

After simple manipulations, the stepping scheme reads

$$\begin{aligned} \begin{bmatrix} 1 & -1 \\ 1 & \frac{4\varepsilon}{h^2} \end{bmatrix}\begin{bmatrix} v_{i+1} \\ h\lambda \end{bmatrix} &= \begin{bmatrix} v_i \\ -\frac{4}{h}x_i - v_i \end{bmatrix} \\ x_{i+1} &= x_i + \frac{h}{2}(v_{i+1}+v_i). \end{aligned} \tag{5}$$

There are no longer problematic terms of $1/\varepsilon$, and we can in principle set $\varepsilon = 0$. If the spring is one sided, i.e., inactive when $x > 0$ the integration should respect the condition $f_i = 0$ if $x_i > 0$.

If we now consider a more general mechanical system with mass matrix $\mathbf{M}$, coordinates $\mathbf{x}$ and velocity $\mathbf{v}$ subject to a strong potential $U = \frac{1}{2\varepsilon}\mathbf{g}^{\mathrm{T}}\mathbf{g}$, the force is then $\mathbf{f} = -\partial U/\partial \mathbf{x}^{\mathrm{T}} = -\frac{1}{\varepsilon}\mathbf{G}^{\mathrm{T}}\mathbf{g}$, where $\mathbf{G} = \partial \mathbf{g}/\partial \mathbf{x}$. Though the mass matrix $\mathbf{M}$ depends on the coordinates $\mathbf{x}$ in general, as in

DOI: 10.1002/nme

the case of the rigid body, this is neglected for simplicity. Following the same steps as before, the implicit integration can be written as either

$$\left[\mathbf{M}+\frac{h^2}{4\varepsilon}\mathbf{G}^{\mathrm{T}}\mathbf{G}\right]\mathbf{v}_{i+1} \;=\; \left[\mathbf{M}-\frac{h^2}{4\varepsilon}\mathbf{G}^{\mathrm{T}}\mathbf{G}\right]\mathbf{v}_i-\frac{h}{\varepsilon}\mathbf{G}^{\mathrm{T}}\mathbf{g}_i \tag{6}$$

or, by introducing auxiliary variable $\boldsymbol{\lambda}=-\varepsilon^{-1}\mathbf{g}(\mathbf{x})$

$$\begin{bmatrix}\mathbf{M} & -\mathbf{G}^{\mathrm{T}}\\ \mathbf{G} & \frac{4\varepsilon}{h^2}\end{bmatrix}\begin{bmatrix}\mathbf{v}_{i+1}\\ \lambda\end{bmatrix}=\begin{bmatrix}\mathbf{v}_i+h\mathbf{f}_{\mathrm{s}}\\ -\frac{h}{4}\mathbf{g}_i+\mathbf{G}_i\mathbf{v}_i\end{bmatrix}, \tag{7}$$

where we absorbed the h factor in $\boldsymbol{\lambda}$ and added soft forces $\mathbf{f}_{\mathrm{s}}$ in view of what we will do below.

Clearly, if there were conditions on the forces $\boldsymbol{\lambda}$, they would have to be applied globally. That is not the case for the explicit method in which cases rules such as one-sided springs would be used, one force at a time at each step. The perturbation $4\varepsilon/(h^2)$ in the matrix in Equation (7) can be clearly related to a penalty. When it vanishes, however, we have a constrained system with hard contacts. From this perspective, the fundamental difference between SDEM and NDEM is related to explicit and implicit integration methods. The first one computes forces locally, the second globally.

The correspondence to constrained mechanical systems contains a number of subtleties with regards to the convergence of the trajectories, velocities, and forces λ in the limit where $\varepsilon\to 0$. The basic theory is simple; however, if we consider the Legendre transform of a potential $\mathbf{U}=\frac{1}{2\varepsilon}\mathbf{g}^{\mathrm{T}}\mathbf{g}$ where $\mathbf{g}(\mathbf{x})$ is a well-behaved function of $\mathbf{x}$ such that the Jacobian matrix $\partial\mathbf{g}/\partial\mathbf{x}=\mathbf{G}$ has full rank at $\mathbf{g}(\mathbf{x})=0$. Choosing $\boldsymbol{\lambda}$ and the transform variable, the Legendre transform is defined as

$$\tilde{U}(\boldsymbol{\lambda})=-\max_{\mathbf{g}}\left[\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{g}+\bar{U}(\mathbf{g})\right] \tag{8}$$

where $\bar{U}(\mathbf{g})$ is the pullback of $U(\mathbf{x})$. This leads to

$$\begin{aligned} U(\mathbf{x})=\bar{U}(\mathbf{g})&=-\frac{\varepsilon}{2}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{g}\\ \varepsilon\boldsymbol{\lambda}+\mathbf{g}&=0. \end{aligned} \tag{9}$$

Note in particular that $\partial U/\partial\mathbf{x}^T=-\mathbf{G}^{\mathrm{T}}\boldsymbol{\lambda}$. For a simple mechanical system with equations of motion $\mathbf{M}\ddot{\mathbf{x}}+\partial U/\partial\mathbf{x}^T-\mathbf{f}_{\mathrm{s}}=0$, where $\mathbf{f}_{\mathrm{s}}$ are the 'weak' forces, of the augmented equations of motion are then

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{x}}-\mathbf{G}^{\mathrm{T}}\boldsymbol{\lambda}&=\mathbf{f}_{\mathrm{s}}\\ \mathbf{g}(\mathbf{x})+\varepsilon\boldsymbol{\lambda}&=0. \end{aligned} \tag{10}$$

It is possible to show that Equation (10) has a well-behaved limit as $\varepsilon\downarrow 0$ provided there is some dissipation force of the form $\mathbf{f}_d=-\gamma\mathbf{G}^{\mathrm{T}}\dot{\mathbf{g}}=-\gamma\mathbf{G}^{\mathrm{T}}\mathbf{G}\dot{\mathbf{x}}$ [11]. At the limit $\varepsilon=0$, we have the dynamics of a constrained system. Nonsmooth formulations are entirely based on constrained systems, and this leads naturally to Differential Algebraic Equations of motion or DVIs when including impacts and dry friction. Though the general numerical methods for these are computationally expensive [12], the special case of multibody systems allows simpler methods such as RATTLE [13].

In principle, therefore, if we accept a nonzero relaxation, ε, there is no difference between the smooth (local) and nonsmooth (global) formulation other than the time integration method until inequalities are considered. However, it is clear from Equation (1) that inequalities and complementarity conditions, such as $0\leqslant\mathbf{f}\perp\mathbf{g}(\mathbf{x})\geqslant 0$, can be treated one at a time in the smooth formulation, but this is not the case when considering the system in Equation (7).

2.2. *Smooth discrete element method*

In SDEM, the contact normal force is a direct function of the geometric overlap function, $g(\mathbf{x})$, and its time derivative. An example of a common force model is the nonlinear Hertz model that follows

DOI: 10.1002/nme

from the theory of linear viscoelastic materials. Friction is usually modeled as a spring in the tangential direction. The tangential spring extension is computed by integrating the slip velocity, and the force is projected onto the friction cone to obey the Coulomb law. For spherical particles, the normal and tangential contact forces are [2]

$$\vec{f}_n = k_\mathrm{n}\left(g^{3/2} + cg^{1/2}\dot{g}\right)\vec{n} \tag{11}$$

$$\vec{f}_t = \mathrm{proj}_{\mu|\vec{\mathbf{f}}_\mathrm{n}|}\left(-\int k_\mathrm{t}\vec{u}_t\,\mathrm{d}t\right) \tag{12}$$

where $\vec{u}_t$ is the tangential relative velocity at the contact point. From the Hertz model for spheres, the normal spring stiffness and damping coefficients are $k_\mathrm{n} = \mathrm{E}\sqrt{2d}/3\left(1-\nu^2\right)$ and $c = 4\left(1-\nu^2\right)(1-2\nu)\eta/15\mathrm{E}\nu^2$, where E is the Youngs's modulus, ν is the Poisson ratio, and $d = \left(d_{[a]}^{-1} + d_{[b]}^{-1}\right)^{-1}$ is the effective diameter from the interaction between spheres with diameter $d_{[a]}$ and $d_{[b]}$ and the material viscosity constant η [14]. For the friction spring coefficient k_t, there is no such relation to fundamental material parameters, and it must be determined by comparison with experimental results.

For explicit time integration methods, $(\mathbf{x}_i, \mathbf{v}_i) \to (\mathbf{x}_{i+1}, \mathbf{v}_{i+1})$, such as Verlet or symplectic Euler, the time step h must be smaller than the shortest time scale given by the viscoelastic interaction time $\sqrt{m/k_\mathrm{n}}$.

2.3. Nonsmooth discrete element method

In NDEM, the dynamics at short time scales is not resolved. Instead, the velocity is allowed to be discontinuous in time and contact forces, represented by impulses or kinematic constraints, can propagate through the system instantly. This enables large time step integration despite stiff materials. The contact force laws may be derived from the same penalty potentials as for SDEM or simply be determined directly empirically. We denote the impulse by $\vec{r}$, instantly changing the contact velocity from $\vec{u}$ to $\vec{u}^+ = \vec{u}^- + \mathbf{W}\vec{r}$, with a transfer matrix $\mathbf{W}$ consistent with the preservation of total momentum. Over one time step h, the time-averaged contact force is $\vec{f} = \vec{r}/h$. Specifically, the explicit contact force model in Equations (11) and (12) is replaced by the *Signorini-Coulomb law* [3,4] that if $g \geqslant 0$ then $\mathrm{law}_\mathrm{SC}\left[\vec{u}_{i+1}, \vec{r}_{i+1}\right] = \mathrm{true}$:

$$0 \leqslant \vec{u}_n \perp \vec{f}_n \geqslant 0 \tag{13}$$

$$\vec{u}_t = 0 \quad \Longrightarrow \quad |\vec{f}_t| \leqslant \mu|\vec{f}_\mathrm{n}| \tag{14}$$

$$\vec{u}_t \neq 0 \quad \Longrightarrow \quad |\vec{f}_t| = \mu|\vec{f}_n|\,, \;\; \vec{f}_t^{\,T}\vec{u}_t = -|\vec{f}_t||\vec{u}_t| \tag{15}$$

where the last condition is the maximum dissipation principle. We use $\mathrm{law}_\mathrm{SC}\left[\vec{u}_{i+1}, \vec{r}_{i+1}\right] = \mathrm{true}$ in short for the inequalities in (13)–(15). Newton's impact law of restitution, $\vec{u}_\mathrm{n}^+ = -e\vec{u}_\mathrm{n}^-$ with coefficient of restitution $0 \leqslant e \leqslant 1$, is conventionally also included in the contact law. The occurrence of time-discontinuities mediated by constraints or impulses changes the dynamic system from an ODE into a DVI. Solving the set of coupled inequalities (13)–(15) and update equations $(\mathbf{x}_i, \mathbf{v}_i) \to (\mathbf{x}_{i+1}, \mathbf{v}_{i+1})$ for the entire granular system is the main computational task in NDEM simulations, while it is a fast straight forward step in SDEM simulations. Observe that in NDEM, the magnitude of contact forces do not necessarily depend on the magnitude of the overlaps. Also, a local change in the system may propagate instantly through the entire contact network thanks to the implicit nature of the method. This produces the required bulk behavior without need for resolving fast pressure waves. Observe that instant propagation speeds assume perfectly rigid bodies, but also for stiff materials, the propagation length in one time-step may exceed the size of the system.

DOI: 10.1002/nme

There are several methods for numerical integration of DVI systems. The approaches may be divided into iterative (splitting) solvers and direct (pivoting) solvers, but it is also possible to construct hybrid direct-iterative solvers. Often, the time-integration is made with a single-step first order method such as the symplectic (semi-implicit) Euler, resulting in a scheme nearly identical to the RATTLE algorithm [13] for molecular dynamics. In general, because the system is non-smooth, there is no benefit from higher order methods, e.g., multistage or multistep. In Section 2.4, we present an extension of NDEM that give the contact constraint force viscoelastic properties and is based on the linear complementarity formulation of NDEM. The simulation result in the paper are produced with an implementation of this method. Therefore, the linear complementarity formulation is outlined in more detail in Section 2.3.2.

2.3.1. Projected Gauss-Seidel. A natural and common approach is to treat each single contact problem, $\alpha, \beta = 1, 2, \ldots, N_\mathrm{c}$, sequentially and then iterate, $k = 1, 2, \ldots, N_\mathrm{it}$, until all contact laws are fulfilled to a desired error tolerance. A common algorithm for this is the blocked *projected Gauss-Seidel* solver [15]. This is a stationary iterative method for solving Equations (13)–(15) approximately. At the k:th iteration step, each pair of contact velocity and impulse $\left(\vec{\boldsymbol{u}}_k^{(\alpha)}, \vec{\boldsymbol{r}}_k^{(\alpha)}\right)$ is solved for each local contact problem α

$$\vec{\boldsymbol{u}}_k^{(\alpha)} - \boldsymbol{W}_{(\alpha\alpha)}\vec{\boldsymbol{r}}_k^{(\alpha)} = \vec{\boldsymbol{u}}_\mathrm{s}^{(\alpha)} + \Delta\vec{\boldsymbol{u}}_k^{(\alpha)} \tag{16}$$

$$\mathrm{law}_\mathrm{SC}\left[\vec{\boldsymbol{u}}_k^{(\alpha)}, \vec{\boldsymbol{r}}_k^{(\alpha)}\right] = \mathrm{true} \tag{17}$$

where

$$\Delta\vec{\boldsymbol{u}}_k^{(\alpha)} = \sum_{\beta<\alpha} \mathbf{W}_{(\alpha\beta)}\vec{\boldsymbol{r}}_k^{(\beta)} + \sum_{\beta>\alpha} \mathbf{W}_{(\alpha\beta)}\vec{\boldsymbol{r}}_{k-1}^{(\beta)} \tag{18}$$

and $\vec{\boldsymbol{u}}_\mathrm{s}$ is the contact velocity as would be in the presence of only smooth forces, $\mathbf{W}_{(\alpha\beta)} = \mathbf{H}^\mathrm{T}_{(\alpha)[a]}\mathbf{M}^{-1}_{[ab]}\mathbf{H}_{(\beta)[b]}$ is the Delassus operator for contact α and β with $\mathbf{H}^\mathrm{T}_{(\alpha)[a]}$ being the affine transformation of velocity of body a to relative contact velocity in point α and $\mathbf{M}$ the mass matrix of the system. After convergence to a set of impulses and contact velocities fulfilling law_SC to desired tolerance, the body velocities in global coordinates are computed by $\mathbf{v}_{i+1} = \mathbf{v}_\mathrm{s} + \mathbf{Hr}$ with the final net impulse $\mathbf{Hr}$ and $\mathbf{v}_\mathrm{s} = \mathbf{v}_i + h\mathbf{M}^{-1}\mathbf{f}_\mathrm{s}$ is the updated velocity in presence of smooth forces. Finally, position is updated by $\mathbf{x}_{i+1} = \mathbf{x}_i + h\mathbf{v}_{i+1}$. In $\theta = 1/2$ integration methods, the positions are first update on half time step with the velocities from smooth forces before contacts are computed.

2.3.2. Mixed linear complementarity problem. Another popular method follows from linearization of the Coulomb friction law, by approximating the friction cone with a (scaled) box or a polyhedral cone. The constrained equations of motion may then be put in the form of a *mixed linear complementarity problem* (MLCP). In terms of body velocity $\mathbf{v}$ and Lagrange multiplier $\boldsymbol{\lambda}$, the MLCP reads [16]

$$\begin{aligned} &\mathbf{Hz} + \mathbf{b} = \mathbf{w}_+ - \mathbf{w}_- \\ &0 \leqslant \mathbf{z} - \mathbf{l} \perp \mathbf{w}_+ \geqslant 0 \\ &0 \leqslant \mathbf{u} - \mathbf{z} \perp \mathbf{w}_- \geqslant 0 \end{aligned} \tag{19}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{M} & -\mathbf{G}_\mathrm{n}^\mathrm{T} & -\mathbf{G}_\mathrm{t}^\mathrm{T} \\ \mathbf{G}_\mathrm{n} & 0 & 0 \\ \mathbf{G}_\mathrm{t} & 0 & 0 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{v}_{i+1} \\ \boldsymbol{\lambda}_{\mathrm{n},i+1} \\ \boldsymbol{\lambda}_{\mathrm{t},i+1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\mathbf{M}\mathbf{v}_\mathrm{s} \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

$\mathbf{G}_\mathrm{n}$ and $\mathbf{G}_\mathrm{t}$ are the normal and tangential constraint Jacobians for the Signorini–Coulomb law and $\mathbf{w}_\pm$ are (temporary) slack variables, $\mathbf{l}$ and $\mathbf{u}$ are the upper and lower limits on the solution implied by the linearized Signorini–Coulomb law. The normal and tangential contact constraint force is

DOI: 10.1002/nme

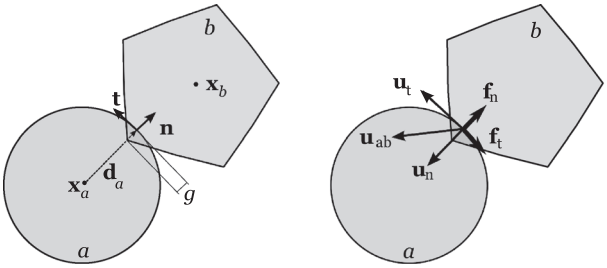

Figure 1. Overlapping discrete elements with notations.

$\mathbf{f}_\mathrm{n} = \mathbf{G}_\mathrm{n}^\mathrm{T}\boldsymbol{\lambda}_\mathrm{n}/h$ and $\mathbf{f}_\mathrm{t} = \mathbf{G}_\mathrm{t}^\mathrm{T}\boldsymbol{\lambda}_\mathrm{t}/h$, respectively. The h^{-1} factor is due to a convenient normalization of the Lagrange multiplier such that it has the dimension of impulse.

The MLCP formulation with saddle-point matrix $\mathbf{H}$ in Equation (20) is common in the realm of *multibody system dynamics* for modeling of linked mechanism using kinematic constraints for describing various joints, their actuations, and geometric limits. One significant advantage of this formulation of NDEM is therefore that it automatically provides a unified formulation for modeling and simulating of granular material strongly coupled with mechatronic systems, such as vehicles, conveying systems, and articulated manipulators [17].

A generalization of the MLCP formulation to include also viscoelastic properties in the contact dynamics is presented in Section 2.4 and used in simulation in Section 4. Therefore, we provide some further details on the MLCP formulation here. The normal constraint force acts to prevent penetration, i.e., to maintain the constraint $g^{(\alpha)} \leqslant 0$, for each contact candidate α between body a and b. The overlap is computed

$$g^{(\alpha)} = \vec{\boldsymbol{n}}_{(\alpha)}^\mathrm{T}\left(\vec{\boldsymbol{x}}_{[a]} + \vec{\boldsymbol{d}}_{[a]}^{(\alpha)} - \vec{\boldsymbol{x}}_{[b]} - \vec{\boldsymbol{d}}_{[b]}^{(\alpha)}\right) \tag{21}$$

with the contact normal $\vec{\boldsymbol{n}}_{(\alpha)}$ directed outwards from a, and $\vec{\boldsymbol{d}}_{[a]}^{(\alpha)}$ is the position of the contacting point on the surface of body a relative to its center of mass position $\vec{\boldsymbol{x}}_{[a]}$, see Figure 1. The nonpenetration constraint implies that the relative contact normal velocity should be zero or separating, i.e., $\mathbf{G}_\mathrm{n}^{(\alpha)}\mathbf{v} \geqslant 0$, with the normal Jacobian $\mathbf{G}_\mathrm{n}^{(\alpha)} = \partial g^{(\alpha)}/\partial\mathbf{x}$. The nonzero blocks of the normal Jacobian are

$$\mathbf{G}_{\mathrm{n}[a]}^{(\alpha)} = \begin{bmatrix} -\vec{\boldsymbol{n}}_{(\alpha)}^\mathrm{T} & -\left(\vec{\boldsymbol{d}}_{[a]}^{(\alpha)} \times \vec{\boldsymbol{n}}_{(\alpha)}\right)^\mathrm{T} \end{bmatrix} \tag{22}$$

$$\mathbf{G}_{\mathrm{n}[b]}^{(\alpha)} = \begin{bmatrix} \vec{\boldsymbol{n}}_{(\alpha)}^\mathrm{T} & \left(\vec{\boldsymbol{d}}_{[b]}^{(\alpha)} \times \vec{\boldsymbol{n}}_{(\alpha)}\right)^\mathrm{T} \end{bmatrix} \tag{23}$$

Friction is introduced as a constraint of vanishing relative contact velocity $\vec{\boldsymbol{u}}_{[ab]}^{(\alpha)}$ in the tangent plane, i.e., $\mathbf{G}_\mathrm{t}^{(\alpha)}\mathbf{v} = 0$, unless the force reach the friction bounds. With the *box friction* approximation, the tangent plane is spanned with two orthogonal vectors $\vec{\boldsymbol{t}}_1^{(\alpha)\mathrm{T}}$ and $\vec{\boldsymbol{t}}_2^{(\alpha)\mathrm{T}}$, and each friction multiplier has two components $\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)} = \begin{bmatrix}\lambda_{\mathrm{t}1}^{(\alpha)} & \lambda_{\mathrm{t}2}^{(\alpha)}\end{bmatrix}^\mathrm{T}$. The nonzero blocks of the tangent Jacobian are

$$\mathbf{G}_{\mathrm{t}[a]}^{(\alpha)} = \begin{bmatrix} -\vec{\boldsymbol{t}}_1^{(\alpha)\mathrm{T}} & -\left(\vec{\boldsymbol{d}}_{[a]}^{(\alpha)} \times \vec{\boldsymbol{t}}_1^{(\alpha)}\right)^\mathrm{T} \\ -\vec{\boldsymbol{t}}_2^{(\alpha)\mathrm{T}} & -\left(\vec{\boldsymbol{d}}_{[a]}^{(\alpha)} \times \vec{\boldsymbol{t}}_2^{(\alpha)}\right)^\mathrm{T} \end{bmatrix} \tag{24}$$

$$\mathbf{G}_{\mathrm{t}[b]}^{(\alpha)} = \begin{bmatrix} \vec{\boldsymbol{t}}_1^{(\alpha)\mathrm{T}} & \left(\vec{\boldsymbol{d}}_{[b]}^{(\alpha)} \times \vec{\boldsymbol{t}}_1^{(\alpha)}\right)^\mathrm{T} \\ \vec{\boldsymbol{t}}_2^{(\alpha)\mathrm{T}} & \left(\vec{\boldsymbol{d}}_{[b]}^{(\alpha)} \times \vec{\boldsymbol{t}}_2^{(\alpha)}\right)^\mathrm{T} \end{bmatrix} \tag{25}$$

DOI: 10.1002/nme

The Jacobian blocks have dimension $\dim(\mathbf{G}^{(\alpha)}_{\mathrm{n}[a]}) = 1 \times 6$ and $\dim\left(\mathbf{G}^{(\alpha)}_{\mathrm{t}[a]}\right) = 2 \times 6$. The assembled constraint vector $\mathbf{g} = \left[g^{(1)}\ g^{(2)}\ \ldots\ g^{(N_\mathrm{c})}\right]^\mathrm{T}$ has dimension $N_\mathrm{c} \times 1$, the assembled Jacobians $\dim(\mathbf{G}_\mathrm{n}) = N_\mathrm{c} \times 6N_\mathrm{p}$ and $\dim(\mathbf{G}_\mathrm{t}) = 2N_\mathrm{c} \times 6N_\mathrm{p}$, and the Lagrange multipliers $\dim(\boldsymbol{\lambda}_\mathrm{n}) = N_\mathrm{c} \times 1$ and $\dim(\boldsymbol{\lambda}_\mathrm{t}) = 2N_\mathrm{c} \times 1$. The resulting MLCP thus consist of a sparse saddle-point matrix of size $\dim(\mathbf{H}) = (6N_\mathrm{p} + 3N_\mathrm{c}) \times (6N_\mathrm{p} + 3N_\mathrm{c})$ and $\dim(\mathbf{z}) = 6N_\mathrm{p} + 3N_\mathrm{c}$ variables. Improving the approximation of the friction cone from a box to a polyhedron result in similar system but with more auxiliary variables.

Contacts are separated into *continuous* and *impacting contacts*. Impacting contacts are those not occurring with last time step. The effect of impacts are treated in an *impact stage*, solving an MLCP based on the Newton impact law $\mathbf{G}_\mathrm{n}\mathbf{v}^+_{[ab]} = -e\mathbf{G}_\mathrm{n}\mathbf{v}^-_{[ab]}$ with restitution coefficient e, before proceeding with the main solve and time integration. See Appendix A for further details.

There are three type of solvers for MLCPs: pivoting methods, Newton (line search) methods, and iterative methods with different requirements on the matrix $\mathbf{H}$, e.g., being positive-definite. Regularization is needed for handling ill-posed or ill-conditioned problems, e.g., due to existence of multiple solutions to the contact problem, contact constraint degeneracy, and occurrence of large mass ratios. Regularization terms are entered as positive diagonal perturbations in $\mathbf{H}$ and stabilization terms on the form $\alpha\mathbf{g} + \beta\mathbf{G}_\mathrm{n}\mathbf{v}$ are introduced in the $\mathbf{b}$ vector which are required for restoring constraint violations if they should occur. These terms bring solvability and numerical stability and corresponds to a solving a slightly different physical system than the original one. The systems converge to (12)-(14) weakly in the limit of zero regularization. See, e.g., Ref. [18].

2.4. Semi-smooth discrete element method

In this section, we present a *semi-smooth DEM* that shares the features of large time step integration with NDEM and the viscoelastic interaction forces of SDEM. Elastic contact models for NDEM by regularization corresponding to linear springs were recently presented and examined in Ref. [19]. In the present paper, we describe a regularization that maps to nonlinear springs according to the Hertz contact law. This is realized by considering the NDEM as the stiff limit of SDEM with particular energy potentials and dissipation functions, for generating the constraint regularization and stabilization terms to the MLCP form in Section 2.3.2. The theoretical basis is thoroughly described in Ref. [18] and has previously been applied to large time step simulation of large-scale granular flows for geometric design of pelletizing drum [20].

First, observe that the normal contact force in Equation (11) follows from $\boldsymbol{f}_\mathrm{n} = -\nabla_x U_\mathrm{n} - \nabla_v \mathcal{R}_\mathrm{n}$ with the following potential energy function and Rayleigh dissipation functions

$$U_\mathrm{n} = \frac{1}{2\varepsilon_\mathrm{n}} \bar{\mathbf{g}}^\mathrm{T} \bar{\mathbf{g}} \tag{26}$$

$$\mathcal{R}_\mathrm{n} = \frac{1}{2\gamma_\mathrm{n}} \left(\bar{\mathbf{G}}_\mathrm{n}\mathbf{v}\right)^\mathrm{T} \left(\bar{\mathbf{G}}_\mathrm{n}\mathbf{v}\right) \tag{27}$$

with $\bar{g}_{(\alpha)} = g^{e_\mathrm{H}}_{(\alpha)}$, $\bar{\mathbf{G}}^{(\alpha)}_\mathrm{n} = e_\mathrm{H} g^{e_\mathrm{H}-1}_{(\alpha)} \mathbf{G}^{(\alpha)}_\mathrm{n}$, $\varepsilon^{-1}_\mathrm{n} = k_\mathrm{n}/e_\mathrm{H}$, $\gamma^{-1}_\mathrm{n} = k_\mathrm{n}c/e_\mathrm{H}$ and exponent $e_\mathrm{H} = 5/4$ for the nonlinear Hertz model with stiffness and damping parametrized as in Section 2.2. Note the relation $\dot{\bar{\mathbf{g}}} = \bar{\mathbf{G}}\mathbf{v}$. Second, instead of treating contact forces explicitly as gradients of potentials, these are introduced as constraint forces $\bar{\mathbf{G}}^\mathrm{T}\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}$ are introduced as auxiliary variables obeying $\varepsilon\boldsymbol{\lambda} + \bar{\mathbf{g}} + \tau\bar{\mathbf{G}}\mathbf{v} = 0$, with regularization and damping parameters ε and τ. Third, employing a time-discrete formulation of the variational principle leads to the following first order fixed time step integration method, coined SPOOK [18],

$$\begin{aligned} &\bar{\mathbf{H}}\mathbf{z} + \bar{\mathbf{b}} = \mathbf{w}_+ - \mathbf{w}_- \\ &0 \leqslant \mathbf{z} - \mathbf{l} \perp \mathbf{w}_+ \geqslant 0 \\ &0 \leqslant \mathbf{u} - \mathbf{z} \perp \mathbf{w}_- \geqslant 0 \end{aligned} \tag{28}$$

DOI: 10.1002/nme

where

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{M} & -\bar{\mathbf{G}}_{\mathrm{n}}^{\mathrm{T}} & -\bar{\mathbf{G}}_{\mathrm{t}}^{\mathrm{T}} \\ \bar{\mathbf{G}}_{\mathrm{n}} & \boldsymbol{\Sigma} & 0 \\ \bar{\mathbf{G}}_{\mathrm{t}} & 0 & \boldsymbol{\Gamma} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{v}_{i+1} \\ \boldsymbol{\lambda}_{\mathrm{n},i+1} \\ \boldsymbol{\lambda}_{\mathrm{t},i+1} \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} -\mathbf{M}\mathbf{v}_{\mathrm{s}} \\ \frac{4}{h}\boldsymbol{\Upsilon}\bar{\mathbf{g}} - \boldsymbol{\Upsilon}\bar{\mathbf{G}}_{\mathrm{n}}\mathbf{v}_i \\ 0 \end{bmatrix} \tag{29}$$

and regularization and stabilization coefficients

$$\boldsymbol{\Sigma} = \frac{4}{h^2}\frac{\varepsilon_{\mathrm{n}}}{1+4\frac{\tau_{\mathrm{n}}}{h}}\mathbf{1}_{N_c\times N_c}, \quad \boldsymbol{\Gamma} = \frac{\gamma_{\mathrm{t}}}{h}\mathbf{1}_{2N_c\times 2N_c}, \quad \boldsymbol{\Upsilon} = \frac{1}{1+4\frac{\tau_{\mathrm{n}}}{h}}\mathbf{1}_{N_c\times N_c} \tag{30}$$

The parameter τ_{n} controls the dissipation rate in the normal force. We use $\tau_{\mathrm{n}} = \max\left(n_{\mathrm{s}}h, \gamma_{\mathrm{n}}^{-1}\right)$ with $n_{\mathrm{s}} = 2$ and the clamping is enforced to guarantee numerical stability at large time step integration. Similarly, the Coulomb friction is modeled as the stiff limit of a Rayleigh dissipation function for the relative velocity in the contact tangent plane with and dissipation rate γ_{t} and limits on the tangential Lagrange multiplier and $\bar{\mathbf{G}}_{\mathrm{t}} = \mathbf{G}_{\mathrm{t}}$. An algorithm for the semi-smooth DEM with projected Gauss-Seidel is presented in Appendix B.

Both the SDEM and NDEM can be recovered from the semi-smooth DEM. NDEM follows directly from taking the stiff limit $\varepsilon, \gamma \to 0$. This implies $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{\Upsilon} \to 0$, and Equations (28)–(29) reduces exactly to Equations (19)–(20). SDEM, on the other hand, is recovered in the limit of time step $h \to 0$ with $\varepsilon, \gamma, \tau$ fix. Observe that in this limit

$$\boldsymbol{\Sigma} \to \frac{\varepsilon_{\mathrm{n}}}{\tau_{\mathrm{n}}h}\mathbf{1}_{N_c\times N_c}, \quad \boldsymbol{\Upsilon} \to \frac{h}{4\tau_{\mathrm{n}}}\mathbf{1}_{N_c\times N_c}, \quad \frac{4}{h}\boldsymbol{\Upsilon}\bar{\mathbf{g}} - \boldsymbol{\Upsilon}\bar{\mathbf{G}}_{\mathrm{n}}\mathbf{v}_i \to \tau_{\mathrm{n}}^{-1}\bar{\mathbf{g}} \tag{31}$$

Assuming smoothness in $\mathbf{v}$, it is possible to eliminate $\boldsymbol{\lambda}_{\mathrm{n}}$ and $\boldsymbol{\lambda}_{\mathrm{t}}$ in the MLCP in Equation (28) whereby it simplifies to the stepping rule $\mathbf{M}\mathbf{v}_{i+1} = \mathbf{M}\mathbf{v}_{\mathrm{s},i+1} + \bar{\mathbf{G}}_{\mathrm{n}}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{n}} + \bar{\mathbf{G}}_{\mathrm{t}}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{t}}$ where

$$\bar{\mathbf{G}}_{\mathrm{n}}^{(\alpha)\mathrm{T}}\lambda_{\mathrm{n}}^{(\alpha)}/h \approx \bar{\mathbf{G}}_{\mathrm{n}}^{(\alpha)\mathrm{T}}\left(\frac{\tau h}{\varepsilon}\bar{\mathbf{G}}_{\mathrm{n}}^{(\alpha)}\mathbf{v}_{i+1} + \frac{\tau h}{\varepsilon\tau}\bar{g}_{(\alpha)}\right)/h \approx k_{\mathrm{n}}\left(g_{(\alpha)}^{3/2} + cg_{(\alpha)}^{1/2}\dot{g}_{(\alpha)}\right)\vec{\boldsymbol{n}}^{(\alpha)} \equiv \boldsymbol{f}_{\mathrm{n}}^{(\alpha)} \tag{32}$$

$$\bar{\mathbf{G}}_{\mathrm{t}}^{(\alpha)\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{t}}^{(\alpha)}/h \approx \mathrm{proj}_{\mu|\mathbf{f}_{\mathrm{n}}^{(\alpha)}|}\left(\bar{\mathbf{G}}_{\mathrm{t}}^{(\alpha)\mathrm{T}}\left(\gamma^{-1}\bar{\mathbf{G}}_{\mathrm{t}}^{(\alpha)}\mathbf{v}^i\right)\right) \approx \mathrm{proj}_{\mu|\mathbf{f}_{\mathrm{n}}^{(\alpha)}|}\left(k_{\mathrm{t}}\vec{\boldsymbol{u}}_{\mathrm{t},i}^{(\alpha)}\right) \equiv \boldsymbol{f}_{\mathrm{t}}^{(\alpha)} \tag{33}$$

in which we have approximated $\bar{\mathbf{G}}_{\mathrm{n}}\mathbf{v}_{i+1} \approx \dot{\bar{\mathbf{g}}}(t_i)$ and $\bar{\mathbf{G}}_{\mathrm{t}}^{\mathrm{T}}\bar{\mathbf{G}}_{\mathrm{t}}\mathbf{v}_{i+1} \approx -\vec{\boldsymbol{u}}_{\mathrm{t},i}$ and used that $k_{\mathrm{n}} = e_{\mathrm{H}}\varepsilon^{-1}$, $k_{\mathrm{t}} = \gamma_{\mathrm{t}}^{-1}$, $c = \tau e_{\mathrm{H}}$ and $e_{\mathrm{H}} = 5/4$. Hence, the multipliers can be removed by substitution, and we obtain an identical time stepping scheme as for SDEM when applying a semi-implicit Euler discretization. We thus conclude that the results of semi-smooth DEM and SDEM coincide in the limit of small time steps.

The elastic properties are present also for large time steps. Specifically, in the quasi-static regime, the regularization corresponds directly to the Hertz model of elastic contacts. Assume $\bar{\mathbf{G}}_{\mathrm{n}}\mathbf{v}_{i+1} = \bar{\mathbf{G}}_{\mathrm{n}}\mathbf{v}_i \approx 0$ and Equations (28)–(29) imply that $\boldsymbol{\Sigma}\boldsymbol{\lambda}_{\mathrm{n}} = (4/h)\boldsymbol{\Upsilon}\bar{\mathbf{g}}$ and the constraint force, $\vec{\boldsymbol{f}}_{\mathrm{n}}^{(\alpha)} = \bar{\mathbf{G}}_{\mathrm{n[b]}}^{(\alpha)\mathrm{T}}(1,1:3)\lambda_{\mathrm{n}}^{(\alpha)}/h$ on body b from contact α with body a become

$$\vec{\boldsymbol{f}}_{\mathrm{n}}^{(\alpha)} = \Sigma^{-1}\left(4/h^2\right)\Upsilon\bar{\mathbf{G}}_{\mathrm{n[b]}}^{\alpha\mathrm{T}}(1,1:3)\bar{g}_{(\alpha)} = k_{\mathrm{n}}g_{(\alpha)}^{3/2}\vec{\boldsymbol{n}}^{(\alpha)} \tag{34}$$

which equals the smooth normal force in Equation (11) in the case of zero relative contact velocity.

3. COMPUTATIONAL PROPERTIES

In this section, we make a theoretical comparison between of the computational properties of SDEM and NDEM. The semi-smooth DEM has identical computational complexity as the nonsmooth and is therefore not covered separately. In particular, we consider how the computational effort scale with the number of particles and dependency on the dynamical state and geometric shape of the systems. The computational time, τ_{DEM}, for simulating a process lasting for τ_{real} units of real time is

$$\tau_{\mathrm{DEM}} = \frac{\Omega_{\mathrm{DEM}}}{h_{\mathrm{DEM}}}\tau_{\mathrm{real}} \tag{35}$$

DOI: 10.1002/nme

where Ω_{DEM} is the required computational time for advancing the simulation one time step of size h_{DEM}. The computational effort is $\tau_{\mathrm{DEM}}/\tau_{\mathrm{real}}$. The main steps of the simulation loop are *contact detection*, *solve*, and *simulation management and I/O*. The contact detection step involves a *broad phase*, where contact candidates are found, and a *narrow phase*, where contact positions and overlap magnitudes are computed. The *solve phase* refer to the numeric integration of the equations of motion.

3.1. Smooth discrete element method

For SDEM with explicit or semi-implicit time integration, the solve stage consists of simple evaluation and summation of forces and update of velocities and positions involving only a few multiplication and additions per particle. The computational bottleneck lies in the contact detection stage. A typical SDEM simulation spends roughly 80% of the computational time on contact detection [21] and, assuming no parallelization, scales at best linearly with N_{p} and as worst as $N_{\mathrm{p}} \log(N_{\mathrm{p}})$ if temporal coherence cannot be exploited. For numerical stability, the time step size is limited by $h_{\mathrm{SDEM}} \lesssim \sqrt{m/k}$ and the computational time is thus

$$\tau_{\mathrm{SDEM}} = \sqrt{\frac{k}{m}} K_{\mathrm{SDEM}} N_{\mathrm{p}} \tau_{\mathrm{real}} \tag{36}$$

where we have split $\Omega_{\mathrm{SDEM}} = K_{\mathrm{SDEM}} N_{\mathrm{p}}$, and K_{SDEM} is the average computational time per time step and particle. The factor K_{SDEM} depends on hardware, software, geometric shape of particles, and collision detection algorithm and its implementation. With a conventional desktop computer (specified in Appendix C) running the SDEM software library LIGGGTHS [22] with spherical particles, it was measured [23] $K_{\mathrm{SDEM}} \sim 10^{-6}$.

3.2. Nonsmooth discrete element method

For nonsmooth DEM, the solve stage dominates the computational time, e.g., 88% of the time was reported in [24]. Typically, the computational time for one time step is much larger than for SDEM. The benefit instead comes from integration with fewer and larger time steps.

The time step limit in NDEM is set by the characteristic relative velocities or gravity acceleration. In one integration time step, these should not cause an impact overlap larger than some fraction ϵ of the particle diameter, i.e., $h_{\mathrm{NDEM}} \leqslant \epsilon d/v_{\mathrm{n}}$ and $h_{\mathrm{NDEM}} \leqslant \sqrt{2\epsilon d/g}$. This introduce ϵ as an error tolerance. The computational time become

$$\tau_{\mathrm{NDEM}} = \frac{\Omega_{\mathrm{NDEM}}}{\min\left(\epsilon \frac{d}{v_{\mathrm{n}}}, \sqrt{\frac{2\epsilon d}{g}}\right)} \tau_{\mathrm{real}} \tag{37}$$

Contrary to SDEM, the computational time per time step Ω_{NDEM} is in general not a linear function of the number of particles (or number of contacts). The equations of motions for NDEM are DVI whose computational properties are largely open questions [25], lacking proof of existence and uniqueness of solutions as well as of general proof of convergence and numerical stability of most solution algorithms. Both theoretical and empirical analyses are complicated by the fact that the solutions are in general discontinuous with respect to change in initial data. The computational scaling depends ultimately on the choice of numerical solver as well as of the dynamical and geometric state of the system.

A theoretical upper limit of the computational efficiency can be found by considering the linear system $Hz + b = 0$ of the MLCP in Equation (19) and ignoring the complementarity conditions. The matrix size is of the order $3N_{\mathrm{c}} \times 3N_{\mathrm{c}}$. The number of contacts is related to the number of particles as $N_{\mathrm{c}} \sim (n_{\mathrm{c}}/2)N_{\mathrm{p}}$, assuming each particle on average has n_{c} neighbors, which has a number ranging between $2, 3-8$, and $6-12$ depending on the system dimensionality ranging between 1D (chain), 2D (plane), and 3D (bulk) and on the packing density. The matrix H is block-sparse and the linear system can in many cases be solved efficiently using direct or iterative methods exploiting the block-sparseness. The best theoretical scaling for direct solvers is provided by algorithms using

DOI: 10.1002/nme

nested dissection for reordering and factorization (the computationally most expensive part). This has time complexity of $\Omega_{\text{NDEM}} \sim \mathcal{O}\left(N_{\text{c}}^{(1+n_{\text{d}})/2}\right)$, for $n_{\text{d}} = 1-3$ dimensionality. For the case of a 1D column of particles (chain), the computational time scales linearly with number of contacts and particles, $\Omega_{\text{NDEM}} \sim \mathcal{O}(N_{\text{c}})$, but for higher dimensionality, the scaling is in general superlinear. Using the AgX Multiphysics Toolkit [26], we measured $\Omega_{\text{NDEM}} = K_{\text{NDEM}} N_{\text{c}}$ with $K_{\text{NDEM}} = 10^{-5}$ s for the 1D column using a block-sparse LU solver. When machine precision is not required, an iterative solver may give the solution to a given tolerance faster. *Multigrid* methods and *conjugate gradient* [27, 28] method may scale almost linearly with N_{c}, but many issues remain in transferring these results to irregular contact networks with complementarity conditions.

3.2.1. Efficiency of the Gauss-Seidel solver. The Gauss-Seidel algorithm (GS) is commonly considered a poor choice for solving linear systems. Still, it is of common use for integrating NDEM simulations. Although the asymptotic convergence is slow[‡] the initial convergence may be fast, and the algorithm allows changes in the active set without restarting, as the conjugate gradient does. The projected block Gauss-Seidel algorithm solves the local contact problem well but approaches to the global solution in a diffusive manner with increasing number of iterations. The residual from truncating at finite number of iterations result in *numeric elasticity* [29] with an effective sound velocity $v_{\text{GS}} = \sqrt{N_{\text{it}}} d / h_{\text{NDEM}}$. Consequently, for a system with side length l and particle size d, the required number of iterations for establishing the contact force network required for maintaining perfect rigidity scales as $N_{\text{it}} \propto (l/d)^2 \sim N_{\text{c}}^{2/n_{\text{d}}}$. Each iteration step involves solving each of the N_{c} two-body contact problem. The resulting computational time complexity thus become

$$\Omega_{\text{NDEM}}^{\text{GS}} = K_{\text{NDEM}}^{\text{GS}} N_{\text{it}} N_{\text{c}} \tag{38}$$

where $K_{\text{NDEM}}^{\text{GS}}$ is the average computational time for solving each two-body problem. Consequently, projected Gauss-Seidel scales $\Omega_{\text{NDEM}} \sim \mathcal{O}\left(N_{\text{c}}^{1+2/n_{\text{d}}}\right)$. In our implementation, described in Appendix B, we measure the scaling coefficient to $K_{\text{NDEM}}^{\text{GS}} = 10^{-6}$ s with a conventional desktop computer (specified in Appendix C).

3.3. Comparing smooth and nonsmooth discrete element methods

To compare the difference in computational efficiency between the SDEM and NDEM, we consider the ratio of their computational time from Equations (36) and (37)

$$\frac{\tau_{\text{NDEM}}}{\tau_{\text{SDEM}}} = \sqrt{\frac{\max\left(\epsilon^{-2} m v_{\text{n}}^2, 2\epsilon^{-1} m g d\right)}{k d^2}} \frac{\Omega_{\text{NDEM}}}{K_{\text{SDEM}} N_{\text{p}}} \tag{39}$$

When this ratio is smaller than one, the NDEM is more efficient and vice versa. The square root term is the effect of different time step size h_{SDEM} and h_{NDEM}. The relative efficiency of the NDEM increase with particle size and stiffness and decrease with increasing relative velocity, and mass. The second term on the right hand side is the effect of different computational scaling of the solvers for SDEM and NDEM.

Example. Consider particles of size $d = 0.01$ m, mass density 2500 kg/m, Young's modulus E = 10 GPa, and average normal contact velocity $v_{\text{n}} \lesssim 0.1 m/s$. This means $m \sim 10^{-3}$ kg and $k \sim 10^9 \text{N/m}^{3/2}$. The resulting smooth and nonsmooth time step sizes become $h_{\text{SDEM}} \lesssim \sqrt{m/k} = 10^{-6}$s and $h_{\text{NDEM}} = 10^{-3}$ s for impact overlap tolerance set to $\epsilon = 0.01$. The time step ratio (the square root in Equation (39)) becomes $h_{\text{NDEM}} / h_{\text{SDEM}} \approx 10^{-3}$. In the case of 1D dimensionality, e.g., a column of particles, and employing a direct solver for the DDEM with $K_{\text{NDEM}} = 10^{-5}$ (measured from AgX) and a SDEM implementation with $K_{\text{SDEM}} \sim 10^{-6}$ (measured from LIGGGHTS), we obtain the time ratio $\frac{\tau_{\text{NDEM}}}{\tau_{\text{SDEM}}} \approx 10^{-2}$. Hence, for the column configuration, the NDEM is 100 times

[‡]For the Gauss-Seidel algorithm, the residual decays asymptotically as the logarithm of the spectral radius of the iteration matrix.

DOI: 10.1002/nme

faster than the SDEM irrespective of the number of particles in the column. If we instead consider an iterative Gauss-Seidel solver, with scaling $K_{\mathrm{NDEM}}^{\mathrm{GS}} = 10^{-6}$ for the NDEM, we get the time ratio $\frac{\tau_{\mathrm{NDEM}}}{\tau_{\mathrm{SDEM}}} \approx 10^{-3} N_{\mathrm{it}}$, which depend on the required number of iterations and thus indirectly on the number of particles and error tolerance. For small systems and large error tolerance, such that $N_{\mathrm{it}} < 10^3$ is sufficient, the NDEM will be the faster simulation method, while the SDEM will be faster for tall columns and small error tolerance.

4. SIMULATIONS

The computational properties presented in Section 3 need to be verified and complemented with empirical data from numerical simulations with NDEM. It is beyond the scope of this paper to make a thorough analysis for a range of solvers and the study is limited to projected Gauss-Seidel solver. In particular, we need to determine how the required number of iterations N_{it} for an error tolerance ϵ depend on the geometric shape and dynamic state of the system. For this purpose, we perform simulations for a number of different test systems.

We use the semi-smooth DEM as outlined in Section 2.4 and implemented according to the algorithm in Appendix B. The reason for this choice is that the computational complexity is identical to NDEM and that it supports nonsmooth impact and frictional stick slip phenomena as well as elastic contact properties, in accord with Hertz law. The resulting forces and flows can therefore be compared with SDEM at the same time as the result in terms of computational efficiency are representative for NDEM.

4.1. Metrics

Specific metrics are chosen for analyzing the quality of different DEM for granular matter. On a microscopic level, we consider individual contacts. For any quantity, say f, we denote the mean value by $\langle \epsilon_f \rangle$ and the corresponding standard deviation by σ_f. In particular, we study the normalized contact *overlap* ϵ_{c}, *slide error* ϵ_{sl} and *slide direction error* ϵ_{dir}. For a contact α between two bodies a and b with diameter d, these are defined

$$\epsilon_{\mathrm{c}}^{(\alpha)} = g^{(\alpha)}/d \tag{40}$$

and if the contact is in slide mode $\epsilon_{\mathrm{sl}}^{(\alpha)} = \left(\left| \vec{\boldsymbol{f}}_{\mathrm{t}}^{(\alpha)} \right| - \mu \left| \vec{\boldsymbol{f}}_{\mathrm{n}}^{(\alpha)} \right| \right) / \mu \left| \vec{\boldsymbol{f}}_{\mathrm{n}}^{(\alpha)} \right|$ and $\epsilon_{\mathrm{dir}} = (\theta_{\mathrm{dir}} - \pi)/\pi$, where $\theta_{\mathrm{dir}}^{(\alpha)} = \arccos\left(\vec{\boldsymbol{u}}_{\mathrm{t}}^{(\alpha)}, \vec{\boldsymbol{f}}_{\mathrm{t}}^{(\alpha)} \right)$ is the angle between the directions of sliding and friction force. In the NDEM and semi-smooth DEM, the forces on particle a are $\vec{\boldsymbol{f}}_{\mathrm{n}}^{(\alpha)} = \mathbf{G}_{\mathrm{n,[a]}}^{(\alpha)\mathrm{T}}(:, 1:3)\lambda_{\mathrm{n}}^{(\alpha)}/h$ and $\vec{\boldsymbol{f}}_{\mathrm{t}}^{(\alpha)} = \mathbf{G}_{\mathrm{t,[a]}}^{(\alpha)}(:, 1:3)^{\mathrm{T}} \boldsymbol{\lambda}_{\mathrm{t}}^{(\alpha)}/h$. We use the velocity threshold $u_{\mathrm{thr}} \equiv 0.01\sqrt{gd/2}$ to separate between stick and slide modes $\left| \vec{\boldsymbol{u}}_{\mathrm{t}}^{(\alpha)} \right| \geqslant u_{\mathrm{thr}}$. We also discard the slide errors for the weakest contacts where $\left| \vec{\boldsymbol{f}}_{\mathrm{n}}^{\alpha} \right| \leqslant f_{\mathrm{thr}} \equiv 0.01\, mg$

The contact forces form *force networks*. These are weighted graphs with the particles as nodes and contact forces as edges. We use the normal force magnitude for the edge weight. The topology and force distributions of the networks are analyzed.

Macroscopic fields of distribution of mass, stresses, and strains are computed using *coarse graining* (or *homogenization*) [30]. In particular, we compute the mass density field $\rho(\vec{\mathbf{x}})$, velocity vector field $\vec{\mathbf{v}}(\vec{\mathbf{x}})$, strain rate tensor field $\dot{\boldsymbol{\gamma}}(\vec{\mathbf{x}})$ with norm $\dot{\gamma} = \sqrt{\mathrm{tr}\left(\dot{\boldsymbol{\gamma}}^{\mathrm{T}}\dot{\boldsymbol{\gamma}}\right)}$, stress tensor field $\boldsymbol{\sigma}(\vec{\mathbf{x}})$, and pressure field $p(\vec{\mathbf{x}}) = \frac{1}{3}\mathrm{tr}(\boldsymbol{\sigma})$. From these, we compute the *inertial number field* $I(\vec{\mathbf{x}}) \equiv \dot{\gamma}(\vec{\mathbf{x}})d/\sqrt{p(\vec{\mathbf{x}})/\rho(\vec{\mathbf{x}})}$. The inertial number is a measure of whether a granular system is in quasistatic resting regime ($0 \leqslant I \ll 1$), dense flow regime ($I \lesssim 1$) or gaseous regime ($I \gtrsim 1$).

DOI: 10.1002/nme

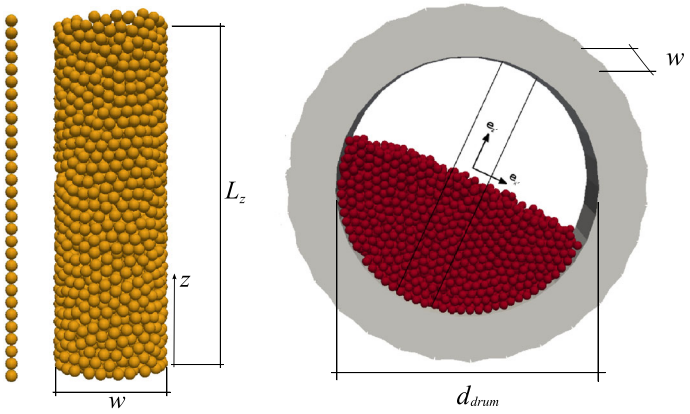

Figure 2. The test systems are cylindrical container and rotating drum.

4.2. *Test systems*

Two class of test systems are considered: *cylindrical containers* of different size with resting granular material and *rotating drums* of different speed with dense material flow. These systems, depicted in Figure 2, are chosen as they represent the different dynamic regimes for which both SDEM and NDEM are applicable and because they are common both in the scientific literature and in real world applications. We use spherical particles with diameter $d = 0.01$m, mass m $= 10^{-3}$kg corresponding to a mass density of 2500kg/m^3 and Young's modulus E $= 5 \cdot 10^6$ Pa. The equivalent spring coefficient is $k_{\mathrm{n}} = 0.5 \cdot 10^6N/m^{3/2}$. We use time step $h_{\mathrm{NDEM}} = 10^{-2}$s, which is 500 times larger than what is required for SDEM $h_{\mathrm{SDEM}} = 5 \cdot 10^{-5}$s. Gravity is set to g $= 9.8$m/s^2. With these parameters, the particles are practically rigid in the performed tests and any particle overlaps larger than $0.01d$ will be due to errors and not due to material elasticity. The friction coefficient is set to $\mu = 0.7$ and we use zero restitution coefficient $e = 0$ for both particle-particle contacts and for particle-surface contacts. Simulations are run with fixed number of iterations ranging as N_{it} ranging between 10 and 500.

4.2.1. Cylinder container. Simulation is performed with particles in cylindrical container with diameter $\Phi = 1, 3, 6, 9,$ and$15d$. The cylinders are filled with different number of particles N_{p} ranging between 5 and 100 in the $\Phi = 1d$ case, 100–30K in the $\Phi = 3 - 15d$ cases. The cylindrical geometry is modeled by 20 rectangular faces. The states are initialized as follows. The particles are placed in a regular cubic grid with slight perturbation. The particles are left to relax with gravity acceleration g gradually increasing from 0.01 to $9.82m/\mathrm{s}^2$ over 60 s, sufficient to reach a stationary state. The simulation is then run for 5 s. Measurements of position, velocity, and contact data are made. Post-analysis is performed to obtain the metrics in Section 4.1. For cylindrical containers, we also compute the pressure on the walls as function of height for verification of the Janssen effect. The wall pressure $p(z)$ as function of height z is computed as the average contact pressure on a cylindrical strip of width Δz $=$ 5d of the container wall. This procedure is repeated for 440 combinations of N_{p}, N_{it}, and Φ. We treat the case $\Phi = 1d$ somewhat special. The particles are initialized in a perfect 1D column, and the contacts with walls are deactivated.

4.2.2. Rotating drum. For the rotating drum tests, the geometric shape is kept fixed with drum diameter $\Phi = 80d$ and length $8d$. Simulations are run with drum rotation speeds $\omega = 0, 0.03, 0.06, 0.13, 0.3, 0.63, 2.5$rad/s corresponding to dimensionless Froude number Fr $= 0, 0.005, 0.013, 0.025, 0.13, 0.51$, with the definition Fr $= \omega\sqrt{\Phi/2g}$. The rotation is about the symmetry axis, which is orthogonal to the direction of gravity. Simulations are run with different number of particles N_{p} ranging from 100 to 7.5K. To make the flow less dependent on the wall friction force, the drum surface is given a structural shape that of a sawtooth with 64 teeth of height $1.2d$, length

DOI: 10.1002/nme

$4d$, and tooth attack angle of 50 degrees. For each combination of ω, N_p and N_{it}, the state is initialized into a stationary flow by running a 2s simulation at $N_{\text{it}} = 500$. When changing N_{it}, this is followed by a simulation lasting $1/3$ of an evolution plus 3.5 s. Measurements are then made during 10 consecutive time steps.

5. RESULTS AND DISCUSSION

In this section, we present results of the simulation tests outlined in Section 4. From this, we deduce a formula for the required number of iterations for projected Gauss-Seidel solver for NDEM (and semi-smooth DEM) depending on a desired error tolerance and system geometric shape and dynamics state. Supplementary data of mean penetrations, slide and friction direction, number of contacts, force networks, stress fields and velocity fields depending on number of particles, and different number of iterations are provided on the web page http://umit.cs.umu.se/granular/dem/.

5.1. Cylinder container

Sample force networks from the cylindrical container test with diameter $\Phi = 9d$ is shown in Figure 3 for $N_{\text{it}} = 10, 100$, and 500 and $N_{\text{p}} = 7.5$K. Two main observations can be made. First, too few iterations cause artificial compression of the material. For $N_{\text{it}} = 10$, the column collapses to a height less than 40 % of the $N_{it} = 500$ solution.

Second, with too few iterations, the strong force chains that are a characteristic feature of granular materials do not appear. Instead, the force distributes as the hydrostatic pressure in a fluid, i.e., increases linearly with depth from the top surface. When increasing the number of iterations, strong force chain structures emerge, and with this, the pressure force saturate and become independent of depth in the column. This is the well-known *Janssen effect* of granular materials, which is due to an arching effect of the force chains whereby the container walls carry part of the weight of the material [31]. The Janssen force profile along the center axis for different N_{it} is shown in Figure 4 for the $9d$ container with 7.5K particles. The pressure is normalized by the pressure $p_0 = 990$ Pa at zero height with $N_{\text{it}} = 500$. It suffices with $N_{\text{it}} = 50$ to capture the force saturation effect but at least $N_{\text{it}} = 100$ for the correct force level of saturation.

The required number of iterations for keeping the mean penetration $\langle \epsilon_{\text{c}} \rangle$ error below a given threshold is estimated from simulations and denoted by N_{it}^{ϵ}. The result for the $\Phi = 1d$ container is shown in Figure 5 for different values of $\langle \epsilon_{\text{c}} \rangle$ and for the cylinder size $\Phi = 3, 6, 9$, and $15d$ in Figure 6, for $\langle \epsilon_{\text{c}} \rangle \leqslant 0.05$ only. In the $1d$ case, the required number of iterations grow roughly linearly with the number of particles, and we estimate the rate of convergence to $N_{\text{it}}^{\epsilon_1} / N_{\text{it}}^{\epsilon_2} = \epsilon_2 / \epsilon_1$. In the $\Phi = 3, 6, 9$, and $15d$ cases in Figure 6, we observe a *saturation effect* in the required number of iterations. Once the number of particles in the container reach above a certain number, the required number of iterations stop to increase. We discuss this effect further in Section 5.3. For the $3d$ container with $N_{\text{p}} \geqslant 1$K, we failed to create stable initial states using the described procedure.

5.2. Rotating drum

Also for the rotating drum, the material behaves more as a compressive fluid than a granular material with too few iterations. Sample contact force networks are displayed in Figure 7 for $N_{\text{p}} = 7.5$K and $\omega = 0.63$ rad/s. The velocity field in the drum cross-section is presented in Figure 8 and the flow profile $v_{x'}(z')$ in Figure 9 in the coordinates indicated in Figure 8. With too few iteration, $N_{\text{it}} \leqslant 25$, the solutions show significant artificial compression ($> 10\%$ decrease in height), and the velocity profile deviates significantly from the ones from high iterations. For notational clarity, we denote the radius by $R/(\Phi/2)$. For high iterations, the flow has the expected two phases: a *plug flow* zone [32], which is a thick layer of material co-rotating rigidly with the drum wall, and a *shear flow* zone. The thickness of the plug flow layer is around 0.3 for $N_{\text{it}} \geqslant 150$ and then decrease continuously with decreasing number of iterations until all material is in shear flow. The variations in the velocity profile at high iterations, $N_{\text{it}} \geqslant 50$, is presumably because the flow is not entirely stationary but

DOI: 10.1002/nme

has a pulsation due to series of avalanches on the surface. For reference, the flow profile for rigid co-rotation with the drum is included as well as a solution computed with time step $h = 10^{-4}$ s and $N_{\text{it}} = 500$.

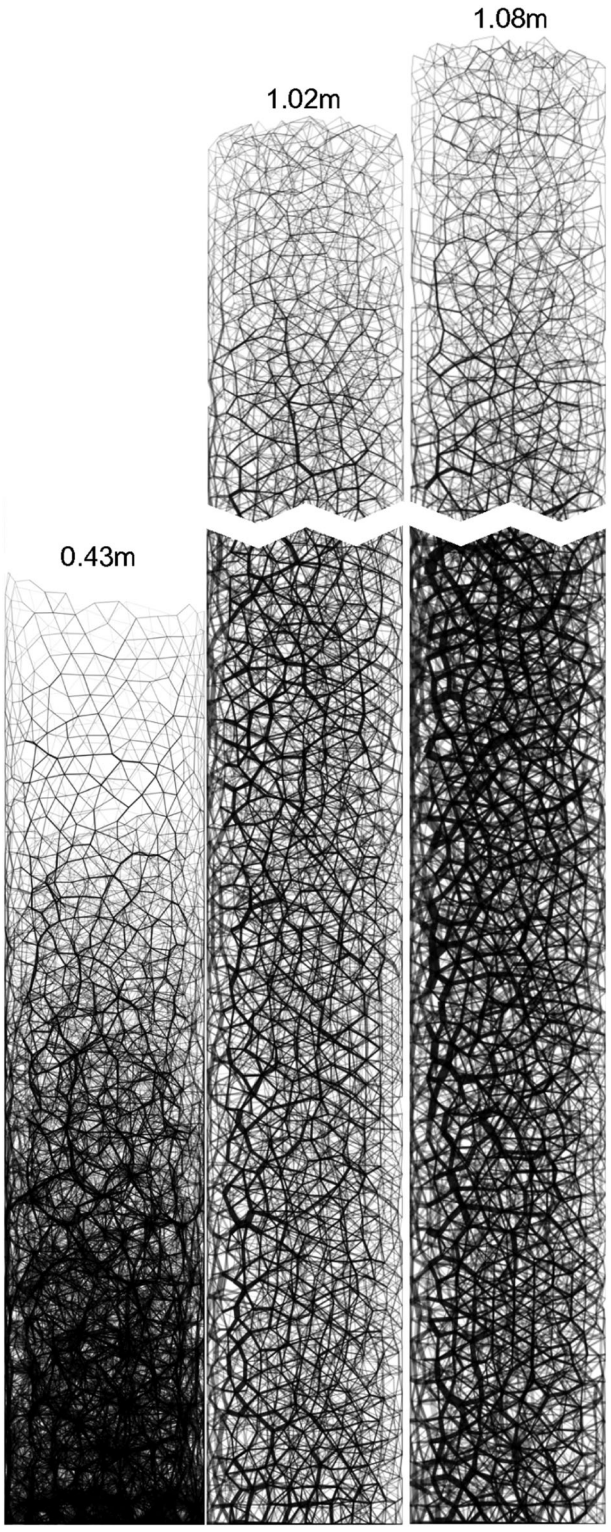

Figure 3. Sample force networks in $9d$ columns for $N_{\text{it}} = 10, 100$, and 500.

DOI: 10.1002/nme

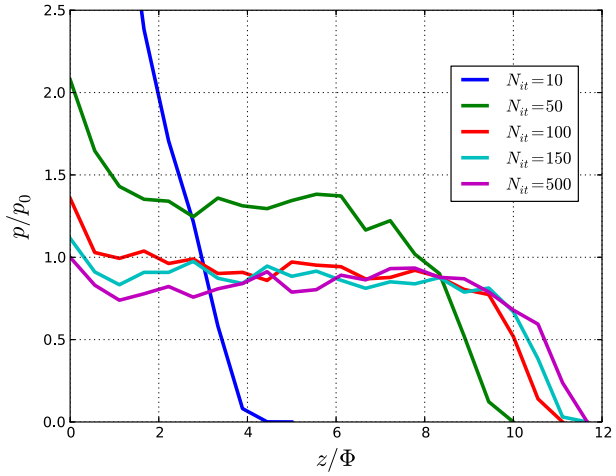

Figure 4. The Janssen force profile in a $9d$ column with 7.5K particles and different number of iterations.

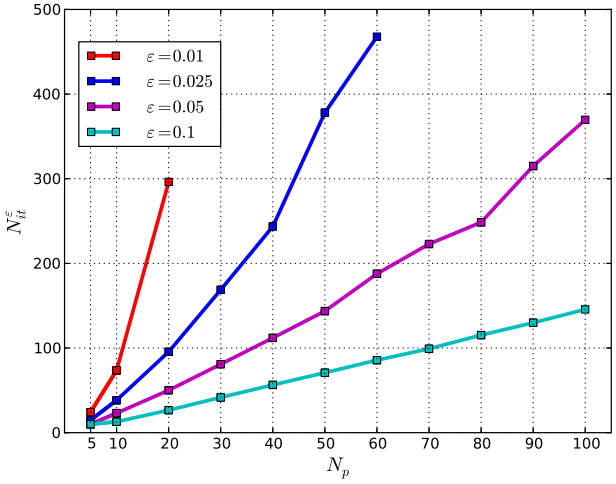

Figure 5. The required number of iterations depending on error and number of particles in a $1d$ column.

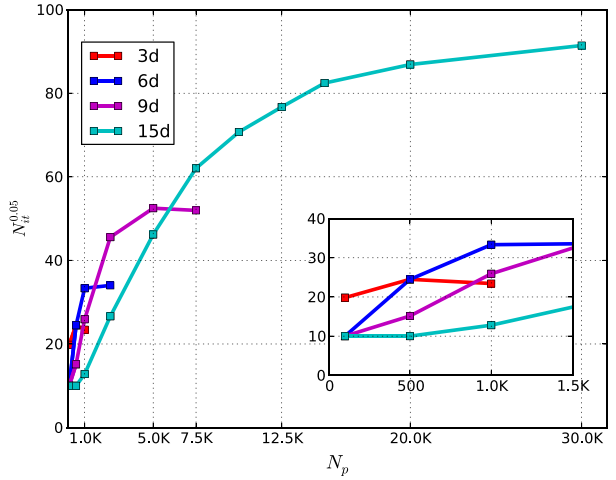

Figure 6. The required number of iterations for $\langle \epsilon_c \rangle \leqslant 0.05$ in the $3, 6, 9,$ and, $15d$ containers.

DOI: 10.1002/nme

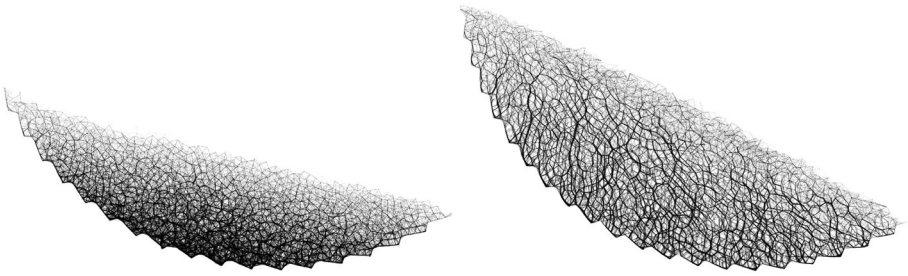

Figure 7. Force network in drum rotating with $\omega = 0.63$ rad/s, $N_{\mathrm{p}} = 7.5$K for $N_{\mathrm{it}} = 10$ (left) and 500 (right).

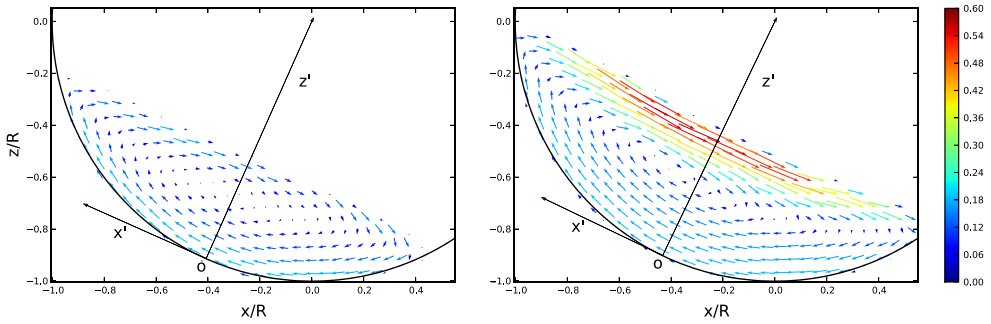

Figure 8. The velocity field in the drum rotating with $\omega = 0.63$ rad/s for $N_{it} = 10$ (left) and 500 (right).

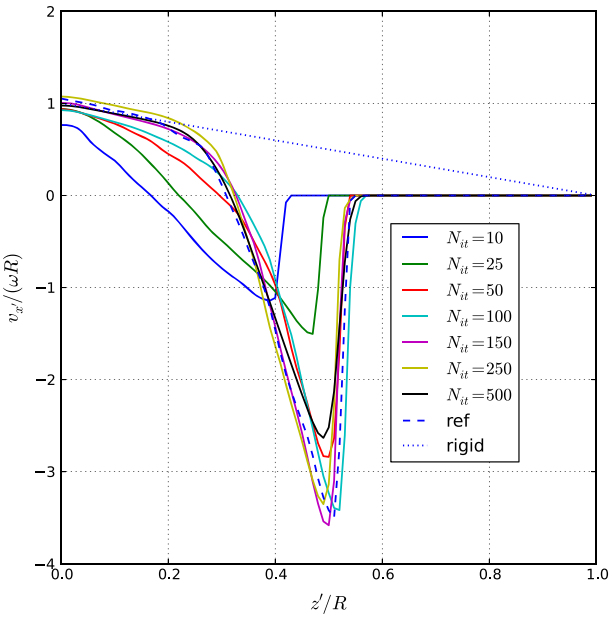

Figure 9. Velocity flow profile in the radial coordinate z' in a drum rotating with $\omega = 0.63$ rad/s and $N_{\mathrm{p}} = 7.5$K particles for different N_{it}. Rigid rotation and $h = 10^{-4}$ s reference solution are included.

The required number of iterations $N_{\mathrm{it}}^{0.05}$ for keeping the mean penetration error below the threshold $\langle \epsilon_{\mathrm{c}} \rangle \leqslant 0.05$ is shown in Figure 10 for drum speeds $\omega = 0, 0.13, 0.34,$ and0.63 rad/s and different number of particles. The required number of iterations scale linearly with the number of particles. Increasing drum velocity increases the required number of iterations. The fast drum

DOI: 10.1002/nme

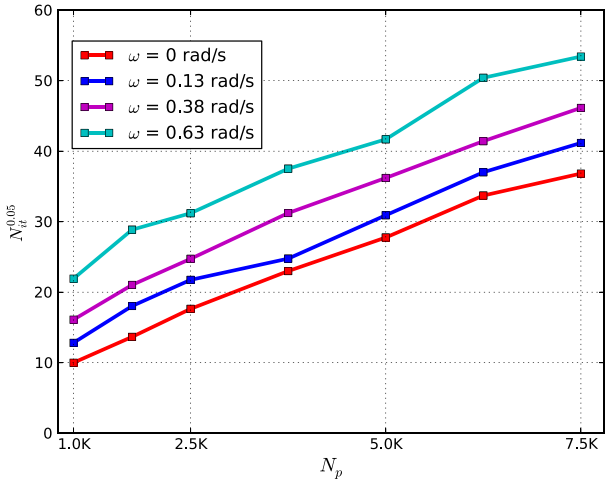

Figure 10. Number of iterations for drum with rotation speed ω and different number of particles.

$\omega = 2.5$ rad/s could not be made to converge below the error threshold $\langle \epsilon_c \rangle \leqslant 0.05$. The total number of contacts is nearly independent of drum rotation speed but are redistributed from continuous contacts to impacting contacts as drum rotation speed increases. For the case of 7.5K particles and the given time-step $h_{\text{NDEM}} = 0.01$, the fraction of impacting contacts is almost 50% for $\omega = 2.5$ rad/s and 10% for $\omega = 0.13$rad/s. For $\omega = 2.5$rad/s, the impact overlaps are too large to be reduced by any number of iterations.

5.3. Computational scaling as function of geometry and state

As is clear from the results in Sections 5.1 and 5.2, the required number of GS iterations does not depend on the number of particles or contacts alone but on the geometric shape and size and of the dynamic state. We conjecture that the shape and size can be characterized by the system length l in the direction of gravity (or more dominant external force) and on the width w of the cross section area. We represent these in units of particle size d. Furthermore, we use the inertial number I for characterizing the dynamic state of the system ranging from static to rapid flow. The cylindrical container results in Figure 6 suggest that the number of Gauss-Seidel iterations $N_{\text{it}}^{\epsilon}(l, w, I)$ scale linearly with l until it saturates at a certain length to width ratio l/w to a value proportional to width w. Supposedly, also this saturation is due to the emergence of strong force chains forming arching structure over between the container walls and thereby reducing the distance over which contact force need to be communicated by the iterative solver. In the Janssen effect, the wall pressure saturates to a value corresponding to supporting an *apparent mass* according (p. 270 [31]) to the formula $m_{\text{app}} = m_{\text{sat}}\,(1 - \exp\,[-m_{\text{fill}}/m_{\text{sat}}])$, with actual fill mass $m_{\text{fill}} = \rho l A$, cross-section area $A = \pi w^2$ and saturation mass $m_{\text{sat}} = \rho A w/2\mu_{\text{w}} K$ with particle-wall friction coefficient μ_{w} and Janssen coefficient $0 < K < 1$. We note that $m_{\text{app}} = m_{\text{sat}}\,(1 - \exp\,[-2\mu_{\text{w}} K l/w])$ and assume functional dependency on l/w holds for the saturation of $N_{\text{it}}^{\epsilon}(l, w, I)$ unless for systems too narrow, $w \lesssim 5$, for arching to occur. We further assume that the convergence rate, $N_{\text{it}}^{\epsilon_1}/N_{\text{it}}^{\epsilon_2} = \epsilon_2/\epsilon_1$, as found for $1d$ container is a general result. The rotating drum results in Figure 10 suggest linear dependency on rate of change. Based on these observations and assumptions, we make the following ansatz for a general system

$$N_{\text{it}}^{\epsilon}(l, w, I) = \frac{c_0(1 + c_1 I)}{\epsilon} \begin{cases} w\left(1 - \exp\left[-\frac{c_2 l}{w}\right]\right) & \text{, if } w \gtrsim 5 \\ c_2 l & \text{, if } w < 5 \end{cases} \tag{41}$$

DOI: 10.1002/nme

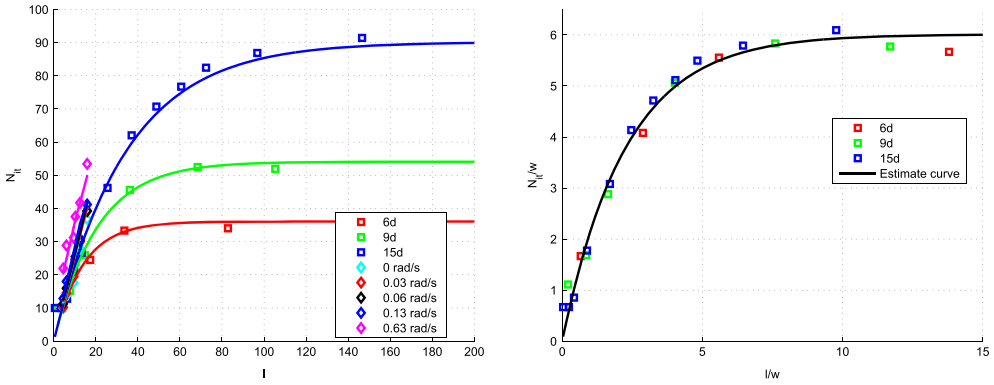

Figure 11. Number of iterations as function of system height l (left) and of shape factor l/w for containers only (right). The fitted formula is indicated as a solid line.

with coefficients c_0, c_1, and c_2 to be determined. Observe that in the limit where $w \gg l$ this approximates to $N_{\text{it}}^{\langle\epsilon\rangle}(l, w, I) \approx \frac{c_0(1+c_1 I)}{\epsilon} c_2 l$. In the container tests, we simply identify l by the material height and w by the cylinder diameter, i.e., $w = \Phi/d$. In the drum tests we estimate l by the averaged thickness of the material in radial direction and width by w by the drum diameter. Using nonlinear least square regression for fitting Equation (41) with the results in Figures 6 and 10, we find the parameter values $c_0 = 0.3(0.005)$, $c_1 = 2.0(1.1)$ and $c_2 = 0.44(0.02)$ with standard error in the parenthesis. The residual of the fitting is $r^2 = 1.25$. The matching of the function compared with the test systems is displayed in Figure 11.

With knowledge of $N_{\text{it}}^{\epsilon}(l, w, I)$, it is now possible to compare the computational efficiency of SDEM and NDEM without running any simulations but instead evaluating Equations (36), (37), (38), (39), and (41) with $K_{NDEM}^{\text{GS}} = 10^{-6}$ from our implementation and $K_{\text{SDEM}} = 10^{-6}$ measured from running LIGGGHTS on the same computer (Appendix C). Observe that these scale factors are implementation and hardware dependent. To illustrate the computational differences, we calculate the computational effort for 33 different cases. The results are found in Table I. For convenience, a calculator for the formula has been made available on web: http://umit.cs.umu.se/granular/dem/. In Table I, we observe the SDEM to be considerably faster in the examples of rows 13, 15, 18, 25, 27, 29, and 30, which are characterized by softer material, higher inertial number, and tall systems. The NDEM is considerably faster in the examples of rows 1, 4–10, and 20–24, which are characterized by stiffer material, low inertial number, and shallow systems. Larger error tolerance make the NDEM more efficient. Real-time performance, i.e., when the computational effort τ/τ_{real} is smaller or equal than one, can be achieved with NDEM in the examples of rows 5, 14, 20, 26, 31, and 31 and for SDEM in examples 26 and 31 and with systems up to roughly 1000 contacting rigid bodies.

For the highest material stiffness, E = 10 GPa, the regularization for semi-smooth DEM become too small for the system to remain well-posed and numerical instability appears as small vibrations. We compensate the high E by reducing the time step to make the regularization Σ factor in Equation (28) remain large enough. Specifically, for E = 10 GPa, we clamp the time step to $h_{\text{NDEM}} = 0.004$ s for $d = 0.01$ m and to $h_{\text{NDEM}} = 0.001$ s for $d = 0.001$ m. The relative velocity was estimated from the inertial number as $v_{\text{n}} \approx 0.2 I \sqrt{p/\rho}$.

5.4. *Effect of contact model and solver settings*

The simulation results in Section 5 are based on the semi-smooth DEM introduced in Section 2.4 integrate using the projected Gauss-Seidel solver implemented following the algorithm described in Appendix B. The elastic Hertz contact model was verified in tests with 1D columns of particles deforming elastically under its weight with Young's modulus ranging from 10^5 to 10^{13} Pa and

DOI: 10.1002/nme

Table I. Computational time of nonsmooth discrete element method and smooth discrete element method for 32 different systems. The quantities are nonsmooth discrete element method computational time τ_N and step size h_N, smooth discrete element method computational τ_S and step size h_S, number of GS iterations N_{it}, error tolerance ϵ, system width w, system height l, number of particles N_p, inertial number I, characteristic normal velocity v_n Young's modulus E, and particle diameter d.

	τ_N/τ_S	τ_N/τ_{real}	τ_S/τ_{real}	h_N/ms	h_S/ms	N_{it}	ϵ	w	l	N_p	I	v_n/mm/s	E/GPa	d/mm
1	0.02	2	100	3	0.003	19	0.05	6	10	203	0	0	0.01	1
2	0.02	10	625	3	0.003	23	0.05	15	10	1.8K	0	0	0.01	1
3	1.0	6e+3	6e+3	1	0.003	511	0.01	15	100	18K	0.1	6.1	0.01	1
4	0.02	2e+4	7e+5	3	0.003	27	0.05	500	10	2M	0.01	0.34	0.01	1
5	0.01	1	112	1	9e-5	132	0.01	1	10	10	0	0	10	1
6	0.002	5	3e+3	1	9e-5	19	0.05	6	10	280	0.01	0.29	10	1
7	0.01	200	2e+4	1	9e-5	117	0.01	15	10	1.8K	0.01	0.32	10	1
8	0.002	40	2e+4	1	9e-5	23	0.05	15	10	1.8K	0.01	0.32	10	1
9	0.05	9e+4	2e+6	1	9e-5	540	0.01	15	1000	180K	0.1	6.3	10	1
10	0.01	3e+5	2e+7	1	9e-5	131	0.01	500	10	2M	0	0	10	1
11	2.6	6e+9	2e+9	0.4	9e-5	10534	0.01	500	1000	200M	0.1	28	10	1
12	0.2	4e+8	2e+9	1	9e-5	2107	0.05	500	1000	200M	0.1	28	10	1
13	150	3e+3	20	5	0.05	13200	0.01	1	1000	1K	0	0	0.01	10
14	0.09	0.5	6	10	0.05	19	0.05	6	10	280	0	0	0.01	10
15	5.7	2e+3	400	5	0.05	511	0.01	15	100	18K	0.1	19	0.01	10
16	0.13	5e+3	4e+4	10	0.05	26	0.05	500	10	2M	0	0	0.01	10
17	0.16	6e+3	4e+4	10	0.05	32	0.05	500	10	2M	0.1	11	0.01	10
18	460	2+9	4e+6	1	0.05	10534	0.01	500	1000	200M	0.1	88	0.01	10
19	0.52	30	60	4	0.002	1320	0.01	1	100	100	0	0	10	10
20	0.007	1	200	4	0.002	19	0.05	6	10	280	0	0	10	10
21	0.07	100	2e+3	4	0.002	180	0.01	6	100	2.8K	0	0	10	10
22	0.04	500	1e+4	4	0.002	102	0.05	15	100	18K	0.1	19	10	10
23	0.01	1e+4	1e+6	4	0.002	27	0.05	500	10	2M	0.01	1.1	10	10
24	0.01	2e+4	1e+6	4	0.002	32	0.05	500	10	2M	0.1	11	10	10
25	15	2e+9	1e+8	1	0.002	10534	0.01	500	1000	200M	0.1	88	10	10
26	0.63	0.2	0.3	30	0.9	22	0.05	6	10	280	0.1	29	0.01	100
27	34	7e+3	200	10	0.9	540	0.01	15	1000	180K	0.1	63	0.01	100
28	3.0	600	200	30	0.9	108	0.05	15	1000	180K	0.1	63	0.01	100
29	2600	6e+8	2e+5	4	0.9	10534	0.01	500	1000	200M	0.1	280	0.01	100
30	100	2e+7	2e+5	20	0.9	2107	0.05	500	1000	200M	0.1	280	0.01	100
31	0.1	0.008	0.08	30	0.1	26	0.05	1	10	10	0	0	0.5	100
32	0.09	0.2	2	30	0.1	22	0.05	6	10	280	0.1	29	0.5	100
33	0.1	2	10	30	0.1	27	0.05	15	10	1.8K	0.1	32	0.5	100

DOI: 10.1002/nme

with $N_{\text{it}} = 500$. In a column of 20 particles, the overlap g agrees with the Hertz contact law in Equation (11) with mean error of less than 0.001 of a particle diameter. We also implemented and tested the NDEM in Section 2.3 with regularization and constraint stabilization following to the SPOOK scheme [18]. No significant difference in convergence rate or stability was found in the comparison between this and the semi-smooth DEM.

We also investigated the effect of applying *solver warmstarting* on the convergence rate, i.e., setting the initial guess on the Lagrange multipliers equal to, or some fraction of, the value from the previous time step. This accelerated the convergence of 1D columns but had no significant effect on other systems. Also the effect of using sequential or random order of iterations was found to be of no significance.

6. CONCLUSIONS

The SDEM and NDEM can both be used for computing the motion and interaction forces in granular matter. The NDEM formulation can be modified with constraints nonlinear in the gap function and mapped to the Hertz contact law with regularization and constraint stabilization terms from conventional to the viscoelastic parameters, e.g., the Young's modulus and Poisson ratio. This model, presented in Section 2.4 and referred to as the *semi-smooth DEM*, thus combines the potential speed-up of nonsmooth DEM with the ability of SDEM to accurately model viscoelastic contacts. Both the SDEM and NDEM follow as special cases in certain limits. Hence, the main difference between SDEM and NDEM lies in the required computational time for a given accuracy.

The time step size for NDEM (and semi-smooth DEM) is several order in magnitude larger than SDEM but each integration step is more computationally intense. It is difficult to judge a priori which of the methods is the most beneficial one for a given system and desired accuracy without actually testing the alternatives.

To remedy this, we provide formulas for estimating the required computational time. For NDEM with projected Gauss-Seidel solver, it is given by Equations (37) and (41). The formula takes as input system size and shape (l and w), dynamic state (I) that may range from static to rapid flow, material properties (ρ, E, ν, d), and error tolerance (ϵ) as arguments as well as the scale factors K_{NDEM}^{GS} and K_{SDEM} that depend on implementation and hardware. The relative computational efficiency of SDEM and NDEM becomes

$$\frac{\tau_{\text{NDEM}}}{\tau_{\text{SDEM}}} = \sqrt{\frac{\max\left(\frac{mv_{\text{n}}^2}{\epsilon^2}, \frac{2mgd}{\epsilon}\right)}{kd^2}} \frac{K_{\text{NDEM}}^{\text{GS}}}{K_{\text{SDEM}}} \frac{N_{\text{c}}}{N_{\text{p}}} \frac{c_0(1+c_1 I)}{\epsilon} \cdot \begin{cases} w\left(1 - e^{-\frac{c_2 l}{w}}\right) & w \gtrsim 5 \\ c_2 l & w < 5 \end{cases} \tag{42}$$

The number of particles is estimated $N_{\text{p}} \approx l \cdot w_x \cdot w_y$, $w = \max(w_x, w_y)$ and the number of contacts $N_{\text{c}} = n_{\text{p}} N_{\text{p}}$, with $n_{\text{p}} \lesssim 10$ is the average number of contact neighbors. The general trends are as follows. SDEM is computationally more efficient for wide and tall systems, rapid flows and soft materials. The NDEM with a projected Gauss-Seidel solver is more beneficial for shallow systems, static or slow flow, stiff materials, and with increasing error tolerance. Examples are provided in Table I. Observe that also in the large-scale limit, $N_{\text{p}} \to \infty$, it is not evident which method is the fastest, as opposed to the formula given in Ref. [1]. The NDEM may be faster for large-scale systems if they are shallow enough. The saturation effect, in Figure 6, on iterations at $l/w \gg c_2$ due to arching over the container walls make the projected Gauss-Seidel solver scale much better than previously thought and can make the NDEM competitive also for tall systems.

It should be emphasized that the results are limited to the use of projected Gauss-Seidel solver for the NDEM. This is the solver most often reported to be used for NDEM. The scaling of the square

DOI: 10.1002/nme

root bracket in Equation (42) is due to the time step size and is a general feature for the NDEM making computational efficiency increase with material stiffness and approaching the quasi-static limit. It is left for future investigations to provide empirical data for computational scaling and error analysis for other solvers that have more promising scalability (Section 3). Also, the effect of parallelization of SDEM [33] and NDEM [34] should be included in future work.

A third interesting topic for future research is the possibility of extending the semi-smooth DEM to an adaptive hybrid DEM, where different domains of the system are solved with smooth and nonsmooth methods and adjusted adaptively based on which method is most computational efficient based on current geometric and dynamic state. In the semi-smooth setting, it is trivial to replace a contact constraint by the corresponding nonlinear spring force, as in Equation (34). The computational effort becomes that of the nonsmooth DEM with $N_{\text{it}} = 1$.

APPENDIX A: IMPACT STAGE MIXED LINEAR COMPLEMENTARITY PROBLEM

In large-scale simulations, it becomes too inefficient to locate the exact time of each impact and solve for the entire contact network. Instead, fixed time step is used and impacts are detected *post facto* and solved simultaneously in a separate *impact stage* while preserving also the previously existing contact constraints.

In the impact stage, we split the contact set into two. One set of impacting contacts $\mathcal{N}_{\text{i}}$, arising during the last time step, and one set of continuous contacts $\mathcal{N}_{\text{c}}$. The contact Jacobians are split correspondingly from $\mathbf{G}_{\text{n}}$ into $\mathbf{G}_{\text{in}}$ and $\mathbf{G}_{\text{cn}}$ such that $\mathbf{G}_{\text{in}}\mathbf{v}^- < 0$.

Impacts are instantaneous impulse transfers where the velocity changes discontinuously from $\mathbf{v}^-$ to $\mathbf{v}^+$ as

$$\mathbf{M}\mathbf{v}^+ = \mathbf{M}\mathbf{v}^- + \mathbf{G}_{\text{in}}^{\text{T}}\boldsymbol{\lambda}_{\text{in}} + \mathbf{G}_{\text{cn}}^{\text{T}}\boldsymbol{\lambda}_{\text{cn}} \tag{A.1}$$

due to the impact impulse $\mathbf{G}_{\text{in}}^{\text{T}}\boldsymbol{\lambda}_{\text{in}}$ and response $\mathbf{G}_{\text{cn}}^{\text{T}}\boldsymbol{\lambda}_{\text{cn}}$ to preserve the Signorin Coulomb law in the continuous contact network. The impacts impulses should satisfy the Newton impact law, $\mathbf{G}_{\text{in}}\mathbf{v}^+ = -e\mathbf{G}_{\text{in}}\mathbf{v}^-$, with restitution coefficient e between 0 and 1 that corresponds to completely inelastic and perfectly elastic collision, respectively. The propagation of the impulse through the network of continuous contacts should not create constraint violations, i.e., should satisfy $\mathbf{G}_{\text{cn}}\mathbf{v}^+ = 0$ and act only repulsive, $\boldsymbol{\lambda}_{\text{in}} \geqslant 0$. Adding tangential friction impulses satisfying the Coulomb law is straight forward as is the extension to semi-smooth DEM with regularization. Collecting the impulse equation (A.1) and the Newton impact law plus constraint preservation, we see that they constitute an MLCP of the same form as in Equation (19) with

$$\mathbf{H} = \begin{bmatrix} \mathbf{M} & -\bar{\mathbf{G}}_{\text{in}}^{\text{T}} & -\bar{\mathbf{G}}_{\text{cn}}^{\text{T}} & -\bar{\mathbf{G}}_{\text{it}}^{\text{T}} & -\bar{\mathbf{G}}_{\text{ct}}^{\text{T}} \\ \bar{\mathbf{G}}_{\text{in}} & \boldsymbol{\Sigma}_{\text{i}} & 0 & 0 & 0 \\ \bar{\mathbf{G}}_{\text{cn}} & 0 & \boldsymbol{\Sigma}_{\text{c}} & 0 & 0 \\ \bar{\mathbf{G}}_{\text{it}} & 0 & 0 & \boldsymbol{\Gamma}_{\text{i}} & 0 \\ \bar{\mathbf{G}}_{\text{ct}} & 0 & 0 & 0 & \boldsymbol{\Gamma}_{\text{c}} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{v}^+ \\ \boldsymbol{\lambda}_{\text{in}} \\ \boldsymbol{\lambda}_{\text{cn}} \\ \boldsymbol{\lambda}_{\text{it}} \\ \boldsymbol{\lambda}_{\text{ct}} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\mathbf{M}\mathbf{v}^- \\ e\mathbf{G}_{\text{in}}\mathbf{v}^- \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{A.2}$$

APPENDIX B: PROJECTED GAUSS-SEIDEL SOLVER

Consider the system of equations for the nonsmooth DEM

$$\begin{bmatrix} \mathbf{M} & -\mathbf{G}^{\text{T}} \\ \mathbf{G} & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \tag{B.1}$$

DOI: 10.1002/nme

with friction cone conditions $\boldsymbol{\lambda} \in \mathcal{C}\,(\preceq \boldsymbol{\lambda}_{\mathrm{n}})$

$$\mathcal{C}\,(\mu\boldsymbol{\lambda}_{\mathrm{n}}) \equiv \left\{\boldsymbol{\lambda} = \left[\boldsymbol{\lambda}_{\mathrm{n}}, \boldsymbol{\lambda}_{\mathrm{t}_1}, \boldsymbol{\lambda}_{\mathrm{t}_1}\right] : \boldsymbol{\lambda}_{\mathrm{n}} \geqslant 0,\ |\boldsymbol{\lambda}_{\mathrm{t}}| \leqslant \mu\,|\boldsymbol{\lambda}_{\mathrm{n}}|\right\}$$

Let the submatrices $\mathbf{M}$ and $\boldsymbol{\Sigma}$ be block diagonal and $\mathbf{G}$ block sparse. Split the linear system on Schur complement form such that

$$\left(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}} + \boldsymbol{\Sigma}\right)\boldsymbol{\lambda} = \mathbf{q} - \mathbf{G}\mathbf{M}^{-1}\mathbf{p} \tag{B.2}$$

$$\mathbf{v} = \mathbf{M}^{-1}\mathbf{p} + \mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{\lambda} \tag{B.3}$$

Split the Schur matrix $\mathbf{S} = \mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}} + \boldsymbol{\Sigma}$ as $\mathbf{S} = \mathbf{L} + \mathbf{D} + \mathbf{L}^{\mathrm{T}}$, where $\mathbf{D}$ is block diagonal and $\mathbf{L}$ is strictly lower triangular. On block form Equation (B.6), with block indices α and β, Equation (B.2) can be solved iteratively $k = 0, 1, 2, \ldots$ by

$$\mathbf{D}_{(\alpha\alpha)}\boldsymbol{\lambda}^{(\alpha)}_{k+1} + \sum_{\beta<\alpha}\mathbf{L}_{(\alpha\beta)}\boldsymbol{\lambda}^{(\beta)}_{k+1} + \sum_{\beta>\alpha}\mathbf{L}_{(\alpha\beta)}\boldsymbol{\lambda}^{(\beta)}_{k} = \mathbf{q}_{(\alpha)} - \mathbf{G}_{(\alpha)}\mathbf{M}^{-1}\mathbf{p}_{(\alpha)} \tag{B.4}$$

with $\mathbf{D}_{(\alpha\alpha)} = \sum_a \mathbf{G}^{(\alpha)}_{[a]}\mathbf{M}^{-1}_{[aa]}\mathbf{G}^{(\alpha)\mathrm{T}}_{[a]} + \boldsymbol{\Sigma}_{(\alpha\alpha)}$. Adding and subtracting $\mathbf{D}_{(\alpha\alpha)}\boldsymbol{\lambda}^{(\alpha)}_{k}$ to this we obtain the following update formula for $\boldsymbol{\lambda}^{(\alpha)}_{k+1}$

$$\mathbf{D}_{(\alpha\alpha)}\boldsymbol{\lambda}^{(\alpha)}_{k+1} + r^{(\alpha)}_{k} - \mathbf{D}_{(\alpha\alpha)}\boldsymbol{\lambda}^{(\alpha)}_{k} = 0 \tag{B.5}$$

where the solution must satisfy the complementarity condition $\boldsymbol{\lambda}^{(\alpha)}_{k+1} \in \mathcal{C}\left(\mu\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{n},k+1}\right)$ and the residual vector is

$$\mathbf{r}^{(\alpha)}_{k} = \mathbf{S}_{(\alpha\alpha)}\boldsymbol{\lambda}^{(\alpha)}_{k} + \mathbf{G}_{(\alpha)}\mathbf{M}^{-1}\mathbf{p}_{(\alpha)} - \mathbf{q}_{(\alpha)} = \mathbf{G}_{(\alpha)}\mathbf{v}' - \mathbf{q}_{(\alpha)} \tag{B.6}$$

where $\mathbf{v}' \equiv \mathbf{M}^{-1}\mathbf{p} + \mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}}_{(\alpha)}\boldsymbol{\lambda}_{k}$. In the projected Gauss-Seidel we solve first the normal component. Then, if $\lambda^{(\alpha)}_{\mathrm{n},k+1} > 0$, we solve for the tangential components and project it onto the cone surface if it was outside

$$\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{t},k+1} \leftarrow \mathrm{proj}_{\mu\lambda^{(\alpha)}_{\mathrm{n},k+1}}\left(\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{t},k+1}\right) = \min\left(\frac{\mu\lambda^{(\alpha)}_{\mathrm{n},k+1}}{\left|\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{t},k+1}\right|}, 1\right)\cdot\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{t},k+1} \tag{B.7}$$

The algorithm is valid also for semi-smooth DEM by substituting the nonlinear normal constraints and Jacobians. For the projection onto the friction cone, replace $\mu\lambda^{(\alpha)}_{\mathrm{n},k+1}/\left|\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{t},k+1}\right| \rightarrow \mu\left|\vec{\boldsymbol{f}}^{(\alpha)}_{\mathrm{n},k+1}\right|/\left|\vec{\boldsymbol{f}}^{(\alpha)}_{\mathbf{t},k+1}\right|$. with $\vec{\boldsymbol{f}}^{(\alpha)}_{\mathrm{n}} = \bar{\mathbf{G}}^{(\alpha)}_{\mathrm{n}[\mathrm{a}]}(:, 1:3)^{\mathrm{T}}\lambda^{(\alpha)}_{\mathrm{n}}/h$ and $\vec{\boldsymbol{f}}^{(\alpha)}_{\mathrm{t}} = \mathbf{G}^{(\alpha)}_{\mathrm{t}[\mathrm{a}]}(:, 1:3)^{\mathrm{T}}\boldsymbol{\lambda}^{(\alpha)}_{\mathrm{t}}/h$.

The algorithm for the projected Gauss-Seidel solver for the semi-smooth DEM then is:

Pseudocode for the algorithm is available at http://umit.cs.umu.se/granular/dem/.

APPENDIX C: COMPUTER SPECIFICATION

The simulations where performed with a desktop computer with Intel(R) Core(TM) Xeon X5690, 3.46 GHz, 48 GB RAM on a Linux 64 bit system.

DOI: 10.1002/nme

Projected GS MLCP solver

Input: State $(\mathbf{x}_i, \mathbf{v}_i)$, smooth force $\mathbf{f}_\mathrm{s}$, contacts $\mathcal{N}_\mathrm{c}$

Constants: Particle properties (m, d)

Output: New state $(\mathbf{x}_{i+1}, \mathbf{v}_{i+1}, \boldsymbol{\lambda}_{i+1})$

1. Initialization

$(\mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}) = (\mathbf{x}_i, \mathbf{v}_i, 0)$

2. Contact constraint data

For every contact $\alpha \in \mathcal{N}_\mathrm{c}$ compute:

$g_{(\alpha)}, \mathbf{G}_\mathrm{n}^{(\alpha)}, \mathbf{G}_\mathrm{t}^{(\alpha)}$ as in Equation (21), (22)-(25)

$\bar{g}_{(\alpha)} = g_{(\alpha)}^{e_\mathrm{H}}, \bar{\mathbf{G}}_\mathrm{n}^{(\alpha)} = e_\mathrm{H} g_{(\alpha)}^{e_\mathrm{H}-1} \mathbf{G}_\mathrm{n}^{(\alpha)}$

$\bar{\mathbf{G}}_\mathrm{t} = \mathbf{G}_\mathrm{t}$

$\boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{\Upsilon}$ by Equation (30)

D_n^{-1}, D_t1^{-1} and D_t2^{-1}

3. Solve impacts

for $k = [0 : N_\mathrm{it}]$ or until error small:

for all $\alpha \in \mathcal{N}_\mathrm{c}$ solve normal then tangent

if α is impact then $c = 1 + e$ else $c = 1$

$r_{\mathrm{n},k}^{(\alpha)} = \Sigma_{(\alpha\alpha)} \lambda_{\mathrm{n},k}^{(\alpha)} - c \sum_a \bar{\mathbf{G}}_{\mathrm{n},k[a]}^{(\alpha)} \mathbf{v}_{[a]}$

$\lambda_{\mathrm{n},k+1}^{(\alpha)} = \lambda_{\mathrm{n},k}^{(\alpha)} + D_{\mathrm{n}(\alpha\alpha)}^{-1} r_{\mathrm{n},k}^{(\alpha)}$

$\lambda_{\mathrm{n},k+1}^{(\alpha)} = \max(0, \lambda_{\mathrm{n},k+1}^{(\alpha)})$

$\Delta\lambda_\mathrm{n}^{(\alpha)} = \lambda_{\mathrm{n},k+1}^{(\alpha)} - \lambda_{\mathrm{n},k}^{(\alpha)}$

for bodies a and b in contact α:

$\mathbf{v}_{[a]} = \mathbf{v}_{[a]} + M_{[aa]}^{-1} \bar{\mathbf{G}}_{\mathrm{n}[a]}^{(\alpha)\mathrm{T}} \Delta\lambda_\mathrm{n}^{(\alpha)}$

$\mathbf{v}_{[b]} = \mathbf{v}_{[b]} + M_{[bb]}^{-1} \bar{\mathbf{G}}_{\mathrm{n}[b]}^{(\alpha)\mathrm{T}} \Delta\lambda_\mathrm{n}^{(\alpha)}$

$r_{\mathrm{t}1\alpha}^{k} = \Gamma_{(\alpha\alpha)} \lambda_{\mathrm{t}2,k}^{(\alpha)} - \sum_a \bar{\mathbf{G}}_{\mathrm{t}1,k[a]}^{(\alpha)} \mathbf{v}_{[a]}$

$r_{\mathrm{t}2\alpha}^{k} = \Gamma_{(\alpha\alpha)} \lambda_{\mathrm{t}2,k}^{(\alpha)} - \sum_a \bar{\mathbf{G}}_{\mathrm{t}2,k[a]}^{(\alpha)} \mathbf{v}_{[a]}$

$\lambda_{\mathrm{t}1,k+1}^{(\alpha)} = \lambda_{\mathrm{t}1,k}^{(\alpha)} + D_{\mathrm{t}1(\alpha\alpha)}^{-1} r_{\mathrm{t}1,k}^{(\alpha)}$

$\lambda_{\mathrm{t}2,k+1}^{(\alpha)} = \lambda_{\mathrm{t}2,k}^{(\alpha)} + D_{\mathrm{t}2(\alpha\alpha)}^{-1} r_{\mathrm{t}2,k}^{(\alpha)}$

$\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)} \equiv \left[\lambda_{\mathrm{t}1,k+1}^{(\alpha)}, \lambda_{\mathrm{t}1,k+1}^{(\alpha)}\right]^\mathrm{T}$

if $\left|\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)}\right| >= \mu \left|\lambda_{\mathrm{n},k+1}^{(\alpha)}\right|$ then:

$\boldsymbol{\lambda}_{\mathrm{t},(k+1)}^{(\alpha)} = \mathrm{proj}_{\mu\lambda_{\mathrm{n},k+1}^{(\alpha)}} \left(\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)}\right)$

$\Delta\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)} = \boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)} - \boldsymbol{\lambda}_{\mathrm{t},k}^{(\alpha)}$

for bodies a and b in contact α:

$\mathbf{v}_{[a]} = \mathbf{v}_{[a]} + M_{[aa]}^{-1} \bar{\mathbf{G}}_{\mathrm{t}[a]}^{(\alpha)\mathrm{T}} \Delta\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)}$

$\mathbf{v}_{[b]} = \mathbf{v}_{[b]} + M_{[bb]}^{-1} \bar{\mathbf{G}}_{\mathrm{t}[b]}^{(\alpha)\mathrm{T}} \Delta\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)}$

4. Pre-solve

$\mathbf{q}_\mathrm{n} = -(4/h)\boldsymbol{\Upsilon}\bar{g} + \Upsilon\bar{\mathbf{G}}_\mathrm{n}\mathbf{v}$

$\mathbf{p} = \mathbf{M}\mathbf{v} + h\mathbf{f}_\mathrm{s}$

$\mathbf{v}' = \mathbf{M}^{-1}\mathbf{p}$

5. Continous contacts

for $k = [0 : N_\mathrm{it}]$ or until error small:

for all $(\alpha) \in \mathcal{N}_\mathrm{c}$ solve normal then tangent:

$r_{\mathrm{n},k}^{(\alpha)} = -q_{\mathrm{n},k}^{(\alpha)} + \Sigma_{(\alpha\alpha)} \lambda_{\mathrm{n},k}^{(\alpha)} - \sum_a \bar{\mathbf{G}}_{\mathrm{n},k[a]}^{(\alpha)} \mathbf{v}'_{[a]}$

$\lambda_{\mathrm{n},k+1}^{(\alpha)} = \lambda_{\mathrm{n},k}^{(\alpha)} + D_{\mathrm{n}(\alpha\alpha)}^{-1} r_{\mathrm{n},k}^{(\alpha)}$

$\lambda_{\mathrm{n},k+1}^{(\alpha)} = \max\left(0, \lambda_{\mathrm{n},k+1}^{(\alpha)}\right)$

$\Delta\lambda_\mathrm{n}^{(\alpha)} = \lambda_{\mathrm{n},k+1}^{(\alpha)} - \lambda_{\mathrm{n},k}^{(\alpha)}$

for bodies a and b in contact α:

$\mathbf{v}'_{[a]} = \mathbf{v}'_{[a]} + M_{[aa]}^{-1} \bar{\mathbf{G}}_{\mathrm{n}[a]}^{(\alpha)\mathrm{T}} \Delta\lambda_\mathrm{n}^{(\alpha)}$

$\mathbf{v}'_{[b]} = \mathbf{v}'_{[b]} + M_{[bb]}^{-1} \bar{\mathbf{G}}_{\mathrm{n}[b]}^{(\alpha)\mathrm{T}} \Delta\lambda_\mathrm{n}^{(\alpha)}$

$r_{\mathrm{t}1,k}^{(\alpha)} = \Gamma_{(\alpha\alpha)} \lambda_{\mathrm{t}1,k}^{(\alpha)} - \sum_a \bar{\mathbf{G}}_{\mathrm{t}1,k[a]}^{(\alpha)} \mathbf{v}'_{[a]}$

$r_{\mathrm{t}2,k}^{(\alpha)} = \Gamma_{(\alpha\alpha)} \lambda_{\mathrm{t}2,k}^{(\alpha)} - \sum_a \bar{\mathbf{G}}_{\mathrm{t}2,k[a]}^{(\alpha)} \mathbf{v}'_{[a]}$

$\lambda_{\mathrm{t}1,k+1}^{(\alpha)} = \lambda_{\mathrm{t}1,k}^{(\alpha)} + D_{\mathrm{t}1(\alpha\alpha)}^{-1} r_{\mathrm{t}1,k}^{(\alpha)}$

$\lambda_{\mathrm{t}1,k+1}^{(\alpha)} = \lambda_{\mathrm{t}2,k}^{(\alpha)} + D_{\mathrm{t}2(\alpha\alpha)}^{-1} r_{\mathrm{t}2,k}^{(\alpha)}$

$\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)} \equiv \left[\lambda_{\mathrm{t}1,k+1}^{(\alpha)}, \lambda_{\mathrm{t}2,k+1}^{(\alpha)}\right]^\mathrm{T}$

if $\left|\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)}\right| >= \mu \left|\lambda_{\mathrm{n},k+1}^{(\alpha)}\right|$ then:

$\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)} = \mathrm{proj}_{\mu\lambda_{\mathrm{n},k+1}^{(\alpha)}} (\boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)})$

$\Delta\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)} = \boldsymbol{\lambda}_{\mathrm{t},k+1}^{(\alpha)} - \boldsymbol{\lambda}_{\mathrm{t},k}^{(\alpha)}$

for bodies a and b in contact α:

$\mathbf{v}'_{[a]} = \mathbf{v}'_{[a]} + M_{[aa]}^{-1} \bar{\mathbf{G}}_{\mathrm{t}[a]}^{(\alpha)\mathrm{T}} \Delta\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)}$

$\mathbf{v}'_{[b]} = \mathbf{v}'_{[b]} + M_{[bb]}^{-1} \bar{\mathbf{G}}_{\mathrm{t}[b]}^{(\alpha)\mathrm{T}} \Delta\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)}$

6. Update position and orientation

$\mathbf{v}_{i+1} = \mathbf{v}'$

$\mathbf{x}_{i+1} = \mathbf{x}_i + h\mathbf{v}_{i+1}$

ACKNOWLEDGEMENTS

This project was supported by LKAB, UMIT Research Lab, and Algoryx Simulations. The authors are grateful to Stefan Hedman for creating the initial analysis pipeline and performing experiments with LIGGGHTS.

REFERENCES

1. Brendel L, Unger T, Wolf DE. Contact dynamics for beginners. In *The Physics of Granular Media*, Hinrichsen H, Wolf DE (eds). Wiley-VCH: Weinheim, 2004; 325–343.
2. Pöschel T, Schwager T. *Computational Granular Dynamics, Models and Algorithms*. Springer-Verlag: Berlin, 2005.
3. Jean M. Numerical simulation of granular materials. In *Micromechanics of Granular Materials*, Cambou B, Jean M, Radjai F (eds). John Wiley & Sons: London, 2004; 149–307.

DOI: 10.1002/nme

4. Radjai F, Richefeu V. Contact dynamics as a nonsmooth discrete element method. *Mechanics of Materials* June 2009; **41**(6):715–728.
5. Cundall PA, Strack ODL. A discrete numerical model for granular assemblies. *Geotechnique* 1979; **29**:47–65.
6. Moreau JJ. Numerical aspects of the sweeping process. *Computer Methods in Applied Mechanics and Engineering* July 1999; **177**:329–349.
7. Munjiza A. *The Combined Finite-Discrete Element Method*. John Wiley & Sons, Ltd: Chichester, UK, 2004.
8. Zienkiewicz OC, Taylor RL. *The Finite Element Method for Solid and Structural Mechanics, Sixth Edition*, 6th ed. Butterworth-Heinemann: Oxford, September 2005.
9. Rougier E, Munjiza A, John NWM. Numerical comparison of some explicit time integration schemes used in dem, fem/dem and molecular dynamics. *International Journal for Numerical Methods in Engineering* 2004; **61**(6):856–879.
10. Pang J-S, Stewart DE. Differential variational inequalities. *Mathematical Programming* 2008; **113**(2):345–424.
11. Bornemann FA. *Homogenization in Time of Singularly Perturbed Mechanical Systems*, Lecture Notes in Mathematics, 0075-8434, Vol. 1687. Springer: Berlin, 1998.
12. Hairer E, Wanner G. *Solving Ordinary Differential Equations II, Stiff and Differential Algebraic Problems*, second revised edition, Springer Series in Computational Mathematics, Vol. 14. Springer-Verlag: Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, 1996.
13. Hairer E, Lubich C, Wanner G. *Geometric Numerical Integration*, Spring Series in Computational Mathematics, Vol. 31. Springer-Verlag: Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, 2001.
14. Brilliantov NV, Spahn F, Hertzsch J-M, Pöschel T. Model for collisions in granular gases. *Physical Review E* May 1996; **53**:5382–5392.
15. Jean M. The non-smooth contact dynamics method. *Computer Methods in Applied Mechanics and Engineering* July 1999; **177**:235–257.
16. Lacoursière C. Splitting methods for dry frictional contact problems in rigid multibody systems: preliminary performance results. In *Conference Proceedings from SIGRAD2003, November 20–21, 2003*, Ollila M (ed.). Umeå University: Umeå, Sweden, November 2003; 11–16. SIGRAD.
17. Lacoursière C, Servin M, Backman A. Fast and stable simulation of granular matter and machines. In *Discrete element methods - Simulation of Discontinua: Theory and Applications*, Munjiza A (ed.). Queen Mary, University of London: London, 2010. DEM5 - The Fifth International Conference on Discrete Element Methods, London 25-26 August (2010).
18. Lacoursière C. Regularized, stabilized, variational methods for multibodies. In *The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), 30-31 October, 2007*, Bunus DFP, Führer C (eds). Linköping University Electronic Press: Göteborg (Särö), Sweden, December 2007; 40–48.
19. Krabbenhoft K, Huang J, Silva MV, Lyamin AV. Granular contact dynamics with particle elasticity. *Granular Matter* 2012; **14**:607–619.
20. Wang D, Servin M, Mickelsson K-O. Outlet design optimization based on large-scale nonsmooth dem simulation. *Powder Technology* 2013. DOI: 10.1016/j.powtec.2013.11.046. in press.
21. Williams J, O'Connor R. Discrete element simulation and the contact problem. *Archives of Computational Methods in Engineering* 1999; **6**:279–304. 10.1007/BF02818917.
22. LIGGGHTS. Open source discrete element method particle simulation code, September 2012.
23. Hedman S. Smooth and non-smooth approaches to simulation of granular materal. *Master's thesis*, Department of Physics, Umeå University, Umeå, Sweden, September 2011.
24. Renouf M, Dubois F, Alart P. A parallel version of the non smooth contact dynamics algorithm applied to the simulation of granular media. *Journal of Computational and Applied Mathematics* 2004; **168**(1-2):375–382.
25. Brogliato B, Dam AT, Paoli L, Génot F, Abadie M. Numerical simulation of finite dimensional multibody nonsmooth mechanical systems. *Applied Mechanics Reviews* 2002; **55**(2):107–150.
26. Algoryx Simulations. AGX multiphysics toolkit, September 2012.
27. Renouf M, Acary V, Dumont G. 3D frictional contact and impact multibody dynamics. A comparison of algorithms suitable for real-time applications. In *Multibody Dynamics 2005, ECCOMAS Thematic Conference*, Goicolea JM, Cuadrado J, Garcìa Orden JC (eds): Madrid, Espagne, 2005.
28. Renouf M, Alart P. Conjugate gradient type algorithms for frictional multi-contact problems: applications to granular materials. *Computer Methods in Applied Mechanics and Engineering* 2005; **194**(18-20):2019–2041.
29. Unger T, Brendel L, Wolf D, Kertsz J. Elastic behavior in contact dynamics of rigid particles. *Physical Review E* 2002; **65**(6):7.
30. Goldhirsch I. Stress, stress asymetry and couple stress: from discrete particles to contiuous fields. *Granular Matter* 2010; **12**(3):239–252.
31. Mehta A. *Granular Physics*. Cambridge University Press: New York, 2007.
32. Boateng AA, Barr PV. Granular flow behaviour in the transverse plane of a partially filled rotating cylinder. *Journal of Fluid Mechanics* 1997; **330**:233–249.
33. Walthe JH, Sbalzarini IF. Large-scale parallel discrete element simulations of granular flow. *Engineering Computations: Int J for Computer-Aided Engineering*; **26**(6):688–697.
34. Iglberger K, Rüde U. Massively parallel granular flow simulations with non-spherical particles. *Computer Science - Research and Development* 2010; **25**:105–113.

DOI: 10.1002/nme

Outlet design optimization based on large-scale nonsmooth DEM simulation

D. Wang[a], M. Servin[b], K-O. Mickelsson[c]

[a] *Umeå University, Umeå, Sweden (da.wang@physics.umu.se).*
[b] *Umeå University, Umeå, Sweden (martin.servin@physics.umu.se).*
[c] *LKAB R&D, Malmberget, Sweden.*

Abstract

We consider the application of a nonsmooth discrete element method to geometric design optimization of a balling drum outlet used in production of iron ore balls. The geometric design optimization problem is based on the need for homogeneous flow of balls from the balling drum onto a wide belt conveyor feeding a roller screen (sieve). An outlet with two design variables is investigated and the optimal shape for the given system and production flow is found by exploring the design space with 2000 simulations.

Keywords: Nonsmooth discrete element method, design optimization, mineral processing, balling drum.

1. Introduction

There is big potential in optimizing particulate flow in mineral and processing technology. Determining the optimal design and control parameters for the systems running on-line in plants is typically too time-consuming, impractical and economically infeasible. This calls for time very efficient simulations that allows exploration of the large design space for the optimization variables.

Simulation based design optimization of systems involving particulate flow is uncommon. Supposedly this is due to that simulations of such systems are associated with long computing times which prohibits systematic approach including exploring large design space. With increasing computing power and advances in the modeling and simulation of granular matter design optimization of complex systems is becoming feasible.

One common method for simulating particulate flow is the *discrete element method* (DEM) [1];[7]. It can be used for simulating granular matter in the gaseous, liquid as well as solid regime whereas many of the alternative methods is best suited for a single regime. In the present paper we use a *nonsmooth DEM* (nDEM) [8] approach, also referred to as the *nonsmooth contact dynamics method* [6];[3]. The particular variant of nonsmooth DEM used in the present paper is described in more detail in an accompanying paper [10]. The nonsmooth approach allows for time integration using time-steps much larger than the characteristic elastic response time and considerable speed-up can be achieved as compared to standard, or *smooth*, DEM. An alternative to DEM would be multiphase computational fluid dynamics [12], e.g., using frictional-kinetic stresses model [11, 9]. However, the width of the gaps in the outlet range down to one particle diameter. At this length scale the continuum models of granular material do not apply.

The current paper considers simulation based geometric design optimization of a balling drum outlet. The purpose is to demonstrate the feasibility of nDEM to this problem and provide a systematic methodology for the geometric design optimization for systems involving particulate flow.

Balling drums are used in agglomeration of powdered mineral ore and binding agency into spherical *green balls* [2]. The balling process precedes the *induration process* where the green balls are hardened into pellets by heating them in a sinter machine. For the sake of quality of the pellets, the ideal goal for the balling process is to produce a nearly *steady flow of spherically symmetric, mono-sized homogeneous green balls* from the balling process. The capacity of the entire pelletizing plant is limited to the maximum flow rate from the balling process that secures high quality pellets. The sieving is maximally efficient when the balls flows over the roller screen in even layer covering the entire screen. This requires that the outlet produce a homogeneous bed of balls. A poor outlet design produces in an inhomogeneous pellet bed and thus less efficient sieving. A homogeneous bed of pellets reaching the end of the horizontal belt conveyor is therefore our basis for formulating the objective function for the design optimization.

We consider a particular balling plant design that is used at LKAB [5] iron ore pelletizing plant MK3 in Malmberget, Sweden. The balling circuits, depicted in Fig. 1 and more schematically in Fig. 2, has the following components: an inclined rotating cylindrical drum with an outlet in the lower end, a wide horizontal belt conveyor where the material lands from the outlet, a roller screen where the material is sieved onto conveyors transporting balls of

correct size (about 10 mm in diameter) further to the sinter machine. Undersized and over-sized material is conveyed from the sieve, over-sized balls are crushed, and fed back into the higher end of the balling drum together with powdered ore and binding agency. This full scale industrial balling circuit has drum diameter about 3.6 m and circulate material in the range 200-300 tons/hour. A laboratory balling circuit in scale of order 1:4 is under construction in the nearby research facility *Pelletizing Research Centre* (PRC). The design optimization in this paper is for the laboratory balling drum outlet.

2. Model of the balling plant

We consider design optimization of a laboratory balling circuit in scale 1:4 scale compared to the one in production. Many parameters of the lab system can be changed. We choose to fix the following geometric parameters: drum diameter $d \approx 0.75$ m, drum length $L \approx 2.5$ m, outlet length $l \approx 0.75$ m, drum inclination $\theta \approx 7.5$ degrees, conveyor width $w \approx 1.0$ m and the rotation speed of the drum is $\omega \approx 2.5$ rad/s. The conveyor velocity is set to $v = 0.25$ m/s unless otherwise stated. See, Fig. 2, for a simple illustration of the design and notations. The positions of the outlet projected onto the conveyor belt are indicated by the points $y_1 = 0.15$ m and $y_2 = 0.9$ m.

The current study is restricted to one particular geometric design of the outlet, illustrated in Fig. 3. It has three gaps with bisector inclination angle $\eta = 15^\circ$. The gap width increases with angle α. The inner base width of the gap at the interface to the drum is denoted by β. These α and β are our design variables.

The drum, outlet, belt conveyor and roller screen are considered as rigid kinematic objects. The drum and outlet are represented by a mesh of resolution 0.01 m. The drum and outlet is also given a geometric 'texture' consisting of Gaussian shaped bumps randomly distributed with random height in the range $[0.002, 0.008]$ and random width in the range $[0.005, 0.02]$. The surface density of irregularities is set to 700 dimples per m^2 with random height in the range. The texture has been added to model the presence of agglomerated fine material on the inside of the drum and outlet which causes higher lifting of the balls. The contact force with the rotating drum surface drive the material into a rotational flow in the cross-section plane with an additional flow component in the axial direction by gravity and drum inclination. A close-up from simulation of full scale system is shown in Fig. 4, and an overview is shown in Fig. 5.

The drum is fed with balls at the rate 4.4 kg/s. In the real system the particulate material has a distribution of particle size, mass and moisture content. We model the material as perfectly spherical balls with 10 mm diameter and mass density of 2500 kg/m^3. The elastic modulus has been measured to be roughly $Y = 0.5 \times 10^6$ Pa according to Fig. 26 in [2]. The friction coefficient is set to 0.7. From experiments with real balls the coefficient of restitution has been estimated to be close to zero and the friction angle to 35°.

3. Simulation

3.1. Nonsmooth DEM

We use nonsmooth DEM for simulating the dynamics of the ore green balls. Each ball is represented as a rigid body. The bodies interact by dry frictional contacts modeled by constraints and complementarity conditions for unilateral nonpenetration and friction according to the Coulomb law. Impacting contacts and stick-slip transitions are considered as instantaneous events making the velocities nonsmooth in time. Our nonsmooth DEM approach, outlined in more detail in Servin et al. [10], allows for time-integration with large time-steps as compared to conventional *smooth* DEM. A particular schema for constraint regularization and stabilization [4] brings both numerical stability and possibility to map the nonsmooth simulation parameters to the conventional viscoelastic material parameters. Time integration $(x_i, v_i) \rightarrow (x_{i+1}, v_{i+1})$ of the system position x and velocity v from time t_i to $t_{i+1} = t_i + h$ involve solving a mixed linear complementarity problem (MLCP) of the form

$$\begin{bmatrix} M & -G_i^T \\ G_i & \Sigma \end{bmatrix} \begin{bmatrix} v_{i+1} \\ \lambda_{i+1} \end{bmatrix} = \begin{bmatrix} p_i \\ q_i \end{bmatrix} \tag{1}$$

with friction cone conditions $\lambda \in \mathcal{C}(\mu\lambda_{\mathrm{n}})$

$$\mathcal{C}(\mu\lambda_{\mathrm{n}}) \equiv \{\lambda = [\lambda_{\mathrm{n}}, \lambda_{\mathrm{t}_1}, \lambda_{\mathrm{t}_1}] : \lambda_{\mathrm{n}} \geq 0,\ |\lambda_{\mathrm{t}}| \leq \mu|\lambda_{\mathrm{n}}|\}$$

on the Lagrange multiplier,λ, for the constraint force, $G^T\lambda$, responsible for maintaining non-penetration in the contact normal direction n and Coulomb dry friction in the contact tangent plane spanned by t_1 and t_2. The submatrices M and Σ are block diagonal matrices of body mass and contact regularization, respectively. The submatrix G is block sparse contact Jacobian built of by normal and tangent vectors. The constraints regularization,

Σ and stabilization terms q_i are based on the energy and dissipation potentials from Hertz contact law and thus link the parameters of the numerical integration scheme directly to measurable or tabulated viscoelastic material parameters. For spheres, the Hertz contact law reads $f = k(g^{3/2} + cg^{1/2}\dot{g})$, where f is the force, g the contact overlap, k the stiffness coefficient and c the damping coefficient. We use constraint regularization and stabilization terms [10] that map to $k = 25$ kN/m$^{3/2}$ and $c = 0.02$ s^{-1}. For impacting contacts, we instead apply zero restitution, which corresponds to setting $q_i = 0$ in Eq. (1). The time-step h should be a fraction of R/v_n not to cause too big contact overlaps, where v_n is the characteristic relative normal velocity in the contact points and R is the ball radius. In the drum, the material shear rate is roughly $\dot{\gamma} \approx \Delta v/\Delta x \approx \omega d/h_\mathrm{p}$, where $h_\mathrm{p} \approx 20R$ is the height of the material in the drum and we assume to have a dense flow with velocity scaling linearly from the drum surface to the top surface. The relative contact velocity between two particle layers is thus estimated to $\dot{\gamma}2R \approx \omega d/10 \approx 0.2$ m/s such that $R/v_\mathrm{n} \approx 0.25$ s. We therefore choose time step $h = 0.01$ s. The time integration includes solving a mixed linear complementarity problem (MLCP) condition with $3N_\mathrm{p} + (3/2)Nc \sim 6N_\mathrm{p}$ variables assuming the N_p particles has on average 6 contacting neighbours summing to N_c contacts in total. The simulation involves approximately 80k particles, which means 500k variables. The contact constraint forces and resulting velocity changes are computed using a projected Gauss-Seidel solver for the MLCP. The number of iterations are set to 25.

The computational time for simulation of one drum evolution (2.6s) with 80k particles is measured to be 585s, i.e., average computational speed for 1k particle is 2.8s computing time for 1s real time. The simulation are r-un single threaded and on a desktop computer Intel(R) Core(TM) Xeon X5690 3.46 GHz processor. Sample simulations with full scale drum and up to 1.8M particles, shown in Fig. 4 and Fig. 5, runs on a desktop computer with 16 GB RAM memory.

Video from simulation is available on the web: `http://umit.cs.umu.se/granular/outlet`.

3.2. State intialization

An initial state close to stationary flow is produced by running a simulation for 40 s, (approximately 15 drum evolutions) with iron ore balls added in the rate of 4.4 kg/s at the upper end of the drum. Particles reaching the belt

conveyor are deleted during the simulation. This state – consisting of N_p position vectors, velocity vector and angular momentum vector plus drum and outlet orientation – is used for each simulation in the design optimization. To let the flow adapt to a new outlet geometry each simulation is initialized by another 10.4 s (4 drum evolutions) of simulation before recording starts.

3.3. Recording

The position of each particle reaching the belt conveyor is stored together with the time of impact. The particle is then removed from the simulation in order to save computational time. Recording is made during a time-interval of two drum evolutions (5.2 s). The positions and time are transformed into a coordinate system co-moving with the belt conveyor, i.e., $(x, y, t) \rightarrow (x - vt, y)$. A sample recording of particle positions on conveyor from one simulation is shown in Fig. 6. The flow inside the drum was observed to be close to stationary after more than 10 s of initialization, i.e., there was no sign of flow instabilities or pulsating phenomena. The characteristic and expected striped pattern on the conveyor belt contact data is due to the outlet gaps and the time duration of the recording is six times of the gap periodicity.

3.4. Surface reconstruction

The particles impacting the belt conveyor forms a bed with some height surface shape $h(x - vt, y)$ m. We use *cellular automata* [7, Ch. 6] to reconstruct this surface from the particle scatter data. The number of particles on the conveyor data in Fig. 9 is roughly 200 k, compared to roughly 80 k in the drum. Hence, simulating all particles would take 3.5 times longer time. Also, stable pile formation would require adding constraint based rolling resistance, which would add additional $2 \times N_c$ equations to the MLCP solved at each time-step. The net effect would be an increase in CPU time by more than 5 times. The cellular automata operates on a regular square grid with cell size of 2.5 particle diameters, friction angle 35° (measured in tests) and packing ratio of 0.7. The surface reconstruction of the particle data in Fig. 6 is shown in Fig. 7.

4. Design optimization

4.1. Objective function

The goal is to determine the values of the design variables $(\alpha, \beta) \in \mathcal{D}$ for the outlet geometry that produces an even flow of iron ore balls onto the

sinter screen. By even flow we mean that the height of the iron ore ball bed reaching the end of the belt conveyor has constant cross-section when time-averaged over one drum evolution $t_\mathrm{d} = 2\pi/\omega$. The rotating outlet typically produces stripes of pellets on a the conveyor belt, see Fig. 7. If the stripes are perfectly uniform it is possible to find a velocity of the belt conveyor that compress the stripes into a planar surface. The time averaged pellet bed height is computed:

$$h(t, y) \equiv \tfrac{1}{t_\mathrm{d}} \int_0^{t_\mathrm{d}} h(x_0 - v(t + \tau), y)\mathrm{d}\tau \tag{2}$$

If the flow of iron ore balls inside the drum is stationary when reaching the outlet the time-averaged height of the iron ore ball bed will be constant in time and we denote it $h(y)$. Any discrete height profile $h(y)$ can be represented by its fast Fourier transform (FFT) with complex amplitudes a_n, $n = 0, \pm1, \pm2, \ldots, \pm N$. Each mode contributes to variations of the heigth profile with magnitude $|a_n|/N$ and wave length $l_\mathrm{c}/\pi|n|$, where l_c is the width of the conveyor. A constant profile would correspond to $h(y) = a_0/N$ and $a_n = 0$ for all $n \neq 0$. The low mode number variations ($n \neq 0$) are the most negative ones for the sintering performance. Higher modes correspond to short wave length fluctuations that disperse more easily on the sinter screen. Therefore we introduce a decaying weight factor $w_n = 2^{|n|-1}$ and chose the following objective function

$$f(\alpha, \beta) = \frac{1}{N} \sum_{n=1}^{N} w_n \left(\|a_{-\mathrm{n}}\| + \|a_n\| \right) \tag{3}$$

We use $N = 20$ modes in the design optimization. The optimal design parameter pair is the solution to $\arg\min_{(\alpha,\beta)\in\mathcal{D}} f(\alpha, \beta)$, where the design space $\mathcal{D}$ is the domain of objective function. Since the object function is based on the time averaged height profile the result depends very weakly on the choice of conveyor speed. The effect on the value of the objective function at the optimum by reducing the speed by half or doubling it was found to be 4%. The conveyor velocity is also limited from below and above to avoid high pellet beds and high impact velocities at the roller screen.

4.2. Optimization procedure

The objective function cannot be assumed convex and local minima might exist. This calls for heuristic strategy in finding optimal solution. In the first

stage a large design space, $\mathcal{D}_1$, is chosen. The space is covered by a regular grid of $\sim$ 200 nodes, each corresponding to a simulation with particular value for (α, β). Potential regions of optimum are investigated further in successive stages. If the objective function appears smooth various local optimization techniques may be applied. We set the procedure is terminated when we found a solution of which all amplitudes $a_n \leq \frac{1}{10} a_0$ for all $N \geq 1$. That suggests the objective function deceeds the chosen tolerance threshold, $\|f(\alpha, \beta)\| < \epsilon_h = 0.01$ m.

5. Results

5.1. Optimal design

The first stage analysis of the objective function in the design space start with a series of simulations in region $\alpha \in [0°, 0.6°]$ and $\beta \in [0.01, 0.2]$ m with a coarse grid. That give us a first rough view of the objective function in the design space.

At the second stage we zoomed in on the region $\mathcal{D}_1$ with $\alpha \in [0.1°, 0.4°]$ and $\beta \in [0.03, 0.07]$ m. The result from 768 simulations, which cost about one hour each to run, provides us an convex surface with some noise.

The third stage is carried out on a smaller region with higher resolution grid, $\mathcal{D}_2$ with $\alpha \in [0.12°, 0.33°]$ and $\beta \in [0.034, 0.06]$ m contains 864 simulations.

The optimum solution is found $(\alpha_I, \beta_I) = (0.20°, 0.048 \text{ m})$ with error tolerance $(\Delta\alpha, \Delta\beta) = (0.01°, 0.003 \text{ m})$. We show the contour plot of combined 2nd stage and 3rd stage grids in Fig. 8. Note that the bold line is the border of the chosen tolerance threshold $\|f(\alpha, \beta)\| < \epsilon_h$.

5.2. Sample data

We provide more detailed simulation results for four sample points listed in table 1. For each data point we present the particle position scatter plots on belt conveyor, post-processed surface, the time-integrated height profile and the corresponding FFT spectrum that builds up the objective function value. These data are found in Fig. 9.

5.3. Observations

We make the following observations. A too wide angle and too wide base width results in too big outflow between the gaps and formation of a heap

Table 1: Sample points

Point	α [degree]	β [m]	$f(\alpha, \beta)$ [m]	Comment
I	0.20	0.048	0.006	optimum
II	0.17	0.051	0.010	on the tolerance
III	0	0.01	0.077	too narrow gaps
IV	0.6	0.2	0.045	too large gaps

under the edge between the end of the drum and the outlet (data point IV). A too narrow angle and too narrow base width produces too small outflow between the gaps resulting in a heap under the end of the outlet. There is no point in the design space that produces a planar bed of ore balls. The optimal solution (I) has a constant rate of outflow between the gaps along the axial direction (y-axis) but the outlet is too short for all particles to spread evenly. The 'excessive' particles form a heap at the end of the outlet.

6. Discussion

There are a number of uncertainties in the model and simulation that must be remedied before the presented design optimization method have fully known predictive power. These include the choice of material and model parameters (size distribution, friction, geometric shape, etc.) and simulation parameters (time-step, number of iterations). The latter is thoroughly considered in a separate publication [10]. Validation of remaining model parameters, foremostly the contact model of the drum wall, is left for future work when the lab drum system is installed and in operation. The result, that the particular design principle considered in this work cannot produce a planar height profile, is likely to remain. Furthermore, it is improbable that this design principle will be optimal for a range of different input mass flow. Alternative design principles are called for and the presented method can be applied to design exploration and sensitivity analysis and finally optimization of the design.

7. Conclusions

Design optimization of balling plant outlet geometry based on nonsmooth DEM simulation has been shown feasible. The relevant design space was

covered by approximately 2000 simulations of 15.6 s material flow each with 80×10^3 particles. The total computational time was 7.6×10^6 CPU seconds (2100 hours or 88 days). We used a computer with 12 CPUs in which case the total time for the design optimization procedure was roughly seven days of computation. With this performance it is possible to cover alternative designs that requires a larger design space. The conclusion from the design analysis is that the particular outlet design considered has no solution for which a planar bed can be produced. One solution is found that produces a even outflow but the outlet form a heap at the end. Future steps involve sensitivity analysis of model and simulation parameters as well as of the realization of stationary flow. Alternative outlet design will be considered that avoids the formation of a heap at the end.

8. Acknowledgements

This project was supported by LKAB, ProcessIT Innovations, UMIT Research Lab and by Algoryx Simulations.

9. References

[1] Cundall, P. A., Strack, O. D. L., 1979. A discrete numerical model for granular assemblies. Geotechnique 29, 47–65.

[2] Forsmo, S., 2007. Influence of green pellet properties on pelletizing of magnetite iron ore. Ph.D. thesis, Luleå University of Technology, Luleå.

[3] Jean, M., 1999. The non-smooth contact dynamics method. Computer Methods in Applied Mechanics and Engineering 177, 235–257.

[4] Lacoursière, C., 2007. Regularized, stabilized, variational methods for multibodies. In: Peter Bunus, D. F., Führer, C. (Eds.), The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), 30-31 October, 2007, Göteborg (Särö), Sweden. Linköping Electronic Conference Proceedings. Linköping University Electronic Press, pp. 40–48.

[5] LKAB, 2011. URL `http://www.lkab.com`

[6] Moreau, J. J., 1999. Numerical aspects of the sweeping process. Computer Methods in Applied Mechanics and Engineering 177, 329–349.

[7] Pöschel, T., Schwager, T., 2005. Computational Granular Dynamics, Models and Algorithms. Springer-Verlag.

[8] Radjai, F., Richefeu, V., 2009. Contact dynamics as a nonsmooth discrete element method. Mechanics of Materials 41 (6), 715–728.

[9] Shuyan, W., Xiang, L., Huilin, L., Long, Y., Dan, S., Yurong, H., Yonglong, D., 2009. Numerical simulations of flow behavior of gas and particles in spouted beds using frictional-kinetic stresses model. Powder Technology 196 (10), 184193

[10] Servin, M., Bodin, K., Lacoursière, C., Wang, D., 2013. Examining the smooth and nonsmooth discrete element approach to granular matter. Accepted for publication in International Journal for Numerical Methods in Engineering (2013).
URL `http://umit.cs.umu.se/granular/dem`

[11] Srivastava, A., Sundaresan, S., 2003. Analysis of a frictionalkinetic model for gasparticle flow. Powder Technology 129 (1), 72-85.

[12] van Wachem, B. G. M., Almstedt, A. E., 2003. Methods for multiphase computational fluid dynamics. Chemical Engineering Journal, 96 (13), 81-98.
12

Fig. 1: A balling plant showing part of the balling drum, outlet, belt conveyor and roller screen. Picture courtesy of LKAB.

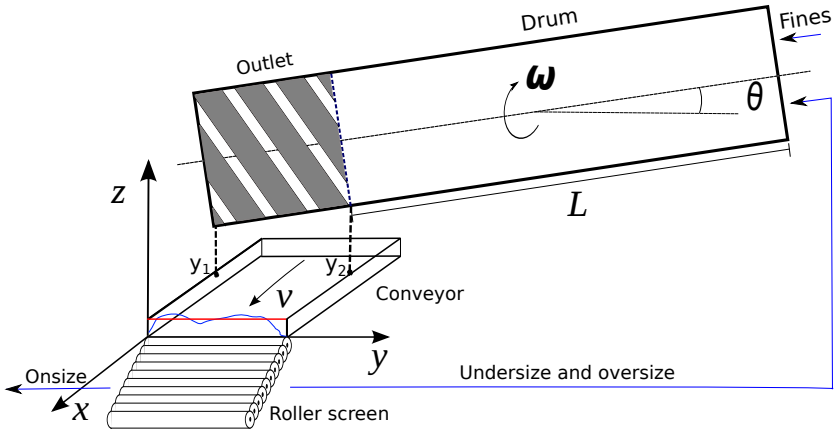

Fig. 2: Drum configuration.

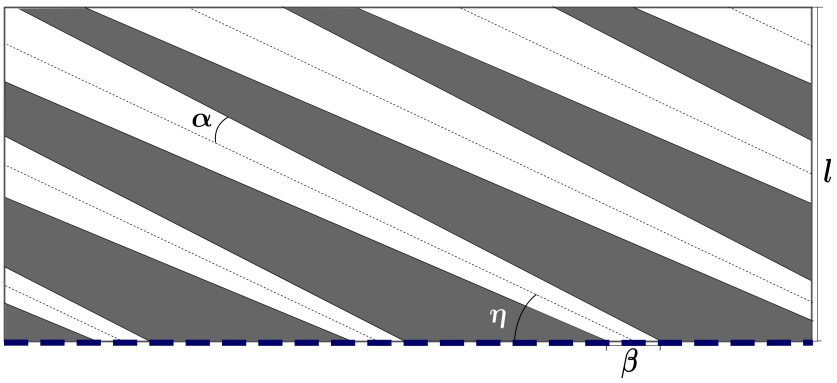

Fig. 3: 2D-projection of the outlet clarifying the design parameters angle and width.

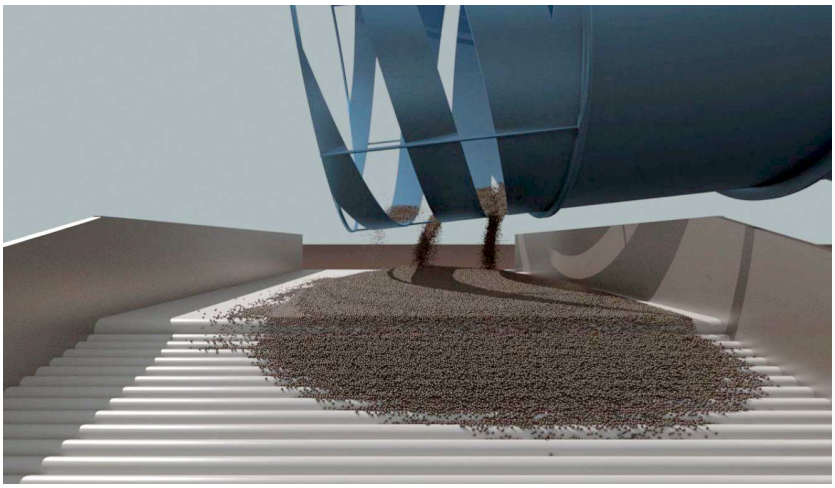

Fig. 4: Capture from simulation showing material flow from the outlet onto the conveyor and roller screen.

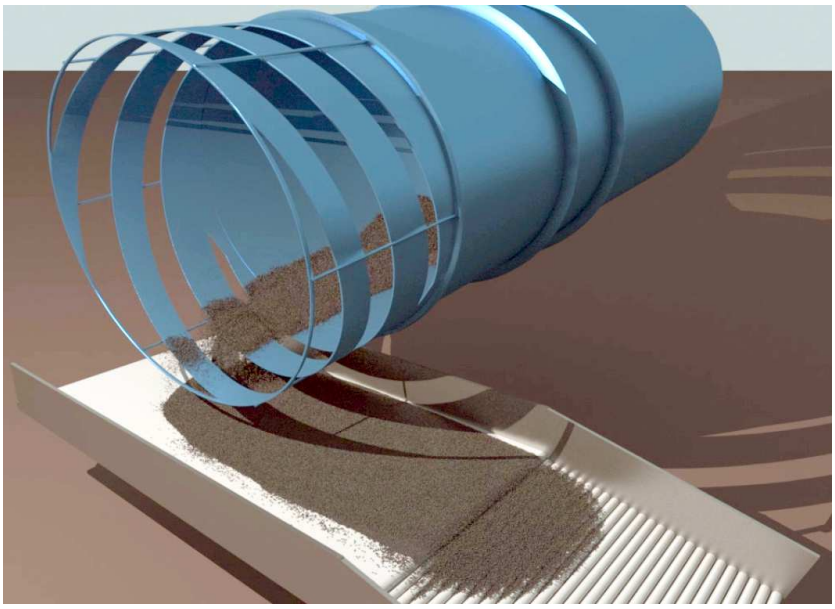

Fig. 5: Capture from simulation showing overview of the balling plant.

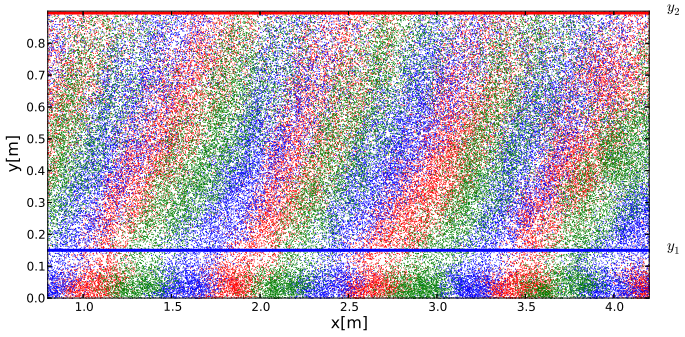

Fig. 6: The particles hitting the conveyor is stored as point scatter in a 2D surface. Particles are colored in red, blue or green depending of which of the three gaps it exited. The projected position of the outlet is at y_1 and y_2.

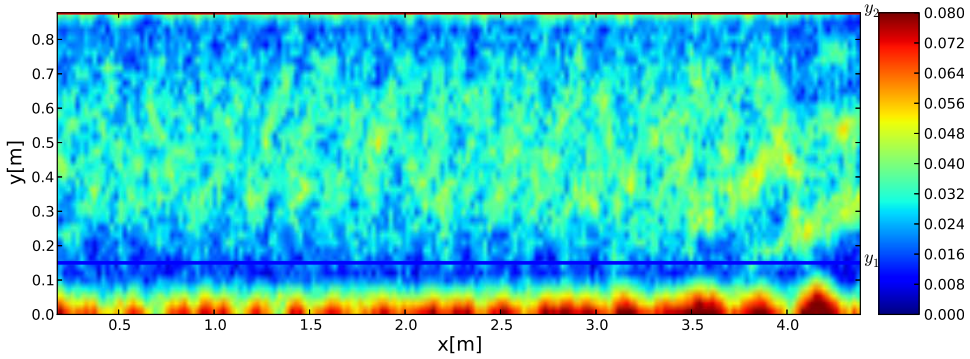

Fig. 7: Particle 2D height surface on conveyor reconstructed from particle impacts. The color codes the height in units m. The projected position of the outlet is at y_1 and y_2.

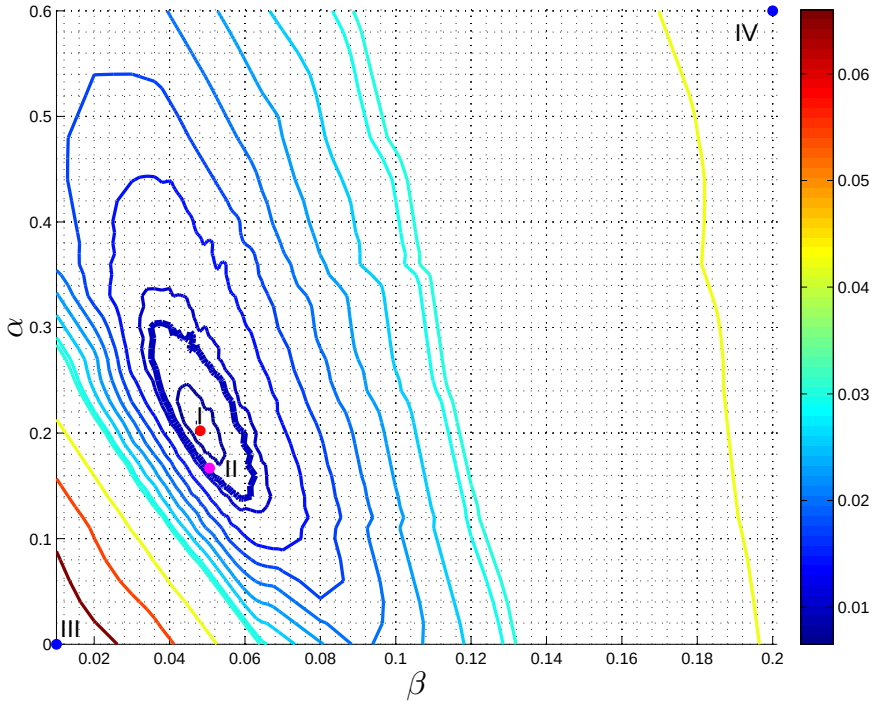

Fig. 8: Contour plot of objective function $f(\alpha, \beta)$ from simulations. The color codes the value of the objective function in unit m.

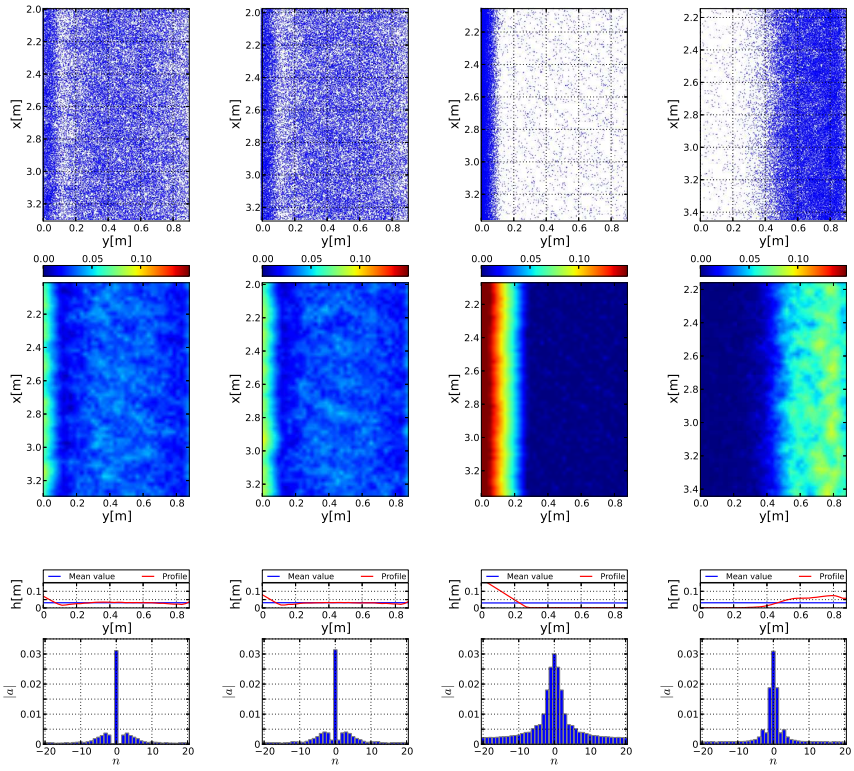

Fig. 9: Simulation sample data from the points I-IV (left to right) in the design space, as described in Table 1. The subplots show raw particle scatter data on the belt conveyor from one drum evolution (first row), reconstructed height surface (second row) with the color indicating the height in units m, time-averaged height profile (third row) and the corresponding FFT spectrum $\frac{1}{N}||a_n||$ (fourth row).

Parametrization and validation of a nonsmooth discrete element method for simulating flows of iron ore green pellets

D. Wang[a], M. Servin[b], T. Berglund[c], K-O. Mickelsson[d], S. Rönnbäck[e]

[a] *Umeå University, Umeå, Sweden.*
[b] *Umeå University, Umeå, Sweden (martin.servin@physics.umu.se).*
[c] *Algoryx Simulation AB, Umeå, Sweden.*
[d] *LKAB R&D, Malmberget, Sweden.*
[e] *Optimation AB, Luleå, Sweden.*

Abstract

The nonsmooth discrete element method (NDEM) have the potential of high computational efficiency for rapid exploration of large design space of systems for processing and transportation of mineral ore. We present parametrization, verification and validation of a simulation model based on NDEM for iron ore green pellet flow in balling circuits. Simulations are compared with camera based measurements of individual pellet motion as well as bulk behaviour of pellets on conveyors and in rotating balling drum. It is shown that the NDEM simulation model is applicable for the purpose of analysis, design and control of iron ore pelletizing systems. The sensitivity to model and simulation parameters is investigated. It is found that: the errors associated with large time-step integration do not cause statistically significant errors to the bulk behaviour; rolling resistance is a necessary model component; and the outlet flow from the drum is sensitive to fine material adhering to the outlet creating a thick coating that narrows the outlet gaps.

Keywords: granular materials; discrete element method; validation; iron ore pellets; pelletizing; balling circuit
2010 MSC: 00-01, 99-00

1. Introduction

Numerical simulation of granular materials is an important tool both for advancing the fundamental understanding of many natural phenomena in material science and geophysics, and for the design, control and optimization of systems for processing, manufacturing, storage and transportation of granular materials, e.g., grains, corn, pharmaceuticals pills, pellets, soil and minerals. In the mineral processing industry, experiments and *in situ* measurements are many times prohibitive for practical and economical reasons, and in these cases,

modeling and simulation play an essential role in finding deeper understanding of the process, making radical improvements and innovating entirely new solutions.

Parametrization, verification and validation are critical steps for making sure that the simulation model provides a sufficiently accurate representation of the real system. By parametrization we mean the process of identifying numerical values of the model parameters from observations of the real system. By the verification it is established that the computer simulation reproduces the mathematical model. A failure indicates either a flaw in the numerical method or in the software implementation. Validation is testing the agreement between the simulated model and the real system. This determines the predictive power of the simulated model to some given degree of accuracy of a selected set of observables. A significant disagreement implies that the model is not useful for describing the systems behaviour.

We consider the use of large-scale granular matter simulation based on the nonsmooth discrete element method (NDEM) [1, 2] for the design of balling drum outlets [3] used in iron ore pelletizing [4]. The NDEM have the potential of high computational efficiency compared to conventional (smooth) DEM. This enables rapid exploration of the design space. The NDEM is on the other hand not as well tested as conventional DEM for industry applications and scarcely put to validation tests. In this paper we present procedure and results for parametrization of the properties of green iron ore pellets and validation of the macroscopic bulk behaviour by comparing the numerical simulations with camera based measurements. The measurements include tracking of individual iron ore green pellets and characterization of bulk behaviour in an industrial pelletizing system. The goal is to establish the predictive power of NDEM simulations for the purpose of design and control of pelletizing systems, including the sensitivity of the flow characteristics with respect to certain model parameters. The NDEM method in [2] is also extended to include a constraint based rolling resistance which is shown to be crucial for the material distribution of iron ore green pellets.

2. Background

2.1. Iron ore pelletizing

The iron ore pelletizing process usually has the following main stages [4]. Comminuted fine size ore, *fines*, is first mixed with binder material. Agglomeration into soft ore balls, *green ore pellets*, occur in balling circuits where fines, water and undersized pellets are fed into rotating drums. In the drum flow the green pellets are mixed with fine material and grow by layering and coalescence. New pellets are formed by nucleation. The drum is slightly inclined to produce an axial flow. The green pellets leave the drum through an outlet and are size distributed on a roller sieve, see Fig. 2. Under-sized particles are fed back to the drum. Over-sized pellets are crushed and mixed with the fines. On-sized pellets (9 to 16 mm in diameter) are conveyed to the induration furnace where

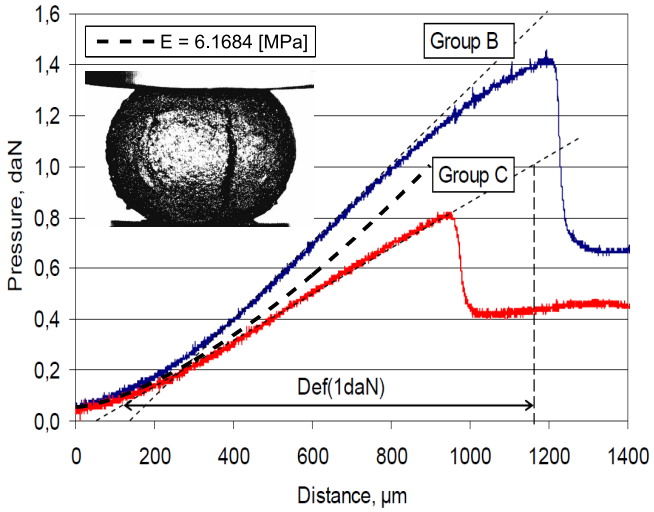

Figure 1: Measurement of elasticity and strength of iron ore green pellets from Forsmo *et al* in Ref. [20] Fig. 1 and 13c.

they form hard pellets by oxidation and sintering. After this stage the cooled pellets are ready for transportion to distant steelmills. A typical iron ore balling circuit may have drum diameter ranging between $3-5$ m and $8-10$ m long and circulate about $400-1200$ ton/h producing $100-300$ ton/h on-size pellets.

The mathematical modeling of granulation systems was reviewed in Ref. [5]. A smooth DEM simulation model of iron ore granules in a continuous drum mixer was developed in [6] to analyse the flow dependence on drum design (angle and length). In [3] a methodology based on the nonsmooth discrete element method (NDEM) was presented for simulation based design of drum outlets, for even flow profile of ore green pellets on to the roller sieve. Fig. 2 show an image from outlet analysis using NDEM simulation. The simulation demonstrate that the original outlet design was far from optimal as the material distribution on the wide-belt conveyor is inhomogeneous. As an effect, the roller sieve cannot be used efficiently. Furthermore, the green pellets may be damaged by the pressure from building a too thick pellet bed. A simulation model for the analysis and design of the balling process must be able to predict the flow and distribution of material both inside the balling drum and on the conveyor belt below the outlet.

2.2. Nonsmooth discrete element methods

In the conventional discrete element method (DEM) the granules are modeled as rigid bodies interacting by contact forces modeled as linear or non-linear damped springs. We refer to this as *smooth* DEM as it involves the numerical integration of smooth (but usually stiff) ordinary differential equations. The computational aspects of smooth DEM is covered in Ref. [7]. In the *nonsmooth*

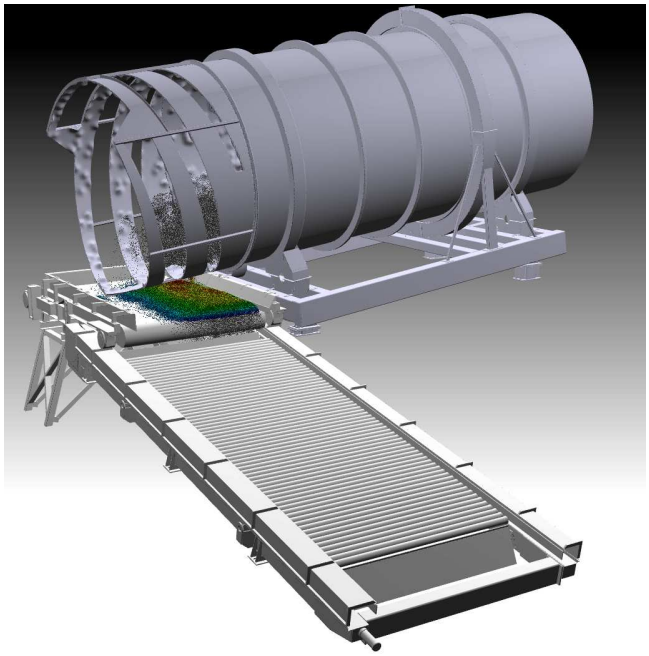

Figure 2: Image from simulation of balling drum with green ore pellets flowing through the outlet gaps onto the wide-belt conveyor feeding the roller sieve.

DEM [8, 9, 1], impacts and frictional stick-slip transitions are considered as instantaneous events making the velocities discontinuous in time. The contact forces and impulses are modeled in terms of kinematic constraints and complementarity conditions between constraint forces and contact velocities, e.g., by the Signorini-Coulomb law for unilateral non-penetration and dry friction. The contact network become strongly coupled and any dynamic event may propagate through the system instantaneously. The benefit of nonsmooth DEM is that it allows integration with much larger simulation step-size than for smooth DEM and is thus potentially faster.

We use a regularized version of nonsmooth DEM referred to as *semi-smooth* DEM in Ref. [2], which combines the numerical stability at large step-size with the possibility of modeling the viscoelastic nature of the contact forces and mapping the simulation parameters to the conventional material parameters. The constrained equations of motion, between impacts, are

$$\begin{aligned} \mathbf{M}\dot{\mathbf{v}} + \dot{\mathbf{M}}\mathbf{v} &= \mathbf{f}_{\text{ext}} + \mathbf{G}_{\text{n}}^{\text{T}}\boldsymbol{\lambda}_{\text{n}} + \mathbf{G}_{\text{t}}^{\text{T}}\boldsymbol{\lambda}_{\text{t}} && (1)\\ 0 \le \varepsilon_{\text{n}}\boldsymbol{\lambda}_{\text{n}} + \mathbf{g}_{\text{n}}(\mathbf{x}) &\perp \boldsymbol{\lambda}_{\text{n}} \ge 0 && (2)\\ \gamma_{\text{t}}\boldsymbol{\lambda}_{\text{t}} + \mathbf{G}_{\text{t}}(\mathbf{x})\mathbf{v} &= 0 && (3)\\ |\boldsymbol{\lambda}_{\text{t}}^{(\alpha)}| &\le \mu|\mathbf{G}_{\text{n}}^{(\alpha)\text{T}}\lambda_{\text{n}}^{(\alpha)}| && (4) \end{aligned}$$

where $\mathbf{x}$, $\mathbf{v}$ and $\mathbf{f}_{\text{ext}}$ are global vectors of position, velocity and external force, and $\mathbf{M}$ is the system mass matrix. Rotational degrees of freedom are included such that $\mathbf{v}$ and $\mathbf{f}_{\text{ext}}$ are vectors of dimension $6N_{\text{p}}$ including components of angular velocity and torque. The constraint forces for maintaining the non-penetration constraint and Coulomb friction are $\mathbf{G}_{\text{n}}^{\text{T}}\boldsymbol{\lambda}_{\text{n}}$ and $\mathbf{G}_{\text{t}}^{\text{T}}\boldsymbol{\lambda}_{\text{t}}$, where $\boldsymbol{\lambda}_{\text{n}}$ and $\boldsymbol{\lambda}_{\text{t}}$ are the Lagrange multipliers for the normal (n) and tangential (t) directions of each contact plane. The corresponding contact Jacobians are $\mathbf{G}_{\text{n}}$ and $\mathbf{G}_{\text{t}}$. In absence of regularization the constraints express non-penetration, $\delta \ge 0$, where δ is the gap function, and no-slip, $\mathbf{G}_{\text{t}}\mathbf{v} = 0$. The constraints can be regarded as the limit of infinitely strong potentials and dissipation functions $U_\varepsilon(\boldsymbol{x}) = \frac{1}{2\varepsilon}\boldsymbol{g}^T\boldsymbol{g}$ and $\mathcal{R}_\gamma(\boldsymbol{x}, \dot{\boldsymbol{x}}) = \frac{1}{2\gamma}\bar{\boldsymbol{g}}^T\bar{\boldsymbol{g}}$ with $\varepsilon, \gamma \to 0$ [10, 12]. Eq. (2) and (3) are the regularized versions of non-penetration and no-slip, with regularization parameters ε_{n} and γ_{t}. We use finite regularization and map the non-penetration constraint, for each contact α with gap function $\delta_{(\alpha)}$, to the Hertz contact force law, $f_{(\alpha)} = k_{\text{n}}\delta_{(\alpha)}^{3/2}$, by defining $g_{\text{n}}^{(\alpha)} \equiv \delta_{(\alpha)}^{e_{\text{H}}}$ with exponent $e_{\text{H}} = 5/4$. This maps the regularization parameter to the Hertz spring coefficient and conventional material parameters as $\varepsilon_{\text{n}} = e_{\text{H}}/k_{\text{n}} = 3e_{\text{H}}(1-\nu^2)/E\sqrt{r^*}$, where E is the Young's modulus, ν the Poisson ratio and r^* is the effective contact radius. Similarly, a dissipation term, $\gamma_{\text{n}}\mathbf{G}_{\text{n}}\mathbf{v}$, may also be added to the normal constraint multiplier condition (2). This produce a viscous damping force term, $f_{\text{d}} = k_{\text{n}}c\sqrt{\delta}\dot{\delta}$, in Hertz contact law with the damping parameter defined as $\gamma_{\text{n}} = e_{\text{H}}^2/k_{\text{n}}c$, which relates it also to the physical viscosity constant η by $c = 4(1-\nu^2)(1-2\nu)\eta/15E\nu^2$ [11]. See Ref. [2] for details on the mapping of regulization parameters. The friction constraint impose zero tangential contact velocity, $\mathbf{G}_{\text{t}}^{(\alpha)}\mathbf{v}_{(\alpha)} = 0$, unless the tangent force reach the friction bounds set by the Coulomb law, Eq. (4). The

tangent plane is spanned with two orthogonal vectors $\mathbf{t}_1^{(\alpha)}$ and $\mathbf{t}_2^{(\alpha)}$ resulting in friction multiplier having two components $\boldsymbol{\lambda}_{\rm t}^{(\alpha)} = [\boldsymbol{\lambda}_{\rm t_1}^{(\alpha)}\ \boldsymbol{\lambda}_{\rm t_2}^{(\alpha)}]^{\rm T}$. When impacts occur, the equations of motion are supplemented by the Newton impact law, $\mathbf{G}_{\rm n}^{(\alpha)}\mathbf{v}_+ = -e\mathbf{G}_{\rm n}^{(\alpha)}\mathbf{v}_-$, with coefficient of restitution e.

For numerical time integration we use the SPOOK stepper [12] derived from discrete variational principle for the augmented system $(\mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}, \dot{\boldsymbol{\lambda}})$. Stepping the system position and velocity, $(\mathbf{x}_i, \mathbf{v}_i) \to (\mathbf{x}_{i+1}, \mathbf{v}_{i+1})$, from time t_i to $t_{i+1} = t_i + \Delta t$ involve solving a mixed linear complementarity problem (MLCP) [13] of the form

$$\begin{gathered} \mathbf{H}\mathbf{z} + \mathbf{b} = \mathbf{w}_l - \mathbf{w}_u \\ 0 \le \mathbf{z} - \mathbf{l} \perp \mathbf{w}_l \ge 0 \\ 0 \le \mathbf{u} - \mathbf{z} \perp \mathbf{w}_u \ge 0 \end{gathered} \tag{5}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{M} & -\mathbf{G}_{\rm n}^{\rm T} & -\mathbf{G}_{\rm t}^{\rm T} \\ \mathbf{G}_{\rm n} & \boldsymbol{\Sigma}_{\rm n} & 0 \\ \mathbf{G}_{\rm t} & 0 & \boldsymbol{\Sigma}_{\rm t} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{v}_{i+1} \\ \boldsymbol{\lambda}_{{\rm n},i+1} \\ \boldsymbol{\lambda}_{{\rm t},i+1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\mathbf{M}\mathbf{v}_i - \Delta t\mathbf{M}^{-1}\mathbf{f}_{\rm ext} \\ \frac{4}{\Delta t}\boldsymbol{\Upsilon}_{\rm n}\mathbf{g}_{\rm n} - \boldsymbol{\Upsilon}_{\rm n}\mathbf{G}_{\rm n}\mathbf{v}_i \\ 0 \end{bmatrix} \tag{6}$$

and the solution vector $\mathbf{z}$ contains the new velocities and the Lagrange multipliers $\boldsymbol{\lambda}_{\rm n}$ and $\boldsymbol{\lambda}_{\rm t}$. For notational convenience, a factor Δt has been absorbed in the multipliers such that the constraint force reads $\mathbf{G}^T\boldsymbol{\lambda}/\Delta t$. The diagonal matrices $\boldsymbol{\Sigma}_{\rm n}$, $\boldsymbol{\Sigma}_{\rm t}$ and $\boldsymbol{\Upsilon}_{\rm n}$ are given in Appendix A in terms of the viscoelastic material parameters. The upper and lower limits, u and l in Eq. (6), follow from Signorini-Coulomb law including $0 \le \lambda_{\rm n}^{(\alpha)}$ and $|\boldsymbol{\lambda}_{\rm t}^{(\alpha)}| \le \mu_{\rm s}|\mathbf{G}_{\rm n}^{(\alpha)\rm T}\boldsymbol{\lambda}_{\rm n}^{(\alpha)}|$ with the friction coefficient $\mu_{\rm s}$. w_l and w_u are temporary slack variables. Impacts are treated *post facto*. After stepping the velocities and positions an impact stage follows. This include solving a MLCP similar to Eq. (5) but with the Newton impact law, $\mathbf{G}_{\rm n}^{(\alpha)}\mathbf{v}_+ = -e\mathbf{G}_{\rm n}^{(\alpha)}\mathbf{v}_-$, replacing the normal constraints for the contacts with normal velocity larger than an impact velocity threshold $v_{\rm imp}$. The remaining constraints are maintained by imposing $\mathbf{G}\mathbf{v}_+ = 0$.

We use a projected Gauss-Seidel (PGS) algorithm, as described in Ref. [2] and summarized in Appendix A, for solving the MLCP (5). The method is implemented in the software AgX Dynamics [14]. The time-step Δt should be chosen

$$\Delta t \lesssim \min(\epsilon d/v_{\rm n}, \sqrt{2\epsilon d/g}) \tag{7}$$

for contact error threshold ϵ, where $v_{\rm n}$ is the characteristic relative normal contact velocity and $g_{\rm acc} = 9.82\ {\rm m/s^2}$ is the gravitational acceleration. We set the impact velocity threshold to $v_{\rm imp} = \epsilon d/\Delta t$. The required number of PGS iterations depend on the size and configuration of the contact network. For bulk systems a rough rule is $N_{\rm it} = 0.1 \times n/\epsilon$, where n is the length of the contact network (number of contacts) in the direction of gravity [2]. The computational time $t_{\rm comp}$ per simulated time $t_{\rm real}$ is

$$t_{\rm comp} = \frac{\Omega}{h} t_{\rm real} \tag{8}$$

where $\Omega = K_{\text{cpu}} N_{\text{it}} \alpha_{\text{p}} N_{\text{p}} / N_{\text{cpu}}$, α_{p} is the average number of contacts per particle, N_{cpu} the number of cpu cores and K_{cpu} the average computational time for a single PGS update. We measure $K_{\text{cpu}} \approx 1^{-6}$ s with a desktop computer with Intel(R) Core(TM) Xeon X5690, 3.46 GHz, 48 GB RAM on a Linux 64 bit system. The PGS implementation parallelizes well up to 8 cores but saturates beyond that.

2.3. Rolling resistance constraint

There are several physical causes for rolling resistance, see, e.g., Ref. [15]. These include the effect of particle shape deviating from a spherical idealization, plastic or viscous deformations of the object itself or in the contact interface, frictional slippage in the contact interface and surface adhesion. In the idealization of rigid bodies, rolling resistance may be modeled as a torque, $\boldsymbol{\tau}_{\text{r}}$, on the contacting bodies counteracting their relative rolling motion. Similarly to Coulomb friction, the rolling resistance torque is limited in magnitude by $|\boldsymbol{\tau}_{\text{r}}| \leq \mu_{\text{r}} r^*_{ab} |\mathbf{f}_{\text{n}}|$, where $0 \leq \mu_{\text{r}}$ is the *rolling resistance coefficient*, $r^*_{ab} = r_a r_b / (r_a + r_b)$ is the *effective contact radius* of two contacting geometries a and b. When the source of rolling resistance is purely geometric the rolling resistance coefficient can be derived from the shape. For octagon shape $\mu_{\text{r}} = 0.1$ [16]. Rolling resistance is important to include for correct prediction of single particle motion as well as for the collective behaviour of granular materials, e.g., formation of stable piles with accurate angle of repose, stress and strain relationships in dense packings and the shear rate in flowing systems. An overview of the conventional smooth DEM models and their agreement with experiments is found in Ref. [15]. Some rolling resistance models in smooth DEM work well for quasi-static systems while poorly for flowing systems and vice versa. In nonsmooth DEM, where the contact forces and dynamics are computed implicitly, e.g., as kinematic constraints, a single model can be used for both regimes. Only a few models of rolling resistance for nonsmooth DEM can be found in literature [17, 18, 19] and no reported results concerning parametrization and validation with experimental data.

We extend the nonsmooth DEM model by including additional rolling resistance constrains in Eq. (5) of the form $\mathbf{G}_{\text{r}}\mathbf{v} = 0$ and constraint force (torque) $\boldsymbol{\tau}_{\text{r}} = \mathbf{G}_{\text{r}}^T \boldsymbol{\lambda}_{\text{r}} / h$. Consider a contact α between two granules, a and b, with linear and angular velocity vectors denoted $\mathbf{u}$ and $\boldsymbol{\omega}$. Let the contact plane have unit normal $\mathbf{n}$ and orthonormal tangents $\mathbf{t}_1$ and $\mathbf{t}_2$. This is illustrated in Fig. 3. The condition for zero relative rolling velocity can be expressed

$$\mathbf{0}_{3\times 1} = \begin{bmatrix} \mathbf{t}_1^{\text{T}}(\boldsymbol{\omega}_a - \boldsymbol{\omega}_b) \\ \mathbf{t}_2^{\text{T}}(\boldsymbol{\omega}_a - \boldsymbol{\omega}_b) \\ \mathbf{n}^{\text{T}}(\boldsymbol{\omega}_a - \boldsymbol{\omega}_b) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{0}_{1\times 3} & \mathbf{t}_1^{\text{T}} & \mathbf{0}_{1\times 3} & -\mathbf{t}_1^{\text{T}} \\ \mathbf{0}_{1\times 3} & \mathbf{t}_2^{\text{T}} & \mathbf{0}_{1\times 3} & -\mathbf{t}_2^{\text{T}} \\ \mathbf{0}_{1\times 3} & \mathbf{n}^{\text{T}} & \mathbf{0}_{1\times 3} & -\mathbf{n}^{\text{T}} \end{bmatrix}}_{\mathbf{G}_{\text{r}}} \begin{bmatrix} \mathbf{u}_a \\ \boldsymbol{\omega}_a \\ \mathbf{u}_b \\ \boldsymbol{\omega}_b \end{bmatrix} \tag{9}$$

from which we identify the constraint Jacobian for rolling resistance. By the geometry of the Jacobian it is clear that this constraint force is indeed a torque.

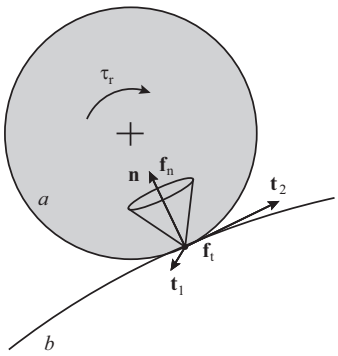

Figure 3: Illustration of two contacting granular geometries a and b. Rolling resistance constraint produce a torque, $\boldsymbol{\tau}_{\mathrm{r}}$, limited in magnitude relative to the normal contact force $\boldsymbol{f}_{\mathrm{n}}$ in a similar way as the Coulomb friction force $\boldsymbol{f}_{\mathrm{t}}$.

Table 1: Identified iron ore green pellet parameters.

ρ	3700 kg/m^3	mass density
d	12.7 ± 3 mm	diameter
E	6.2 ± 0.7 MPa	Young's modulus
e	0.18 ± 0.04	coefficient of restitution
μ_{s}	0.91 ± 0.04	surface friction coefficient
μ_{r}	0.32 ± 0.02	rolling resistance coefficient

We denote the components of the multiplier $\boldsymbol{\lambda}_{\mathrm{r}} = (\lambda_{\mathrm{r}t_1}, \lambda_{\mathrm{r}t_2}, \lambda_{\mathrm{r}n})^T$. The third constraint $0 = \mathbf{n}^{\mathrm{T}}(\boldsymbol{\omega}_a - \boldsymbol{\omega}_b)$ oppose relative twisting and produce torsional force. We linearize the limits on the torque, $|\boldsymbol{\tau}_{\mathrm{r}}| \leq \mu_{\mathrm{r}} r^*_{ab} |\mathbf{f}_{\mathrm{n}}|$, and obtain additional multiplier conditions to Eq. (5)

$$-\mu_{\mathrm{r}} r^*_{ab} \lambda_{\mathrm{n}} \leq \quad \lambda_{\mathrm{r}t_1} \leq \quad \mu_{\mathrm{r}} r^*_{ab} \lambda_{\mathrm{n}} \tag{10}$$

$$-\mu_{\mathrm{r}} r^*_{ab} \lambda_{\mathrm{n}} \leq \quad \lambda_{\mathrm{r}t_2} \leq \quad \mu_{\mathrm{r}} r^*_{ab} \lambda_{\mathrm{n}} \tag{11}$$

$$-\alpha_{\mathrm{r}n} \mu_{\mathrm{s}} r^*_{ab} \lambda_{\mathrm{n}} \leq \quad \lambda_{\mathrm{r}n} \leq \quad \alpha_{\mathrm{r}n} \mu_{\mathrm{s}} r^*_{ab} \lambda_{\mathrm{n}} \tag{12}$$

where $\alpha_{\mathrm{r}n}$ is a factor depending on shape with value zero for perfect spheres and up to 1 for shapes with contact surface area as for a cube. A regularization term $\boldsymbol{\Sigma}_{\mathrm{r}}$ is added to the new diagonal block in $\mathbf{H}$, see Appendix A, and the corresponding components in $\boldsymbol{b}$ are set to zero.

3. Identification of iron ore green pellet parameters

The identified parameters for onsize iron ore green pellets are summarized in Table 1.

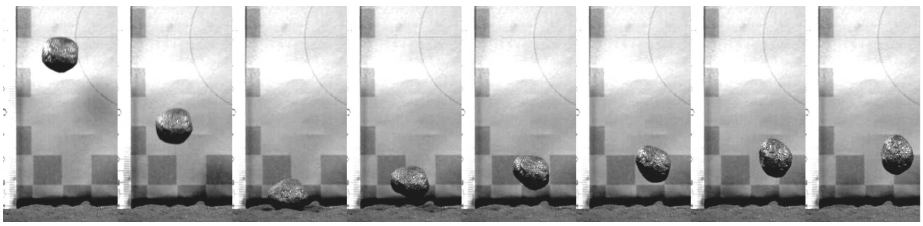

Figure 4: Measurement of coefficient of restitution by impacting ore green pellet captured at 100 Hz.

3.1. Mass and geometry

Iron ore green pellet have mass density of about 3700 kg/m^3. The shape is approximately spherical with diameter ranging between 9 and 16 mm and *shape factor*[1] in the range $0.7 - 0.95$.

3.2. Elasticity

The elasticity and strength of iron ore green pellets was investigated by Forsmo *et al* [20]. Assuming a relation between pressure force f_n and compression δ of the form of Hertz contact law, $f_\mathrm{n} = k_\mathrm{n}\delta^{3/2}$, we identify $k_\mathrm{n} = (0.35 \pm 0.05) \times 10^6$, which translates to Young's modulus $E = 3k_\mathrm{n}(1-\nu^2)/\sqrt{2d} = 6.2 \pm 0.7$ MPa , see Fig. 1. We assume Poisson ratio $\nu = 0.25$.

3.3. Restitution

The coefficient of restitution is identified from drop tests where ore green pellets impact on a surface coated with a 10 mm thick layer of ore material packed to similar density as of ore green pellets, see Fig. 4. The green pellet was released from height 0.45 m and bounces to a height of 12 ± 4 mm, which implies an impact velocity $v^- = 2.97$ m/s and a post impact velocity $v^+ = 0.49 \pm 0.08$ m/s. The coefficient of restitution $e = -v^+/v^-$ is thus found to be $e = 0.18 \pm 0.04$.

3.4. Surface friction

The friction coefficient, μ_s, between two ore surfaces is identified by measuring the required force f for pulling a block of packed ore over a surface of packed ore, see Fig. 5. From the Coulomb law, $f = \mu_\mathrm{s} m g_\mathrm{acc}$, we find $\mu_\mathrm{s} = 0.91 \pm 0.04$

3.5. Rolling resistance

The rolling resistance coefficient is determined from observing the angle ϕ_r at which the ore green pellet starts to roll down an inclined plane. We observe $\phi_\mathrm{r} = 17.8 \pm 0.1°$. With $\boldsymbol{\tau}_\mathrm{r} = (d/2)\sin(\phi_\mathrm{r}) f_\mathrm{n}$ we thus identify the rolling resistance coefficient to $\mu_\mathrm{r} = 0.32 \pm 0.02$.

[1]The shape factor of a cross-section of area A and perimeter length L is $4\pi A/L^2$. A sphere has shape factor 1 and a square has roughly 0.78.

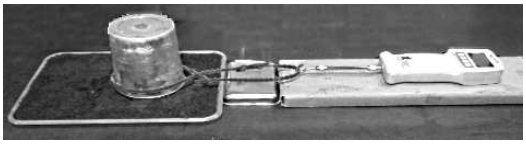


Figure 5: Measurement of surface friction (left) and rolling resistance (right).

Table 2: Summary of results from verification tests.

Test	**Quantity**	**Result**	**Comment**
Elasticity	$\|k_{\text{eff}} - k\|/k$	$0 - 2\%$	error 2% at $\rho/d \gtrsim 0.1$ and $\Delta t \gtrsim 5$ ms
Restitution	$\|e_{\text{eff}} - e\|/e$	1%	$\Delta t \lesssim \min(\epsilon d/v_{\text{n}}, \sqrt{2\epsilon d/g})$ and $\epsilon = 0.01$
Friction	$\|\mu_{\text{eff}} - \mu\|/\mu$	0%	fulfilled to machine precision
Rolling resistance	$\|\mu_{\text{r,eff}} - \mu_{\text{r}}\|/\mu_{\text{r}}$	$0 - 1\%$	1% error at rolling onset.

4. Verification of simulated pellets

Material parameters in Table 1 are translated to simulation parameters of NDEM according to Sec. 2.2 and verified in elementary tests described below. The results are summarized in Table 2. Simulation parameters are set to $\Delta t = 0.01$ s, $N_{\text{it}} = 150$ if nothing else is stated.

4.1. Elasticity

The elasticity model is verified in simulation by compressing a pellet between a moving piston and a static plane. The piston moves in 0.02 mm/s towards to the plane. The measured constraint force $\mathbf{G}_{\text{n}}^{\text{T}}\boldsymbol{\lambda}_{\text{n}}/\Delta t$ coincide with the Hertz model with an effective elasticity coefficient $k_{\text{eff}} = f_{\text{n}}/\delta^{3/2}$ deviating from k_{n} with maximally 2% at $\delta/d = 0.1$ for time step $\Delta t = 5$ ms. The deviation decrease for smaller overlap and with decreasing time step. In the case of a single particle compressed towards to the static ground by external force, the result match to machine precision.

4.2. Restitution

The impact model is verified by measuring the re-bounce height h_{b} from dropping particles from height h_{d} and computing the effective coefficient of restitution $e_{\text{eff}} = \sqrt{h_{\text{b}}/h_{\text{d}}}$. The impact stage reproduce the impact law $v_+ = -ev_-$ to machine precision. However, finite time-stepping cause impact overlaps of magnitude $\delta \leq \Delta t v_+$. This produce errors in the particle trajectories of the same order. Furthermore, the division of contacts into an impact stage and continuous contact stage can produce residual energy at contact separation that cause violation of the impact law. Thus, the effective dissipation can be smaller than predicted by the impact law, i.e., the effective restitution become bigger. The effect is more notable for low-restitution materials and appears as an

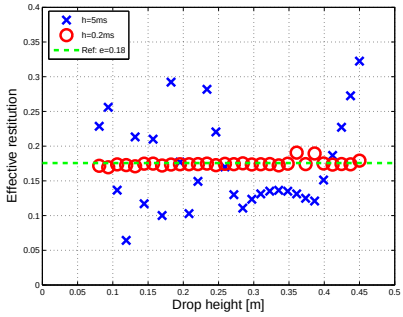

Figure 6: Verification of the impact model by measuring the effective restitution.

effective restitution larger than e. The test is performed for drop height ranging up to 0.45 m and time step 5 ms and 0.2 ms. The results are shown in Fig. 6. The effective restitution become 0.17 ± 0.07 and 0.175 ± 0.004, respectively. Note that given the drop-height 0.45 m and an error tolerance of $\epsilon = 2\%$, the time-step rule $\Delta t \lesssim \epsilon d/v_{\mathrm{n}}$ imply $\Delta t = 0.4$ ms. The time-step 5 ms, on the other hand, correspond to an error tolerance of $\epsilon = 100\%$. The verification results are thus in good agreement with these error estimates but it is clear that using too large time-step may cause significant errors in energy dissipation at impacts.

4.3. Surface friction

The surface friction model is verified by simulating a pellet being pressed towards a static plane and pulled horizontally until it starts to slide. The effective friction coefficient is computed as the ratio of the horizontally applied force at slide onset over applied normal pressure, $\mu_{\mathrm{s,eff}} = f_{\mathrm{t}}/f_{\mathrm{n}}$. The result agree with μ_{s} to machine precision.

4.4. Rolling resistance

As verification of rolling resistance we measure the maximum angle ϕ_{r} where a simulated ore green pellet doesn't start rolling on an inclined plane and compute the effective rolling resistance coefficient $\mu_{\mathrm{r,eff}} = \sin(\phi_{\mathrm{r}})$. The result is $\phi_{\mathrm{r}} = 17.84^{\circ}$ and $\mu_{\mathrm{r,eff}} = 0.31$, to be compared to the corresponding values $17.8 \pm 1^{\circ}$ and 0.32 from experiment. The discrepancy is due to truncation error of the residual in the Gauss-Seidel solver but is of no practical significance to the results in the paper. As a complementary verification test, we simulate the deceleration of a fast rolling pellet on a horizontal plane assuming no-slip. The result agree with the analytical solution $\dot{v} = -\frac{5}{7}\mu_{\mathrm{r}} g_{\mathrm{acc}}$ to machine precision.

5. Observed bulk behaviour of iron ore green pellets

We use two on-line production balling circuits at LKAB pelletizing plant in Malmberget, Sweden, for observation and validation of iron ore green pellet

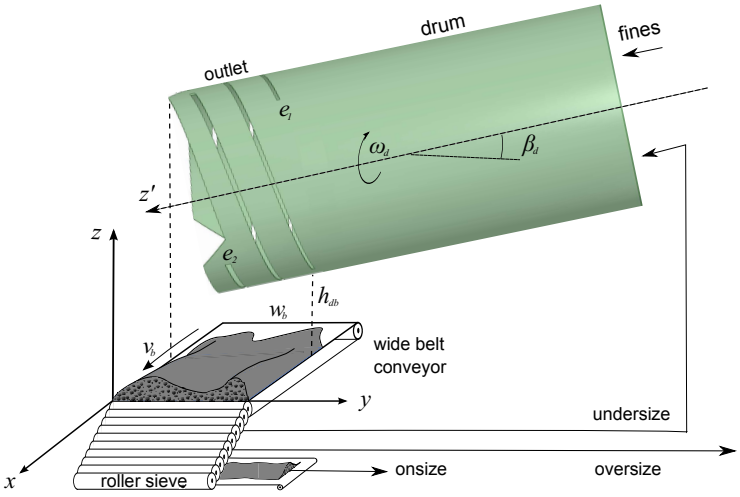

Figure 7: Illustration of the balling circuit.

Table 3: Specification of balling circuit parameters.

Notation	**Value**	**Parameter**
ω_d	0.53 rad/s	drum rotation speed
β_d	7°	drum inclination
D	3.7 m	drum inner diameter
L	8.1 m	drum length
$\{e_1 : e_2\}$	$\{4.7 : 7.9\}$ d	inner : outer gap width
w_b	2.4 m	width of wide-belt conveyor
v_b	0.19 m/s	speed of wide-belt conveyor
h_db	0.45 m	distance drum to conveyor
$\dot{M}_\mathrm{RK1}$	340 ton/h	mass flow rate in RK1
$\dot{M}_\mathrm{RK5}$	275 ton/h	mass flow rate in RK5

bulk behaviour. The circuits, referred to as RK1 and RK5, are identical but run with different feed rate. The balling process was described in Sec. 2.1 and the balling circuit is illustrated in Fig. 7. The key parameters are given in table 3. We observe and validate three bulk properties: the angle of repose of static piles on the conveyor of on-size pellets, the properties of the flow inside the balling drum and the spatial distribution of material on the wide-belt conveyor that results from the interaction of the flow with the outlet geometry.

5.1. Pile shape

The resting angle of repose, θ_r, is measured on the conveyor belt transporting on-size green pellets from the balling circuit to the induration furnace. An elongated pile is formed on the conveyor by feeding material from another conveyor, aligned perpendicularly to the first. The drop height is 0.3 m and feed

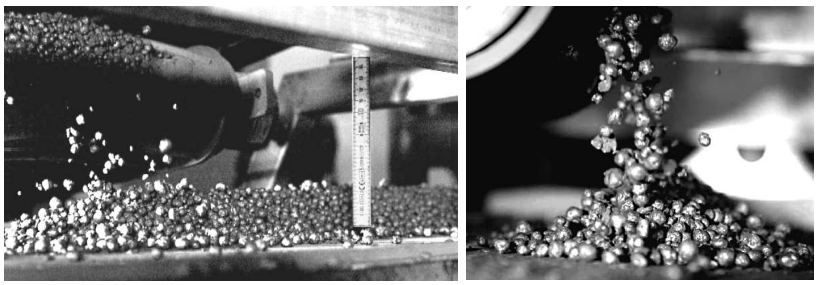

Figure 8: Photos from angle of repose measurement.

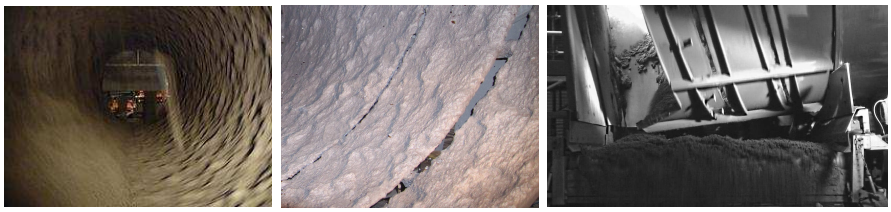

Figure 9: Sample images of drum interior (left and middle) and distribution on wide-belt conveyor (right). The bulk flow of ore pellets circulate in the lower left section of the drum. The fine material adhere to the drum walls and form an irregular surface coating. The middle image show a close-up of the irregular drum surface and coating of fine material around the outlet gaps.

rate 14.4 ton/h. The pile formation is filmed with high-speed cameras, see Fig.8. The average angle of repose of the pile is found to be $\theta_{\mathrm{r}} = 34 \pm 3°$.

5.2. Flow in an inclined drum

The flow inside balling drum RK1 was observed during 30 minutes of stable production of green ore pellets. A camera was placed at the end of the outlet to capture the flow inside the drum , see. Fig. 9. The drum was fed with mass rate $\dot{M}_{\mathrm{RK1}} = 340$ ton/h divided in 110 ton/h iron fines mixed with binding agents and 230 ton/h return feed of undersized material. The drum rotation produce a circulating flow that is nearly stationary and in the rolling or cascading regime [21]. At the bottom of the drum the material form a *plug zone* where ore pellets co-move rigidly with the drum rotation. The material is lifted up to some maximal angle θ_1 where particles begin to slide and form a *shear zone* of a gravity driven flow on top of the plug zone down to the drum bottom at angle θ_2, Fig. 9. The dynamic angle of repose is identified by the surface inclination, i.e., $\theta'_{\mathrm{r}} = 180° - \frac{1}{2}(\theta_1 + \theta_2)$.

From camera measurements it is found $\theta_1 = 120 \pm 2°$, $\theta_2 = 167 \pm 2°$ and $\theta'_{\mathrm{r}} = 35 \pm 5°$. The inclination of the drum also lead to an axial transportation flow, presumably localized to the shear zone. Cloth tracers are dropped into the drum and tracked by camera in order to measure the surface velocity of the

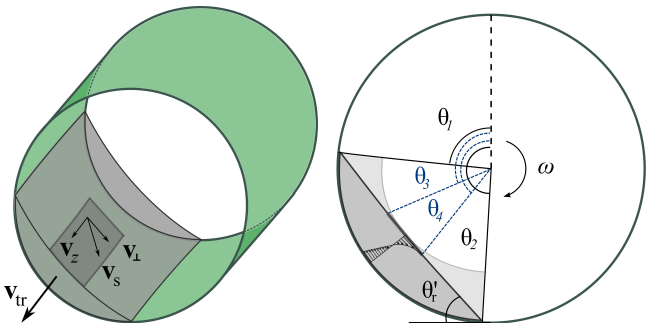

Figure 10: Pellet flow in rotating inclined drum.

bulk flow, v_s, and its axial and cross-sectional components, v_{sz} and $v_{s\perp}$. The measurement region is limited by the angles θ_3 and θ_4 and between the drum center and beginning of the outlet, as indicated in Fig. 10. The measurement results are found in Table 4. The measurement of θ_1 and θ_2 is based on a 100 s recording. The surface velocity is computed by time-of-flight from 10 passages of cloth tracers over the measurement region. We compute the axial bulk transportation velocity as $v_{tr} = \dot{M}/\rho A\chi$, where $A \in [A_{min}, A_{max}]$ is the bulk cross-section area and χ is the packing ratio. The upper and lower bounds of the cross-section area are determined by assuming either the shape of circle sector or of an annulus sector, both limited by θ_1 and θ_2, see Fig, 10. This give $A_{min} = 0.18$ m^2 and $A_{max} = 0.27$ m^2. The packing ratio is assumed $\chi = 0.7$.

The video material also reveal that the drum interior is not perfectly cylindrical but has a structure of bumps and dimples formed by fine material adhering and loosening from drum interior surface, see Fig. 9. This *drum texture* presumably lead to an increased effective surface friction, higher lifting of the material and induce more flow disturbances.

5.3. Material distribution on wide-belt conveyor

The spatial distribution of material on the wide-belt conveyor depends on the flow structure inside the drum and of the geometric shape of the outlet. The outlet is 2.3 m long and has three spiral shaped gaps. The inner and outer width of the gaps are $e_1 = 4.7d$ and $e_2 = 7.9d$ in the z' direction. We describe the resulting material height profile by $h(y, t - x/v_b)$. The goal is a constant height profile $h(y, t - x/v_b) = h_0$ which is presumed to maximize the efficiency of the roller sieve. The flow of material at the end of the wide-belt conveyor, $x = 0$, is captured using video camera over a time period of 16 s. The height profile is extracted by image analysis using feature matching to localize the conveyor belt and color gradient for tracking the material surface. A sample is shown in

Table 4: Validation of simulated bulk behaviour by comparing with observations in balling circuits.

Test	**Quantity**	**Observation**	**Simulation**
Pile shape	θ_{r}	$34 \pm 3^{\circ}$	$36 \pm 2^{\circ}$
Drum flow	θ'_{r}	$35 \pm 5^{\circ}$	$34 \pm 2^{\circ}$
	v_{tr}	0.20 ± 0.03 m/s	0.22 ± 0.02 m/s
	v_{s}	1.31 ± 0.06 m/s	1.27 ± 0.09 m/s
	v_{sz}	0.58 ± 0.05 m/s	0.49 ± 0.03 m/s
	$v_{s\perp}$	1.18 ± 0.07 m/s	1.17 ± 0.09 m/s
Bed profile RK1	σ_h	0.41	
RK1-o			0.83
RK1-m			0.66
RK1-c			0.46
RK1-ic			0.44
	$\sigma_{h-\bar{h}}$	0	
RK1-o			1.01
RK1-m			0.81
RK1-c			0.54
RK1-ic			0.27
Bed profile RK5	σ_h	0.33	
RK5-o			0.83
RK5-m			0.68
RK5-c			0.42
RK5-ic			0.42
	$\sigma_{h-\bar{h}}$	0	
RK5-o			0.71
RK5-m			0.47
RK5-c			0.35
RK5-ic			0.11

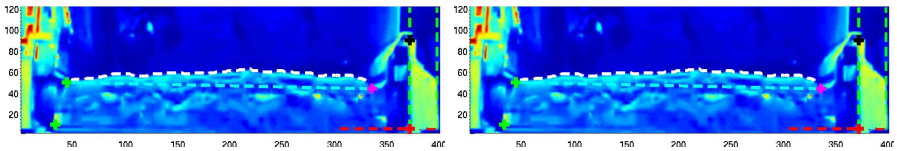

Figure 11: Sample images from extraction of height profile on wide-belt conveyor in RK5.

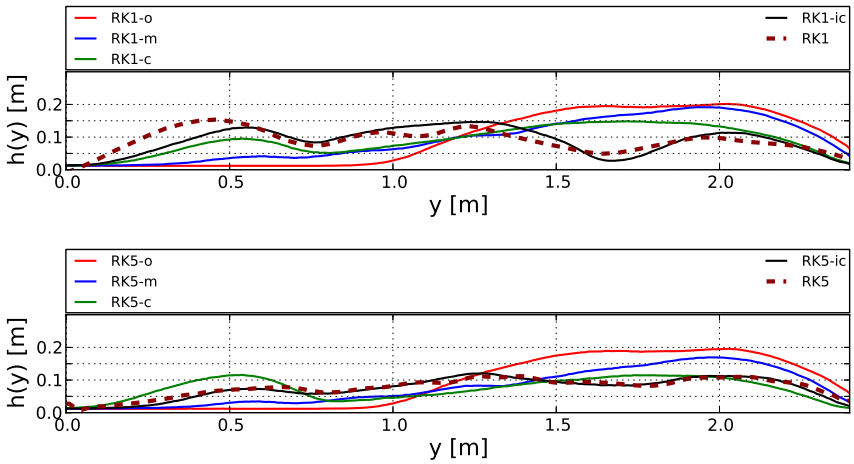

Figure 12: The time averaged height profile on the wide-belt conveyor from RK1 and RK5. The dashed line is the result from video measurement. The solid lines are from simulations using the original outlet (o) , modified outlet (m), with uniform gap coating (c) and with inhomogenous gap coating (ic).

Fig. 11. A 2D height profile $h(y, t - x/v_\mathrm{b})$ is reconstructed for comparison with simulations in Sec. 6. The time averaged profile at $x = 0$ is computed as

$$h(y) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} h(y, t) \mathrm{d}t \tag{13}$$

The result from RK1 and RK5 are found in Fig. 12. The coefficient of variation of the height profile is

$$\sigma_h = \sqrt{\frac{1}{w_\mathrm{b}} \int_0^{w_\mathrm{b}} \left[\frac{h(y) - \langle h \rangle}{\langle h \rangle} \right]^2 \mathrm{d}y} \tag{14}$$

where $\langle h \rangle = \frac{1}{w_\mathrm{b}} \int_0^{w_\mathrm{b}} h(y) \mathrm{d}y$ is the average height. The observed values are $\sigma_h^{\mathrm{RK1}} = 0.41$ and $\sigma_h^{\mathrm{RK5}} = 0.33$, which are acceptable although not optimal. The design objective of producing a uniform profile of pellets correspond to $\sigma_h = 0$.

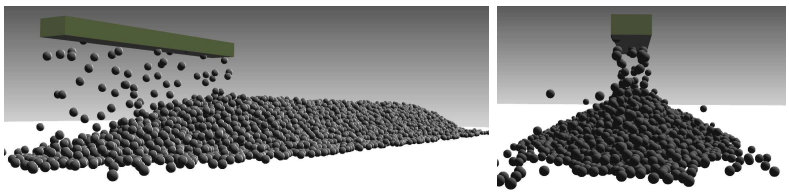

Figure 13: Images from simulation of pile formation.

6. Validation of the simulation model

The observations of ore green pellet flow in the balling circuit described in Sec. 5 are used for validation of simulated bulk behaviour. The validation results are summarized in Table 4. The simulation model is parametrized by the values in Table 1 found from experiments. Monosized particles with diameter d are used. All simulations are performed with time-step $\Delta t = 5$ ms, $N_{\text{it}} = 150$ PGS iterations and $v_{\text{imp}} = 0.05$ m/s, unless otherwise is mentioned. See Sec. 6.4 for an analysis and discussion regarding the choice of time-step. Videos from simulations are found at `http://umit.cs.umu.se/granular/video/`.

6.1. Angle of repose

A long pile is formed by dropping particles at rate 14.4 ton/h from a 0.05×0.4 m^2 source elevated 0.3 m over a planar surface moving 0.1 m/s relative to the source, see Fig. 13. The cross-section profile is measured as the average from 20 segments between $10d$ and $60d$ from the source, see Fig. 14. The average angle of repose is found to be $\theta_{\text{r}} = 36 \pm 2°$. To examine the sensitivity to rolling resistance, the simulation is also performed with rolling resistance coefficient μ_{r} decreased to 10%. The piles disperse more and the resulting angle of repose become 22°. With zero rolling resistance coefficient the initial pile formation has angle of 15° but quickly melt to 0°. Rolling resistance is clearly needed for the formation of piles with correct shape. A simulation with time-step $\Delta t = 0.2$ ms was also made to investigate the effect of errors from large time-step integration. The resulting angle of repose is $33 \pm 2°$. The difference is believed to reside from larger overlap errors occuring at impact in the $\Delta t = 5$ ms simulations. Both results are within the standard variation of the observed angle of repose $34 \pm 3°$.

6.2. Flow in an inclined drum

A virtual replica of the inclined rotating drum is constructed from CAD drawings. The interior drum texture is modeled by perturbing the cylindrical surface with a random distribution of gaussian shaped bumps of width 50 mm and height in the range 30 to 50 mm. The average density of bumps is 20 per m^2. The texture cover also the outlet interior. A particle source is placed at the center of the drum emitting particles of diameter d at a rate $\dot{M}_{\text{RK1}} = 340$ ton/h. The simulation is first run for five evolutions (60 s) to create a nearly

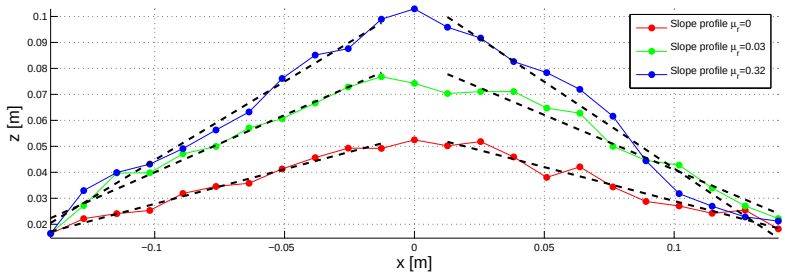

Figure 14: The average pile profile and its linear interpolation for different values of rolling resistance. For $\mu_\mathrm{r} = 0$ the pile quickly disperse to zero angle of repose.

stationary flow, see Fig. 16. The particle dynamics is then captured during one drum evolution, 12 s. The analysis is limited to a 1 m long section centred between the source and the outlet and, for the surface flow, between angles $\theta_3 = 135°$ and $\theta_4 = 160°$. Sample plot of the velocity field and mass density field from a cross-section is shown in Fig. 15 computed by coarse graining on a grid with mesh size of $1d$. The plug zone where material co-rotate rigidly with the drum is clearly visible. The axial transportation occur in the shear zone layer above the plug zone. The surface shape and flow is slightly irregular and nonstationary. The time averaged mass distribution is shown in Fig. 15. The average dynamic angle of repose is found to be $\theta'_\mathrm{r} = 34 \pm 2°$.

The average bulk transportation velocity in the measurement region is 0.22 ± 0.02 m/s. The surface between angles θ_3 and θ_4 is tracked over time and the average surface velocity is found to be $v_\mathrm{s} = 1.27 \pm 0.09$ with cross-sectional and axial components $v_{\mathrm{s}\perp} = 1.17 \pm 0.09$ m/s and $v_{\mathrm{s}z} = 0.49 \pm 0.03$ m/s.

To test the sensitivity to rolling resistance, the simulation is also performed with rolling resistance coefficient μ_r decreased from 0.33 to 0.03. The effect on the flow is significant. The dynamic angle of repose becomes $25 \pm 2°$ and the surface velocity $v_\mathrm{s} = 0.6 \pm 0.1$ with components $v_{\mathrm{s}\perp} = 0.43 \pm 0.1$ m/s and $v_{\mathrm{s}z} = 0.41 \pm 0.03$ m/s. Hence, rolling resistance is a necessary model component also for the simulated drum flow to agree with observations. The effect on the flow by variations of the surface friction, elasticity and particle size was also investigated and found to be small. The time-averaged cross-sectional flow velocities and dynamic angle of repose was affected by roughly 5 % by the changes $\mu'_\mathrm{s} = 0.9\mu_\mathrm{s}$, $E' = 0.5E$, $E' = 2E$ and $d' = 0.8d$.

6.3. Material distribution on wide-belt conveyor

The third validation test is the distribution of ore pellets on the wide-belt conveyor below the drum outlet. This tests the predictive power of NDEM simulation to capture the non-stationary granular flow created by the interaction with a moving irregular geometry. Sample images from simulation is shown in Fig. 16. Simulations are performed both with the original outlet design (RK1-o

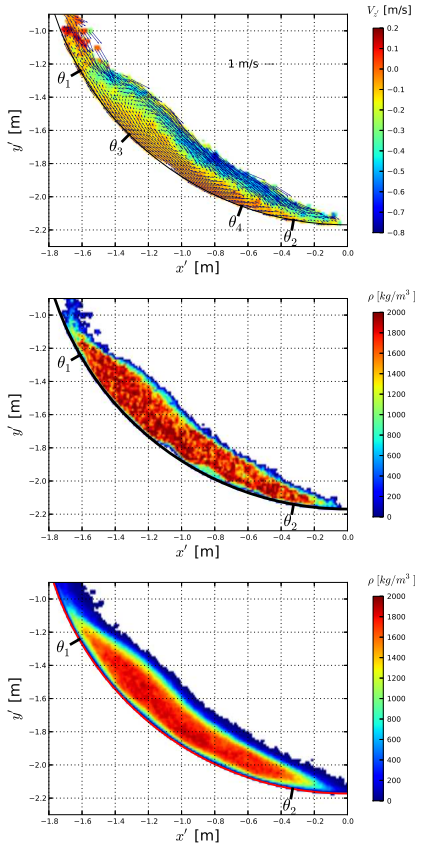

Figure 15: Time instants of flow velocity field (top), mass density field (middle) and time-averaged mass density field (bottom) in a drum cross-section.

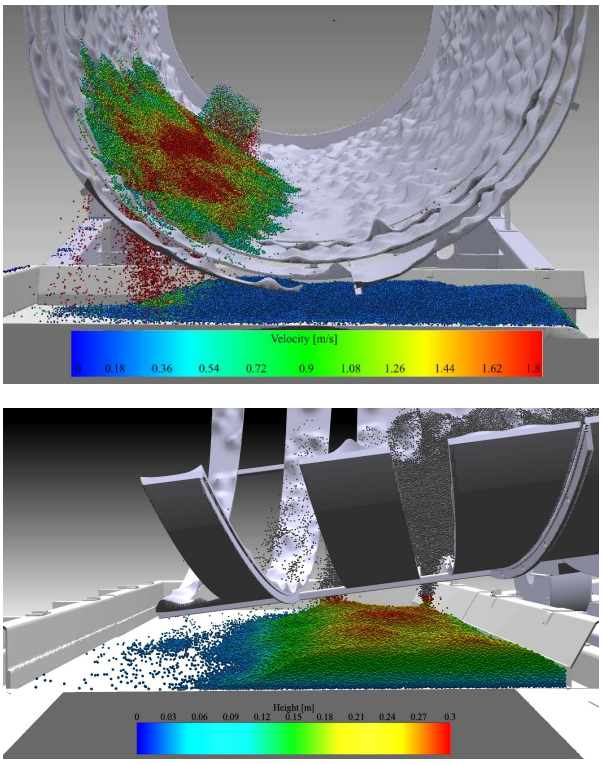

Figure 16: Image from simulation showing material distribution inside drum (top) and on the wide-belt conveyor (bottom). The original outlet deisgn is used. Particles are color coded by velocity and height, respectively.

and RK5-o), in Fig. 2 and 16, with gap width $\{e_1^{\mathrm{o}} : e_2^{\mathrm{o}}\} = \{11.8 : 39.4\}d$, the CAD models of modified outlet that are in operation (RK1 and RK5), see Fig. 7, with $\{e_1^{\mathrm{m}} : e_2^{\mathrm{m}}\} = \{4.7 : 7.9\}d$ as well as outlet geometry models that include coating effect of fine material that make the effective gap width smaller (RK1-c and RK5-c), $\{e_1^{\mathrm{c}} : e_2^{\mathrm{c}}\} = \{3.1 : 6.3\}d$. The gap models are illustrated in Fig. 17.

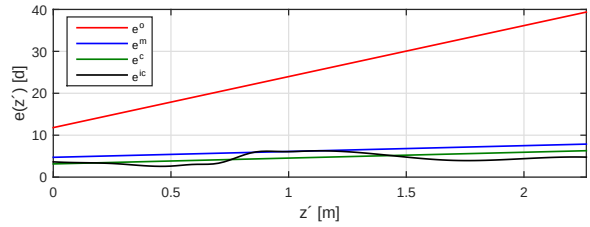

Figure 17: The different outlet gap functions $e(z')$ in RK5 simulations.

First, a stationary flow through the drum is established from a feed of rate $\dot{M}_{\mathrm{RK1}}$. The simulations are then run for three drum evolutions, $t = 36s$, while recording the material distribution on the wide-belt conveyor. The simulations involved nearly 1 M particles for which the total computational time on a 12 cpu machine become of the order 10 hours, see Eq. (8). A sample height surface from the RK5 simulation is found in Fig. 18. The time-averaged profile from RK5 and RK1 is found in Fig. 12 and the coefficients of variation are found in Table 4. To compare the simulated and experimentally measured profiles we compute also the relative coefficient of variation

$$\sigma_{h-\bar{h}} = \sqrt{\frac{4}{w_{\mathrm{b}}} \int_0^{w_{\mathrm{b}}} \left[\frac{h(y) - \bar{h}(y)}{\langle h \rangle + \langle \bar{h} \rangle} \right]^2 \mathrm{d}y} \tag{15}$$

where $h(y)$ is a profile from simulation and $\bar{h}(y)$ is the profile from experimental observation. The simulations confirm that the original outlet model, RK5-o, was indeed a very poor design as it produce a nonuniform profile with almost all material distributed on the right hand side of the wide-belt conveyor and $\sigma_h^{\mathrm{RK5}-o} = 0.83$. But also the simulation with the modified CAD model, RK5-m, distribute a substantial excess of material on the right-hand side. Much more than the experimental observation from balling drums as is seen both in Fig. 18 and 12, and by the value of the relative coefficient of variation $\sigma_{h-\bar{h}}^{\mathrm{RK5-m}} = 0.47$. The clogged outlet geometry, RK5-c, agree better with observation, $\sigma_{h-\bar{h}}^{\mathrm{RK5-c}} = 0.35$, but not entirely. In the region $y \in [1.5, 2.0]$ m the experimental profile show a material depletion that has no correspondence in the simulated profile.

We hypothesize that the material depletion is due to the coating being inhomogeneous and time-dependent. Supposedly, the coating increase gradually as material adheres until it reaches a critical thickness and become too heavy to support its own weight and drop from the outlet. We test this by modifying the gap geometry to an inhomogeneous coating, $e^{\mathrm{ic}}(z')$, as illustrated in

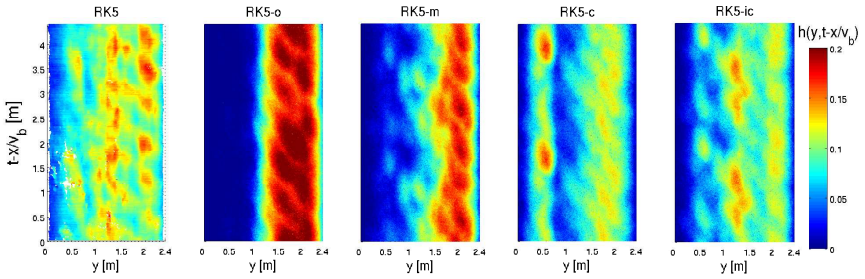

Figure 18: The height surface of the material distribution on the wide-belt conveyor in RK5. Left image is the result from video measurement and the others from simulation. The second to fifth images are from simulations using the original outlet (o) , modified outlet (m), with uniform gap coating (c) and with inhomogenous gap coating (ic).

Fig. 17. The results from the simulations with inhomogeneous gap coating, RK5-ic, match the experimental observations fairly well, $\sigma_{h-\bar{h}}^{\mathrm{RK5-ic}} = 0.11$.

Variations in surface friction, elasticity and particle size were tested to rule out that the deviation in material distribution is mainly due to too imprecise material parameters. The time-averaged bed profile was affected by roughly 5 % by the changes $\mu'_{\mathrm{s}} = 0.9\mu_{\mathrm{s}}$ and $E' = 2E$. The changes $E' = 2E$ and $d' = 0.8d$ affect the bed profile by roughly 15 %. As can be expected, with smaller particles the bed is shifted more to the right. The effect is significant but not enough to explain the deviation from the observed profile. The sensitivity of time-step size is considered in the next subsection.

6.4. Dependency on time-step size

The balling circuit simulations are run with time-step $\Delta t = 5$ ms. This choice is based on the formula Eq. (7) and an assumed impact normal velocity $v_{\mathrm{n}} \sim 0.02$ m/s and error tolerance $\epsilon = 0.01$. This assumed impact velocity is characteristic for flow in a drum with rotation speed $\omega_{\mathrm{d}} = 0.53$ rad/s, causing a characteristic shear rate $\dot{\sigma} \sim 2\omega/[1-cos(\theta/2)]$, where the circular sector angle is $\theta = \theta_2 - \theta_1$. There are impacts with higher contact velocity in the system but we assume the statistical occurrence of these are small and their error contribution to the overall bulk behaviour is insignificant. On the other hand, as found in Sec. 4, using too large time-step may lead to significant errors in the energy dissipation for impacts. Adapting the time-step for high velocity would have severe effects for the computational time. Particles impacting with the pellet bed on the wide-belt conveyor, for instance, have velocity up to $\sqrt{2g_{\mathrm{acc}}h_{\mathrm{db}}} \sim 3$ m/s. The required time-step for maintaining an error tolerance of $\epsilon = 0.01$ for these contacts is 0.04 ms.

The distribution of impact velocity and contact overlap from a simulation with $\Delta t = 5$ ms of material flowing from the drum onto the belt conveyor are presented in histograms in Fig. 19. Analysis show that 7% of the contacts are impacts, i.e., occur with relative normal velocity higher than $v_{\mathrm{imp}} = 0.05$ m/s

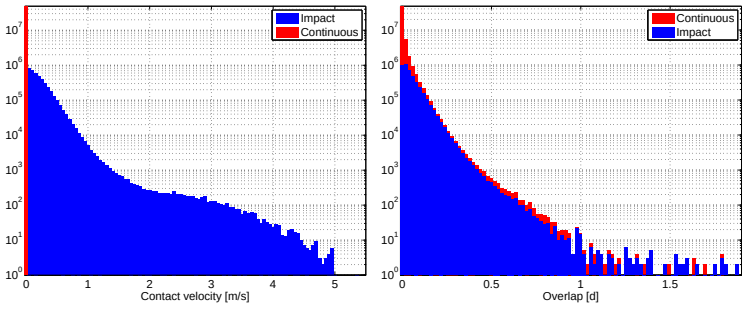

Figure 19: Distribution of normal contact velocity (left) and contact overlap (right) in balling drum and conveyor system from simulation using $\Delta t = 5$ ms. The contacts are divided into continuous contacts (red) and impacts (blue).

and less than 0.01 % has velocity higher than 1 m/s. The majority of contact overlap are below the error tolerance $\epsilon = 0.02$ but 17 % of the contacts have larger overlap. The overlap range up to $2d$, which is consistent with the impact velocity between particles and drum, or conveyor, ranging up to 5 m/s.

To verify the assumption that $\Delta t = 5$ ms is indeed a valid time-step and that the errors from high-velocity contacts do not have a significant contribution to the bulk behaviour, a simulation was also run with time-step $\Delta t = 0.1$ ms. The drum flow characteristics are $\theta'_{\mathrm{r}} = 33 \pm 2°$, $v_{\mathrm{tr}} = 0.23 \pm 0.03$ m/s, $v_{\mathrm{s}} = 1.0 \pm 0.1$, $v_{\mathrm{sz}} = 0.45 \pm 0.03$ m/s and $v_{s\perp} = 0.9 \pm 0.1$ m/s for the RK1. This is in good agreement with both the experimental observation and with the $\Delta t = 5$ ms simulations in Table 4. Histograms of the contact normal velocity and overlap from the $\Delta t = 0.1$ ms simulation is found in Fig. 20. At finer time-discretization more contact events can be resolved in time. The impact threshold $v_{\mathrm{imp}} = \epsilon d/\Delta t$ become roughly 2.5 m/s, i.e., essentially all contacts are resolved as continuous contacts. Furthermore, in this regime the time-step is small enough for the normal contact dissipation to be resolved with the physical viscosity from Hertz contact law, i.e., $\tau_{\mathrm{n}} = \max(n_{\mathrm{s}}\Delta t, \varepsilon_{\mathrm{n}}/\gamma_{\mathrm{n}})$ become $\varepsilon_{\mathrm{n}}/\gamma_{\mathrm{n}} = c/e_{\mathrm{H}}$. We identified $c \approx 1$ ms, from the high-speed camera measurements in Fig. 4. These adjustments of v_{imp} and τ_{n} with time-step are important. Otherwise the small time-step simulation model become too dissipative and produce a flow that does not agree well with observations.

7. Conclusions

A successful parameterization, verification and validation of a NDEM model for iron ore green pellets for the design and control of balling circuits has been demonstrated. The parameterization consists in the direct identification of individual ore green pellet physical parameters. The procedure involves no parameter calibration. The simulated bulk behaviour in the formation of piles and flow in a rotating inclined drum agrees with camera-based measurements in the

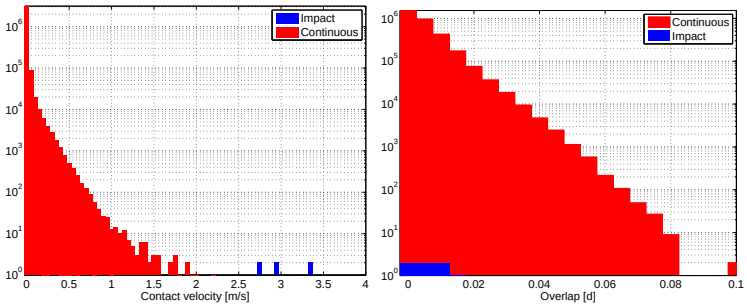

Figure 20: Distribution of normal contact velocity (left) and contact overlap (right) in balling drum and conveyor system from simulation using $\Delta t = 0.1$ ms. The contacts are divided into continuous contacts (red) and impacts (blue).

pelletizing plant. The angle of repose agrees within 5% and the flow velocity within 10 %. The pellet distribution on the wide-belt conveyor from the drum outlet show a more significant discrepancy between simulation and real system. The proposed explanation is that the simulated and actual outlet geometry do not agree although they are based on the same CAD model. Observations reveals that fine material adheres to the inside of the drum and outlet, creating a thick coating that alters the geometry. In particular, the outlet gaps become more narrow. Simulations confirm that the outlet flow is sensitive to this effect and that the material distribution produced by outlet geometries where this is included agree better with observation. The coating is believed to be dynamic in nature, gradually increasing in thickness until it breaks and drop, making the outlet gap narrowing variable and inhomogeneous. This has the consequence that even if a stationary flow inside the drum can be achieved, the material distribution on wide-belt conveyor and roller sieve will nevertheless have variations. The conclusion is that the outlet should be designed with materials and geometric shape which minimize the amount of coating or at least minimize the variability and effect on the flow.

The sensitivity of the simulation model to parameters is also investigated. It is shown that the rolling resistance is a necessary component of the model to obtain stable piles and the rolling resistance coefficient significantly affect the shape of piles as well as the flow characteristics in the rotating drum. The drum flow is found not to be sensitive to particle size. For an accurate conveyor bed profile beneath the outlet the detailed outlet geometry and rolling resistance are the critical parameters, but next to this the particle size was also found to be important.

It is also demonstrated that using time-step as large as 5 ms do not cause any statistically significant errors to the bulk behaviour as compared to using 0.1 ms although the larger time-step occasionally produce large errors in contacts between individual particles. As contrast, a conventional DEM simulation would require a time-step of size $\Delta t_{\mathrm{DEM}} \leq 0.17\sqrt{m/k_{\mathrm{n}}}$ [22], which evaluates to 0.02

ms for the given material parameters. Hence, NDEM simulation provide a time-efficient and reliable tool for exploring and optimizing the design and control of iron ore pellet balling drums and of similar systems. Future work should include extension to nonuniform and variable size distribution of ore green pellets and modeling of the mixing with ore slurry and the agglomeration process inside the drum.

Acknowledgements

This project was supported by Algoryx Simulations, LKAB, UMIT Research Lab and VINNOVA (dnr 2012-01235, 2014-01901).

Appendix

A. Simulation algorithm

The algoritm for simulating a system of granular material using NDEM with PGS solver is given in Algorithm 1. The projection on line 14 limit the

Algorithm 1 NDEM simulation with PGS solver

1: constants and parameters
2: initialization: $(\mathbf{x}_0, \mathbf{v}_0)$
3: **for** $i = 0, 1, 2, \ldots, t/\Delta t$ **do** ▷ Time stepping
4: contact detection
5: compute $\mathbf{g}, \mathbf{G}, \boldsymbol{\Sigma}, \boldsymbol{\Upsilon}, \mathbf{D}$, see Eq. (18)
6: impact stage PGS solve $\mathbf{v}_i \to (\mathbf{v}_i^+, \boldsymbol{\lambda}_i^+)$
7: $\mathbf{b}_\mathrm{n} = -(4/\Delta t)\boldsymbol{\Upsilon}_\mathrm{n}\mathbf{g}_\mathrm{n} + \boldsymbol{\Upsilon}_\mathrm{n}\mathbf{G}_\mathrm{n}\mathbf{v}_i^+$
8: pre-step $\mathbf{v} = \mathbf{v}_i^+ + \Delta t \mathbf{M}^{-1}\mathbf{f}_\mathrm{ext}$
9: **for** $k = 0, 1, \ldots, N_\mathrm{it} - 1$ or $|\mathbf{r}| \leq r_\mathrm{min}$ **do** ▷ PGS iteration
10: **for** each contact $\alpha = 0, 1, \ldots, N_\mathrm{c} - 1$ **do**
11: **for** each constraint n of contact α **do**
12: $\mathbf{r}_{n,k}^{(\alpha)} = -\mathbf{b}_{n,k}^{(\alpha)} + \mathbf{G}_n^{(\alpha)}\mathbf{v}$ ▷ residual
13: $\boldsymbol{\lambda}_{n,k}^{(\alpha)} = \boldsymbol{\lambda}_{n,k-1}^{(\alpha)} + \mathbf{D}_{n,(\alpha)}^{-1}\mathbf{r}_{n,k}^{(\alpha)}$ ▷ multiplier
14: $\mathrm{proj}(\boldsymbol{\lambda}_{n,k}^{(\alpha)}, \mathbf{v}) \to \boldsymbol{\lambda}_{n,k}^{(\alpha)}$ ▷ project
15: $\Delta\boldsymbol{\lambda}_{n,k}^{(\alpha)} = \boldsymbol{\lambda}_{n,k}^{(\alpha)} - \boldsymbol{\lambda}_{n,k-1}^{(\alpha)}$
16: $\mathbf{v} = \mathbf{v} + \mathbf{M}^{-1}\mathbf{G}_{n,(\alpha)}^T \Delta\boldsymbol{\lambda}_{n,k}^{(\alpha)}$
17: **end for**
18: **end for**
19: **end for**
20: $\mathbf{v}_{i+1} = \mathbf{v}$ ▷ velocity update
21: $\mathbf{x}_{i+1} = \mathbf{x}_i + \Delta t \mathbf{v}_{i+1}$ ▷ position update
22: **end for**

multipliers to the Signorini-Coulomb law. Each contact α between body a and

b add contributions to the constraint vector and normal and friction Jacobians according to

$$\begin{aligned}
\delta_{(\alpha)} &= \mathbf{n}_{(\alpha)}^{\mathrm{T}}(\mathbf{x}_a + \mathbf{d}_a^{(\alpha)} - \mathbf{x}_b - \mathbf{d}_b^{(\alpha)}) \\
g_{(\alpha)} &= \delta_{(\alpha)}^{e_{\mathrm{H}}} \\
\mathbf{G}_{\mathrm{n}a}^{(\alpha)} &= e_{\mathrm{H}} g_{(\alpha)}^{e_{\mathrm{H}}-1} \left[\begin{array}{cc} -\mathbf{n}_{(\alpha)}^{\mathrm{T}} & -(\mathbf{d}_a^{(\alpha)} \times \mathbf{n}_{(\alpha)})^{\mathrm{T}} \end{array} \right] \\
\mathbf{G}_{\mathrm{n}b}^{(\alpha)} &= e_{\mathrm{H}} g_{(\alpha)}^{e_{\mathrm{H}}-1} \left[\begin{array}{cc} \mathbf{n}_{(\alpha)}^{\mathrm{T}} & (\mathbf{d}_b^{(\alpha)} \times \mathbf{n}_{(\alpha)})^{\mathrm{T}} \end{array} \right] \\
\mathbf{G}_{\mathrm{t}a}^{(\alpha)} &= \left[\begin{array}{cc} -\mathbf{t}_1^{(\alpha)\mathrm{T}} & -(\mathbf{d}_a^{(\alpha)} \times \mathbf{t}_1^{(\alpha)})^{\mathrm{T}} \\ -\mathbf{t}_2^{(\alpha)\mathrm{T}} & -(\mathbf{d}_a^{(\alpha)} \times \mathbf{t}_2^{(\alpha)})^{\mathrm{T}} \end{array} \right] \\
\mathbf{G}_{\mathrm{t}b}^{(\alpha)} &= \left[\begin{array}{cc} \mathbf{t}_1^{(\alpha)\mathrm{T}} & (\mathbf{d}_b^{(\alpha)} \times \mathbf{t}_1^{(\alpha)})^{\mathrm{T}} \\ \mathbf{t}_2^{(\alpha)\mathrm{T}} & (\mathbf{d}_b^{(\alpha)} \times \mathbf{t}_2^{(\alpha)})^{\mathrm{T}} \end{array} \right]
\end{aligned} \tag{16}$$

The diagonal matrices and Schur complement matrix $\mathbf{D}$ are

$$\begin{aligned}
\mathbf{\Sigma}_{\mathrm{n}} &= \frac{4}{\Delta t^2} \frac{\varepsilon_{\mathrm{n}}}{1 + 4\frac{\tau_{\mathrm{n}}}{\Delta t}} \mathbf{1}_{N_c \times N_c} \\
\mathbf{\Sigma}_{\mathrm{t}} &= \frac{\gamma_{\mathrm{t}}}{\Delta t} \mathbf{1}_{2N_c \times 2N_c} \\
\mathbf{\Sigma}_{\mathrm{r}} &= \frac{\gamma_{\mathrm{r}}}{\Delta t} \mathbf{1}_{3N_c \times 3N_c} \\
\mathbf{\Upsilon}_{\mathrm{n}} &= \frac{1}{1 + 4\frac{\tau_{\mathrm{n}}}{\Delta t}} \mathbf{1}_{N_c \times N_c} \\
\mathbf{D} &= \mathbf{G}\mathbf{M}^{-1}\mathbf{G}^T + \mathbf{\Sigma}
\end{aligned} \tag{17}$$

The mapping between regularization parameters and material parameters are

$$\begin{aligned}
\varepsilon_{\mathrm{n}} &= e_{\mathrm{H}}/k_{\mathrm{n}} = 3e_{\mathrm{H}}(1-\nu^2)/E\sqrt{r^*} \\
\tau_{\mathrm{n}} &= \max(n_{\mathrm{s}}\Delta t, \varepsilon_{\mathrm{n}}/\gamma_{\mathrm{n}}) \\
\gamma_{\mathrm{n}}^{-1} &= k_{\mathrm{n}} c/e_{\mathrm{H}}^2
\end{aligned} \tag{18}$$

and we use $\gamma_{\mathrm{t}} = \gamma_{\mathrm{r}} = 10^{-6}$, $n_{\mathrm{s}} = 2$.

References

[1] F. Radjai, V. Richefeu, Contact dynamics as a nonsmooth discrete element method, Mechanics of Materials 41 (6) (2009) 715–728.

[2] M. Servin, D. Wang, C. Lacoursière, K. Bodin, Examining the smooth and nonsmooth discrete element approach to granular matter, Int. J. Numer. Meth. Engng. 97 (2014) 878–902.

[3] D. Wang, M. Servin, K.-O. Mickelsson, Outlet design optimization based on large-scale nonsmooth DEM simulation, Powder Technology 253 (0) (2014) 438–443.

[4] S. Forsmo, Influence of green pellet properties on pelletizing of magnetite iron ore, Ph.D. thesis, Luleå University of Technology, Luleå (2007).

[5] I. Cameron, F. Wang, C. Immanuel, F. Stepanek, Process systems modelling and applications in granulation: A review, Chemical Engineering Science 60 (14) (2005) 3723–3750.

[6] R. Soda, A. Sato, J. Kano, E. Kasai, F. Saito, M. Hara, T. Kawaguchi, Analysis of granules behavior in continuous drum mixer by DEM, ISIJ International 49 (5) (2009) 645–649.

[7] T. Pöschel, T. Schwager, Computational Granular Dynamics, Models and Algorithms, Springer-Verlag, 2005.

[8] J. J. Moreau, Numerical aspects of the sweeping process, Computer Methods in Applied Mechanics and Engineering 177 (1999) 329–349.

[9] M. Jean, The non-smooth contact dynamics method, Computer Methods in Applied Mechanics and Engineering 177 (1999) 235–257.

[10] F. A. Bornemann, C. Schütte, Homogenization of Hamiltonian systems with a strong constraining potential, Phys. D 102 (1-2) (1997) 57–77.

[11] N. V. Brilliantov, F. Spahn, J.-M. Hertzsch, T. Pöschel, Model for collisions in granular gases, Phys. Rev. E 53 (1996) 5382–5392.

[12] C. Lacoursière, Regularized, stabilized, variational methods for multibodies, in: D. F. Peter Bunus, C. Führer (Eds.), The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), Linköping University Electronic Press, 2007, pp. 40–48.

[13] K. G. Murty, Linear Complementarity, Linear and Nonlinear Programming, Helderman-Verlag, Heidelberg, 1988.

[14] Algoryx Simulations. AGX Dynamics, December 2014.

[15] J. Ai, J.-F. Chen, J. M. Rotter, J. Y. Ooi, Assessment of rolling resistance models in discrete element simulations, Powder Technology 206 (3) (2011) 269–282.

[16] N. Estrada, E. Azéma, F. Radjai, A. Taboada, Identification of rolling resistance as a shape parameter in sheared granular media, Phys. Rev. E 84 (2011) 011306.

[17] N. Estrada, A. Taboada, F. Radjaï, Shear strength and force transmission in granular media with rolling resistance, Phys. Rev. E 78 (2008) 021301.

[18] J. Huang, M. V. da Silva, K. Krabbenhoft, Three-dimensional granular contact dynamics with rolling resistance, Computers and Geotechnics 49 (0) (2013) 289–298.

[19] A. Tasora, M. Anitescu, A complementarity-based rolling friction model for rigid contacts, Meccanica 48 (7) (2013) 1643–1659.

[20] S. Forsmo, P.-O. Samskog, B. Björkman, A study on plasticity and compression strength in wet iron ore green pellets related to real process variations in raw material fineness, Powder Technology 181 (3) (2008) 321–330.

[21] H.-T. Chou, C.-F. Lee, Cross-sectional and axial flow characteristics of dry granular material in rotating drums, Granular Matter 11 (1) (2008) 13–32.

[22] C. O. J. Bray, Selecting a suitable time step for discrete element simulations that use the central difference time integration scheme, Engineering Computations 21 (2/3/4) (2004) 278–303.

Computational Particle Mechanics manuscript No.
(will be inserted by the editor)

Warm starting the projected Gauss-Seidel algorithm for granular matter simulation

Da Wang · Martin Servin · Tomas Berglund

Received: date / Accepted: date

Abstract The effect on the convergence of warm starting the projected Gauss-Seidel solver for nonsmooth discrete element simulation of granular matter are investigated. It is found that the computational performance can be increased by a factor 2 to 5.

Keywords Discrete elements · Nonsmooth contact dynamics · Convergence · Warm starting · Projected Gauss-Seidel

1 Introduction

In simulations of granular matter using the nonsmooth discrete element method (NDEM) [1–3] the computational time is dominated by the solve stage, where the contact forces and velocity updates are computed. Conventionally this involves solving a mixed complementarity problem or a quasi-optimization problem that arises from implicit integration of the rigid multibody equations of motion in conjunction with set-valued contact laws and impulse laws, usually the Signorini-Coulomb law and Newton impulse law. The computational properties of the solution algorithms for these problems are largely open questions, lacking general proof of existence and uniqueness of solutions as well as of general proof of convergence and numerical stability [4]. The projected Gauss-Seidel (PGS) algorithm is widely used. The popularity of PGS is likely due to having low computational cost per iteration, small memory footprint and produce smooth distribution of errors that favour stable simulation. In many cases PGS require few iterations to identify the active set of constraints. This make PGS a natural choice for fast simulations of large-scale rigid multibody systems with frictional contacts. The asymptotic convergence, however, is slow. The PGS algorithm solves each local two-body contact problem accurately but approaches to the global solution in a diffusive manner with iterations. This has a smoothing effect both on errors and on the solution and limits the practical use of PGS for simulations of high accuracy. The residual error appear as artificial elasticity [5], with an effective sound velocity $v_{\mathrm{PGS}} = \sqrt{N_{\mathrm{it}}}d/\Delta t$, where N_{it} is the number of iterations, d is the particle size and Δt is the timestep. To accurately resolve the impulse propagation of stiff materials require large number of iterations or small timestep. The required number of iterations for a given error tolerance increase with the size of the contact network, particularly with the number of contacts in direction of gravity or applied stress, but may also saturate by arching phenomena as in Janssen's law for silos [14]. Contrary to many authors claim, the PGS algorithm is parallelizable for hardware with distributed memory using domain decomposition methods [6,7].

Warm starting is to start PGS with an initial guess, $\boldsymbol{\lambda}_0^{\mathrm{w}}$, that presumably is closer to the exact solution, $\boldsymbol{\lambda}$, than starting with the nominal choice of $\boldsymbol{\lambda}_0 = 0$. The idea, illustrated in Fig. 1, is that the warm started PGS reach an approximate solution, $\boldsymbol{\lambda}_{k'}$, with fewer iterations than the solution, $\boldsymbol{\lambda}_k$, starting from nominal value. In other words, $|\boldsymbol{\lambda} - \boldsymbol{\lambda}_{k'}^{\mathrm{w}}| \lesssim |\boldsymbol{\lambda} - \boldsymbol{\lambda}_k| < \varepsilon$ with $k' < k$. The effective increase in convergence should be most significant for static or nearly static configu-

D. Wang
Umeå University

M. Servin
Umeå University
Tel.: +46-90-7866508
E-mail: martin.servin@umu.se

T. Berglund
Algoryx Simulation AB

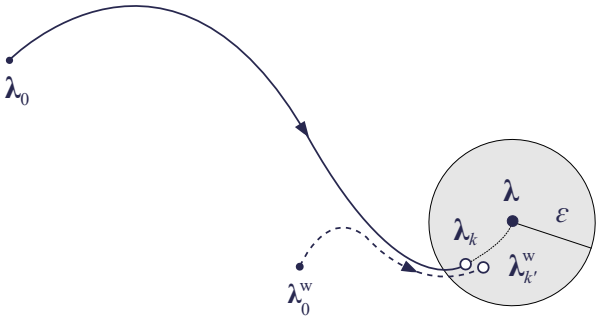

Fig. 1 Illustration of improved convergence by warm starting.

rations. For rapid granular flows the solution change rapidly with time and no or little effect on convergence is expected. There have been several reports on improved convergence by using warm starting [3,8–12] but to the best of our knowledge no quantitative analysis has previously been presented.

2 PGS for nonsmooth discrete element simulation

The mixed complementarity problem (MCP) for computing the update of the velocity from $\mathbf{v}_{\mathrm{old}} \equiv \boldsymbol{v}(t-\Delta t)$ to $\mathbf{v} \equiv \boldsymbol{v}(t)$ and the Lagrange multiplier $\boldsymbol{\lambda}$ of the contact constraints and take the form

$$\begin{bmatrix} \mathbf{M} & -\mathbf{G}^{\mathrm{T}} \\ \mathbf{G} & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \tag{1}$$

$$\boldsymbol{\lambda}^{(\alpha)} \in \mathcal{C}_\mu(\lambda_{\mathrm{n}}^{(\alpha)}) \ \ , \ \alpha = 1,2,\ldots,N_{\mathrm{c}} \tag{2}$$

where $\boldsymbol{M}$ is the mass matrix and $\boldsymbol{G}$ the Jacobian of contact constraints. The contact force, $\boldsymbol{G}^T\boldsymbol{\lambda}$, is restricted by a friction cone condition that we represent $\boldsymbol{\lambda}^{(\alpha)} \in \mathcal{C}_\mu(\lambda_{\mathrm{n}}^{(\alpha)})$, where α indexes the contacts. The diagonal perturbation $\boldsymbol{\Sigma}$ regularize the problem and allow modeling of contact elasticity. The vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ on the right hand side depend on particle inertia, external force and constraint violations on position and velocity level. As friction cone condition we use the Signorini-Coulomb law including $0 \leq \lambda_{\mathrm{n}}^{(\alpha)}$ and $|\boldsymbol{\lambda}_{\mathrm{t}}^{(\alpha)}| \leq \mu_{\mathrm{s}}|\mathbf{G}_{\mathrm{n}}^{(\alpha)\mathrm{T}}\lambda_{\mathrm{n}}^{(\alpha)}|$ with the friction coefficient μ_{s} for each contact α divided in one normal (n) and two tangential (t) components. The constraint forces act to prevent contact overlap, $\boldsymbol{g} \leq 0$, and contact sliding, $\boldsymbol{G}_{\mathrm{t}}\boldsymbol{v} = 0$. Similarly, rolling resistance (r) is imposed by a constraint $\boldsymbol{G}_{\mathrm{r}}\boldsymbol{v} = 0$ with a Coulomb like law: $|\boldsymbol{\lambda}_{\mathrm{r}}^{(\alpha)}| \leq \mu_{\mathrm{r}} r^* |\mathbf{G}_{\mathrm{n}}^{(\alpha)\mathrm{T}}\lambda_{\mathrm{n}}^{(\alpha)}|$, where r^* is the effective radius. See Appendix A for further details. For a system with N_{p} particles represented as rigid bodies and N_{c} contacts with normal and tangential force and rolling and twisting resistance the vectors and matrices in Eq. (1) have the following dimensions $\dim(\boldsymbol{M}) = 6N_{\mathrm{p}} \times 6N_{\mathrm{p}}$, $\dim(\boldsymbol{G}) = 6N_{\mathrm{c}} \times N_{\mathrm{p}}$, $\dim(v) = \dim(\boldsymbol{p}) = 6N_{\mathrm{p}}$ $\dim(\boldsymbol{\lambda}) = \dim(\boldsymbol{q}) = 6N_{\mathrm{c}}$. The matrices are however very sparse. $\boldsymbol{M}$ and $\boldsymbol{\Sigma}$ are block diagonal and $\boldsymbol{G}$ is block sparse. The blocks have dimension 6×6. The main steps of the PGS iteration are

$$\boldsymbol{\lambda}_{k+1}^{(\alpha)} = \boldsymbol{\lambda}_k^{(\alpha)} + D_{(\alpha\alpha)}^{-1}\mathbf{r}_k^{(\alpha)} \tag{3}$$

$$\boldsymbol{\lambda}_{k+1}^{(\alpha)} \leftarrow \mathrm{proj}_{\mathcal{C}_\mu}(\boldsymbol{\lambda}_{k+1}^{(\alpha)}) \tag{4}$$

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \mathbf{M}^{-1}\mathbf{G}_{(\alpha)}^{\mathrm{T}}\Delta\boldsymbol{\lambda}_{k+1}^{(\alpha)} \tag{5}$$

with iteration index $k = 0,1,2,\ldots,N_{\mathrm{it}}-1$, change in multiplier $\Delta\mathrm{g}\lambda_{k+1}^{(\alpha)} = \boldsymbol{\lambda}_{k+1}^{(\alpha)} - \boldsymbol{\lambda}_k^{(\alpha)}$ and residual

$$\mathbf{r}_k^{(\alpha)} = \mathbf{G}_{(\alpha)}\mathbf{v}_k - \mathbf{q}_{(\alpha)} \tag{6}$$

where $\mathbf{v}_0 = \mathbf{M}^{-1}\mathbf{p}$ and $\mathbf{D}$ is the block diagonal part of the Schur complement matrix $\mathbf{S} = \mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}} + \boldsymbol{\Sigma}$. The details of the vectors $\mathbf{p}$ and $\mathbf{q}$ depend on the stepping scheme and constraint stabilization method. When integrating with fix timesteps Δt using the SPOOK stepper [13] one has $\mathbf{p} = \mathbf{M}\mathbf{v}_{\mathrm{old}} + \Delta t\mathbf{f}_{\mathrm{ext}}$, with smooth external forces $\mathbf{f}_{\mathrm{ext}}$, and $\mathbf{q} = (\mathbf{q}_{\mathrm{n}}^T, \mathbf{q}_{\mathrm{t}}^T, \mathbf{q}_{\mathrm{r}}^T)^T$ with $\mathbf{q}_{\mathrm{n}} = -(4/\Delta t)\boldsymbol{\Upsilon}\bar{\mathbf{g}} + \Upsilon\mathbf{G}_{\mathrm{n}}\mathbf{v}_{\mathrm{old}}$, $\mathbf{q}_{\mathrm{t}} = \mathbf{0}$ and $\mathbf{q}_{\mathrm{r}} = \mathbf{0}$. The projection $\boldsymbol{\lambda}_{k+1}^{(\alpha)} \leftarrow \mathrm{proj}_{\mathcal{C}_\mu}(\boldsymbol{\lambda}_{k+1}^{(\alpha)})$ is made by simply clamping $\boldsymbol{\lambda}_{k+1}^{(\alpha)}$ to the friction or rolling resistance limit if exceeded. After stepping the velocities and positions an impact stage follows. This include solving a MCP similar to Eq. (1) but with the Newton impact law, $\mathbf{G}_{\mathrm{n}}^{(\alpha)}\mathbf{v}_+ = -e\mathbf{G}_{\mathrm{n}}^{(\alpha)}\mathbf{v}_-$, replacing the normal constraints for the contacts with normal velocity larger than an impact velocity threshold v_{imp}. The remaining constraints are maintained by imposing $\mathbf{G}\mathbf{v}_+ = 0$. An algorithm of NDEM simulation with PGS is given in Appendix A together with details on the Jacobians and relation between the solver parameters and material parameters.

3 PGS warm starting

By default the PGS algorithm is initialized with $\boldsymbol{\lambda}_0^{(\alpha)} = \mathbf{0}$. We refer to this as *cold starting*. In a stationary state the contact force $\mathbf{G}^T\boldsymbol{\lambda}$ is constant in time. In a nearly stationary state we expect the multipliers to remain almost constant between two timestep. Therefore it is reasonable to use the solution from last timestep as an initial guess, $\boldsymbol{\lambda}(t) \approx \boldsymbol{\lambda}(t-\Delta t)$. We use a fraction $\beta = 0.85$ of the solution from last timestep

$$\boldsymbol{\lambda}_0(t) = \beta\boldsymbol{\lambda}_{N_{\mathrm{it}}}(t-\Delta t) \tag{7}$$

It is important to also apply the corresponding impulse to the particles and update the velocity

$$\mathbf{v}_0 = \mathbf{M}^{-1}\mathbf{p} + \mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{\lambda}_0 \tag{8}$$

such it become consistent with the initial guess for the multiplier. We refer to warm starting based on the last

solution as *history based warm starting*. For any new contact we set $\boldsymbol{\lambda}_0^{(\alpha)} = \mathbf{0}$. Warm starting is not applied at the impact stage and we assume that the contact network is not fundamentally rearranged by the impacts and use the solution from last timestep despite the occurrence of impacts.

An alternative method for warm starting a nearly stationary state is to estimate each local contact force and assign this to the contact multipliers. When using regularized NDEM the local contact force can be estimated from the overlaps and relative contact velocities much as in conventional smooth DEM. We refer to this approach as *model based warm starting*. For normal forces we use the Hertz contact law $f_\mathrm{n} = k_\mathrm{n} g_\mathrm{n}^{3/2}$, with overlap function g_n and based on $\mathbf{f}_\mathrm{n} \approx \mathbf{G}_\mathrm{n}^T \boldsymbol{\lambda}_\mathrm{n}/\Delta t$ we estimate

$$\lambda_{\mathrm{n},0}^{(\alpha)} \approx \tfrac{5}{4}\Delta t k_\mathrm{n} g_{\mathrm{n}(\alpha)}^{5/4} \tag{9}$$

Similarly the regularized tangent friction force and rolling resistance force can be estimated via the Rayleigh dissipation functions to

$$\boldsymbol{\lambda}_{\mathrm{t},0} \approx \gamma_\mathrm{t}^{-1}\Delta t(\mathbf{G}_\mathrm{t}\mathbf{v})^\mathrm{T}\mathbf{G}_\mathrm{t} \tag{10}$$

$$\boldsymbol{\lambda}_{\mathrm{r},0} \approx \gamma_\mathrm{r}^{-1}\Delta t(\mathbf{G}_\mathrm{r}\mathbf{v})^\mathrm{T}\mathbf{G}_\mathrm{r} \tag{11}$$

Note that the friction and rolling resistance should, if large, be clamped to obey the conditions $|\boldsymbol{\lambda}_\mathrm{t}^{(\alpha)}| \leq \mu_\mathrm{s}|\mathbf{G}_\mathrm{n}^{(\alpha)\mathrm{T}}\lambda_\mathrm{n}^{(\alpha)}|$ and $|\boldsymbol{\lambda}_\mathrm{r}^{(\alpha)}| \leq \mu_\mathrm{r} r^* |\mathbf{G}_\mathrm{n}^{(\alpha)\mathrm{T}}\lambda_\mathrm{n}^{(\alpha)}|$.

4 Numerical experiments

Numerical simulation of different systems were performed to analyse the effect of warm starting on the convergence of NDEM simulations. The following systems were studied: a 1D column, formation of a pile, dense flow in a rotating drum and a triaxial shear cell. The main material and simulation parameters are listed in Table 1. The method was implemented in the software AgX Dynamics [15] in the module for NDEM simulation with optimized data structures and support for collision detection and PGS using parallel computing on multicore processors. The simulations were run on a desktop computer with Intel(R) Core(TM) i7 CPU, 2.8 GHz, 8 GB RAM on a Windows 64 bit system. Videos from simulations are found at `http://umit.cs.umu.se/granular/warmstarting/`.

4.1 Column

Particles of diameter $d = 13$ mm are initiated on top of another with zero overlap. The system compress slightly under its weight. The simulation is run until the 1D column have come to rest. Sample images from simulation with and without warm starting and for different number of iterations are shown in Fig. 2. Warm starting clearly improve the convergence. To make a quantitative convergence analysis we study the deviation of the simulated column height, l_{N_it}, from the theoretical height, l, computed using the Hertz contact law

$$\varepsilon_\mathrm{l} = \frac{l - l_{N_\mathrm{it}}}{l} \tag{12}$$

A series of simulations are run with number of particles, N_P, ranging from 5 and 100, number of iterations, N_it, ranging from 10 to 500. The required number of iterations, $N_\mathrm{it}^\varepsilon$, to reach a solution with error tolerance $\varepsilon_\mathrm{l} = 0.1\%, 1\%$ and 5% are presented in Fig. 3. It scales almost linearly with the number of particles and increase with decreasing error tolerance ε_l. History based warm starting is on average three times as efficient as cold starting. Also model based warm starting improve the convergence at low error tolerance. The performance gain from model based warm starting decrease with increasing error tolerance and for $\varepsilon_\mathrm{l} = 5\%$ model based warm starting require twice as many iterations as cold starting. Figure 4 show the evolution of the mean residual, see Eq. (6), for the normal force constraint during a PGS solve for a column with $N_\mathrm{p} = 25$. The convergence rates are similar but warm starting clearly has the advantage of starting closer to the solution.

Table 1 Main material and simulation parameters

Notation	Value	Comment
$[d, d_2]$	$[13, 10]$ m	bi-disperse particle diameter
ρ	3700 kg/m^3	particle mass density
E	6 MPa [1]	Young's modulus
e	0.18	restitution coefficient
μ_s	0.91	surface friction coefficient
μ_r	0.32	rolling resistance coefficient
Δt	5 ms	timestep
v_imp	0.05 m/s	impact threshold

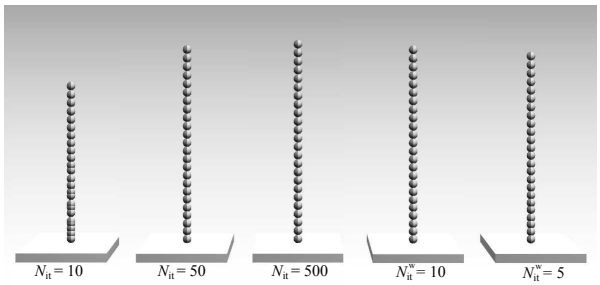

Fig. 2 Samples of five columns simulated, from left to right, with cold starting $N_\mathrm{it} = 10, 50$ and 500, and history based warm starting $N_\mathrm{it}^\mathrm{w} = 10$ and 5.

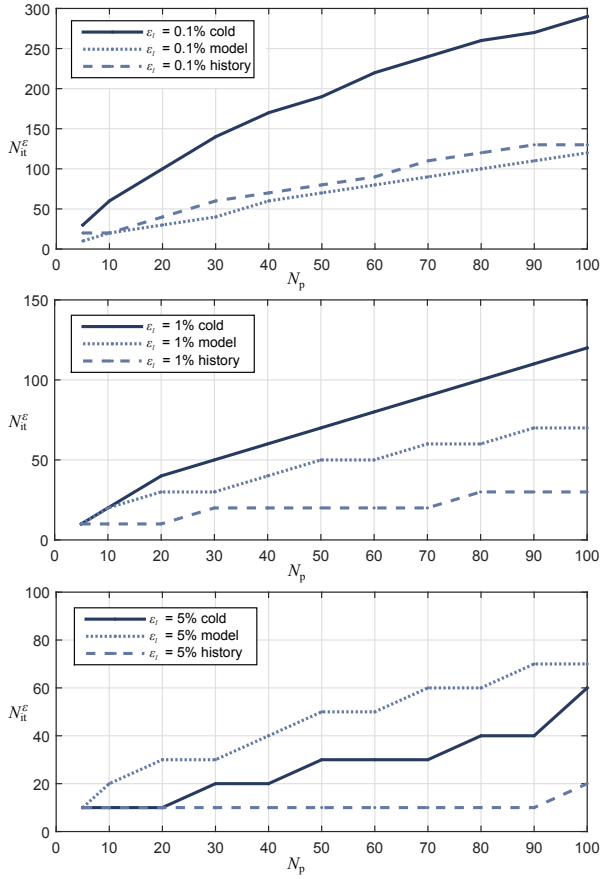

Fig. 3 The required number of iterations for a 1D column simulation for error tolerance $\varepsilon_{1\mathrm{D}} = 0.1\%$ (top), 1% (middle) and 5% (bottom) depending on the number of particles and warm starting method.

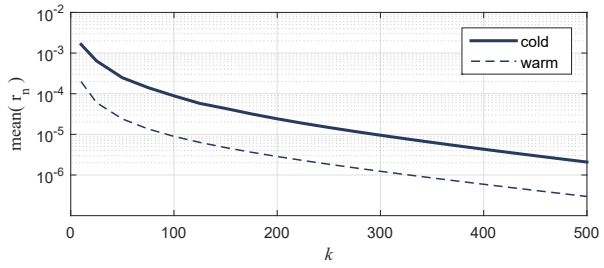

Fig. 4 The evolution of the mean normal force residual during a PGS solve for a $N_{\mathrm{p}} = 25$ column using cold starting and history based warm starting.

4.2 Pile formation

A pile is formed by continuously emitting particles of diameter $d = 13$ mm from a $3d$ wide source placed $15d$ above a ground plane. The number of particles in the pile is $N_{\mathrm{p}} = 3363$. Again we use the relative height, ε_{l} in Eq. (12), as error measurement. The reference height about 15 d is measured from a pile constructed using small time-step $\Delta t = 0.2$ ms and $N_{\mathrm{it}} = 500$. Pile formation is then simulated using time-step $\Delta t = 5$ ms for different number of iterations and warm starting methods. Sample images from simulations are shown in Fig. 5. The angle of repose is an alternative mea-

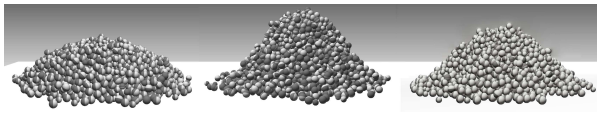

Fig. 5 Samples from simulations of pile formation. From left to right is the cold started pile ($\Delta t = 5$ ms, $N_{\mathrm{it}} = 50$), a reference pile ($\Delta t = 0.2$ ms, $N_{\mathrm{it}} = 500$) and a warm started pile ($\Delta t = 5$ ms, $N_{\mathrm{it}} = 50$)

sure but was found to give less precise result. The pile height is measured 10 s after the last emitted particle has come to approximate rest. Simulations are run with warm starting applied both to normal forces, friction and rolling resistance and to normal forces only. The historical warm starting is tested with and without the velocity update associated with the warm start in Eq. (7). The required number of iterations for a given error threshold are given in Fig. 6. With few iterations the piles experience artificial compression and contact sliding such that the pile gradually melt down to a singe particle layer. The pile stability increase with the number of iterations. History based warm starting, applied to both normals, friction and rolling resistance, give the best result and require roughly half the number of iterations of cold starting. If the warm start velocity is not applied the result is worse than cold starting. Model based warm starting is only marginally better than cold starting and is from further experiments here on excluded. Applying warm starting to the normal constraints only does not improve the convergence significantly. The convergence is also analysed by studying the evolution of the Lagrange multiplier and the residual. The relative error of the normal force multiplier is computed as

$$\varepsilon_{\lambda_k} = \left\langle \frac{|\boldsymbol{\lambda}_{500}^{\mathrm{n}(\alpha)} - \boldsymbol{\lambda}_{k}^{\mathrm{n}(\alpha)}|}{|\boldsymbol{\lambda}_{500}^{\mathrm{n}(\alpha)}|} \right\rangle \tag{13}$$

The evolution of ε_{λ_k} during a solve of a stationary pile is shown in Fig. 7. The multiplier error for history based warm starting is roughly five times smaller than for cold starting and remain more accurate indefinitely. A more careful analysis can be made by studying the evolution of the residual, defined in Eq. (6), and how it is distributed over the constraints. To get comparable states a stationary pile is prepared by using 500 iterations from which the cold and warm started simulations are started and run for 1 s before the measurement. The evolution of the mean residual during a PGS solve is shown in Fig. 8. The convergence rates are similar but

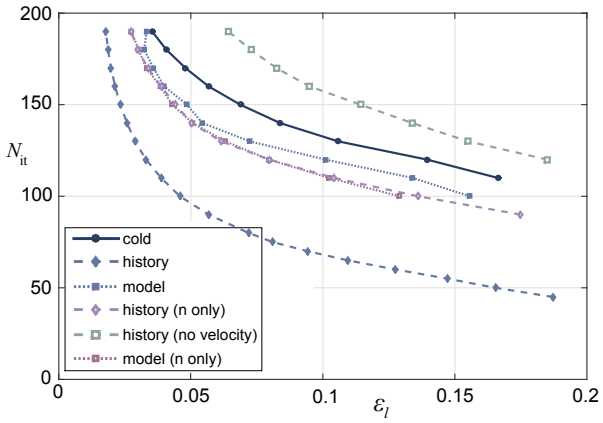

Fig. 6 The required number of iterations versus pile height error for different warm starting methods.

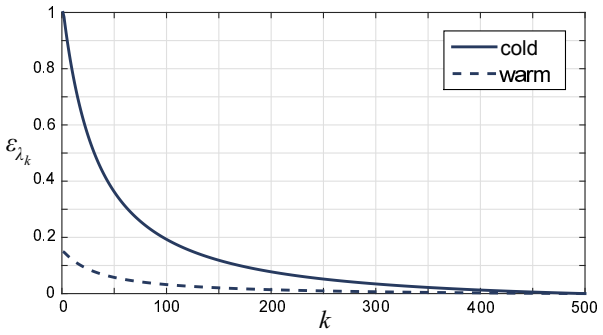

Fig. 7 The evolution of the relative multiplier during a PGS solve for a resting pile using cold starting and history based warm starting.

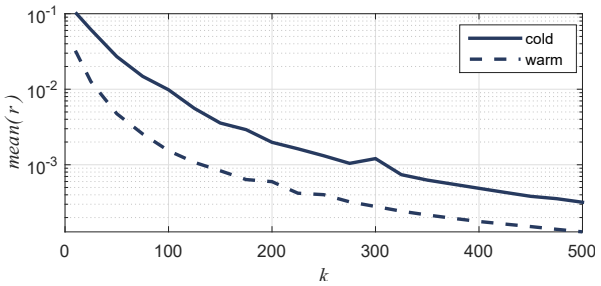

Fig. 8 The mean residual dependency on the number of iterations when simulating a resting pile for 1 s using cold starting and history based warm starting.

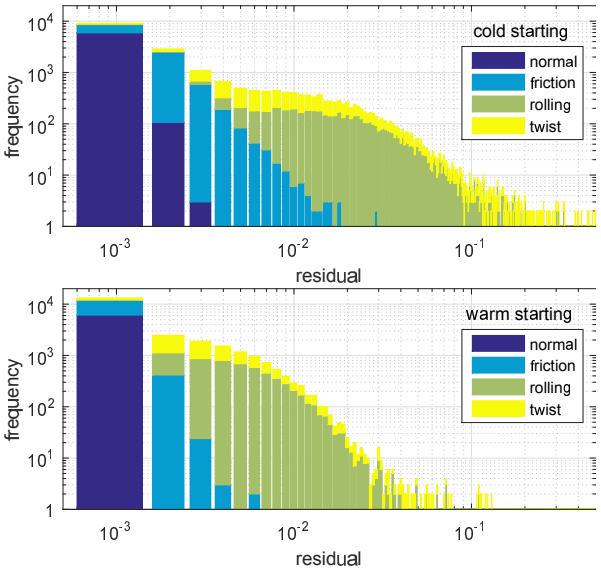

Fig. 9 The residual distribution for a resting pile after 1 s using $N_{\text{it}} = 100$ iterations, cold starting (top) and history based warm starting (bottom).

the initial lead of history based warm starting over cold starting by roughly a factor 5 remains throughout the 500 PGS iterations. Comparing the residual histograms from using cold and warm starting in Fig. 9 it is clear that the solutions differs primarily in the errors for the rolling resistance and friction constraints and less so for normal force constraints. This is consistent with the faster melting of the piles simulated with cold starting.

4.3 Rotating drum

A cylindrical drum with diameter $D = 40d$ and width $w = 7d$ is rotated with angular velocity $\Omega = 0.25$ rad/s. This corresponds to the Froude number $\text{Fr} \equiv D\Omega^2/2g \sim 10^{-3}$ which corresponds to the dense rolling flow regime. A nearly stationary flow of $N_{\text{p}} = 4864$ particles with bi-disperse size distribution d and d_2. At this low Froude number a large plug-zone is developed where particles co-rotate rigidly with the drum. A convergence analysis is made of the plug-zone number fraction, $N_{\text{plug}}/N_{\text{p}}$, and the dynamic angle of repose, θ'. These are measured for different number of iterations on a flow averaged over 2 s for cold starting and historical warm starting. The sample trajectories in Fig. 10 illustrate the general trend that the dynamic angle of repose and the size of the plug zone decrease with decreasing number of iterations but less so using warm starting. The normalized particle flow velocity relative the plug flow is computed $v_{\text{r}}^i \equiv |\mathbf{v}^i - \mathbf{r}^i \times \boldsymbol{\Omega}|/R\Omega$ and sample plots are shown in Fig. 11. As threshold for the plug zone flow we set $v_{\text{r}} \leq 0.15$, which is fulfilled by $N_{\text{plug}}^{500}/N_{\text{p}} = 58\% \pm 5\%$ particles where the variations reflect the slightly pulsating nature of the flow, due to sequential onset of avalanches. The plug zone fraction

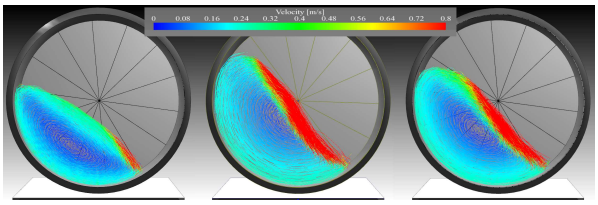

Fig. 10 A sample of particle trajectories from simulation of a rotating drum with $\Omega = 0.25$ rad/s, $\Delta t = 5$ ms and $N_{\mathrm{it}} = 10$ (left), $N_{\mathrm{it}} = 500$ (middle) and warm starting $N_{\mathrm{it}} = 10$ (right).

number error is defined

$$\varepsilon_{\mathrm{plug}} = \frac{N^{500}_{\mathrm{plug}} - N^{N_{\mathrm{it}}}_{\mathrm{plug}}}{N_{\mathrm{p}}} \tag{14}$$

and the relation to the required number of iterations is found in Fig. 12. The warm starting solution approach the solution faster but seems to have larger variations at high iteration numbers. The convergence analysis of

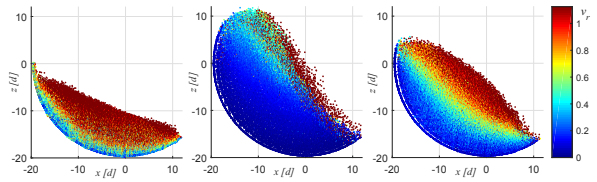

Fig. 11 A sample of cross-section flow from a simulation of a rotating drum with $\Omega = 0.25$ rad/s, $\Delta t = 5$ ms and $N_{\mathrm{it}} = 10$ (left), $N_{\mathrm{it}} = 500$ (middle) and warm starting $N_{\mathrm{it}} = 10$ (right). The colour coding show the particle velocities relative to rigid co-motion with the drum.

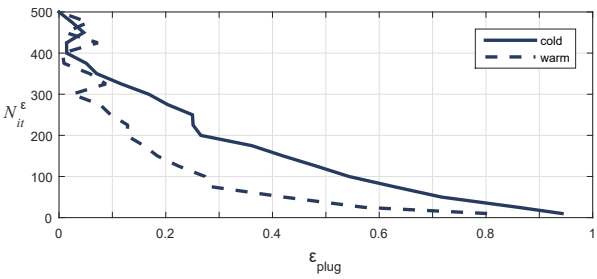

Fig. 12 The convergence of the plug zone fraction number for cold starting and history based warm starting.

the dynamic angle of repose also show that warm starting converges faster although to a slightly higher angle $\theta^{500}_{\mathrm{w,plug}} = 50°$ compared to $\theta^{500}_{\mathrm{plug}} = 48°$, see Fig. 13. The dynamic angle of repose is measured as the displacement of the material centre of mass from the z-axis which is more robust than tracking the surface.

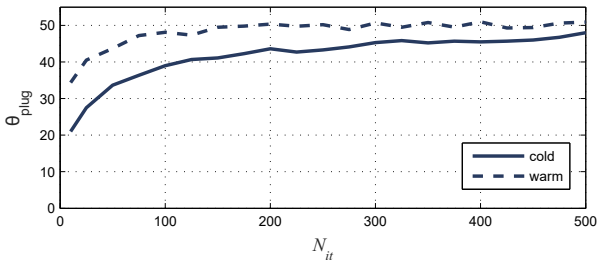

Fig. 13 The dynamic angle of repose as function of number of iterations for cold starting and history based warm starting.

4.4 Triaxial shear

The triaxial shear test is constructed by six dynamic rigid walls of mass 100 kg each that are driven with prismatic motors to apply a specific stress $\sigma_i = f_i/A_i$, where A_i is the cross-section area and f_i the applied motor force in the coordinate direction $i = x, y, z$, see Fig. **??** First, a hydrostatic pressure of $\sigma^{\mathrm{h}} = 100$ Pa is

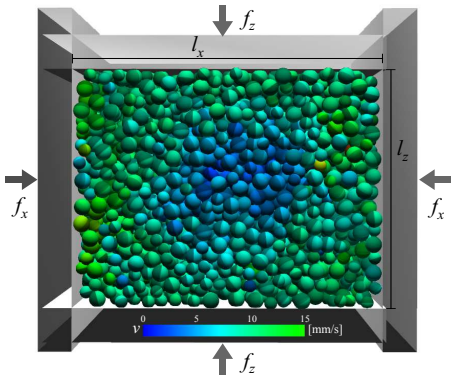

Fig. 14 Sample image from the triaxial test.

applied on all sides. Then the top and bottom walls are driven inwards at 0.01 m/s by regulating σ_z and maintaining a constant side wall pressure at $\sigma_x = \sigma_y = \sigma^{\mathrm{h}}$. At some critical deviator stress $\sigma^{\mathrm{c}}_z - \sigma^{\mathrm{h}}$ the material fail to sustain further increase in stress and starts to shear indefinitely. The transition is more or less sharp depending on the initial packing ratio, hydrostatic pressure and applied shear rate. In this test the Young's modulus is set to the stiffer value of $E = 60$ MPa to get a sharper transition between compression and shear. The critical axial stress σ^{c}_z is computed as the averaged σ_z in the shear phase between lateral strain $\varepsilon = 10\%$ to $\varepsilon = 25\%$. The critical stress deviator depending on the number of iterations for cold starting and history based warm starting is shown in Fig. 15 Both curves converge to about 1 ± 0.2 kPa. With warm starting the

stress levels out at $N_{\text{it}} \gtrsim 200$ while cold starting require $N_{\text{it}} = 1000$. Sample curves of the stress deviator

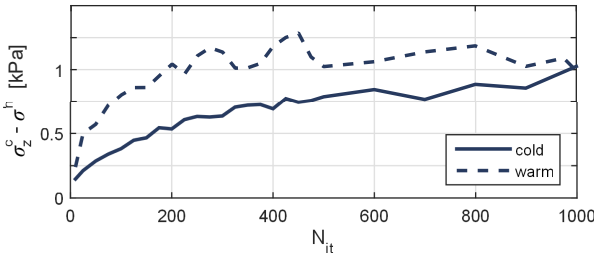

Fig. 15 Critical yield stress as function of number of iterations for cold starting and history based warm starting.

as function of lateral strain are shown in Fig. 16. These confirm the faster convergence when warm starting but also show higher stress fluctuations in the shear phase. Whether this is an artefact of the warm starting or an actual feature of the triaxial test has not been pursued.

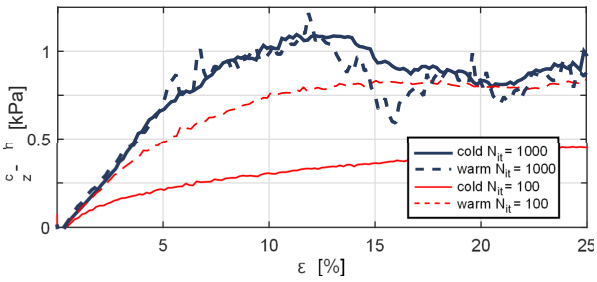

Fig. 16 Sample stress curves in triaxial test for 100 and 1000 iterations.

5 Application example

The effect of using warm starting in practical simulation applications is illustrated with two application examples. The first example is part of a balling drum circuit used in ore pelletizing systems [16], see Fig. 17. Simulations are used for the purpose of process control and for finding a design of the drum outlet that maximizes an evenly distributed throughput on the roller sieve where material is size distributed. Three distinctive subsystems with different dynamics can be identified. Firstly, there is the drum with an almost stationary flow. Secondly, material is distributed onto quasistationary piles on a wide-belt conveyor. Thirdly, the particles disperse over a roller sieve with increasing gap size downwards to achieve a size separation. The design problem is fore mostly a geometric flow problem and the material distribution need to be computed with sufficient accuracy. We assume 5 % is a required accuracy for dynamic and static angle of repose. From Fig. 13 we estimate that warm starting is roughly three times more computationally efficient in computing the drum flow and, according to Fig. 6, twice as efficient for pile formation on the conveyor. The flow on the roller sieve is more disperse and collision dominated requiring only few iterations ($N_{\text{p}} < 25$) and it can be expected that warm and cold starting are equally efficient. The overall computational speed-up by applying warm starting is thus estimated to a factor 2.

The second example is an excavator. A rectangular trench is filled with roughly 10^5 spherical particles of uniform size distribution between 25 and 100 mm and particle mass density of 2500 kg/m^3. The excavator is modeled as a rigid multibody system of total mass 50 ton divided in 10 bodies, 8 joints and 3 linear actuators (hydraulic cylinders) and one rotational motor. The full system of granulars and vehicle take the mathematical form of Eq. (1) and is solved using a split solver where the vehicle part is solved using a direct block-sparse pivoting method [15] and the granular material with a PGS solver as described in this paper. Simulations were run with time-step $h = 2.5$ ms, which allow for a low number of iterations. The machine perform an excavation cycle by a pre-programmed control signal to the actuators. The resulting actuator forces are measured and these include the back reaction from the resistance and inertia of the granular material. Two simulations, with and without warm starting, are run with $N_{\text{it}} = 25$. Sample images from the simulations are shown in Fig. 18. Observe the difference in height surface of the granular material due to artificial compression and frictional slippage due to numerical errors in the PGS solve. The undisturbed height in the two simulations differ by 10 % and volume of displaced material differ by at least 30 %. The difference in granular dynamics also affect the measured force response. The force trajectory of the middle actuator is provided in Fig. 19. In the phase between $8-10$ s, when the bucket is dragged through the material the force when using warm starting is almost 50% larger because more material is set in motion and stronger resistance to shear motion. Whether $h = 2.5$ ms, $N_{\text{it}} = 25$ and the improvement by using warm starting give sufficiently accurate force response depend on the intended use of the data and require further convergence analysis. On a desktop computer[2] with the given NDEM settings the computational time is roughly 100 s per realtime second.

[2] Performance measurement are made on a desktop computer with Intel(R) Core(TM) Xeon X5690, 3.46 GHz, 48 GB RAM on a Linux 64 bit system.

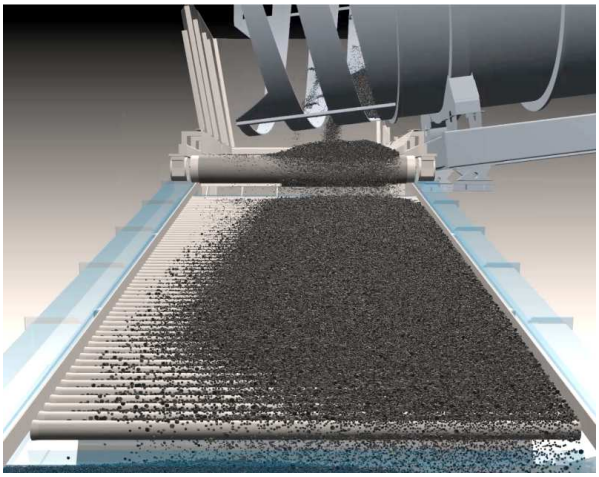

Fig. 17 A balling drum circuit with granular material in different states.

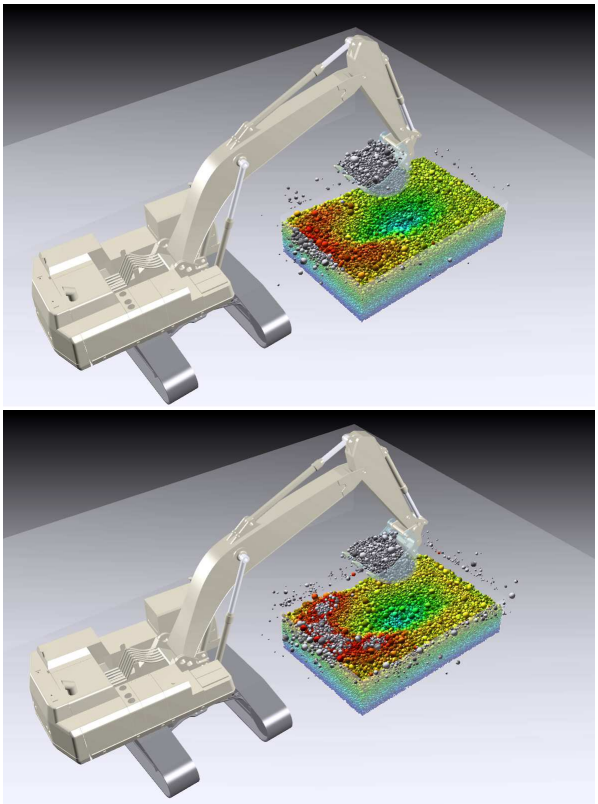

Fig. 18 An excavator digging in trench with 10^5 particles, $h = 2.5ms$, $N_{\text{it}} = 25$ and using cold starting (top) and warm starting (bottom). The colour codes the particle height with red to blue ranging from 0 m to −2 m. Gray particles are above 0 m.

6 Conclusions

The convergence of the projected Gauss-Seidel algorithm for NDEM simulation is increased by warm starting with the solution from previous time-step. The computational speed-up by warm starting is demonstrated

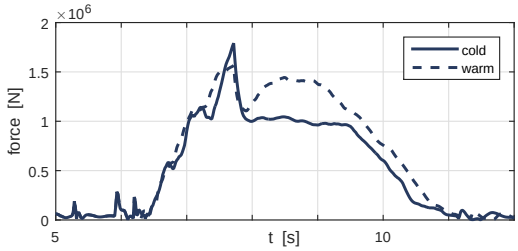

Fig. 19 The force trajectory of the middle link pistons of the excavator while digging with $N_{\text{it}} = 25$ using cold starting and warm starting.

to be about $2-5$ for a wide range of systems including pile formation, granular drum flow and triaxial shear. An examination of the residual distribution show that convergence improvement primarily improve on the velocity constraints - friction and rolling resistance - and less so on the normal force constraints. Warm starting the Lagrange multiplier based on 85 % of the value from last time-step was found to give best results. Warm starting based on an explicit contact force model give only marginal speed-up, for example 20 % for a pile formation. This is not surprising since the damping coefficients in the dissipation models for sliding and rolling are not physics based and can only predict the value of contact forces in slide mode but not of stick mode inside the friction and rolling resistance limits. For materials under high stress, compared to the stress produced by the materials own weight, warm starting show larger fluctuations in the stress when shearing. Whether this is an artefact or correct behaviour has not been established. A more in depth analysis of systems under large stress should be made considering also alternative size of time step, shear rate, hydrostatic load stress and particle stiffness.

Acknowledgements This project was supported by Algoryx Simulations, LKAB, UMIT Research Lab and VINNOVA (dnr 2014-01901).

Appendix

A. Simulation algorithm

The algorithm for simulating a system of granular material using NDEM with PGS solver with warm starting is given in Algorithm 1. Based on the Hertz contact law, each contact α between body a and b add contributions to the constraint vector and normal and friction Jaco-

Algorithm 1 NDEM simulation with warm started PGS solver

set constants and parameters
initial state: $(\mathbf{x}_0, \mathbf{v}_0)$
for $i = 0, 1, 2, \ldots, t/\Delta t$ **do** ▷ Time stepping
  contact detection
  compute $\mathbf{g}, \mathbf{G}, \boldsymbol{\Sigma}, \mathbf{D}$
  impact stage PGS solve $\mathbf{v}_i \rightarrow (\mathbf{v}_i^+, \boldsymbol{\lambda}_i^+)$ ▷ impacts
  compute $\mathbf{q}_\mathrm{n} = -(4/\Delta t)\boldsymbol{\Upsilon}_\mathrm{n}\mathbf{g}_\mathrm{n} + \boldsymbol{\Upsilon}_\mathrm{n}\mathbf{G}_\mathrm{n}\mathbf{v}_i^+$
  pre-step $\mathbf{v} = \mathbf{v}_i^+ + \Delta t\mathbf{M}^{-1}\mathbf{f}_\mathrm{ext}$
  $\boldsymbol{\lambda}_{k_0} = \mathbf{0}$ or warm start $\boldsymbol{\lambda}_{k_0}$
  warm-step $\mathbf{v} = \mathbf{v} + \mathbf{M}^{-1}\mathbf{G}^\mathrm{T}\boldsymbol{\lambda}_{k_0}$
  − − − ▷ PGS solve for continuous contacts− − −−
  for $k = 1, \ldots, N_\mathrm{it}$ and **while** *criteria*$(\mathbf{r})$ **do**
    for each contact $\alpha = 0, 1, \ldots, N_\mathrm{c} - 1$ **do**
      for each constraint n of contact α **do**
        $\mathbf{r}_{n,k}^{(\alpha)} = -\mathbf{q}_{n,k}^{(\alpha)} + \mathbf{G}_n^{(\alpha)}\mathbf{v}$ ▷ residual
        $\boldsymbol{\lambda}_{n,k}^{(\alpha)} = \boldsymbol{\lambda}_{n,k-1}^{(\alpha)} + \mathbf{D}_{n,(\alpha)}^{-1}\mathbf{r}_{n,k}^{(\alpha)}$ ▷ multiplier
        $\boldsymbol{\lambda}_{n,k}^{(\alpha)} \leftarrow \mathrm{proj}_{\mathcal{C}_\mu}(\boldsymbol{\lambda}_k^{(\alpha)})$ ▷ project
        $\Delta\boldsymbol{\lambda}_{n,k}^{(\alpha)} = \boldsymbol{\lambda}_{n,k}^{(\alpha)} - \boldsymbol{\lambda}_{n,k-1}^{(\alpha)}$
        $\mathbf{v} = \mathbf{v} + \mathbf{M}^{-1}\mathbf{G}_{n,(\alpha)}^T\Delta\boldsymbol{\lambda}_{n,k}^{(\alpha)}$
      end for
    end for
  end for
  $\mathbf{v}_{i+1} = \mathbf{v}$ ▷ velocity update
  $\mathbf{x}_{i+1} = \mathbf{x}_i + \Delta t\mathbf{v}_{i+1}$ ▷ position update
end for

bians according to

$$
\begin{aligned}
\delta_{(\alpha)} &= \mathbf{n}_{(\alpha)}^\mathrm{T}(\mathbf{x}_a + \mathbf{d}_a^{(\alpha)} - \mathbf{x}_b - \mathbf{d}_b^{(\alpha)}) \\
g_{(\alpha)} &= \delta_{(\alpha)}^{e_\mathrm{H}} \quad , \; e_\mathrm{H} = 5/4 \\
\mathbf{G}_{\mathrm{n}a}^{(\alpha)} &= e_\mathrm{H} g_{(\alpha)}^{e_\mathrm{H}-1}\begin{bmatrix} -\mathbf{n}_{(\alpha)}^\mathrm{T} & -(\mathbf{d}_a^{(\alpha)} \times \mathbf{n}_{(\alpha)})^\mathrm{T} \end{bmatrix} \\
\mathbf{G}_{\mathrm{n}b}^{(\alpha)} &= e_\mathrm{H} g_{(\alpha)}^{e_\mathrm{H}-1}\begin{bmatrix} \mathbf{n}_{(\alpha)}^\mathrm{T} & (\mathbf{d}_b^{(\alpha)} \times \mathbf{n}_{(\alpha)})^\mathrm{T} \end{bmatrix} \\
\mathbf{G}_{ta}^{(\alpha)} &= \begin{bmatrix} -\mathbf{t}_1^{(\alpha)\mathrm{T}} & -(\mathbf{d}_a^{(\alpha)} \times \mathbf{t}_1^{(\alpha)})^\mathrm{T} \\ -\mathbf{t}_2^{(\alpha)\mathrm{T}} & -(\mathbf{d}_a^{(\alpha)} \times \mathbf{t}_2^{(\alpha)})^\mathrm{T} \end{bmatrix} \\
\mathbf{G}_{tb}^{(\alpha)} &= \begin{bmatrix} \mathbf{t}_1^{(\alpha)\mathrm{T}} & (\mathbf{d}_b^{(\alpha)} \times \mathbf{t}_1^{(\alpha)})^\mathrm{T} \\ \mathbf{t}_2^{(\alpha)\mathrm{T}} & (\mathbf{d}_b^{(\alpha)} \times \mathbf{t}_2^{(\alpha)})^\mathrm{T} \end{bmatrix} \\
\mathbf{G}_{ra}^{(\alpha)} &= \begin{bmatrix} \mathbf{0}_{1\times3} & \mathbf{t}_1^{(\alpha)\mathrm{T}} & \mathbf{0}_{1\times3} & -\mathbf{t}_1^{(\alpha)\mathrm{T}} \\ \mathbf{0}_{1\times3} & \mathbf{t}_2^{(\alpha)\mathrm{T}} & \mathbf{0}_{1\times3} & -\mathbf{t}_2^{(\alpha)\mathrm{T}} \\ \mathbf{0}_{1\times3} & \mathbf{n}^{(\alpha)\mathrm{T}} & \mathbf{0}_{1\times3} & -\mathbf{n}^{(\alpha)\mathrm{T}} \end{bmatrix} \\
\mathbf{G}_{rb}^{(\alpha)} &= \begin{bmatrix} \mathbf{0}_{1\times3} & -\mathbf{t}_1^{(\alpha)\mathrm{T}} & \mathbf{0}_{1\times3} & \mathbf{t}_1^{(\alpha)\mathrm{T}} \\ \mathbf{0}_{1\times3} & -\mathbf{t}_2^{(\alpha)\mathrm{T}} & \mathbf{0}_{1\times3} & \mathbf{t}_2^{(\alpha)\mathrm{T}} \\ \mathbf{0}_{1\times3} & -\mathbf{n}^{(\alpha)\mathrm{T}} & \mathbf{0}_{1\times3} & \mathbf{n}^{(\alpha)\mathrm{T}} \end{bmatrix}
\end{aligned}
\tag{15}
$$

where $\mathbf{d}_a^{(\alpha)}$ and $\mathbf{d}_b^{(\alpha)}$ are the positions of the contact point α relative to the particle positions $\mathbf{x}_a$ and $\mathbf{x}_b$. The orthonormal contact normal and tangent vectors are $\mathbf{n}^{(\alpha)}$, $\mathbf{t}^{(\alpha)_1}$ and $\mathbf{t}^{(\alpha)_2}$.

The diagonal matrices and Schur complement matrix $\mathbf{D}$ are

$$
\begin{aligned}
\boldsymbol{\Sigma}_\mathrm{n} &= \frac{4}{\Delta t^2}\frac{\varepsilon_\mathrm{n}}{1 + 4\frac{\tau_\mathrm{n}}{\Delta t}}\mathbf{1}_{N_c\times N_c} \\
\boldsymbol{\Sigma}_\mathrm{t} &= \frac{\gamma_\mathrm{t}}{\Delta t}\mathbf{1}_{2N_c\times 2N_c} \\
\boldsymbol{\Sigma}_\mathrm{r} &= \frac{\gamma_\mathrm{r}}{\Delta t}\mathbf{1}_{3N_c\times 3N_c} \\
\boldsymbol{\Upsilon}_\mathrm{n} &= \frac{1}{1 + 4\frac{\tau_\mathrm{n}}{\Delta t}}\mathbf{1}_{N_c\times N_c} \\
\mathbf{D} &= \mathbf{G}\mathbf{M}^{-1}\mathbf{G}^T + \boldsymbol{\Sigma}
\end{aligned}
\tag{16}
$$

The mapping between MCP parameters and material parameters are

$$
\begin{aligned}
\varepsilon_\mathrm{n} &= e_\mathrm{H}/k_\mathrm{n} = 3e_\mathrm{H}(1-\nu^2)/E\sqrt{r^*} \\
\tau_\mathrm{n} &= \max(n_\mathrm{s}\Delta t, \varepsilon_\mathrm{n}/\gamma_\mathrm{n}) \\
\gamma_\mathrm{n}^{-1} &= k_\mathrm{n}c/e_\mathrm{H}^2
\end{aligned}
\tag{17}
$$

where $r^* = (r_a + r_b)/r_a r_b$ is the effective radius and we use $\gamma_\mathrm{t} = \gamma_\mathrm{r} = 10^{-6}$, $n_\mathrm{s} = 2$.

References

1. J. J. Moreau. Numerical aspects of the sweeping process. *Computer Methods in Applied Mechanics and Engineering*, 177:329–349, July 1999.
2. M. Jean. The non-smooth contact dynamics method. *Computer Methods in Applied Mechanics and Engineering*, 177:235–257, July 1999.
3. Farhang Radjai and Vincent Richefeu. Contact dynamics as a nonsmooth discrete element method. *Mechanics of Materials*, 41(6):715–728, June 2009.
4. B Brogliato, Aa Ten Dam, L Paoli, F Génot, and M Abadie. Numerical simulation of finite dimensional multibody nonsmooth mechanical systems. *Applied Mechanics Reviews*, 55(2):107–150, 2002.
5. T. Unger, L. Brendel, D. Wolf, and J. Kertsz. Elastic behavior in contact dynamics of rigid particles. *Physical Review E*, 65(6):7, 2002.
6. T. Precklik, U. Rude Ultrascale simulations of non-smooth granular dynamics, *Comp. Part. Mech.*, DOI 10.1007/s40571-015-0047-6, 2015.
7. V. Visseq, P. Alart, D. Dureisseix, High performance computing of discrete nonsmooth contact dynamics with domain decomposition. *Int J Numer Methods Eng*, 96(9):584598, 2013.
8. D. Kaufman, S. Sueda, D. James, D. Pai, Staggered Projections for Frictional Contact in Multibody Systems, ACM Transactions on Graphics 27 (5) (2008) 164:1–164:11.
9. D. Kaufman, Coupled Principles for Computational Frictional Contact Mechanics, Dissertation (2009).
10. K. Erleben, Numerical Methods for Linear Complementarity Problems in Physics-based Animation, ACM SIGGRAPH 2013 Courses, 8:1–8:42 (2013).
11. A. Moravnszky, P. Terdiman, Fast Contact Reduction for Dynamics Simulation, in Game Programming Gems 4 ed. A. Kirmse, Charles River Media (2004) 253–263.
12. E. Todorov, Implicit nonlinear complementarity: A new approach to contact dynamics, 2010 IEEE International Conference on Robotics and Automation (ICRA) (2010) 2322–2329.

13. C. Lacoursière, Regularized, stabilized, variational methods for multibodies, in: D. F. Peter Bunus, C. Führer (Eds.), The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), Linköping University Electronic Press, 2007, pp. 40–48.
14. M. Servin, D. Wang, C. Lacoursière, K. Bodin, Examining the smooth and nonsmooth discrete element approach to granular matter, Int. J. Numer. Meth. Engng. 97 (2014) 878–902.
15. Algoryx Simulations. AGX Dynamics, December 2014.
16. D. Wang, M. Servin, T. Berglund, K-O. Mickelsson, S. Rnnbck, Parametrization and validation of a nonsmooth discrete element method for simulating flows of iron ore green pellets, Powder Technology, 283 (2015) 475–487.

Adaptive model reduction for nonsmooth discrete element simulation

Martin Servin Da Wang

December 1, 2015

Abstract

A method for adaptive model order reduction for nonsmooth discrete element simulation is developed and analysed in numerical experiments. Regions of the granular media that collectively move as rigid bodies are substituted with rigid bodies of the corresponding shape and mass distribution. The method also support particles merging with articulated multibody systems. A model approximation error is defined and used to derive conditions for when and where to apply reduction and refinement back into particles and smaller rigid bodies. Three methods for refinement are proposed and tested: prediction from contact events, trial solutions computed in the background and using split sensors. The computational performance can be increased by 5 - 50 times for model reduction level between 70 - 95 %.

1 Introduction

Simulation of granular matter is important for increased understanding of the nature of granular media and as an engineering tool for design, control and optimization of processing and transportation systems [1]. With the discrete element method (DEM) the material is modeled as a system of contacting rigid bodies, referred to as particles in this text. This provides detailed information about force structures and particle kinematics on a microscopic level. DEM accurately capture many of the characteristic phenomena of granular media - for instance jamming, dilatancy, emergence of strong force chains, strain localization, avalanches and size-segregation upon fluidization - that are difficult or even impossible to model with continuum based methods. The required computational time increase with the number of bodies and this limit the practical use of DEM for exhaustive simulation studies of large-scale systems and large parameter spaces. One strategy to remedy this is to increase the computational performance by use of parallel algorithms and dedicated hardware. Another strategy is the use of implicit integration with large time-step using the nonsmooth DEM (NDEM) [2], also known as the contact dynamics method [3,4], where velocities may be time-discontinuous and impulses can propagate instantly through the system.

A third strategy, that is pursued in the paper, is to reduce the computational complexity by identifying regions in the granular media where the particles may be substituted by approximate models with less degrees of freedom.

1.1 Previous work

Model order reduction is well established and widely used for reducing the computational complexity in solid and fluid mechanics, dynamical systems and control theory [5, 6]. In multibody dynamics, it is often used for reducing the degrees of freedom of flexible bodies [7] while the number of multibodies are preserved. There are few examples that resemble model order reduction for the discrete element method. Glössmann [8] applied the Karhunen-Loéve transform to clusters of discrete elements that show dynamic coherence to reduce the order of generalized coordinates. In the combined finite-discrete element method (FDEM) each body is represented as a discrete element that is also discretized by a finite element method [9]. The bodies may deform, fracture and fragment indefinitely into smaller elements based on the internal stresses. An inversion of this is the hierarchical multiscale modeling of granular media discretized by a coarse mesh of finite elements in combination with assemblies of fine grained discrete elements for numerical computation of the local constitutive law for the finite element computations [10, 11]. A two-scale and two-method approach for modeling granular materials is presented in [12], where DEM is used for domains of large and discontinuous deformations and as an elastoplastic solid using FEM in continuous domains. Automatic simplification algorithms of articulated multibody systems have been developed and shown to increase computational performance by two orders in magnitude on large-scale linkage systems [13, 14]. Dynamic formation of both rigid aggregates (clumped particles) and elastic aggregates (clustered particles), are supported by several discrete element codes and is used for modeling grains in brittle rock [15].

1.2 Outline of the idea and the challenges

The idea is to identify regions of the granular media that collectively move as rigid bodies and substitute each of these regions with rigid aggregates of the corresponding shape and mass distribution. Particles and rigid aggregates may also merge with rigid bodies or kinematic geometries that do not represent granular media, for instance the particles in an excavator bucket may merge with the bucket into one single rigid body. The aggregated rigid bodies still contribute to the system dynamics but require only a few degrees of freedom. When merged material is disturbed, by a change in external forces or boundary contacts, it may split into smaller constituents that are either rigid aggregates of fewer particles or single particles.

The complexity of systems with granular media in the solid state is thus largely reduced and the computational performance increase correspondingly while the macroscopic dynamics may be preserved. For granular media in the

gaseous or liquid state, on the other hand, a model reduction into rigid aggregates will cause significant errors. This paper is limited to model reduction into rigid bodies. The extension to elastoplastic bodies can be imagined but is beyond the current scope. The ultimate goal is to achieve optimal trade-off between: maximum system reduction; minimum errors on the macroscopic dynamics; minimum computational overhead.

The main challenge is to predict when and where merged material should be refined, or split. Most real systems are in a combination of the three states of granular media: solid, fluid and gas [16]. If the split conditions are too restrictive, or the merge condition too progressive, the bulk properties become wrong. This may appear as incorrect angle of repose, artificial resistance to compression and shearing forces and erroneous rheology in the fluid state. If the split conditions are too permissive, or the merge conditions too strict, most particle will remain free and the there is no computational gain. Also, the computational overhead of model reduction must be small in comparison to the computational time for the fully resolved system.

The idea is illustrated in Fig. 1 with an excavator digging in a bed of granular material. Only a finite domain of the material around the bucket is displaced and need to be simulated dynamically. The remaining part is static and contribute merely with supporting contact pressure. When the bucket is filled and starts to lift, most material co-move rigidly with the bucket. If the purpose of the simulation is to compute the dynamics of the excavator and the load forces in the mechanism, the material and the bucket can be approximated by a single rigid body. When the bucket accelerates or rotate slowly the force distribution in the granular material change and it might start to flow. Several methods for predicting splitting of the rigid aggregates are proposed and tested in numerical experiments. The method are based either on contact events or estimating force distribution or particle motion by computations in the background.

2 Particles, rigid aggregate and multibodies

This section is devoted to the mathematical representation of contacting particles and rigid bodies as multibody systems with nonsmooth dynamics and kinematic constraints.

2.1 Global system variables

The variables for particles, rigid aggregates and other rigid bodies are components of the *global system variables* that we denote $\mathbf{x}, \mathbf{v}, \boldsymbol{f}$ and $\mathbf{M}$ and refer to as generalized position, velocity, force and mass although they are concatenations of linear and rotational degrees of freedom. Quaternions are used for representing orientations. The matrix dimension of the global quantities are $\dim(\mathbf{x}) = 7N_\mathrm{b}$, $\dim(\mathbf{v}) = \dim(\mathbf{f}) = 6N_\mathrm{b}$, $\dim(\mathbf{M}) = 6N_\mathrm{b} \times 6N_\mathrm{b}$, where N_b is the total number of bodies.

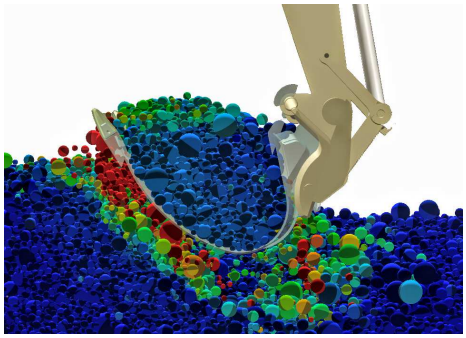

Figure 1: Illustration of model reduction in an excavation scenario simulated without model reduction. The particles are color coded by velocity from blue (stationary) to red. Most particles in the bed are in relative rest and would be well described by a single rigid body. The particles in the excavator bucket may be aggregated with the bucket into a single rigid body. The main challenge is to predict when and where the merged material should split.

2.2 Nonsmooth multibody dynamics

The following equations of motion for nonsmooth multibody dynamics are assumed:

$$\mathbf{M}\dot{\mathbf{v}} + \dot{\mathbf{M}}\mathbf{v} = \mathbf{f}_{\text{ext}} + \mathbf{G}_{\text{n}}^{\text{T}}\boldsymbol{\lambda}_{\text{n}} + \mathbf{G}_{\text{t}}^{\text{T}}\boldsymbol{\lambda}_{\text{t}} + \mathbf{G}_{\text{r}}^{\text{T}}\boldsymbol{\lambda}_{\text{r}} + \mathbf{G}_{\text{j}}^{\text{T}}\boldsymbol{\lambda}_{\text{j}}, \quad (1)$$

$$0 \leq \varepsilon_{\text{n}}\boldsymbol{\lambda}_{\text{n}} + \mathbf{g}_{\text{n}} + \tau_{\text{n}}\mathbf{G}_{\text{n}}\mathbf{v} \perp \boldsymbol{\lambda}_{\text{n}} \geq 0, \quad (2)$$

$$\gamma_{\text{t}}\boldsymbol{\lambda}_{\text{t}} + \mathbf{G}_{\text{t}}\mathbf{v} = 0, \quad |\boldsymbol{\lambda}_{\text{t}}^{(\alpha)}| \leq \mu_{\text{t}}|\mathbf{G}_{\text{n}}^{(\alpha)\text{T}}\lambda_{\text{n}}^{(\alpha)}|, \quad (3)$$

$$\gamma_{\text{r}}\boldsymbol{\lambda}_{\text{r}} + \mathbf{G}_{\text{r}}\mathbf{v} = 0, \quad |\boldsymbol{\lambda}_{\text{r}}^{(\alpha)}| \leq \mu_{\text{r}} r|\mathbf{G}_{\text{n}}^{(\alpha)\text{T}}\lambda_{\text{n}}^{(\alpha)}|, \quad (4)$$

$$\varepsilon_{\text{j}}\boldsymbol{\lambda}_{\text{j}} + \eta_{\text{j}}\mathbf{g}_{\text{j}} + \tau_{\text{j}}\mathbf{G}_{\text{j}}\mathbf{v} = 0. \quad (5)$$

The first equation is the Newton-Euler equation of motion for rigid bodies with external (smooth) forces $\mathbf{f}_{\text{ext}}$ and constraint force $\mathbf{G}^{\text{T}}\boldsymbol{\lambda}$ with Lagrange multiplier $\boldsymbol{\lambda}$ and Jacobian $\mathbf{G}$, divided into normal (n), tangential (t), rolling (r) and articulated and possibly motorized joints (j). Details are found in the Appendix. Equations (2)-(3) are the Signorini-Coulomb conditions with constraint regularization and stabilization terms ε_{n}, τ_{n} and γ_{t}. With $\varepsilon_{\text{n}} = \tau_{\text{n}} = 0$, Eq. (2) state that bodies should be separated or have zero overlap, $\mathbf{g}_{\text{n}}(\mathbf{x}) \geq 0$, and if so the normal force should be non-cohesive, $\boldsymbol{\lambda}_{\text{n}} \geq 0$. With $\gamma_{\text{t}} = 0$, Eq. (3) state that contacts should have zero relative slide velocity, $\mathbf{G}_{\text{t}}\mathbf{v} = 0$, provided that the friction force remain bounded by the Coulomb friction law with friction coefficient μ_{t}. Eq. (4) similarly constrains relative rotation of contacting bodies provided the constraint torque do not exceed the rolling resistance law with rolling resistance coefficient μ_{r} and radius r. The constraint force, $\mathbf{G}_{\text{j}}^{\text{T}}\boldsymbol{\lambda}_{\text{j}}$, arise for articulated rigid bodies jointed with kinematic links and motors represented with the generic constraint equation (5). With $\varepsilon_{\text{j}}, \tau_{\text{j}} = 0$ and $\eta_{\text{j}} = 1$, it become an ideal holonomic constraint $\mathbf{g}(\mathbf{x}) = 0$. For $\varepsilon, \eta = 0$ and $\tau = 1$, it

become an ideal Pfaffian constraint $\mathbf{G}\dot{\mathbf{x}} = 0$. With $\varepsilon, \eta, \tau \neq 0$ it can represent a generic constraint with compliance and damping. The set of equations (1)-(5) may thus model granular materials strongly coupled with mechatronic systems, e.g., vehicles, robots and mechanical processing units.

The Lagrange multiplier become an auxiliary variable to solve for in addition to position and velocity. The regularization and stabilization terms, ε and γ, introduce compliance and dissipation in motion orthogonal to the constraint manifold. In the absence of the inequality and complementarity conditions, the regularized constraints may be viewed as Legendre transforms of a potential and Rayleigh dissipation function of the form $U_\varepsilon(\mathbf{x}) = \frac{1}{2\varepsilon}\mathbf{g}^T\mathbf{g}$ and $\mathcal{R}_\gamma(\mathbf{x}, \mathbf{v}) = \frac{1}{2\gamma}(\mathbf{G}\mathbf{v})^T(\mathbf{G}\mathbf{v})$ [17, 18]. This enable modeling of arbitrarily stiff elastic and viscous interactions in terms of constraint forces with direct mapping between the regularization and stabilization terms to physical material parameters. This is applied to map the stiffness and damping terms from the nonlinear Hertz contact law, or linear spring and dashpot, from conventional (smooth) discrete element method to the constraint based and nonsmooth discrete element method. The detailed constraints, Jacobians and parameters are found in Ref. [19] and summarized in Appendix A together with the numerical integration scheme used in this paper.

The dynamics is allowed to be nonsmooth which means that velocities may change discontinuously in time. Impacts and frictional stick-slip transitions may thus be considered as instantaneous events and propagate immediately through the entire system by an impulse transfer altering the velocities from $\mathbf{v}_-$ to $\mathbf{v}_+$. The contacts are divided into impacts and continuous contacts, depending on the magnitude of the incoming relative normal velocities $\mathbf{G}_\mathrm{n}\mathbf{v}_-$. The impulse transfer through the system should satisfy the Newton impact law, $\mathbf{G}_\mathrm{n}^{(n)}\mathbf{v}_+ = -e\mathbf{G}_\mathrm{n}^{(n)}\mathbf{v}_-$, with coefficient of restitution e for the impacts (n), as well as preserve all remaining constraints (m) on velocity level, $\mathbf{G}^{(m)}\mathbf{v}_+ = 0$.

2.3 Particles

Each elementary granule is referred to as a *particle* and is modeled as a rigid body with solid geometry. A particle is either free or part of a rigid aggregate of particles. Each free particle is represented by a *dynamic* discrete element obeying the equation of motions in Sec. 2.2 and may interact with other particles and rigid aggregate bodies via contacts. Each particle that is part of a rigid aggregate is a *kinematic* discrete element that co-move rigidly with the aggregate body. No forces are applied to aggregated particles. For simplicity, particles are assumed spherical but can easily be extended to more general geometric shapes. Reference to a specific particle is made by latin indices, e.g., $a, b, \ldots = 1, 2, \ldots, N_\mathrm{p}$, and we use square brackets to emphasize particle index. We use the notations $\vec{\boldsymbol{x}}_{[a]}, \vec{\boldsymbol{v}}_{[a]}, \vec{\boldsymbol{f}}_{[a]}$, $m_{[a]}$ and $d_{[a]}$ for position, translational velocity, force, mass and diameter, and $\vec{\boldsymbol{e}}_{[a]}, \vec{\boldsymbol{\omega}}_{[a]}, \vec{\boldsymbol{\tau}}_{[a]}$ and $\boldsymbol{I}_{[a]}$, for orientation, angular velocity, torque and inertia tensor. Spatial components of a vector or matrix is indexed by $\alpha, \beta = 1, 2, 3$ referring to the x, y, z axes in a global Cartesian coordinate

system, e.g., $x_\alpha^{[a]}$ and $I_{\alpha\beta}$. We concatenate these variables into a particle's generalized position, velocity, force and mass, denoted $\mathbf{x}_{[a]}, \mathbf{v}_{[a]}, \mathbf{f}_{[a]}$ and $\mathbf{M}_{[a]}$, with $\mathbf{v}_{[a]} = (\vec{\boldsymbol{v}}_{[a]}^{\mathrm{T}}, \vec{\boldsymbol{\omega}}_{[a]}^{\mathrm{T}})^{\mathrm{T}}$ etc. and $\mathbf{M}_{[a]} = \mathrm{diag}(m_{[a]}\mathbf{1}_{3\times 3}, \boldsymbol{I}_{[a]})$.

2.4 Rigid aggregates

A *rigid aggregate*, or just aggregate, is a rigid body that represent an aggregate of particles co-moving as a single rigid body. The rigidity is an assumed collective result of the contact forces creating a jammed state although no such internal forces are modeled or computed explicitly. The variables of rigid aggregates are represented with similar notations as for particles, e.g., $\mathbf{x}_{[A]}$ and $\mathbf{M}_{[A]}$, but with capital latin indices $A, B, \ldots = 1, 2, \ldots, N_{\mathrm{a}}$. The set of particles that constitute an aggregate A is denoted $\mathcal{N}_A$. The relation between the dynamic variables of the aggregate and the particles are illustrated in Fig. 2 and computed as follows:

$$m_{[A]} = \sum_{a\in\mathcal{N}_A} m_{[a]}, \tag{6}$$

$$\vec{\boldsymbol{x}}_{[A]} = m_{[A]}^{-1} \sum_{a\in\mathcal{N}_A} m_{[a]}\vec{\boldsymbol{x}}_{[a]}, \tag{7}$$

$$\vec{\boldsymbol{v}}_{[A]} = m_{[A]}^{-1} \sum_{a\in\mathcal{N}_A} m_{[a]}\vec{\boldsymbol{v}}_{[a]}, \tag{8}$$

$$I_{\alpha\beta}^{[A]} = \sum_{a\in\mathcal{N}_A} m_{[a]} \left(|\vec{\boldsymbol{r}}_{[aA]}|^2 \delta_{\alpha\beta} - r_\alpha^{[aA]} r_\beta^{[aA]} \right), \tag{9}$$

$$\vec{\boldsymbol{\omega}}_{[A]} = \boldsymbol{I}_{[A]}^{-1} \sum_{a\in\mathcal{N}_A} m_{[a]}\vec{\boldsymbol{r}}_{[aA]} \times \vec{\boldsymbol{v}}_{[aA]}, \tag{10}$$

where $\vec{\boldsymbol{r}}_{[aA]} = \vec{\boldsymbol{x}}_{[a]} - \vec{\boldsymbol{x}}_{[A]}$ and $\vec{\boldsymbol{v}}_{[aA]} = \vec{\boldsymbol{v}}_{[a]} - \vec{\boldsymbol{v}}_{[A]}$. The kinematics of the aggregated particles $a \in \mathcal{N}_A$ is

$$\vec{\boldsymbol{x}}_{[a]} = \vec{\boldsymbol{x}}_{[A]} + \vec{\boldsymbol{r}}_{[aA]}, \tag{11}$$

$$\vec{\boldsymbol{v}}_{[a]} = \vec{\boldsymbol{v}}_{[A]} + \vec{\boldsymbol{\omega}}_{[A]} \times \vec{\boldsymbol{r}}_{[aA]}, \tag{12}$$

$$\vec{\boldsymbol{\omega}}_{[a]} = \vec{\boldsymbol{\omega}}_{[A]}. \tag{13}$$

2.5 Multibodies

By elementary rigid bodies it is meant bodies that represent other than granular bodies, e.g., part of an articulated multibody. Elementary rigid bodies are also included in model reduction and may form reduced aggregates with both elementary rigid bodies and particles. But the additional complexity of merge and split conditions in articulated system is not covered here. For convenience elementary rigid bodies are represented with the same notation as used for aggregates, that is, capital index in square brackets $[A]$.

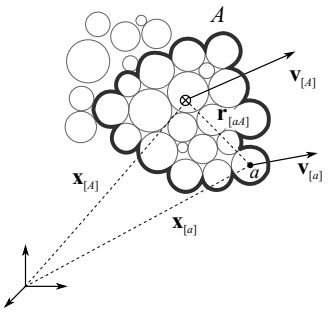

Figure 2: Illustration of a rigid aggregate of contacting particles and some free particles.

2.6 Contacts

The set of contacts is denoted $\mathcal{N}_\mathrm{c}$. Integer $n = 1, 2, \ldots, N_\mathrm{c}$ is used for contact index and this is emphasized by round brackets, e.g., $g^{(n)}$. The gap function $\delta(\mathbf{x})$ measure the magnitude of overlap between two contacting bodies. Contact forces and velocities are sometimes decomposed in the directions of contact normal, $\vec{\mathbf{n}}$ and tangents, $\vec{\mathbf{t}}_1$ and $\vec{\mathbf{t}}_2$. The relative velocity at a contact n between a particle a and a rigid body A can thus be written $\vec{\boldsymbol{u}}^{(n)} = \vec{\boldsymbol{v}}_{[a]} + \vec{\boldsymbol{\omega}}_{[a]} \times \vec{\boldsymbol{d}}^{(n)}_{[a]} - \vec{\boldsymbol{v}}_{[A]} - \vec{\boldsymbol{\omega}}_{[A]} \times \vec{\boldsymbol{d}}^{(n)}_{[A]}$, where $\vec{\boldsymbol{d}}^{(n)}_{[a]}$ is the position of the contact point relative to $\vec{\boldsymbol{x}}_{[a]}$ and $\vec{\boldsymbol{d}}^{(n)}_{[A]}$ relative to $\vec{\boldsymbol{x}}_{[A]}$.

3 Adaptive model order reduction

Let the following equations represent the full system of particles and elementary rigid bodies coupled with constraints

$$\mathbf{M}\ddot{\mathbf{x}} + \dot{\mathbf{M}}\dot{\mathbf{x}} = \mathbf{f}_\mathrm{ext} + \mathbf{G}^\mathrm{T}\boldsymbol{\lambda}, \tag{14}$$

$$\varepsilon\boldsymbol{\lambda} + \eta\mathbf{g}(\mathbf{x}) + \tau\mathbf{G}\dot{\mathbf{x}} = 0, \tag{15}$$

having solution $\mathbf{x} \in \mathbb{R}^n$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$. The constraint equation (15) represent the collection of both position and velocity constraints and it is assumed to be appended with additional inequalities and complementarity conditions for the multiplier. The full system, (14)-(15), is approximated with a reduced system $\tilde{\mathbf{x}} \in \mathbb{R}^{\tilde{n}}$ and $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^{\tilde{m}}$ with less degrees of freedom $\tilde{n} < n$ and $\tilde{m} < m$. The reduced system belong to a subspace of the full system. We define the model order reduction level as $h = 1 - \tilde{n}/n$. When $h \to 1$ the system is maximally reduced to one single rigid body and when $h = 0$ it is fully resolved in all free particles. The approximate relation between the reduced and full system is expressed using subspace transformation matrices $\mathbf{P} \in \mathbb{R}^{n\times\tilde{n}}$ and $\mathbf{Q} \in \mathbb{R}^{m\times\tilde{m}}$

such that

$$\mathbf{x} \approx \mathbf{P}\tilde{\mathbf{x}}, \tag{16}$$
$$\boldsymbol{\lambda} \approx \mathbf{Q}\tilde{\boldsymbol{\lambda}}. \tag{17}$$

Given a rigid aggregate, the transformation matrix $\mathbf{P}$ is easily constructed from the rigid transformations in Eq. (11), that relate the positions of the aggregated particles relative to the aggregate centre. In a rigid aggregate the inter-particle constraints are redundant. The transformation matrix $\mathbf{Q}$ eliminate the redundant equations when a rigid aggregate is substituted for a collection of particles. The reduced multibody system become

$$\tilde{\mathbf{M}}\ddot{\tilde{\mathbf{x}}} = \tilde{\mathbf{f}}_{\mathrm{e}} + \tilde{\mathbf{G}}^{\mathrm{T}}\tilde{\boldsymbol{\lambda}}, \tag{18}$$
$$\varepsilon\tilde{\boldsymbol{\lambda}} + \eta\tilde{\mathbf{g}}(\tilde{\mathbf{x}}) + \tau\tilde{\mathbf{G}}\dot{\tilde{\mathbf{x}}} = 0 \tag{19}$$

with $\tilde{\mathbf{M}} = \mathbf{P}\mathbf{M}\mathbf{P}^T$, $\tilde{\mathbf{f}}_{\mathrm{e}} = \mathbf{P}\mathbf{f}_{\mathrm{ext}} - \dot{\tilde{\mathbf{M}}}\dot{\tilde{\mathbf{x}}}$, $\tilde{\mathbf{G}} = \mathbf{Q}^T\mathbf{G}\mathbf{P}$, $\tilde{\lambda} = \mathbf{Q}\boldsymbol{\lambda}$, $\tilde{\mathbf{g}} = \mathbf{Q}\mathbf{g}$. Note that $\dot{\tilde{\mathbf{M}}}\dot{\tilde{\mathbf{x}}}$ is included as an explicit force in $\tilde{\mathbf{f}}_{\mathrm{e}}$ as is common and often referred to as the gyroscopic force. If the complexity of the reduced system is much smaller than the original, $\tilde{n} \ll n$ and $\tilde{m} \ll m$, the computational efficiency can increase dramatically.

The reduced system can, however, expected to deviate more or less from the original system. The model reduction may be applied adaptively to keep the approximation error below a specified error tolerance. The approximation error is defined

$$\mathcal{E}(t) = \mathbf{x}(t) - \mathbf{P}\tilde{\mathbf{x}}(t) \tag{20}$$

and can be decomposed in two orthogonal terms $\mathcal{E}(t) = \mathcal{E}_{\perp}(t) + \mathcal{E}_{\parallel}(t)$, where

$$\mathcal{E}_{\perp}(t) = \left[\mathbf{1} - \mathbf{P}\mathbf{P}^T\right]\mathbf{x}(t), \tag{21}$$
$$\mathcal{E}_{\parallel}(t) = \mathbf{P}\left[\mathbf{P}^T\mathbf{x}(t) - \tilde{\mathbf{x}}(t).\right] \tag{22}$$

The orthogonal error, $\mathcal{E}_{\perp}$, arise when the trajectory of the full system is not strictly within the subspace of the projection and do not move entirely like a single rigid body. The error parallel to the projection, $\mathcal{E}_{\parallel}$, means that the motion of the reduced system behave different from the original system although they represent equivalent rigid systems, if $\mathcal{E}_{\perp} = 0$. The parallel error can only be computed a posteriori. This may be practically infeasible as it requires solving both the full and reduced system and involve explicit projections between the spaces. The orthogonal error, on the other hand, can be estimated a priori. By substituting Eq. (15) into (14) and multiplying by $\mathbf{P}_{\perp} \equiv \mathbf{1} - \mathbf{P}\mathbf{P}^T$ one obtain an evolution equation for the orthogonal approximation error

$$\ddot{\mathcal{E}}_{\perp} = \underline{\mathbf{M}}^{-1}\left[\underline{\mathbf{f}}_{\mathrm{e}} - \frac{\eta}{\varepsilon}\underline{\mathbf{G}}^T\mathbf{g} - \frac{\tau}{\varepsilon}\underline{\mathbf{G}}^T\underline{\mathbf{G}}\dot{\underline{\mathbf{x}}}\right] \tag{23}$$

where $\underline{\mathbf{M}} = \mathbf{P}_{\perp}\mathbf{M}\mathbf{P}_{\perp}^T$, $\underline{\mathbf{f}}_{\mathrm{e}} = \mathbf{P}_{\perp}\mathbf{f}_{\mathrm{ext}} - \dot{\underline{\mathbf{M}}}\dot{\underline{\mathbf{x}}}$ and $\underline{\mathbf{G}} = \mathbf{G}\mathbf{P}_{\perp}$. The rigid aggregate is a good approximation only if the error is small. When the error is large, or

growing rapidly, the reduced approximation should not be applied. When particles co-move rigidly, the third term vanishes, $\underline{\mathbf{G}}\dot{\underline{\mathbf{x}}} = 0$, since the relative contact velocity is zero. The first and second term cancel when force balance occur in the subspace, $\underline{\mathbf{f}}_{\mathrm{e}} = \frac{\eta}{\varepsilon}\underline{\mathbf{G}}^T\mathbf{g}$. This is equivalent to zero relative acceleration in the contact points, $\underline{\mathbf{G}}\ddot{\underline{\mathbf{x}}} = 0$. Observe that $\underline{\mathbf{M}}^{-1}\underline{\mathbf{f}}_{\mathrm{e}} = 0$ in the case of uniform gravity, since this cause no relative acceleration. We thus identify the following conditions for a rigid aggregate to be a good approximation

$$-\boldsymbol{\xi}_v^- < \underline{\mathbf{G}}^{(n)}\dot{\underline{\mathbf{x}}} < \boldsymbol{\xi}_v^+ \tag{24}$$

and

$$\frac{\left|\underline{\mathbf{f}}_{\mathrm{e}[a]} - \frac{\eta}{\varepsilon}\underline{\mathbf{G}}_{[a]}^T\mathbf{g}\right|}{\left|\underline{\mathbf{f}}_{\mathrm{e}[a]}\right| + \left|\frac{\eta}{\varepsilon}\underline{\mathbf{G}}_{[a]}^T\mathbf{g}\right|} \leq \xi_f \tag{25}$$

for each contact n and particle a of the aggregate with upper and lower thresholds for relative contact velocity $\xi_v^{\pm}$ and force balance bound ξ_f. The inequalities in Eq. (24) should be understood component wise for normal, tangent and rolling. The conditions (24) and (25) provide a starting point for adaptive model order reduction and refinement by merging and splitting particles into and from rigid aggregates. Identifying wether particles should merge is a simple examination of Eq. (24) and (25). Predicting if and which particle should split is non-trivial since it requires some form of estimation of the unknown dynamics of the fully resolved system.

An algorithm for numerical simulation of nonsmooth multibody systems of the form of Eqs. (1)-(5) and with adaptive model reduction is given in Algorithm 1. The algorithm is based on the SPOOK stepper [18] using fix timestep, Δt, and involve solving a mixed complementarity problem (MCP) with matrix $\mathbf{H}$, vector $\mathbf{b}$ and regularization and stabilization matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Upsilon}$ that are found in the Appendix. A popular choice of MCP solver for NDEM is the projected Gauss-Seidel (PGS) method, which is also listed in the Appendix. It should be straightforward to modify the algorithm to other time-integration schemes and solver methods for nonsmooth dynamical systems. The test for model reduction is done directly after the continuous MCP solve when the new velocities are known. If this instead is placed after the position update, some contacts that fulfil condition (24) may be lost due to infinitesimal geometric separation. This would make the model reduction unnecessarily sensitive to solver truncation errors. Particles with contacts that fulfil the test are merged. The test for model refinement is done after contact detection and before solving the impact stage MCP. This way the rigid aggregates can be split before the impact impulses are computed and transferred. Otherwise the granular matter will behave overly rigid. The merge and split processes are described in more detail below as well as a number of methods for predicting model refinement.

Algorithm 1 `main_algorithm`

define constants and parameters
initialil state: $(\mathbf{x}_0, \mathbf{v}_0)$
for $i = 0, 1, 2, \ldots, t/\Delta t - 1$ **do** ▷ time-stepping
  $[\mathbf{g}, \mathcal{N}_\mathrm{c}] = \texttt{contact\_detection}(\mathbf{x}_i, \mathbf{v}_i)$
  $[\mathbf{G}, \mathbf{\Sigma}, \mathbf{\Upsilon}] = \texttt{compute\_contact\_data}(\mathbf{x}_i, \mathbf{v}_i, \mathcal{N}_\mathrm{c})$
  $\texttt{split}(\mathbf{x}_i, \mathbf{v}_i, \mathbf{g}, \mathbf{G}, \mathcal{N}_\mathrm{c})$ ▷ model refinement
  $\mathbf{H} = \texttt{compute\_H}(\mathbf{M}, \mathbf{G}, \mathbf{\Sigma})$
  $\mathbf{b}^- = \texttt{compute\_b\_}(\mathbf{G}, \mathbf{v}_i, e)$
  $[\mathbf{v}_i^+, \boldsymbol{\lambda}_i^+] = \texttt{mcp}(\mathbf{H}, \mathbf{b}^-)$ ▷ impact stage MCP
  $\mathbf{b} = \texttt{compute\_b}(\mathbf{g}, \mathbf{G}, \mathbf{v}_i^+, \mathbf{f}_\mathrm{ext}, \mathbf{\Upsilon})$
  $[\mathbf{v}_{i+1}, \boldsymbol{\lambda}_{i+1}] = \texttt{mcp}(\mathbf{H}, \mathbf{b})$ ▷ continuous stage MCP
  $\mathbf{v}_{i+1} = \texttt{co-move}(\mathbf{x}_i, \mathbf{v}_{i+1})$ ▷ update aggregate particles
  $\texttt{merge}(\mathbf{x}_i, \mathbf{v}_i, \mathbf{g}, \mathbf{G}, \mathcal{N}_\mathrm{c})$ ▷ model reduction
  $\mathbf{x}_{i+1} = \mathbf{x}_i + \Delta t \mathbf{v}_{i+1}$ ▷ position update
end for

3.1 Merge

The model reduction test consist of traversing the contact network $\mathcal{N}_\mathrm{c}$ and testing the condition for rigid motion in Eq. (24). The condition is divided into

$$\mathbf{G}^{(n)}_{\mathrm{n}[aA]}\mathbf{v} \in [-\xi^{\text{i-mrg}}_{\mathrm{n}v}, \xi^{\text{s-mrg}}_{\mathrm{n}v}], \tag{26}$$

$$\mathbf{G}^{(t)}_{\mathrm{t}[aA]}\mathbf{v} \in [-\xi^{\mathrm{mrg}}_{\mathrm{t}v}, \xi^{\mathrm{mrg}}_{\mathrm{t}v}], \tag{27}$$

$$\mathbf{G}^{(r)}_{\mathrm{r}[aA]}\mathbf{v} \in [-\xi^{\mathrm{mrg}}_{\mathrm{r}v}, \xi^{\mathrm{mrg}}_{\mathrm{r}v}], \tag{28}$$

where we separate between incident and separating normal velocity thresholds, $\xi^{\text{i-mrg}}_{\mathrm{n}v}$ and $\xi^{\text{s-mrg}}_{\mathrm{n}v}$, and use symmetric tangential and rolling velocity thresholds $\xi^{\mathrm{mrg}}_{\mathrm{t}v}$ and $\xi^{\mathrm{mrg}}_{\mathrm{r}v}$. The test result in a set of disconnected networks representing rigidly co-moving bodies. Each such network is merged into rigid aggregates. Both particles and elementary rigid bodies are allowed to merge into aggregates. All bodies that are merged into an aggregate body are changed from being a dynamic body to a kinematic body co-moving with the new aggregate body. The aggregate variables $m_{[A]}, \vec{\mathbf{x}}_{[A]}, \vec{\mathbf{v}}_{[A]}, \vec{\mathbf{I}}_{[A]}, \vec{\boldsymbol{\omega}}_{[A]}$ are computed by Eq. (6)-(10). The total mass and momentum is preserved when bodies are merged. The merge procedure is illustrated in Fig. 3. It should be emphasized that the contact network need not a force network, which would require solving Eq. (1)-(5), but merely a connectivity network, and is automatically produced by the contact detection algorithm.

3.2 Split

There are a number of ways to predict if and how the reduced model should be refined by splitting the rigid aggregates into smaller aggregates and free

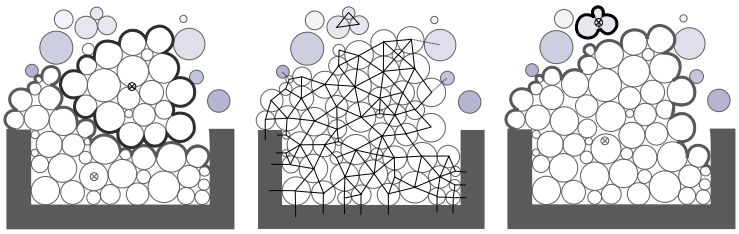

Figure 3: Illustration the merge procedure: detection of contacts involving two aggregates and 11 free particles (left), identification of two disconnected networks fulfilling the merge conditions (middle), creating two new aggregates (right). The colour intensity codes velocity.

particles. One strategy is to rely on contact events, that is, trigger splitting by impacts and separations. Another strategy is to do a fast trial solve in the background, using a more resolved model, and decide splitting from the outcome. A third, more heuristic approach, is to add split sensors in the system. The placement of the sensors can be made automatic, after a posteriori analysis of previous simulations of the same or similar system, or manually, based on experience or perspicacity. The split methods can also be used in combination with each other. The different methods, illustrated in Fig. 4, are outlined in further detail below and tested in numerical experiments. Observe that the split process does not affect on the total mass or momentum. The process of splitting

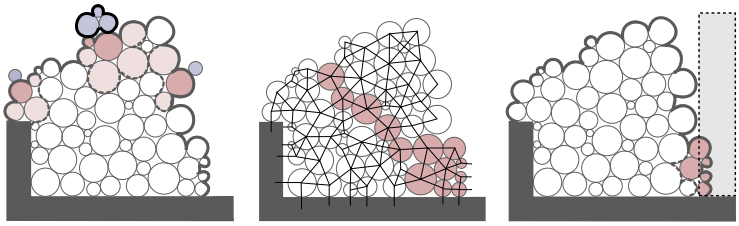

Figure 4: Illustration of different split methods: contact split (left), trial solve split (middle) and sensor split (right). Blue particles have impacting or separating contacts. Red particles are aggregate particles that are to be split from the aggregate. The dashed grey box represent a split sensor.

an aggregate is the same irrespective of the method used for the prediction. The split particles are made dynamic and the new aggregates are determined and created as in the merge process, by analysing the contact network after removing the split particles and applying Eq. (6)-(10).

3.2.1 Contact split

Contacts are either impacting, continuous or separating depending on the sign and magnitude of the relative velocity. For each contact $n \in \mathcal{N}_{\mathrm{c}}$, between particle a and aggregate A, the following conditions for normal, tangential and rolling motion are tested

$$\mathbf{G}^{(n)}_{\mathrm{n}[aA]}\mathbf{v} \notin [-\xi^{\text{i-c}}_{\mathrm{n}v}, \xi^{\text{s-c}}_{\mathrm{n}v}], \tag{29}$$

$$\mathbf{G}^{(n)}_{\mathrm{t}[aA]}\mathbf{v} \notin [-\xi^{\mathrm{c}}_{\mathrm{t}v}, \xi^{\mathrm{c}}_{\mathrm{t}v}], \tag{30}$$

$$\mathbf{G}^{(n)}_{\mathrm{r}[aA]}\mathbf{v} \notin [-\xi^{\mathrm{c}}_{\mathrm{r}v}, \xi^{\mathrm{c}}_{\mathrm{r}v}], \tag{31}$$

with impact split threshold denoted with *i-c* and separation split threshold with *s-c*. If either of the relative contact velocities are found to be outside the valid domain, the aggregate is refined by splitting off particles at the contact node n. The split may be applied to some split depth $N_{\text{c-spl}} \in \mathbb{N}$ along the contact network from the impact (or separation) node in order to capture shock propagation phenomena. Observe that the split action is merely a redefinition of which particles are free and which are kinematically bound to aggregate bodies. This does not immediately alter the position or velocity of any particle. The split is then proceeded with solving the impact stage MCP and the continuous stage MCP, see Algorithm 1. There is therefore no risk, other than unnecessary computations, of splitting too many particles. These will be merged back with the aggregate if the merge condition is fulfilled after solving the impact and continuous MCP. The numerical experiments presented in this paper is limited to splitting aggregate particles that are triggered by impact or separation directly and their contact neighbours, i.e., $N_{\text{c-spl}} = 2$. Impacts between two aggregates are treated the same way.

3.2.2 Trial solve split

A trial solver is run in the background to estimate the dynamics of the full system. The background system state is initialized by projecting the reduced sub-space system back to the full resolution space of Eq. (11)-(13). The purpose of the background solve is not to do precise integration of the particle positions and velocities but to provide sufficient estimate for if and how to split rigid aggregates. Assuming that the state of the reduced system from previous timestep was a good approximation of the full system it is conjectured that doing a PGS solve of the full system MCP with low number of iterations $N^{\mathrm{tr}}_{\mathrm{it}}$ will suffice for this, although the error of such a simulation would increase rapidly over time. Other alternatives can be imagined, e.g., doing the background computation using other solvers or on a partially resolved system. The background trial solution is tested for the following conditions for relative velocity of each contact

$n \in \mathcal{N}_\mathrm{c}$

$$\mathbf{G}^{(n)}_{\mathrm{n}[aA]}\mathbf{v} \notin [-\xi^{\text{i-tr}}_{\mathrm{n}v}, \xi^{\text{s-tr}}_{\mathrm{n}v}], \tag{32}$$

$$\mathbf{G}^{(n)}_{\mathrm{t}[aA]}\mathbf{v} \notin [-\xi^{\mathrm{tr}}_{\mathrm{t}v}, \xi^{\mathrm{tr}}_{\mathrm{t}v}], \tag{33}$$

$$\mathbf{G}^{(n)}_{\mathrm{r}[aA]}\mathbf{v} \notin [-\xi^{\mathrm{tr}}_{\mathrm{r}v}, \xi^{\mathrm{tr}}_{\mathrm{r}v}], \tag{34}$$

and the following conditions for force and torque balance are tested separately for each aggregated particle a

$$\frac{|\vec{\mathbf{f}}_{\mathrm{e}[a]} - \mathbf{G}^T_{f[a]}\boldsymbol{\lambda}|}{|\vec{\mathbf{f}}_{\mathrm{e}[a]}| + |\mathbf{G}^T_{f[a]}\boldsymbol{\lambda}|} < \xi^{\mathrm{tr}}_f, \tag{35}$$

$$\frac{|\vec{\boldsymbol{\tau}}_{\mathrm{e}[a]} - \mathbf{G}^T_{\tau[a]}\boldsymbol{\lambda}|}{|\vec{\boldsymbol{\tau}}_{\mathrm{e}[a]}| + |\mathbf{G}^T_{\tau[a]}\lambda|} < \xi^{\mathrm{tr}}_\tau. \tag{36}$$

where $\mathbf{G}^T_{f[a]}\boldsymbol{\lambda}$ and $\mathbf{G}^T_{\tau[a]}\boldsymbol{\lambda}$ are the force and torque components of the sum of generalized contact forces acting on particle a, $\mathbf{G}^T_{[a]}\boldsymbol{\lambda} = \sum_{b\in\mathcal{N}^a_\mathrm{c}} \mathbf{G}^T_{[ab]}\boldsymbol{\lambda}$. The Jacobians are the blocks for linear and rotational degrees of freedom $\mathbf{G}^T_{[a]} = [\mathbf{G}^T_{f[a]}, \mathbf{G}^T_{\tau[a]}]$. Observe that $\boldsymbol{\lambda}$ has replaced $\frac{\eta}{\varepsilon}\mathbf{g}$ in the force balance condition (25). This is a stronger test as it can detect also acceleration due to impulse force propagation through the system that has not yet resulted in relative contact velocity or particle displacements. The particles that are indicated by the tests are eliminated from the aggregate body and activated as a dynamic particle. In the numerical implementation a small perturbation is added to the denominators to avoid numerical round-off errors.

3.2.3 Sensor split

The split sensor is simply a geometrical shape that triggers model refinement of aggregate bodies that overlap with the sensor geometry. The splitting is applied only to the aggregate particles that overlap the sensor. Observe that the sensors are physically transparent and do not produce any contact forces. Split sensors is a useful tool for when it can be anticipated where model refinement is required without doing a background trial solve. The sensor must be given a size, shape and position, either manually or automated based on data from simulations, models or experiments.

4 Numerical experiments

The described method for adaptive model order reduction is investigated in numerical experiments. The test systems are a conveyor with a continuous formation of a pile on one end and discharge at the other end, a granular collapse and granular flow in a slowly rotating drum. The granular dynamics from using model reduction is compared with reference simulations run in full resolution.

Table 1: Model and NDEM parameters

d	$13, 10$ mm
ρ	3700 kg/m^3
k_n	3 kN/m
e	0.18
μ_t	0.91
μ_r	0.32
Δt	5 ms
N_p	$4 - 90 \cdot 10^3$
N_it	150
$N_\mathrm{it}^\mathrm{ref}$	$150, 500$

The achieved model order reduction level, $h(t) = 1 - \tilde{N}_\mathrm{p}(t)/N_\mathrm{p}$, is monitored. The systems are selected to represent different types of flow and transitions between static and dynamic states. Model reduction can be expected to work well for the formation of piles where the dynamics mainly occur in the surface layers. For granular discharge and collapse there is high risk of approximation errors by to predict when and where the aggregate should split and flow. In a slowly rotating drum the granular flow separate in one layer of rapid flowing material (shear zone) on top of a plug zone that co-rotate with the drum. The plug zone does not rotate as an ideal rigid body, however, but has a small creep shear [20] that may render model reduction into rigid aggregates problematic. The simulations are made using the simulation software AgX Dynamics [21] with a prototype implementation of the adaptive model reduction algorithms in Lua scripts [22]. The prototype implementation is not optimized for speed and memory and the tests are therefore limited to relatively small systems ranging between $N_\mathrm{p} = 4 - 90 \cdot 10^3$ particles. The model and NDEM simulation parameters are listed in Table 1 and the adaptive model order reduction parameters in Table 2. A linear contact model is used, with normal stiffness k_n, and may easily be replaced by the nonlinear Hertz contact law. Mono-sized spherical particles are used, except in the granular collapse where bi-disperse spheres are used. Gravity acceleration is 9.81 m/s^2. Videos from simulations are available as supplementary material at `http://umit.cs.umu.se/modelreduction/`.

The pile formation is performed by emitting particles at a rate of 1000 s^{-1} from 0.1 m above a planar conveyor surface moving with horizontal speed 0.1 m/s. The emitter surface is $15d \times 4d$, in which the particle positions are chosen randomly. The particles quickly come to relative rest on the conveyor, forming an elongated pile, roughly $10d$ high and with static angle repose θ_conv. The angle of repose is computed as the average inclination of the pile surface defined by the surface particles in the mid section of the conveyor, neglecting particles resting directly on the conveyor surface, see Fig. 5. The discharge take place on the end of the $50d$ long conveyor, where the material loose support and flow over the edge. The cross sectional flow distribution in the horizontal plane is

Table 2: Model order reduction parameters

parameter	**value**
$\xi_{nv}^{\text{i-mrg}}$	2.5 mm/s
$\xi_{nv}^{\text{s-mrg}}$	2.5 mm/s
ξ_{tv}^{mrg}	2.5 mm/s
ξ_{rv}^{mrg}	0.5 rad/s
$\xi_{na}^{\text{i-mrg}}$	5 m/s^2
$\xi_{na}^{\text{s-mrg}}$	5 m/s^2
ξ_{ta}^{mrg}	5 m/s^2
ξ_{ra}^{mrg}	∞ rad/s
$N_{\text{c-spl}}$	2
$\xi_{nv}^{\text{i-c}}$	0.15 m/s
$\xi_{nv}^{\text{s-c}}$	∞ m/s
ξ_{tv}^{c}	0.15 m/s
ξ_{rv}^{c}	∞ rad/s
$N_{\text{it}}^{\text{tr}}$	$50, 100$
$\xi_{nv}^{\text{i-tr}}$	2.6 m/s
$\xi_{nv}^{\text{s-tr}}$	0.26 m/s
ξ_{tv}^{tr}	0.26 m/s
ξ_{rv}^{tr}	∞ rad/s
ξ_{f}^{tr}	$0.15, 0.25$
ξ_{τ}^{tr}	∞

measured $15d$ below the conveying surface and the geometric centre of the flow is computed, (x_c, y_c). The result is time averaged over 6 s. The conveyor system involves $90 \cdot 10^3$ particles when the conveyor is filled.

The granular collapse test is made with 4000 particles emitted randomly from above into a frictionless cubic container with side length $15d$. The particles are left to comte to rest before model reduction is applied. One side-wall is raised quickly and the particles are left to collapse by the new and unstable force configuration. Since the side walls are frictionless the raising of the wall does not disturb the particles other than the change in the confining pressure. The granular collapse last for about 0.7 s, after which the particle have come to rest in a semi-pile with well-defined angle of repose, θ_{collapse}. The evolution of the angle of repose is tracked by estimating the motion of the plane defined by the particles on the top surface of the granular cube, discarding any particles that disconnect from the main contact network. Images from simulations are shown in Fig. 6.

The rotating drum has diameter $D = 40d$, width $w = 7d$ and is run with angular velocity $\Omega = 0.5$ rad/s. The corresponding dimensionless Froude number is $\text{Fr} \equiv D\Omega^2/2g \approx 0.01$, which is in the dense rolling flow regime. The side walls are frictionless while the cylinder surface has the sama friction as between particles. A slow dense nearly stationary flow with $N_p = 4864$ particles is established within one drum revolution without applying model reduction and starting from a regular particle distribution. An image from the simulations is shown in Fig. 7. The particles have bi-disperse size distribution $d_1 = 13$ mm and $d_2 = 10$ mm. The mass distribution and cross sectional flow velocity field are measured and the dynamic angle of repose, θ_{drum}, is computed by tracking the a $D/2$ wide section of the surface around the drum centre. When using sensor split, the sensor is placed around the estimated lift height.

5 Results

The simulation results are labeled either *ref-500*, *ref-150*, *cs-150*, *tr-150-50-f* or *ctr-150-50-fv*, where the first prefix refer to reference simulation or the method for model reduction, the first number is the number or PGS iterations and the second number is the number iterations in the background trial solver (tr) using force balance condition only (f) or in conjunction with velocity condition (fv). It was found that separation splitting was very sensitive to parameters and typically lead to either a propagation of splitting over the entire aggregates or not enough splitting. Therefore it is not applied, i.e., $\xi_{nv}^{\text{s-c}} = \infty$, and contact splitting need to be combined with either sensors (cs) or with trial splitting (ctr). The results are summarized in Table 3.

The resting angle of repose of the pile formed on the conveying surface is found to be $40.3 \pm 0.8^\circ$ using contact splitting. This is in good agreement with the references $41.5 \pm 0.7^\circ$ for $N_{\text{it}} = 500$ and $40.6 \pm 0.5^\circ$ for $N_{\text{it}} = 150$. The discharge flow at the end of the conveyor is presented in Fig. 8. The cross-sectional flow distribution is very similar in the three simulations. The model

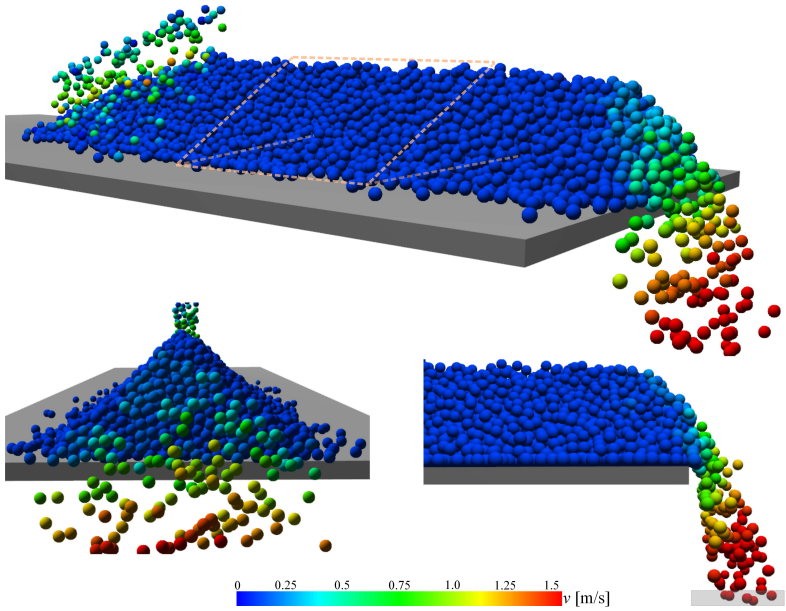

Figure 5: Simulation test for pile formation and discharge on the ends of a conveyor. The colour is coded by particle velocity relative to the conveyor speed and ranges from 0 (blue) to 1.5 m/s or above (red).

Table 3: Simulation results

	ref-500	ref-150	cs-150	tr-150-50-f	tr-150-50-fv
θ_{conv} $h_{\text{conv}}^{\text{mean}}$	$41.5 \pm 0.7°$ 0 %	$40.6 \pm 0.5°$ 0 %	$40.3 \pm 0.8°$ 85 %		
θ_{collapse} h_{collapse}	36° 0 %	30° 0 %		38° $50 - 100$ %	34° $50 - 100$ %
θ_{drum} $h_{\text{drum}}^{\text{mean}}$	$44 \pm 3°$ 0 %	$43 \pm 2°$ 0 %	$49 \pm 5°$ 70 ± 10 %	$44 \pm 3°$ 40 ± 10 %	$45 \pm 2°$ 40 ± 10 %

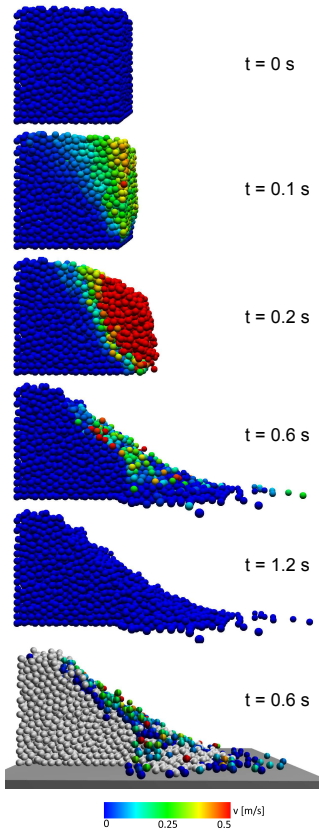

Figure 6: Image sequence from granular collapse. The colour is coded by particle velocity ranging from 0 (blue) to 0.5 m/s or above (red). The top five figures are reference simulation and the bottom figure uses model reduction with background trial solve split.

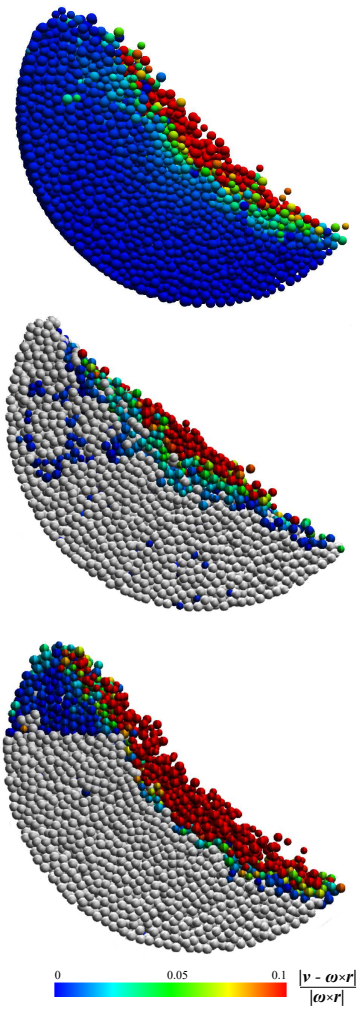

Figure 7: Image from drum flow simulation. The colour is coded by particle relative velocity to rigid co-motion with the drum, ranging from 0 (blue) to 10% or above (red). The top figure show the reference simulation, the middle one uses model reduction with background trial solve split and the bottom one contact event and sensor split.

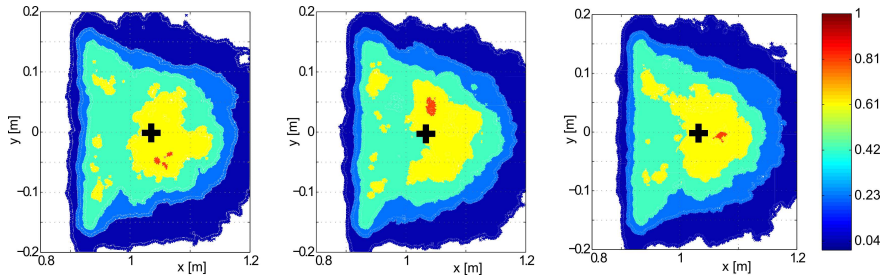

Figure 8: The discharge flow through a horizontal cross-section beneath the end of the conveyor. The contour plots show the accumulated particles distribution from the reference simulations $N_{\text{it}} = 500$ (left), $N_{\text{it}} = 150$ (right) and simulation with model reduction using contact and sensor splitting (right). The crosses mark the geometric centre of the particle distributions.

order reduction level was steady around $h = 85$ % during the simulation.

The distribution of particles after the granular collapse are displayed in Fig. 9. The positions are measured at time $t = 2$ s. The angle of repose, θ_{collapse}, of the $N_{\text{it}} = 500$ reference pile is 36.2°, to be compared with 29.9° for the $N_{\text{it}} = 150$ reference and 37.5° and 34° for model reduction with background trial solve with force (f) and force and velocity (fv) split conditions, respectively. The evolution of the inclination angle of the top surface of the collapsing cube is found in Fig. 10. The initial collapse is similar in all the simulations except the *tr-150-50-f* that evolve somewhat slower initially. The $N_{\text{it}} = 150$ reference simulation reach as highest 34° and then decrease gradually due to insufficient sliding and rolling resistance with that number of iterations. The model reduction simulation become rigid at its maximum angle and fail to resolve the final relaxation of the slope through small avalanches, that are present in the $N_{\text{it}} = 500$ reference simulation. The combination of contact split and background trial split did not resolve this. The evolution of the model reduction level is found in Fig. 11. In the background trial solve split simulation the reduction level varies between 50 and 100 %. This is close to the theoretical maximum for the given thresholds, found by analysing the $N_{\text{it}} = 500$ reference simulation.

Sample states from model reduction of drum flow simulations are presented in Fig. 7. The evolution of the dynamic angle of repose is found in Fig. 12. The time-averaged angle of repose for the reference $N_{\text{it}} = 500$ is $\theta_{\text{drum}} = 44 \pm 3°$. The background trial solve split method produces a flow with angle $44 \pm 3°$ while the contact plus sensor split methods produces a flow with angle $49 \pm 5°$ and with notable artefacts appearing as structures with angle much larger than the angle of repose and high lifting of material in the drum. This is also the reason for the bigger variation on the averaged angle of repose. No thresholds were found for the contact split method alone that showed any significant model reduction but did not produce large approximation errors (overly rigid). No thresholds

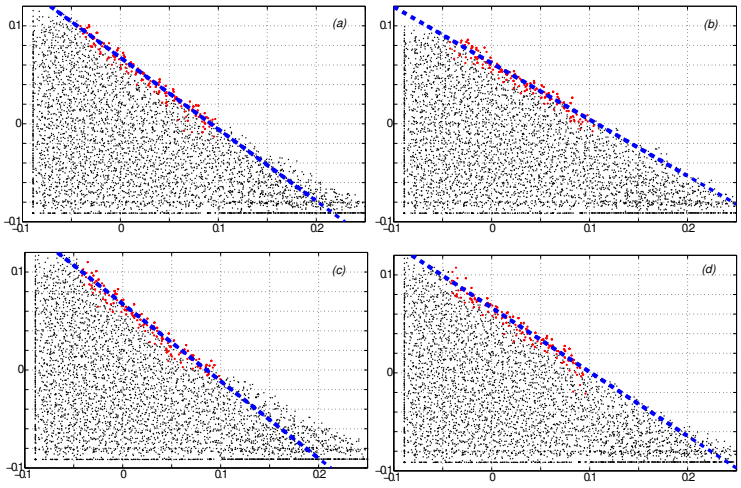

Figure 9: The final particle distribution in the granular collapse simulations: (a) $N_{\text{it}} = 500$ reference; (b) $N_{\text{it}} = 150$ reference; (c) model reduction tr-15-50-f; (d) model reduction tr-15-50-fv. The blue line indicate the angle of repose computed from the surface defined by the red particles.

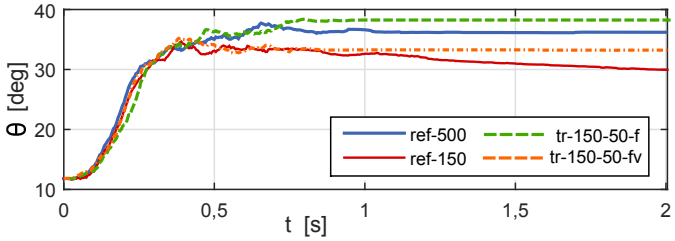

Figure 10: Evolution of the top surface angle during granular collapse.

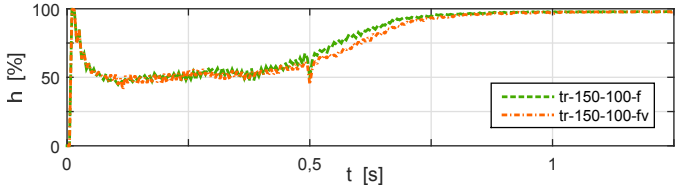

Figure 11: Model reduction level as function of time in granular collapse.

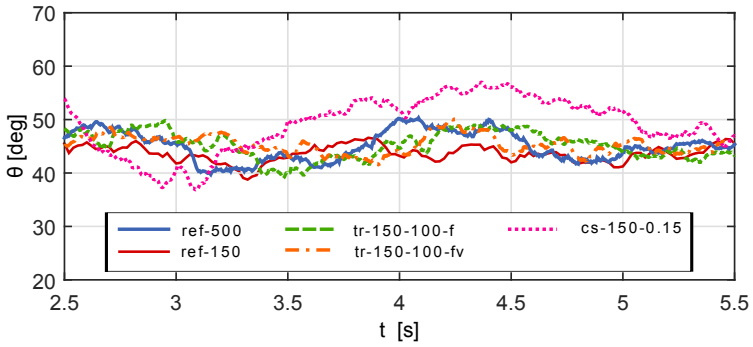

Figure 12: Evolution of the dynamic angle of repose in the rotating drum.

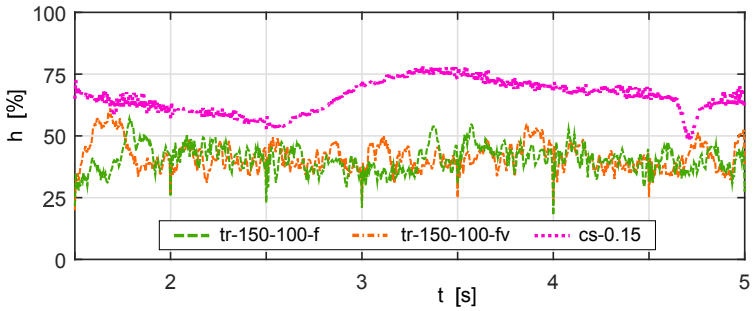

Figure 13: Model reduction level h as function of time in the rotating drum.

were found for the contact split method alone that showed any significant model reduction but did not produce large approximation errors (overly rigid). The model reduction level over time is presented in Fig. 13. For the contact method plus sensor split method it varies between $50-75$ %. The background trial solve split oscillate between 25 and 50%. No parameters were found that gave higher level of reduction without increase of artifacts in the dynamics.

6 Computational acceleration

The potential computational acceleration of using model order reduction in NDEM simulations is estimated and discussed in this section. The time required for simulating t_{real} seconds of evolution is the product of the number of time steps and the computational time for each step

$$t_{\text{comp}} = \frac{t_{\text{real}}}{\Delta t} \cdot [t_{\text{coll}} + t_{\text{mod}} + t_{\text{solve}}] \tag{37}$$

with time serially separated in collision detection, t_{coll}, model order reduction, t_{mod} and solver time, t_{solve}. We define the computational speed-up from model

order reduction as $S \equiv t^0_{\text{comp}}/t_{\text{comp}}(h)$, where t^0_{comp} refer to a simulation of the fully resolved system without adaptive model reduction, while $t_{\text{comp}}(h)$ is the time for a simulation with model reduction level h. It is characteristic for NDEM simulations that $t_{\text{solve}} \gg t_{\text{coll}}$, e.g., 88% of the total time was reported in [23]. We therefore discard collision detection time from here on. When using PGS, the solve time can be estimated by $t_{\text{solve}} = K_{\text{cpu}} \cdot N_{\text{c}} \cdot N_{\text{it}}/S_{\parallel}$, where K_{cpu} is the computational time for doing a single contact constraint solve, $N_{\text{c}} \sim n_{\text{p}}\tilde{N}_{\text{p}}$ is the number of contact constraints assuming on average $2n_{\text{p}}$ contacts per particle, and $S_{\parallel}(N_{\text{cpu}})$ is the parallel speed-up of N_{cpu} cores. Given a spatial error tolerance, ϵ, the required number of iterations scale with the number of particles as $N_{\text{it}} = c\tilde{N}_{\text{p}}^{\gamma}/\epsilon$ [19], for constant $c \approx 0.1$ and exponent $\gamma = 1/n_{\text{D}}$ that depend on the whether the system is close to a linear column ($n_{\text{D}} = 1$) , 2D plane ($n_{\text{D}} = 2$) or a 3D volumetric system ($n_{\text{D}} = 3$). The solve time can thus be written

$$t_{\text{solve}}(h) = cK_{\text{cpu}} \left[(1-h)N_{\text{p}}\right]^{1+\gamma} / \epsilon S_{\parallel}. \tag{38}$$

The computational overhead for doing merge and refinement using contact or sensor based splitting do not involve more than one pass through the contact network. The computational time can thus be estimated by $t_{\text{mod}} = \alpha K_{\text{cpu}} n_{\text{p}} N_{\text{p}}/S_{\parallel}$ for some constant $\alpha < 1$, since the operations do not involve solving the local contact problem. This imply the following speed-up

$$S \approx \frac{1}{\alpha N_{\text{p}}^{-\gamma} + (1-h)^{1+\gamma}}. \tag{39}$$

Using background trial solve split for model reduction is more demanding. If a PGS background solve can be limited to fraction $\beta < 1$ of the full system and run with a larger error tolerance, $\epsilon_{\text{mod}} > \epsilon$, the computational overhead can be estimated to $t_{\text{mod}} = cK_{\text{cpu}} \left[\beta N_{\text{p}}\right]^{1+\gamma} / \epsilon_{\text{mod}} S_{\parallel}$ and the speed-up become

$$S \approx \frac{1}{\beta^{1+\gamma} \cdot \epsilon/\epsilon_{\text{mod}} + (1-h)^{1+\gamma}}. \tag{40}$$

The computational speed-up, depending on model reduction level h and overhead factor $\delta = \alpha N_{\text{p}}^{-\gamma}$ or $\delta = \beta^{1+\gamma} \cdot \epsilon/\epsilon_{\text{mod}}$, is plotted in Fig. 14 for the case of 3D volumetric systems ($\gamma = 1/3$). A significant speed-up of a factor 5 can be achived even at the modest model reduction level $h = 0.7$. The speed-up can reach up to 50 at $h = 0.95$ and overhead $\delta = 0.01$. The potential speed-up is lost almost entirely if the computational overhead cannot be made lower than $\delta = 0.1$. This should be easy to accomplish in the case of contact and sensor based splitting, especially for large systems where $\delta \propto N_{\text{p}}^{-\gamma}$. Achieving a significant speed-up from background trial solve split thus relies on $\beta^{1+\gamma} \cdot \epsilon/\epsilon_{\text{mod}} \lesssim 0.1$. This can be reached if it is sufficient to apply background solve on a fraction of $\beta = 0.2$ of the full system or use a background error tolerance $\epsilon_{\text{mod}} = 10 \cdot \epsilon$. Combining the two possibilities, an overhead of about 0.1 can be reached by $\beta = 0.5$ and $\epsilon_{\text{mod}} = 4 \cdot \epsilon$. The prototype implementation do not scale sufficiently well for large systems to verify these estimates. To

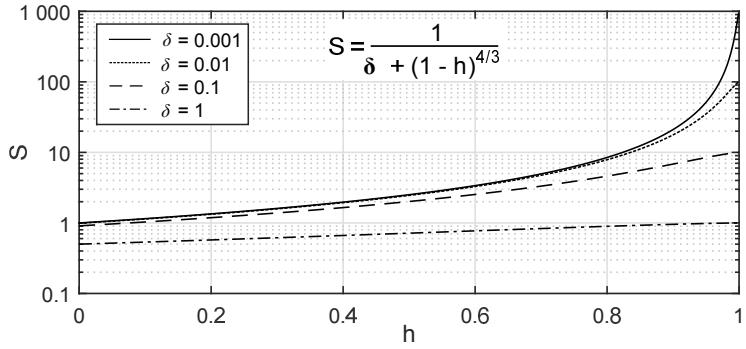

Figure 14: The computational speed-up depending on model reduction level and computational overhead.

ease prototyping and experimentation with different algorithms and parameters for model order reduction these routines were implemented in a Lua scripting framework which lead to an unnecessary amount of data copying of states.

7 Conclusion and discussion

A method for adaptive model order reduction for nonsmooth discrete element simulation has been developed and analysed. In the reduced model rigid aggregate bodies are substituted for collections of contacting particles collectively moving as rigid bodies. Conditions for model reduction and refinement are derived from a model approximation error. The scaling analysis show that the computational performance may be increased by $5 - 50$ times for a model reduction level between $70 - 95$ % given that the computational overhead do not exceed the given scaling conditions. The method is highly applicable for granular systems with large regions in resting or rigidly co-moving state over long periods. It is less efficient and harder to parametrize for systems with sudden and frequent transitions between rigid to liquid or gaseous regime and it is directly inappropriate for systems dominated by shear motion. Furthermore, the presented method is fully compatible with rigid multibody dynamics and can support particles merging with articulated mechanisms, such as the excavator in Fig. 1.

A number of observations were made in the numerical experiments. When doing model refinement based on contact events, it is in general insufficient to split only particles that are impacted directly. The refinement typically need to propagate further into the contact network. The refinement depth $N_c^{\text{spl}} = 2$ was used in the experiments. Refinement based on contact separation events was found to be particularly difficult to parametrize and was eventually not used at all. Small changes in the separation threshold easily altered the behaviour from non-responding to splitting propagating throughout the contact network. As an effect, contact event based refinement is not reliable for simulating quasi-

stable configurations and gravity driven flow. In the absence of impacts, rigid aggregates remain rigid indefinitely. Split sensors can be used to remedy this when there is knowledge on where to place such sensors to guarantee refinement.

When using background trial solve it was found that the force balance condition alone give reliable prediction for model refinement. No improvement was found by adding torque balance or contact velocity conditions. Using background solve, velocity condition alone led to higher fluctuations in the model order reduction level than using both force and velocity conditions.

To increase the applicability of adaptive model order reduction for discrete element methods further, the reduced model need to be extended beyond rigid aggregates to elastic and shearing modes. The application of model order reduction to conventional smooth DEM is also an interesting question to address.

Appendix

A. Numerical integration of NDEM

The numerical time integration scheme is based on the SPOOK stepper [18] derived from discrete variational principle for the augmented system $(\mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}, \dot{\boldsymbol{\lambda}})$ and applying a semi-implicit discretization. The stepper is linearly stable and $\mathcal{O}(\Delta t^2)$ accurate for constraint violations [18]. Stepping the system position and velocity, $(\mathbf{x}_i, \mathbf{v}_i) \rightarrow (\mathbf{x}_{i+1}, \mathbf{v}_{i+1})$, from time t_i to $t_{i+1} = t_i + \Delta t$ involve solving a mixed complementarity problem (MCP) [24]. For a NDEM system the MCP take the following form

$$\begin{aligned} \mathbf{H}\mathbf{z} + \mathbf{b} &= \mathbf{w}_l - \mathbf{w}_u \\ 0 \leq \mathbf{z} - \mathbf{l} &\perp \mathbf{w}_l \geq 0 \\ 0 \leq \mathbf{u} - \mathbf{z} &\perp \mathbf{w}_u \geq 0 \end{aligned} \tag{41}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{M} & -\mathbf{G}_{\mathrm{n}}^{\mathrm{T}} & -\mathbf{G}_{\mathrm{t}}^{\mathrm{T}} & -\mathbf{G}_{\mathrm{r}}^{\mathrm{T}} \\ \mathbf{G}_{\mathrm{n}} & \boldsymbol{\Sigma}_{\mathrm{n}} & 0 & 0 \\ \mathbf{G}_{\mathrm{t}} & 0 & \boldsymbol{\Sigma}_{\mathrm{t}} & 0 \\ \mathbf{G}_{\mathrm{r}} & 0 & 0 & \boldsymbol{\Sigma}_{\mathrm{r}} \end{bmatrix}, \tag{42}$$

$$\mathbf{z} = \begin{bmatrix} \mathbf{v}_{i+1} \\ \boldsymbol{\lambda}_{\mathrm{n},i+1} \\ \boldsymbol{\lambda}_{\mathrm{t},i+1} \\ \boldsymbol{\lambda}_{\mathrm{r},i+1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\mathbf{M}\mathbf{v}_i - \Delta t \mathbf{M}^{-1}\mathbf{f}_{\mathrm{ext}} \\ \frac{4}{\Delta t}\boldsymbol{\Upsilon}_{\mathrm{n}}\mathbf{g}_{\mathrm{n}} - \boldsymbol{\Upsilon}_{\mathrm{n}}\mathbf{G}_{\mathrm{n}}\mathbf{v}_i \\ 0 \\ 0 \end{bmatrix}, \tag{43}$$

and the solution vector $\mathbf{z}$ contains the new velocities and the Lagrange multipliers $\boldsymbol{\lambda}_{\mathrm{n}}$, $\boldsymbol{\lambda}_{\mathrm{t}}$ and $\boldsymbol{\lambda}_{\mathrm{r}}$. For notational convenience, a factor Δt has been absorbed in the multipliers such that the constraint force reads $\mathbf{G}^T\boldsymbol{\lambda}/\Delta t$. The upper and lower limits, u and l in Eq. (41), follow from Signorini-Coulomb and rolling resistance law with the friction and rolling resistance coefficients μ_{t} and μ_{r}. Since the limits depend on the solution this is a partially nonlinear complementarity problem. w_l and w_u are temporary slack variables. Each contact n between

body a and b add contributions to the constraint vector and normal, friction and rolling Jacobians according to

$$
\begin{aligned}
g_{(n)} &= \vec{\mathbf{n}}_{(n)}^{\mathrm{T}}(\vec{\mathbf{x}}_{[a]} + \vec{\mathbf{d}}_{[a]}^{(n)} - \vec{\mathbf{x}}_{[b]} - \vec{\mathbf{d}}_{[b]}^{(n)})_{(\alpha)}^{e_{\mathrm{H}}}, \\
\mathbf{G}_{\mathrm{n}[a]}^{(n)} &= e_{\mathrm{H}} g_{(\alpha)}^{e_{\mathrm{H}}-1} \left[\begin{array}{cc} -\vec{\mathbf{n}}_{(n)}^{\mathrm{T}} & -(\vec{\mathbf{d}}_{[a]}^{(n)} \times \vec{\mathbf{n}}_{(n)})^{\mathrm{T}} \end{array} \right], \\
\mathbf{G}_{\mathrm{n}[b]}^{(n)} &= e_{\mathrm{H}} g_{(\alpha)}^{e_{\mathrm{H}}-1} \left[\begin{array}{cc} \vec{\mathbf{n}}_{(n)}^{\mathrm{T}} & (\vec{\mathbf{d}}_{[b]}^{(n)} \times \vec{\mathbf{n}}_{(n)})^{\mathrm{T}} \end{array} \right], \\
\mathbf{G}_{\mathrm{t}[a]}^{(n)} &= \left[\begin{array}{cc} -\vec{\mathbf{t}}_1^{(n)\mathrm{T}} & -(\vec{\mathbf{d}}_{[a]}^{(n)} \times \vec{\mathbf{t}}_1^{(n)})^{\mathrm{T}} \\ -\vec{\mathbf{t}}_2^{(n)\mathrm{T}} & -(\vec{\mathbf{d}}_{[a]}^{(n)} \times \vec{\mathbf{t}}_2^{(n)})^{\mathrm{T}} \end{array} \right], \\
\mathbf{G}_{\mathrm{t}[b]}^{(n)} &= \left[\begin{array}{cc} \vec{\mathbf{t}}_1^{(n)\mathrm{T}} & (\vec{\mathbf{d}}_{[b]}^{(n)} \times \vec{\mathbf{t}}_1^{(n)})^{\mathrm{T}} \\ \vec{\mathbf{t}}_2^{(n)\mathrm{T}} & (\vec{\mathbf{d}}_{[b]}^{(n)} \times \vec{\mathbf{t}}_2^{(n)})^{\mathrm{T}} \end{array} \right], \\
\mathbf{G}_{\mathrm{r}[a]}^{(n)} &= \left[\begin{array}{cccc} \mathbf{0}_{1\times 3} & \vec{\mathbf{t}}_1^{(n)\mathrm{T}} & \mathbf{0}_{1\times 3} & -\vec{\mathbf{t}}_1^{(n)\mathrm{T}} \\ \mathbf{0}_{1\times 3} & \vec{\mathbf{t}}_2^{(n)\mathrm{T}} & \mathbf{0}_{1\times 3} & -\vec{\mathbf{t}}_2^{(n)\mathrm{T}} \\ \mathbf{0}_{1\times 3} & \vec{\mathbf{n}}^{(n)\mathrm{T}} & \mathbf{0}_{1\times 3} & -\vec{\mathbf{n}}^{(n)\mathrm{T}} \end{array} \right], \\
\mathbf{G}_{\mathrm{r}[b]}^{(n)} &= \left[\begin{array}{cccc} \mathbf{0}_{1\times 3} & -\vec{\mathbf{t}}_1^{(n)\mathrm{T}} & \mathbf{0}_{1\times 3} & \vec{\mathbf{t}}_1^{(n)\mathrm{T}} \\ \mathbf{0}_{1\times 3} & -\vec{\mathbf{t}}_2^{(n)\mathrm{T}} & \mathbf{0}_{1\times 3} & \vec{\mathbf{t}}_2^{(n)\mathrm{T}} \\ \mathbf{0}_{1\times 3} & -\mathbf{n}^{(n)\mathrm{T}} & \mathbf{0}_{1\times 3} & \vec{\mathbf{n}}^{(n)\mathrm{T}} \end{array} \right],
\end{aligned}
\tag{44}
$$

where $\vec{\mathbf{d}}_{[a]}^{(n)}$ and $\vec{\mathbf{d}}_{[b]}^{(n)}$ are the positions of the contact point n relative to the particle positions $\vec{\mathbf{x}}_{[a]}$ and $\vec{\mathbf{x}}_{[b]}$. The orthonormal contact normal and tangent vectors are $\vec{\mathbf{n}}_{(n)}$, $\vec{\mathbf{t}}_{(n)_1}$ and $\vec{\mathbf{t}}_{(n)_2}$. For linear contact model $e_{\mathrm{H}} = 1$ and for the nonlinear Hertz-Mindlin model $e_{\mathrm{H}} = 5/4$. The diagonal matrices $\mathbf{\Sigma}_{\mathrm{n}}$, $\mathbf{\Sigma}_{\mathrm{t}}$, $\mathbf{\Sigma}_{\mathrm{r}}$ and $\mathbf{\Upsilon}_{\mathrm{n}}$ contain the contact material parameters and are as follows

$$
\begin{aligned}
\mathbf{\Sigma}_{\mathrm{n}} &= \frac{4}{\Delta t^2} \frac{\varepsilon_{\mathrm{n}}}{1 + 4\frac{\tau_{\mathrm{n}}}{\Delta t}} \mathbf{1}_{N_c \times N_c}, \\
\mathbf{\Sigma}_{\mathrm{t}} &= \frac{\gamma_{\mathrm{t}}}{\Delta t} \mathbf{1}_{2N_c \times 2N_c}, \\
\mathbf{\Sigma}_{\mathrm{r}} &= \frac{\gamma_{\mathrm{r}}}{\Delta t} \mathbf{1}_{3N_c \times 3N_c}, \\
\mathbf{\Upsilon}_{\mathrm{n}} &= \frac{1}{1 + 4\frac{\tau_{\mathrm{n}}}{\Delta t}} \mathbf{1}_{N_c \times N_c}.
\end{aligned}
\tag{45}
$$

The MCP parameters map to DEM material parameters by $\varepsilon_{\mathrm{n}} = e_{\mathrm{H}}/k_{\mathrm{n}}$, $\gamma_{\mathrm{n}}^{-1} = k_{\mathrm{n}} c / e_{\mathrm{H}}^2$ and $\tau_{\mathrm{n}} = \max(n_{\mathrm{s}}\Delta t, \varepsilon_{\mathrm{n}}/\gamma_{\mathrm{n}})$, with elastic stiffness coefficient k_{n} and viscosity c. For the Hertz-Mindlin contact law, $k_{\mathrm{n}} = e_{\mathrm{H}} E \sqrt{r^*}/3(1-\nu^2)$ where $r^* = (r_a + r_b)/r_a r_b$ is the effective radius, E is the Young's modulus and ν is the Poisson ratio. For small relative contact velocities the normal force approximates $\mathbf{G}_{\mathrm{n}}^{(n)\mathrm{T}} \boldsymbol{\lambda}_{\mathrm{n}}^{(n)} \approx \varepsilon_{\mathrm{n}}^{-1} \mathbf{G}_{\mathrm{n}}^{(n)\mathrm{T}} \mathbf{g}_{\mathrm{n}}^{(n)} = \pm k_{\mathrm{n}} \left[\mathbf{g}_{\mathrm{n}}^{2e_{\mathrm{H}}-1} + c \mathbf{g}_{\mathrm{n}}^{2(e_{\mathrm{H}}-1)} \dot{g}_{\mathrm{n}} \right] \boldsymbol{n}$. High ingoing velocities are treated as impacts and this is done *post facto*. After stepping

the velocities and positions an impact stage follows. This include solving a MCP similar to Eq. (41) but with the Newton impact law, $\mathbf{G}_{\mathrm{n}}^{(n)}\mathbf{v}_{+} = -e\mathbf{G}_{\mathrm{n}}^{(n)}\mathbf{v}_{-}$, replacing the normal constraints for the contacts with normal velocity larger than an impact velocity threshold v_{imp}. The remaining constraints are maintained by imposing $\mathbf{G}^{(n)}\mathbf{v}_{+} = 0$. This can be expressed by a matrix multiplication $\mathbf{G}_{\mathrm{n}}\mathbf{v}_{+} = -\mathbf{E}\mathbf{G}_{\mathrm{n}}\mathbf{v}_{-}$, where the diagonal matrix $\mathbf{E}$ values alternate between e and 0 for impacting and resting contacts, respectively. The MCP is solved using a projected Gauss-Seidel (PGS) algorithm, as described in Ref. [19]. The algorithm is listed in Algorithm 2. The NDEM method with PGS solver is implemented in the software AgX Dynamics [21]. In the present study $\gamma_{\mathrm{t}} = \gamma_{\mathrm{r}} = 10^{-6}$, $n_{\mathrm{s}} = 2$ was used and a linear contact model, $e_{\mathrm{H}} = 1$, for consistency with contacts between elementary rigid bodies and aggregate bodies, for which linear contact constraints are default in AgX.

Algorithm 2 PGS solver for the MCP

if impact stage **then**
  $\mathbf{b}_{\mathrm{n}} = \mathbf{E}\mathbf{G}_{\mathrm{n}}\mathbf{v}$
else if continuous stage **then**
  $\mathbf{b}_{\mathrm{n}} = (4/\Delta t)\boldsymbol{\Upsilon}_{\mathrm{n}}\mathbf{g}_{\mathrm{n}} - \boldsymbol{\Upsilon}_{\mathrm{n}}\mathbf{G}_{\mathrm{n}}\mathbf{v}$
  pre-step $\mathbf{v} = \mathbf{v} + \Delta t\mathbf{M}^{-1}\mathbf{f}_{\mathrm{ext}}$
end if
$\mathbf{q} = [-\mathbf{b}_{\mathrm{n}}^{T}, 0, 0]^{T}$
for $k = 1, \ldots, N_{\mathrm{it}}$ and **while** *criteria*$(\mathbf{r})$ **do**
  for each contact $n = 0, 1, \ldots, N_{\mathrm{c}} - 1$ **do**
    for each constraint α of contact n **do**
      $\mathbf{r}_{\alpha,k}^{(n)} = -\mathbf{q}_{\alpha,k}^{(n)} + \mathbf{G}_{\alpha}^{(n)}\mathbf{v}$ ▷ residual
      $\boldsymbol{\lambda}_{\alpha,k}^{(n)} = \boldsymbol{\lambda}_{\alpha,k-1}^{(n)} + \mathbf{D}_{\alpha,(n)}^{-1}\mathbf{r}_{\alpha,k}^{(n)}$ ▷ multiplier
      $\boldsymbol{\lambda}_{\alpha,k}^{(n)} \leftarrow \mathrm{proj}_{\mathcal{C}_{\mu}}(\boldsymbol{\lambda}_{k}^{(n)})$ ▷ project
      $\Delta\boldsymbol{\lambda}_{\alpha,k}^{(n)} = \boldsymbol{\lambda}_{\alpha,k}^{(n)} - \boldsymbol{\lambda}_{\alpha,k-1}^{(n)}$
      $\mathbf{v} = \mathbf{v} + \mathbf{M}^{-1}\mathbf{G}_{\alpha,(n)}^{T}\Delta\boldsymbol{\lambda}_{\alpha,k}^{(n)}$
    end for
  end for
end for

Acknowledgment

This project was supported by Algoryx Simulations, LKAB (dnr 223-2442-09), UMIT Research Lab and VINNOVA (2014-01901).

References

[1] T. Pöschel, T. Schwager, Computational Granular Dynamics, Models and Algorithms, Springer-Verlag, 2005.

[2] F. Radjai, V. Richefeu, Contact dynamics as a nonsmooth discrete element method, Mechanics of Materials 41 (6) (2009) 715–728.

[3] J. J. Moreau, Numerical aspects of the sweeping process, Computer Methods in Applied Mechanics and Engineering 177 (1999) 329–349.

[4] M. Jean, The non-smooth contact dynamics method, Computer Methods in Applied Mechanics and Engineering 177 (1999) 235–257.

[5] A. Antoulas, Approximation of Large-Scale Dynamical Systems, Society for Industrial and Applied Mathematics (2005).

[6] G. Kerschen, J-C. Golinval, A. Vakakis, L. Bergman, The Method of Proper Orthogonal Decomposition for Dynamical Characterization and Order Reduction of Mechanical Systems: An Overview, Nonlinear Dynamics 41(2005) 147–169.

[7] C. Nowakowski, J. Fehr, M. Fischer, P. Eberhard, Model Order Reduction in Elastic Multibody Systems using the Floating Frame of Reference Formulation, In Proceedings MATHMOD 2012-7th Vienna International Conference on Mathematical Modelling, Vienna, Austria, (2012).

[8] P. Glösmann, Reduction of discrete element models by KarhunenLove transform: a hybrid model approach, Computational Mechanics 45 (4) (2010) 375–385.

[9] A. Munjiza, The Combined Finite-Discrete Element Method, Wiley, (2004).

[10] C. Miehe, J. Dettmar and D. Zah, Homogenization and two-scale simulations of granular materials for different microstructural constraints, Int. J. Numer. Meth. Engng 83 (2010) 1206-1236.

[11] N. Guo, J. Zhao, A coupled FEM/DEM approach for hierarchical multiscale modelling of granular media, Int. J. Numer. Meth. Engng 99 (11) (2014) 789–818.

[12] C. Wellmann, P. Wriggers, A two-scale model of granular materials, Computer Methods in Applied Mechanics and Engineering, 205-208 (2012), 46–58.

[13] S. Redon, M. Lin, An Efficient, Error-Bounded Approximation Algorithm for Simulating Quasi-Statics of Complex Linkages, In Proceedings of ACM Symposium on Solid and Physical Modeling (2005).

[14] S. Redon, N. Galoppo,M. Lin, Adaptive Dynamics of Articulated Bodies, In ACM Transactions on Graphics (SIGGRAPH 2005), 24 (3) (2015).

[15] N. Cho, C. D. Martin, D. C. Sego, A clumped particle model for rock, International Journal of Rock Mechanics & Mining Sciences 44 (2007) 997–1010.

[16] H. Jaeger, S. Nagel, R. Behringer, Granular solids, liquids, and gases, Rev. Mod. Phys. 68 (4) (1996) 1259–1273.

[17] Bornemann F.,Schütte C. Homogenization of Hamiltonian systems with a strong constraining potential. *Phys. D*, 102(1-2):57–77, 1997.

[18] C. Lacoursière, Regularized, stabilized, variational methods for multibodies, in: D. F. Peter Bunus, C. Führer (Eds.), The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), Linköping University Electronic Press, 2007, pp. 40–48.

[19] M. Servin, D. Wang, C. Lacoursière, K. Bodin, Examining the smooth and nonsmooth discrete element approach to granular matter, Int. J. Numer. Meth. Engng. 97 (2014) 878–902.

[20] M. Renouf, D. Bonamy, F. Dubois, P. Alart, Numerical simulation of two-dimensional steady granular flows in rotating drum: On surface flow rheology, Phys. Fluids 17 (2005).

[21] Algoryx Simulations, AgX Dynamics User Guide Version 2.12.1.0, December 2014.

[22] Lua, http://www.lua.org, October 2015.

[23] M. Renouf, F. Dubois P. Alart, A parallel version of the non smooth contact dynamics algorithm applied to the simulation of granular media, Journal of Computational and Applied Mathematics 168(1-2), 2004, pp. 375-382.

[24] K. G. Murty, Linear Complementarity, Linear and Nonlinear Programming, Helderman-Verlag, Heidelberg, 1988.