An Inexact Interior Point Method for the Large-Scale Simulation of Granular Material

Jan Kleinert[a], Bernd Simeon[b], Martin Obermayr[a]

[a] *Fraunhofer ITWM, Fraunhofer Platz 1, 67663 Kaiserslautern, Germany. Email: jan.kleinert@itwm.fraunhofer.de.*
[b] *Felix Klein Zentrum, Postfach 3049, 67653 Kaiserslautern, Germany. Email: simeon@mathematik.uni-kl.de.*

Abstract

Non-smooth contact dynamics provides an increasingly popular simulation framework for granular material. In contrast to classical discrete element methods, this approach is stable for arbitrary time steps and produces visually acceptable results in very short computing time. Yet when it comes to the prediction of draft forces, non-smooth contact dynamics is typically not accurate enough. We therefore propose to combine the method class with an interior point algorithm for higher accuracy. Our specific algorithm is based on so-called Jordan algebras and exploits the relation to symmetric cones in order to tackle the conical constraints that are intrinsic to frictional contact problems. In every interior point iteration a linear system has to be solved. We analyze how the interior point method behaves when it is combined with Krylov subspace solvers and incomplete factorizations. We show that efficient preconditioners and efficient linear solvers are essential for the method to be applicable to large-scale problems. Using BiCGstab as a linear solver and incomplete Cholesky factorizations, we substantially improve the accuracy in comparison to the projected Gauß–Jacobi solver.

Keywords: Non-Smooth Dynamical Systems, Granular Material, Cone Complementarity Problem, Symmetric Cones, Interior Point Method, Jordan Algebra

1. Introduction

Granular material appears in numerous areas of engineering. As examples, we mention the behavior of powder, which is of interest for toner production, process engineering, which benefits from the simulation of silo discharges, and the mining industry, where one is interested in optimizing hoppers and conveyor belts. Overall, there is an increasing demand for realistic large-scale computer simulations for these application fields.

If we look more closely at the product design of excavators, it turns out that draft forces acting on the machine from the interaction with the soil are of great importance. The realistic prediction, however, of draft forces demands the highest standards from simulation methods for granular matter. Currently, two methods are mainly used in this field: the classical Discrete Element Method (DEM) and Non-Smooth Contact Dynamics (NSCD) methods. The DEM has shown its potential to deliver draft forces that are in agreement with experimental data (Obermayr et al., 2011). But a simulation of a few seconds can take up days to weeks of calculation time. NSCD provides a class of methods that are substantially faster and more stable, but due to their low accuracy, they fail to reproduce realistic draft forces. To improve the accuracy of NSCD, we propose here the combination with a specialized Interior Point Method (IPM). While being computationally more expensive than standard NSCD solvers, it does not suffer from instabilities as DEM does and it gives direct control over the trade–off between calculation time and accuracy.

In the DEM (Cundall and Strack, 1979), contact between two particles is modeled locally as a stiff spring that pushes them apart. Particles are allowed to slightly penetrate each other and a reaction force proportional to the overlap is calculated. Combined with frequent changes in contact states, this limits the maximum time step size to maintain a stable simulation. Thus, the classical DEM is computationally very expensive. One way to interpret the DEM is by saying small penetrations emulate microscopic deformations. In this sense, DEM resolves the problem at much smaller time scales than those of interest. From the perspective of larger space and time scales, the granular particles seem perfectly rigid and collisions seem to be resolved instantaneously. One way of eliminating uninteresting time scales from the model is by using a non–smooth formulation for the dynamics.

In NSCD (Moreau, 1988; Moreau and Panagiotopoulos, 1990; Stewart, 2000; Acary and Brogliato, 2008), contacts and collisions between particles are modeled by inequality constraints, e.g., by demanding that the distance between two particles shall always be greater than or equal to zero. Satisfying these equations leads to absolutely continuous trajectories that need not be differentiable at every point in time. This yields a velocity and impulse based scheme, since forces no longer exist as classical functions in time. The non–smooth formulation combined with a time–stepping scheme, where the times of collision are not resolved exactly, yields a numerical method that is stable

for arbitrary time step sizes.

Satisfying the inequality constraints boils down to solving a variational inequality, i.e., a complementarity problem[1], at every time step of the simulation. In many applications, specifically in soil mechanics, a friction model is indispensable and therefore a large Cone Complementarity Problem (CCP) has to be solved per time step (Anitescu and Potra, 1997; Anitescu and Tasora, 2008; Tasora and Anitescu, 2011).

A popular iterative solver for large scale CCPs is the Projected Gauß–Jacobi (PGJ) method, since it is easily parallelizable, it can be implemented in a matrix–free fashion, and one iteration is very cheap (Tasora and Anitescu, 2011; Balzer et al., 2013). Figure 1 shows a snapshot from a simulation with one million particles, computed by PGJ on 48 compute nodes. Its main drawback is that, after good initial convergence, the convergence rate stalls quickly if the problems are large or if large mass ratios are present. Stopping the iteration prematurely leads to artificial compliance in the granular material, incorrect angles of repose and incorrect forces (Kleinert et al., 2013). In practice, a combination of NSCD and PGJ is not necessarily more efficient than classical DEM, as a lot of PGJ iterations are necessary. Sometimes PGJ completely fails to meet the desired accuracy requirements.

A natural next step is to study second–order methods for CCPs to improve the convergence. Noteworthy contributions in this direction are (Daviet et al., 2011), where the authors employ a semi–smooth Newton method to the Fischer–Burmeister Function associated with a CCP; (Krabbenhoft et al., 2012) where the authors consider Quadratic Programming solvers and Interior Point Methods for Linear Complementarity Problems; (Heyn et al., 2013), where Krylov subspace methods are generalized for the solution of variational inequalities; and finally, the Ph.D. thesis (Heyn, 2013), which compares active–set strategies, accelerated gradient–descent methods and IPMs.

In this paper we consider a specialized IPM based on the Jordan–algebraic structure of $\mathbb{R}^3$. IPMs are very promising due to their good theoretical complexity bound. In practice, the required number of iterations to reach a certain accuracy stays bounded even for very large problems. Symmetric cones play an important role in Jordan algebras and hence, exploiting their structure is one way to tackle the conical constraints resulting from the Coulomb friction law.

In each iteration, a linear system of equations has to be solved. In most publications on this topic it is assumed that this system is solved directly using matrix factorizations. In large granular assemblies it is impractical to use a direct solver due to the dimension of the linear system. We favor the idea presented in (Gondzio, 2012) of using an iterative method to obtain inexact search directions in every IPM iteration, while taking precautions to maintain good numerical properties of the linear systems.

We give here a self–contained description of the Interior Point Method for NSCD. Moreover, we test different preconditioners and Krylov methods to solve the linear systems and demonstrate that preconditioning is essential to make this method applicable to large-scale problems. Among the analyzed techniques, we found that regularizing the linear systems, preconditioning with Incomplete Cholesky factorizations and solving them with the Bi–Conjugate Gradient method lead to the most efficient approach. Finally, we also show that the IPM solves the CCP to a higher accuracy than PGJ, even as the problem size increases.

In the following section, we outline the physical model and the equations of motion for non–smooth dynamical systems subject to unilateral contact and friction. Some elementary properties of symmetric cones are outlined in Section 3. The IPM is explained in detail in Section 4. Section 5 delivers numerical tests, and some concluding remarks are given in Section 6.

2. Non–Smooth Dynamical Systems

In this section the equations of motion for a system of rigid bodies with non–smooth trajectories are stated. We start with a simple frictionless contact model, introduce Coulomb Friction and we provide the equations of motion in discrete time.

2.1. Equations of Motion

In NSCD, collisions between particles are resolved instantaneously, and thus the trajectories are not necessarily differentiable everywhere. Nevertheless, if $\mathbf{q}(t) \in \mathbb{R}^{6m}$ is the concatenated vector containing the positions and orientations of all m particles at time t, it is an absolutely continuous function in time and hence we can write it as the anti–derivative of a velocity $\mathbf{v}(t) \in \mathbb{R}^{6m}$,

$$\mathbf{q}(t) = \mathbf{q}(t_0) + \int_{t_0}^{t} \mathbf{v}(s)\mathrm{d}s.$$

The velocity can have jumps and therefore is not the anti–derivate of a classical function in time. However, we can associate it to a differential measure $\mathrm{d}\mathbf{v}(t)$ that can be identified with the weak derivative in the distributional sense,

$$\mathbf{v}(t) = \mathbf{v}(t_0) + \int_{t_0}^{t} \mathrm{d}\mathbf{v}(s).$$

The equations of motion must formally be written as a set of measure differential equations

$$M\mathrm{d}\mathbf{v}(t) = \mathrm{d}\mathbf{p}(t), \tag{1}$$

[1] The complementarity problem results from the Karush–Kuhn–Tucker first order optimality conditions of an energy principle subject to inequality constraints.

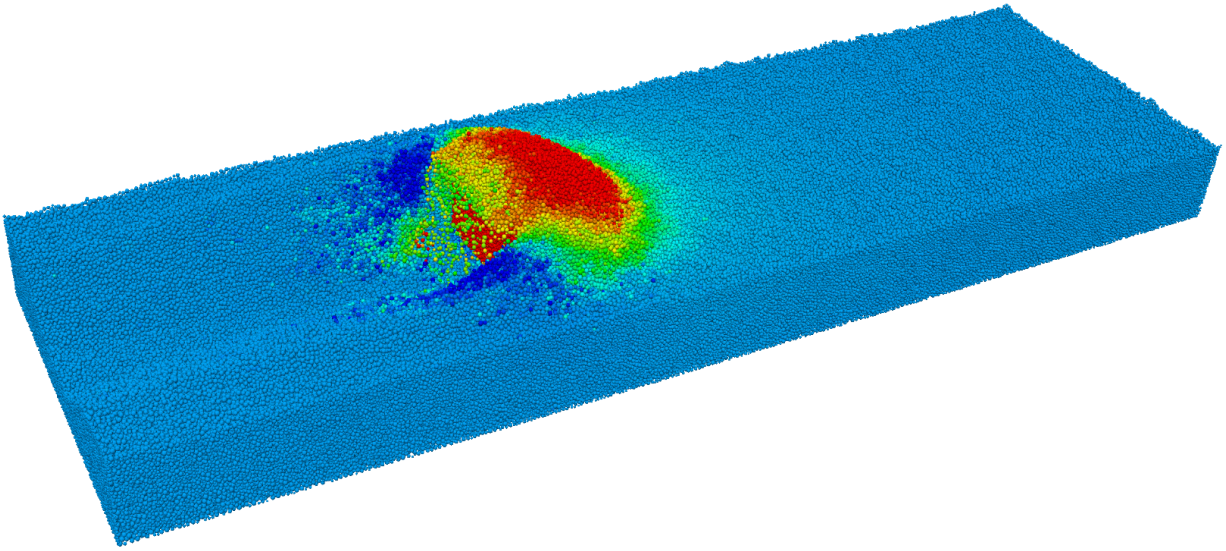


Figure 1: Snapshot from a simulation, where a rectangular blade is moved through a trench filled with one million particles. The color of the particles indicates the magnitude of their velocity.

where $M \in \mathbb{R}^{6m\times 6m}$ is the block–diagonal mass matrix and $\mathrm{d}\mathbf{p}(t)$ is the vector measure associated to the momentum of the system. If the system is only subjected to absolutely continuous external forces $\mathbf{f}_{\text{ext}}(t) \in \mathbb{R}^{6m}$, we can write

$$M\mathrm{d}\mathbf{v}(t) = \mathrm{d}\mathbf{p}(t) = \mathbf{f}_{\text{ext}}(t)\mathrm{d}t.$$

Due to the Lebesgue decomposition theorem (Halmos, 1950), the measure $\mathrm{d}\mathbf{v}(t)$ can be split into a part that has a density $\mathbf{a}(t)$ with respect to the measure $\mathrm{d}t$, a discrete measure $\boldsymbol{\eta}(t)$, that captures velocity jumps, and a negligible Cantor part $\boldsymbol{\zeta}(t)$,

$$\mathrm{d}\mathbf{v}(t) = \mathbf{a}(t)\mathrm{d}t + \boldsymbol{\eta}(t) + \boldsymbol{\zeta}(t) \approx \mathbf{a}(t)\mathrm{d}t + \boldsymbol{\eta}(t).$$

In this sense, (1) can be understood as a weak version of Newton's second law of motion (Stewart, 2000; Acary and Brogliato, 2008).

2.2. Unilateral Frictionless Contact

Unilateral contact means that a contact law only exerts forces in one direction to keep particles separated. Each potential contact i between two particles within the granular material is represented by an inequality constraint

$$u_i : \mathbb{R}^{6m} \to \mathbb{R}, \quad u_i(\mathbf{q}(t)) \geq 0.$$

Under the assumption that $u_i(\mathbf{q}(t))$ is differentiable in a neighborhood of $u_i(\mathbf{q}(t)) = 0$, the constraint can be rewritten on the velocity level as

$$u_i(\mathbf{q}(t)) = 0 \quad \Rightarrow \quad \dot{u}_i(\mathbf{q}(t)) = \nabla u_i(\mathbf{q}(t))\mathbf{v}(t) \geq 0.$$

To each constraint we associate a reaction impulse

$$\lambda_i(t) = \int_{t_0}^{t} \mathrm{d}\lambda_i(s)$$

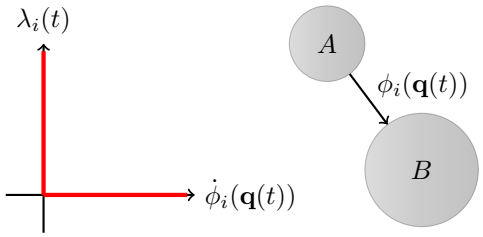


Figure 2: The Signorini Contact Condition: Either the particles are separating or a reaction impulse must be invoked.

via the complementarity condition

$$0 \leq \lambda_i(t) \quad \perp \quad u_i(t) \geq 0 \tag{2}$$

meaning that at least one of the two non–negative values must be zero. Equation (1) turns into

$$M\mathrm{d}\mathbf{v}(t) = \mathbf{f}_{\text{ext}}(t)\mathrm{d}t + \sum_{\text{contacts } i} \nabla u_i(\mathbf{q}(t))^T \mathrm{d}\lambda_i(t). \tag{3}$$

Just as for the measure $\mathrm{d}\mathbf{v}(t)$ associated with the velocity, the measure $\mathrm{d}\lambda_i(t)$ can be split into a continuous part $\alpha_i(t)$ with respect to the Lebesgue measure $\mathrm{d}t$ and a discrete part $\xi_i(t)$,

$$\mathrm{d}\lambda_i = \alpha_i(t)\mathrm{d}t + \xi_i(t).$$

The discrete part resolves collisions and percussions, while the continuous part captures reaction forces from smooth parts of the motion, such as persistent contact in static assemblies of granular materials.

Consider a contact with index i associated to two particles in the system indexed by A and B. Let $\phi_i(\mathbf{q}(t))$ be the signed distance between the two particles. If $\phi_i(\mathbf{q}(t))$

is zero and λ_i is the reaction impulse in normal direction of the contact, the complementarity condition (2) on velocity level amounts to

$$0 \leq u_i(\mathbf{q}(t)) = \dot{\phi}_i(\mathbf{q}(t)) \quad \perp \quad \lambda_i \geq 0, \tag{4}$$

see Figure 2. Either two particles are separating and no reaction impulse is needed, or a reaction impulse must be enforced that keeps them from penetrating. This formulation corresponds to a completely inelastic collision.

Let $\boldsymbol{n}_i(t) \in \mathbb{R}^3$ be the contact normal pointing from body A to body B and let $\mathbf{r}_A(t)$ and $\mathbf{r}_B(t) \in \mathbb{R}^3$ be the vectors pointing from the centers of mass to the contact points on bodies A and B respectively. With

$$\mathbf{v}(t) = \begin{bmatrix} \mathbf{v}_1(t) \\ \vdots \\ \mathbf{v}_m(t) \end{bmatrix} \in \mathbb{R}^{6m}, \quad \mathbf{v}_k(t) = \begin{bmatrix} \boldsymbol{\nu}_k(t) \\ \boldsymbol{\omega}_k(t), \end{bmatrix} \in \mathbb{R}^6$$

where $\boldsymbol{\nu}_k(t)$ is the translational velocity of the k–th particle and $\boldsymbol{\omega}_k(t)$ its angular velocity, it holds

$$D_{i\mathbf{n}}(t) := \nabla u_i(\mathbf{q}(t))^T = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -\mathbf{n}_i(t) \\ -\mathbf{r}_A(t) \times \mathbf{n}_i(t) \\ 0 \\ \vdots \\ 0 \\ \mathbf{n}_i(t) \\ \mathbf{r}_B(t) \times \mathbf{n}_i(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Let there be n contacts at time t. By writing

$$D(t) := [D_{1\mathbf{n}}(t), ..., D_{n\mathbf{n}}(t)] \in \mathbb{R}^{6m \times n}$$

and

$$\boldsymbol{\lambda} := \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \in \mathbb{R}^n, \quad \mathbf{u} := \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$$

the equations of motion (3) together with the complementarity condition (4) are given by

$$\begin{aligned} \mathbf{q}(t) &= \mathbf{q}(t_0) + \int_{t_0}^{t} \mathbf{v}(s)\mathrm{d}s, \\ \mathbf{v}(t) &= \mathbf{v}(t_0) + \int_{t_0}^{t} \mathrm{d}\mathbf{v}(s), \\ M\mathrm{d}\mathbf{v}(t) &= \mathbf{f}_{\text{ext}}(t)\mathrm{d}t + D(t)\mathrm{d}\boldsymbol{\lambda}(t) \end{aligned}$$

$$\mathbf{0} \leq \mathbf{u} \quad \perp \quad \boldsymbol{\lambda} \geq 0,$$

where the inequality has to be understood componentwise.

2.3. Coulomb Friction

The Coulomb friction model relates the tangential reaction impulse $\boldsymbol{\lambda}_{i\mathbf{t}} \in \mathbb{R}^2$ of a contact in the granular system to the normal reaction impulse $\lambda_{i\mathbf{n}}$ via the restriction $\|\boldsymbol{\lambda}_{i\mathbf{t}}\| \leq \mu_i \lambda_{i\mathbf{n}}$, where μ_i is the frictional coefficient. The maximum dissipation principle reformulates this concept as an optimization problem (Stewart, 2000). It states that $\boldsymbol{\lambda}_{i\mathbf{t}}$ and the relative tangential velocity $\mathbf{v}_{i\mathbf{t},\text{rel}}$ in the contact are anti–parallel, and friction maximizes the energy dissipation from the system subject to the constraint $\|\boldsymbol{\lambda}_{i\mathbf{t}}\| \leq \mu_i \lambda_{i\mathbf{n}}$.

In accordance with (DeSaxcé and Feng, 1998), we cast the complementarity condition (4) for the normal reaction impulse together with the maximum dissipation principle into a Cone Complementarity Problem (CCP) of the form

$$K_{\mu_i} \ni \boldsymbol{\lambda}_i = \begin{bmatrix} \lambda_{i\mathbf{n}} \\ \boldsymbol{\lambda}_{i\mathbf{t}} \end{bmatrix} \quad \perp \quad \mathbf{u}_i = \begin{bmatrix} u_{i\mathbf{n}} \\ \mathbf{u}_{i\mathbf{t}} \end{bmatrix} \in K^*_{\mu_i} \tag{5}$$

where

$$\mathbf{u}_i = \begin{bmatrix} \dot{\phi}_i(\mathbf{q}(t)) + \mu_i \|\mathbf{v}_{i\mathbf{t},\text{rel}}\| \\ \mathbf{v}_{i\mathbf{t},\text{rel}} \end{bmatrix} \in \mathbb{R}^3. \tag{6}$$

Figure 3 depicts an example of the CCP associated to a brick sliding down an inclined slope. The cone

$$K_{\mu_i} := \left\{ \begin{bmatrix} \lambda_{\mathbf{n}} \\ \boldsymbol{\lambda}_{\mathbf{t}} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^2 \;\middle|\; \mu_i \lambda_{\mathbf{n}} \geq \|\boldsymbol{\lambda}_{\mathbf{t}}\| \right\}$$

is called the Coulomb Friction Cone and its dual cone is given by

$$\begin{aligned} K^*_{\mu_i} &:= \left\{ \mathbf{u} \in \mathbb{R}^3 \;\middle|\; \mathbf{u}^T \boldsymbol{\lambda} \geq 0 \text{ for all } \boldsymbol{\lambda} \in K_{\mu_i} \right\} \\ &= \left\{ \begin{bmatrix} u_{\mathbf{n}} \\ \mathbf{u}_{\mathbf{t}} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^2 \;\middle|\; u_{\mathbf{n}} \geq \mu_i \|\mathbf{u}_{\mathbf{t}}\| \right\}. \end{aligned}$$

For the contact with index i, let $\mathbf{t}_{i1}(t)$ and $\mathbf{t}_{i2}(t)$ span the contact plane and let

$$D_i(t) = \begin{bmatrix} D_{i\mathbf{n}}(t) & D_{i\mathbf{t}}(t) \end{bmatrix} \in \mathbb{R}^{6m \times 3}$$

be the constraint Jacobian as described in (Anitescu et al., 1995; Tasora and Anitescu, 2011), where $D_{i\mathbf{t}}(t) \in \mathbb{R}^{6m \times 2}$ is given by

$$D_{i\mathbf{t}}(t) := \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ -\mathbf{t}_{i1}(t) & -\mathbf{t}_{i2}(t) \\ -\mathbf{r}_A(t) \times \mathbf{t}_{i1}(t) & -\mathbf{r}_A(t) \times \mathbf{t}_{i2}(t) \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \mathbf{t}_{i1}(t) & \mathbf{t}_{i2}(t) \\ \mathbf{r}_B(t) \times \mathbf{t}_{i1}(t) & \mathbf{r}_B(t) \times \mathbf{t}_{i2}(t) \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}.$$

Then we can write (6) as

$$\mathbf{u}_i = D_i(t)^T \mathbf{v}(t) + \mu_i \begin{bmatrix} \|D_{i\mathbf{t}}(t)^T \mathbf{v}(t)\| \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3. \tag{7}$$

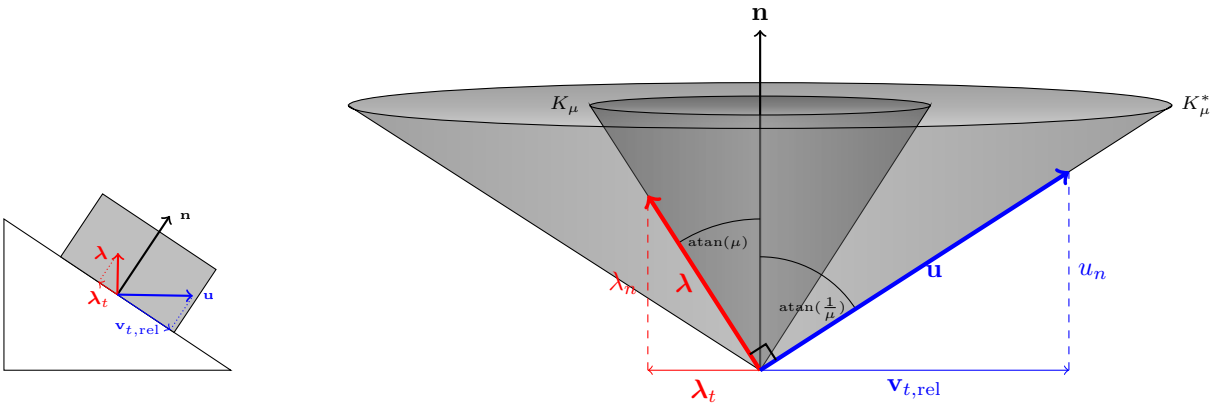


Figure 3: The Cone Complementarity Problem of a contact in sliding mode: Both $\boldsymbol{\lambda}$ and $\mathbf{u}$ are non–zero, and thus they have to be on the boundary of their respective cones to fulfill the complementarity condition.

With

$$D(t) := [D_1(t), ..., D_n(t)] \in \mathbb{R}^{6m\times 3n}$$

and

$$\boldsymbol{\lambda} := \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_n \end{bmatrix} \in \mathbb{R}^{3n}, \quad \mathbf{u} := \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} \in \mathbb{R}^{3n}$$

the equations of motion (3) together with the complementarity condition (5) are given by

$$\begin{aligned} \mathbf{q}(t) &= \mathbf{q}(t_0) + \int_{t_0}^{t} \mathbf{v}(s)\mathrm{d}s, \\ \mathbf{v}(t) &= \mathbf{v}(t_0) + \int_{t_0}^{t} \mathrm{d}\mathbf{v}(s), \\ M\mathrm{d}\mathbf{v}(t) &= \mathbf{f}_{\mathrm{ext}}(t)\mathrm{d}t + D(t)\mathrm{d}\boldsymbol{\lambda}(t) \\ \mathcal{K}_\mu^* \ni \mathbf{u} \quad &\perp \quad \boldsymbol{\lambda} \in \mathcal{K}_\mu, \end{aligned} \tag{8}$$

where

$$\mathcal{K}_\mu := K_{\mu_1} \times ... \times K_{\mu_n}$$

and

$$\mathcal{K}_\mu^* = K_{\mu_1}^* \times ... \times K_{\mu_n}^*.$$

2.4. Time Discretization

When simulating granular material, first order integration methods suffice. In contrast to the classical DEM, we do not require higher order methods to maintain a stable simulation. Due to the non-smoothness of the trajectories, one can only expect first order accuracy in time steps that contain a non–smooth event such as an impact. Thus, given positions $\mathbf{q}(t_j)$, velocities $\mathbf{v}(t_j)$ and external forces $\mathbf{f}_{\mathrm{ext}}(t_j)$ at a time step t_j, we calculate positions at $t_{j+1} = t_j + \Delta t$ using an implicit Euler scheme,

$$\mathbf{q}(t_{j+1}) = \mathbf{q}(t_j) + \Delta t\mathbf{v}(t_{j+1}).$$

Under the assumption that $D(t)$ and $\mathbf{f}_{\mathrm{ext}}(t)$ are approximately constant within one time step $[t_j, t_{j+1}]$, we can discretize equation (8) via

$$\mathbf{v}(t_{j+1}) = \mathbf{v}(t_j) + \Delta t M^{-1}\mathbf{f}_{\mathrm{ext}}(t_j) + M^{-1}D\boldsymbol{\lambda}, \tag{9}$$

where $D = D(t_j)$.

In the spirit of time–stepping schemes, we want to satisfy the inequality constraints $u_i(\mathbf{q}(t)) \geq 0$, $i = 1, ..., n$ only at the boundaries of the intervals $[t_j, t_{j+1}]$. We do not resolve the exact times of impacts and the reaction impulse

$$\boldsymbol{\lambda} = \int_{t_j}^{t_{j+1}} \mathrm{d}\boldsymbol{\lambda}$$

captures all smooth and non–smooth interactions within the time step $[t_j, t_{j+1}]$.

One way of satisfying the constraint $\phi_i(\mathbf{q}(t_{j+1})) \geq 0$ on the velocity level is by demanding

$$\dot{\phi}_i(\mathbf{q}(t_{j+1})) \geq 0$$

for all body pairs with $\phi_i(\mathbf{q}(t_j)) \leq 0$. This is numerically not stable, as initial penetrations increase with time. A better choice is to check for potential contacts in the time interval $[t_j, t_{j+1}]$ at time t_j and requiring

$$\dot{\phi}_i(\mathbf{q}(t_{j+1})) = \nabla\phi_i(\mathbf{q}(t_{j+1}))^T\mathbf{v}(t_{j+1}) \geq -\frac{\phi_i(\mathbf{q}(t_j))}{\Delta t},$$

i.e. the relative normal velocity has to be larger than the constant velocity needed to close the gap between the particles by the time t_{j+1}. This is equivalent to the constraint

$$\begin{aligned} \phi_i(\mathbf{q}(t_{j+1})) &= \phi_i(\mathbf{q}(t_j) + \Delta t\mathbf{v}(t_{j+1})) \\ &\approx \phi_i(\mathbf{q}(t_j)) + \Delta t\nabla\phi_i(\mathbf{q}(t_{j+1}))^T\mathbf{v}(t_{j+1}) \geq 0. \end{aligned}$$

With this stabilization and (9), Equation (7) becomes

$$\begin{aligned} \mathbf{u}_i &= D_i^T\mathbf{v}(t_{j+1}) + \begin{bmatrix} \frac{\phi_i(\mathbf{q}(t_j))}{\Delta t} \\ 0 \\ 0 \end{bmatrix} + \mu_i \begin{bmatrix} \|D_{i\mathbf{t}}^T\mathbf{v}(t_{j+1})\| \\ 0 \\ 0 \end{bmatrix} \\ &= D_i^T M^{-1} D\boldsymbol{\lambda} + \bar{\mathbf{r}}_i + \tilde{\mathbf{r}}_i(\boldsymbol{\lambda}), \end{aligned}$$

where

$$\bar{\mathbf{r}}_i = D_i^T \left(\mathbf{v}(t_j) + \Delta t M^{-1} \mathbf{f}_{\text{ext}}(t_j) \right) + \begin{bmatrix} \frac{\phi_i(\mathbf{q}(t_j))}{\Delta t} \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

and

$$\tilde{\mathbf{r}}_i(\boldsymbol{\lambda}) = \mu_i \begin{bmatrix} \| D_{i\mathbf{t}}^T \mathbf{v}(t_{j+1}) \| \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \mu_i \| D_{i\mathbf{t}}^T M^{-1} D \boldsymbol{\lambda} + D_{i\mathbf{t}}^T \left(\mathbf{v}(t_j) + \Delta t M^{-1} \mathbf{f}_{\text{ext}}(t_j) \right) \| \\ 0 \\ 0 \end{bmatrix}.$$

Finally, by writing

$$\bar{N} = D^T M^{-1} D \in \mathbb{R}^{3n \times 3n}, \tag{11}$$

$$\bar{\mathbf{r}} := \begin{bmatrix} \bar{\mathbf{r}}_1 \\ \vdots \\ \bar{\mathbf{r}}_n \end{bmatrix}, \tilde{\mathbf{r}}(\boldsymbol{\lambda}) := \begin{bmatrix} \tilde{\mathbf{r}}_1(\boldsymbol{\lambda}) \\ \vdots \\ \tilde{\mathbf{r}}_n(\boldsymbol{\lambda}) \end{bmatrix} \in \mathbb{R}^{3n},$$

we obtain

$$\mathbf{u} = \bar{F}(\boldsymbol{\lambda}) = \bar{N}\boldsymbol{\lambda} + \bar{\mathbf{r}} + \tilde{\mathbf{r}}(\boldsymbol{\lambda})$$

with a positive semi–define matrix $\bar{N}$.

The equations that have to be satisfied in every time step of a simulation are

$$\begin{aligned} \mathbf{q}(t_{j+1}) &= \mathbf{q}(t_j) + \Delta t \mathbf{v}(t_{j+1}), \\ \mathbf{v}(t_{j+1}) &= \mathbf{v}(t_j) + \Delta t M^{-1} \mathbf{f}_{\text{ext}}(t_j) + M^{-1} D \boldsymbol{\lambda}, \\ \mathcal{K}_\mu \ni \boldsymbol{\lambda} \quad &\perp \quad \mathbf{u} = \bar{F}(\boldsymbol{\lambda}) \in \mathcal{K}_\mu^*. \end{aligned} \tag{12}$$

Almost the entire computational effort for a time step is spent on solving (12). Note, that without the term $\tilde{\mathbf{r}}(\boldsymbol{\lambda})$, $\bar{F}$ would be a linear function in $\boldsymbol{\lambda}$ and (12) would be equivalent to a quadratic minimization problem subject to conical constraints. The term $\tilde{\mathbf{r}}(\boldsymbol{\lambda})$ complicates numerical methods that make use of the gradient of $\bar{F}(\boldsymbol{\lambda})$, because $\bar{F}$ is not differentiable if the relative contact velocity is zero. Thus we have to make use of the generalized Jacobian

$$\begin{aligned} \partial \bar{F}(\boldsymbol{\lambda}) &= \bar{N} + \partial \tilde{\mathbf{r}}(\boldsymbol{\lambda}) \\ &= \bar{N} + \begin{bmatrix} \partial \tilde{\mathbf{r}}_1(\boldsymbol{\lambda}) \\ \vdots \\ \partial \tilde{\mathbf{r}}_n(\boldsymbol{\lambda}) \end{bmatrix} \end{aligned} \tag{13}$$

with

$$\partial \tilde{\mathbf{r}}_i(\boldsymbol{\lambda}) = \begin{bmatrix} \frac{\mu_i}{\| D_{i\mathbf{t}}^T \mathbf{v}(t_{j+1}) \|} \mathbf{v}(t_{j+1})^T D_{i\mathbf{t}} D_{i\mathbf{t}}^T M^{-1} D \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{3 \times 3n}$$

if $D_{i\mathbf{t}}^T \mathbf{v}(t_{j+1}) \neq \mathbf{0}$; and

$$\partial \tilde{\mathbf{r}}_i(\boldsymbol{\lambda}) = \left\{ \begin{bmatrix} \frac{\mu_i}{\| \boldsymbol{\xi}_i \|} \boldsymbol{\xi}_i^T D_{i\mathbf{t}}^T M^{-1} D \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{3 \times 3n} \;\middle|\; \boldsymbol{\xi}_i \in \mathbb{R}^2 \right\}$$

if $D_{i\mathbf{t}}^T \mathbf{v}(t_{j+1}) = \mathbf{0}$. The gradient $\nabla \bar{F}(\boldsymbol{\lambda})$ would be constant if it wasn't for the term $\tilde{\mathbf{r}}(\boldsymbol{\lambda})$ and would not need to be re–evaluated whenever $\boldsymbol{\lambda}$ changes. In addition, $\nabla \bar{F}(\boldsymbol{\lambda})$ would be symmetric.

Fortunately, μ_i and $D_{i\mathbf{t}}^T \mathbf{v}(t)$ are usually fairly small in the context of granular material, and ignoring $\tilde{\mathbf{r}}(\boldsymbol{\lambda})$ is a valid approximation in a wide range of applications (Anitescu, 2005).

3. Symmetric Cones

The Interior Point Method for the solution of (12) proposed in this paper makes use of the Jordan–algebraic structure defined on $\mathbb{R}^N$, which is directly related to a self–dual cone $C = C^*$. This section gives a definition of cones and dual cones and we revise the most important properties of Jordan algebras and symmetric cones needed in Section 4.

A proper cone $K \in \mathbb{R}^N$ is a closed, non–empty set, such that

- it is closed with respect to positive linear combinations, i.e., for $\mathbf{x}, \mathbf{y} \in K$ and $\alpha, \beta \geq 0$ it holds

$$\alpha \mathbf{x} + \beta \mathbf{y} \in K;$$

- it does not contain a linear space, i.e.,

$$\mathbf{x} \in K \text{ and } -\mathbf{x} \in K \quad \Rightarrow \quad \mathbf{x} = 0.$$

For a proper cone K we define the dual cone as

$$K^* := \left\{ \mathbf{y} \in \mathbb{R}^N \;\middle|\; \mathbf{x}^T \mathbf{y} \geq 0 \text{ for all } \mathbf{x} \in K \right\}.$$

K^* is also a proper cone.

A real Jordan algebra is a vector space A with a vector product *(or Jordan product)*

$$\circ : A \times A \to A$$

satisfying the properties

$$\begin{aligned} &\mathbf{x} \circ \mathbf{y} = \mathbf{y} \circ \mathbf{x} && \textit{(commutativity)}, \\ &(\mathbf{x} \circ \mathbf{x}) \circ (\mathbf{x} \circ \mathbf{y}) = \mathbf{x} \circ (\mathbf{x} \circ (\mathbf{x} \circ \mathbf{y})) && \textit{(power associativity)}, \end{aligned}$$

for all $\mathbf{x}, \mathbf{y} \in A$. The symmetric cone C in A is the set of squares with respect to the Jordan product,

$$C := \left\{ \mathbf{x} \circ \mathbf{x} \;\middle|\; \mathbf{x} \in A \right\}.$$

There are several examples of Jordan algebras, and the Interior Point algorithm presented here can be seen as a generic solver for any Cone Complementarity Problem with a symmetric cone[2]. In this paper however, we are

[2] Important examples of symmetric cones are the cone of positive semi–definite matrices or the positive orthant $\mathbb{R}_+^N$.

concerned specifically with the space $A = \mathbb{R}^N$ for some $N \geq 1$ together with the Jordan product

$$\mathbf{x} \circ \mathbf{y} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{x}^T\mathbf{y} \\ x_{\mathbf{n}}\mathbf{y_t} + y_{\mathbf{n}}\mathbf{x_t} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^{N-1}$$

for all

$$\mathbf{x} = \begin{bmatrix} x_{\mathbf{n}} \\ \mathbf{x_t} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{\mathbf{n}} \\ \mathbf{y_t} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^{N-1}.$$

The symmetric cone in $(\mathbb{R}^N, \circ)$ is

$$\begin{aligned} C &:= \{ \ \mathbf{x} \circ \mathbf{x} \ \mid \ \mathbf{x} \in A \} \\ &= \left\{ \begin{bmatrix} x_{\mathbf{n}} \\ \mathbf{x_t} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^{N-1} \quad \mid \quad x_{\mathbf{n}} \geq \|\mathbf{x_t}\| \right\}. \end{aligned} \tag{14}$$

In the following, we revise the essential properties of the Jordan algebra $(\mathbb{R}^N, \circ)$ needed to derive the IPM. The results from this section are taken from (Faraut and Korányi, 1994; Faybusovich, 2002; Fukushima et al., 2002; Alizadeh and Goldfarb, 2003; Bai et al., 2004; Hayashi, 2004).

(a) C is self–dual, i.e., $C = C^*$ and thus $\mathbf{x}^T\mathbf{y} \geq 0$ for all $\mathbf{x}, \mathbf{y} \in C$.

(b) The unit element

$$\mathbf{e} = \begin{bmatrix} \sqrt{2} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^{N-1}$$

satisfies $\mathbf{x} \circ \mathbf{e} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^N$.

(c) Each $\mathbf{x} \in \mathbb{R}^N$ can be written as

$$\mathbf{x} = \lambda_1 \mathbf{q}_1 + \lambda_2 \mathbf{q}_2$$

with spectral values

$$\lambda_1 := \frac{1}{\sqrt{2}} (x_{\mathbf{n}} - \|\mathbf{x_t}\|) \quad \text{and} \quad \lambda_2 := \frac{1}{\sqrt{2}} (x_{\mathbf{n}} + \|\mathbf{x_t}\|)$$

and spectral vectors

$$\mathbf{q}_1 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -\frac{\mathbf{x_t}}{\|\mathbf{x_t}\|} \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ +\frac{\mathbf{x_t}}{\|\mathbf{x_t}\|} \end{bmatrix}.$$

The spectral vectors are an orthonormal coordinate system of the plane spanned by $\mathbf{e}$ and $\mathbf{x}$. If $\mathbf{x_t}$ is zero, the term $\mathbf{x_t}/\|\mathbf{x_t}\|$ in the definition of $\mathbf{q}_1$ and $\mathbf{q}_2$ can be replaced with any vector of unit length in $\mathbb{R}^{N-1}$.

It holds $\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{e}$; $\mathbf{x} \in C$ if and only if $\lambda_1, \lambda_2 \geq 0$; and $\mathbf{x} \in \operatorname{int} C$, if and only if $\lambda_1, \lambda_2 > 0$.

(d) The trace of a vector $\mathbf{x} \in \mathbb{R}^N$ is defined by

$$\operatorname{tr}(\mathbf{x}) := \lambda_1 + \lambda_2 = \sqrt{2} x_{\mathbf{n}}$$

and the determinant is

$$\det(\mathbf{x}) := \lambda_1 \cdot \lambda_2 = \frac{1}{2} \left(x_{\mathbf{n}}^2 - \|\mathbf{x_t}\|^2 \right).$$

(e) $J = \left[\begin{smallmatrix} 1 & \mathbf{0} \\ \mathbf{0} & -I_{N-1} \end{smallmatrix} \right] \in \mathbb{R}^{N \times N}$ is called the reflection matrix, where I_{N-1} denotes the unit matrix in $\mathbb{R}^{N-1}$. A vector $\mathbf{x} \in \mathbb{R}^N$ is said to be invertible, if $\det(\mathbf{x}) \neq 0$ and its inverse is given by

$$\mathbf{x}^{-1} = \frac{1}{\det(\mathbf{x})} J\mathbf{x}.$$

It holds $\mathbf{x}^{-1} \circ \mathbf{x} = \mathbf{e}$ for all invertible $\mathbf{x} \in \mathbb{R}^N$.

(f) For any function $\psi : \mathbb{R} \to \mathbb{R}$ we define

$$\psi(\mathbf{x}) := \psi(\lambda_1)\mathbf{q}_1 + \psi(\lambda_2)\mathbf{q}_2$$

for $\mathbf{x} \in \mathbb{R}^N$. More specifically, for any $\alpha \in \mathbb{R}$ we can define

$$\mathbf{x}^\alpha = \lambda_1^\alpha \mathbf{q}_1 + \lambda_2^\alpha \mathbf{q}_2,$$

if $\lambda_1^\alpha, \lambda_2^\alpha$ exist.

(g) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, the function $g(\mathbf{x}, \mathbf{y}) = \mathbf{x} \circ \mathbf{y}$ is bilinear and it holds

$$\mathbf{x} \circ \mathbf{y} = L(\mathbf{x}) \cdot \mathbf{y}$$

where $L(x)$ is the arrowhead matrix

$$L(x) = \nabla_{\mathbf{y}} g(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{2}} \begin{bmatrix} x_{\mathbf{n}} & \mathbf{x_t}^T \\ \mathbf{x_t} & I_{N-1} x_{\mathbf{n}} \end{bmatrix}.$$

The eigenvalues of $L(\mathbf{x})$ are

$$\begin{aligned} \lambda_1 &= \frac{1}{\sqrt{2}}(x_{\mathbf{n}} - \|\mathbf{x_t}\|), \\ \lambda_2 &= \frac{1}{\sqrt{2}}(x_{\mathbf{n}} + \|\mathbf{x_t}\|), \\ \lambda_k &= \sqrt{2}\mathbf{x_n} \quad \text{for } k = 3, ..., N. \end{aligned}$$

(h) The matrix

$$P(\mathbf{x}) := 2L(\mathbf{x})^2 - L(\mathbf{x}^2).$$

is called the quadratic representation of $\mathbf{x} \in \mathbb{R}^N$. It has the following properties:

- $P(\mathbf{x}) = \mathbf{x}\mathbf{x}^T - \det(\mathbf{x})J$,
- $P(\mathbf{x})$ is symmetric and positive definite,
- $P(\mathbf{x})$ and $L(\mathbf{x})$ commute and thus share a set of eigenvectors,
- $P(\mathbf{x})\mathbf{e} = \mathbf{x}^2$,
- $P(\mathbf{x})^\alpha = P(\mathbf{x}^\alpha)$, if $\mathbf{x}^\alpha$ exists,
- $P(\mathbf{x}^{-1})\mathbf{y}^{-1} = (P(\mathbf{x})\mathbf{y})^{-1}$ if $\mathbf{x}$ and $\mathbf{y}$ are invertible,
- $P(P(\mathbf{x})\mathbf{y}) = P(\mathbf{x})P(\mathbf{y})P(\mathbf{x})$ and
- $P(\mathbf{x})C = C$ and $P(\mathbf{x})\operatorname{int} C = \operatorname{int} C$ if $\mathbf{x}$ is invertible.

The eigenvalues of $P(\mathbf{x})$ are

$$\lambda_1 = \lambda_1(\mathbf{x})^2 = \frac{1}{2}(x_\mathbf{n} - \|\mathbf{x_t}\|)^2,$$
$$\lambda_2 = \lambda_2(\mathbf{x})^2 = \frac{1}{2}(x_\mathbf{n} + \|\mathbf{x_t}\|)^2,$$
$$\lambda_k = \det(\mathbf{x}) = \frac{1}{2}(x_\mathbf{n}^2 - \|\mathbf{x_t}\|^2) \quad \text{for } k = 3, ..., N.$$

(i) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, let $\mathbf{v} = P(\mathbf{x})^{\frac{1}{2}}\mathbf{y}$. It holds

$$\det(\mathbf{v}) = \det(\mathbf{x}) \cdot \det(\mathbf{y}),$$
$$\mathrm{tr}(\mathbf{v}) = \mathbf{x}^T\mathbf{y}.$$

It follows, that the spectral values of $\mathbf{v}$ are

$$\lambda_1(\mathbf{v}) = \frac{\mathbf{x}^T\mathbf{y}}{2} - \sqrt{\left[\frac{\mathbf{x}^T\mathbf{y}}{2}\right]^2 - \det(\mathbf{x})\det(\mathbf{y})},$$
$$\lambda_2(\mathbf{v}) = \frac{\mathbf{x}^T\mathbf{y}}{2} + \sqrt{\left[\frac{\mathbf{x}^T\mathbf{y}}{2}\right]^2 - \det(\mathbf{x})\det(\mathbf{y})}.$$

(j) There exists a unique automorphism $W = W(\mathbf{x}, \mathbf{y})$ such that for $\mathbf{x}, \mathbf{y} \in \mathrm{int}\, C$ it holds $W\mathbf{x} = W^{-1}\mathbf{y}$. The automorphism is given by

$$W(\mathbf{x}, \mathbf{y}) = P(\mathbf{w}^{\frac{1}{2}})$$

with

$$\mathbf{w} = P(\mathbf{x}^{-\frac{1}{2}})\left(P(\mathbf{x}^{\frac{1}{2}})\mathbf{y}\right)^{\frac{1}{2}}.$$

The vector $\mathbf{w} \in \mathbb{R}^N$ is called the scaling point for $\mathbf{x}, \mathbf{y} \in C$. It holds $\mathbf{w} \in \mathrm{int}\, C$ and thus

$$W\mathrm{int}\, C = W^{-1}\mathrm{int}\, C = \mathrm{int}\, C.$$

(k) The direct product of n symmetric cones

$$\mathcal{C} = C \times ... \times C \in \mathbb{R}^{N \cdot n}$$

is a symmetric cone and $\left(\mathbb{R}^{N \cdot n}, \diamond\right)$ is a Jordan algebra with Jordan product

$$\mathbf{x} \diamond \mathbf{y} = \begin{bmatrix} \mathbf{x}_1 \circ \mathbf{y}_1 \\ \vdots \\ \mathbf{x}_n \circ \mathbf{y}_n \end{bmatrix}.$$

All the properties of $\left(\mathbb{R}^N, \circ\right)$ hold for $\left(\mathbb{R}^{N \cdot n}, \diamond\right)$ and have to be understood componentwise.

To avoid overloading the notation, we use the same names in $\mathbb{R}^N \times ... \times \mathbb{R}^N$ as in $\mathbb{R}^N$ in the remainder of this paper. E.g., both the unit element in $\mathbb{R}^N \times ... \times \mathbb{R}^N$ and the unit element in $\mathbb{R}^N$ are denoted by $\mathbf{e}$ and the respective Jordan Product of $\mathbf{x}$ and $\mathbf{y}$ is written as $\mathbf{x} \circ \mathbf{y}$. We only comment on the dimension if it is not clear from the context.

4. The Interior Point Algorithm

In this section, we illustrate our Interior Point algorithm to solve the CCP (12). The goal is to formulate the IPM analogously to the algorithms presented in (Kojima et al., 1991) for Linear Complementarity Problems[3]. This is done in six steps. Firstly, we interpret the Cone Complementarity Problem as a minimization of a potential function. Secondly, we introduce the idea behind the central path and path–following. As a third step we provide one way of constructing a feasible starting point for the Interior Point method. As a fourth step, we discuss the maximum step length of the algorithm. Then we will introduce Nesterov–Todd scaling to improve the convergence rate and finally, inexact search directions are discussed. This section is concluded with a pseudo–code of the Interior Point Method.

4.1. The Potential Function

To make use of the algebraic structure presented in Section 3, the cones $\mathcal{K}_\mu$ and $\mathcal{K}_\mu^*$ must first be transformed to a symmetric cone $\mathcal{C} = \mathcal{C}^*$. Using the transformations

$$\mathbf{x} = \mathcal{T}_x \cdot \boldsymbol{\lambda} = \begin{bmatrix} T^x_{\mu_1} & & \\ & \ddots & \\ & & T^x_{\mu_n} \end{bmatrix} \cdot \boldsymbol{\lambda},$$
$$\mathbf{y} = \mathcal{T}_y \cdot \mathbf{u} = \begin{bmatrix} T^y_{\mu_1} & & \\ & \ddots & \\ & & T^y_{\mu_n} \end{bmatrix} \cdot \mathbf{u}$$

with

$$T^x_{\mu_i} := \begin{bmatrix} \mu_i & & \\ & 1 & \\ & & 1 \end{bmatrix} \quad \text{and} \quad T^y_{\mu_i} := \begin{bmatrix} 1 & & \\ & \mu_i & \\ & & \mu_i \end{bmatrix}$$

and

$$F(\mathbf{x}) = \mathcal{T}_y \cdot \bar{F}(\mathcal{T}_x^{-1} \cdot \mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^{3n}$$

we can rewrite the CCP (12) as

$$\mathcal{C} \ni \mathbf{x} \quad \perp \quad \mathbf{y} = F(\mathbf{x}) \in \mathcal{C} \tag{15}$$

in terms of the cone $\mathcal{C} = C \times .... \times C$ with

$$C = \left\{ \begin{bmatrix} x_\mathbf{n} \\ \mathbf{x_t} \end{bmatrix} \in \mathbb{R} \times \mathbb{R}^2 \quad | \quad x_\mathbf{n} \geq \|\mathbf{x_t}\| \right\}.$$

In the remainder of this paper we are concerned with the CCP (15).

Proposition 1.

$$\mathbf{x}^T\mathbf{y} = 0, \;\; \mathbf{x}, \mathbf{y} \in \mathcal{C} \quad \Leftrightarrow \quad \mathbf{x} \circ \mathbf{y} = \mathbf{0}, \;\; \mathbf{x}, \mathbf{y} \in \mathcal{C}$$

[3] A Linear Complementarity Problem is a Cone Complementarity Problem where the Jordan product is the componentwise product and the symmetric cone is the positive orthant $\mathcal{C} = \mathbb{R}^n_+$.

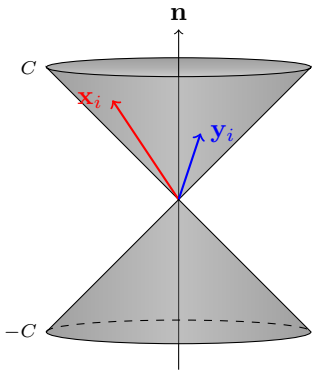


Figure 4: A point $(\mathbf{x}, \mathbf{y}) \in S_{++}$ requires $\mathbf{x}_i, \mathbf{y}_i \in \operatorname{int} C$ for all $i = 1, ..., n$; while $\det(\mathbf{x}_i), \det(\mathbf{y}_i) \geq 0$ only implies $\mathbf{x}_i, \mathbf{y}_i \in C \cup (-C)$.

In (Fukushima et al., 2002) a proof of Proposition 1 is given for the cone C, which can easily be extended for the cone $\mathcal{C} = C \times ... \times C$ by using the fact that for $\mathbf{x}_i, \mathbf{y}_i \in C$ it holds $\mathbf{x}_i^T \mathbf{y}_i \geq 0$ and thus

$$\mathbf{x}^T \mathbf{y} = 0 \quad \Leftrightarrow \quad \mathbf{x}_i^T \mathbf{y}_i = 0, \text{ for } i = 1, ..., n.$$

Let

$$S_{++} := \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{3n} \times \mathbb{R}^{3n} \mid \mathbf{x}, \mathbf{y} \in \operatorname{int} \mathcal{C},\ \mathbf{y} = F(\mathbf{x}) \right\}$$

denote the set of interior points. We can interpret (15) as the minimization problem

$$\min\ \mathbf{x}^T \mathbf{y} \quad \text{subject to} \quad (\mathbf{x}, \mathbf{y}) \in \overline{S_{++}}. \tag{16}$$

It is convenient for the definition of the central path to minimize a potential function with a logarithmic barrier for the set

$$\begin{aligned} &\mathcal{C} \cup (-\mathcal{C}) \\ &= \left\{ \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^{3n} \mid \det(\mathbf{x}_i) \geq 0 \ \forall\ i = 1, ..., n \right\} \end{aligned}$$

rather than using a logarithmic barrier for $\mathcal{C}$, see Figure 4. This gives rise to the minimization of the potential function

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) =& (2n + \rho) \log \mathbf{x}^T \mathbf{y} - 2n \log n \\ &- \sum_{i=1}^{n} \log \left(2 \det(\mathbf{x}_i)\right) - \sum_{i=1}^{n} \log \left(2 \det(\mathbf{y}_i)\right) \end{aligned}$$

where $\rho > 0$ is an arbitrary positive constant. The first two terms are a scaled version of the cost function in (16). The last two terms act as a logarithmic potential that drives $\mathbf{x}_i$ and $\mathbf{y}_i$ away from the manifold defined by $\det(\mathbf{x}_i) = 0, \det(\mathbf{y}_i) = 0, i = 1, ..., n$, i.e. the boundary of the double cone $\mathcal{C} \cup (-\mathcal{C})$. We split the potential function into two parts via

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) =& \rho \log \mathbf{x}^T \mathbf{y} + f_{cen}(\mathbf{x}, \mathbf{y}), \\ f_{cen}(\mathbf{x}, \mathbf{y}) :=& 2n \log \mathbf{x}^T \mathbf{y} - 2n \log n \\ &- \sum_{i=1}^{n} \log \left(2 \det(\mathbf{x}_i)\right) - \sum_{i=1}^{n} \log \left(2 \det(\mathbf{y}_i)\right) \\ =& 2n \log \left(\frac{\mathbf{x}^T \mathbf{y}}{n} \right) \\ &- 2n \log \left(\prod_{i=1}^{n} \left[4 \det(\mathbf{x}_i) \det(\mathbf{y}_i)\right]^{1/2n} \right) \\ =& 2n \log \frac{\mathbf{x}^T \mathbf{y} / n}{\prod_{i=1}^{n} \left[2\sqrt{\det(\mathbf{x}_i) \det(\mathbf{y}_i)}\right]^{1/n}}. \end{aligned} \tag{17}$$

Lemma 1. *Let* $\mathbf{x} = \begin{bmatrix} x_\mathbf{n} \\ \mathbf{x_t} \end{bmatrix} \in \operatorname{int} C$ *and* $\det(\mathbf{y}) \geq 0$. *It holds*

(a) $2\sqrt{\det(\mathbf{x}) \cdot \det(\mathbf{y})} \leq |\mathbf{x}^T \mathbf{y}|$.

(b) $2\sqrt{\det(\mathbf{x}) \cdot \det(\mathbf{y})} = |\mathbf{x}^T \mathbf{y}| \quad \Leftrightarrow \quad \mathbf{x} \circ \mathbf{y} = \alpha \mathbf{e}$ *for some* $\alpha \in \mathbb{R}$.

Proof.

(a) Let $\mathbf{v} = P(\mathbf{x})^{\frac{1}{2}} \mathbf{y}$. Then, because of Section 3(i), the left–hand side in (a) is twice the geometric mean of the spectral values λ_1, λ_2 of $\mathbf{v}$, and the right–hand side is twice their arithmetic mean:

$$\begin{aligned} \sqrt{\det(\mathbf{x}) \cdot \det(\mathbf{y})} = \sqrt{\det(\mathbf{v})} = \sqrt{\lambda_1 \cdot \lambda_2}, \\ \frac{\mathbf{x}^T \mathbf{y}}{2} = \frac{\operatorname{tr}(\mathbf{v})}{2} = \frac{1}{2} (\lambda_1 + \lambda_2). \end{aligned}$$

$\mathbf{x}$ is invertible because of $\mathbf{x} \in \operatorname{int} C$. It follows that $\mathbf{x}^{\frac{1}{2}}$ is invertible:

$$0 \neq \det(\mathbf{x}) = \det(\mathbf{x}^{\frac{1}{2}})^2.$$

With $\det(\mathbf{y}) \geq 0$ we have $\mathbf{y} \in C \cup (-C)$. We distinguish two cases.

- $\mathbf{y} \in C$
  With (h) from Section 3 it follows $\mathbf{v} = P(\mathbf{x})^{\frac{1}{2}} \mathbf{y} \in C$. Thus, the spectral values

  $$\begin{aligned} \lambda_1 &= \frac{1}{\sqrt{2}} (v_\mathbf{n} - \|\mathbf{v_t}\|) \quad \text{and} \\ \lambda_2 &= \frac{1}{\sqrt{2}} (v_\mathbf{n} + \|\mathbf{v_t}\|) \end{aligned}$$

  are non–negative and the arithmetic mean of two non–negative values is always greater or equal to their geometric mean.

- $\mathbf{y} \in -C$
  Just as in the first case we can use Section 3(h) to obtain

$$-\mathbf{v} = P(\mathbf{x})^{\frac{1}{2}}(-\mathbf{y}) \in C \quad \Leftrightarrow \quad \mathbf{v} \in -C,$$

  and thus

$$0 \geq \lambda_1 = \frac{1}{\sqrt{2}}\left(v_{\mathbf{n}} - \|\mathbf{v_t}\|\right) \quad \text{and}$$
$$0 \geq \lambda_2 = \frac{1}{\sqrt{2}}\left(v_{\mathbf{n}} + \|\mathbf{v_t}\|\right).$$

  Let $\tilde{\lambda}_1 = -\lambda_1$ and $\tilde{\lambda}_2 = -\lambda_2$. Analogously to the first case, we conclude

$$\begin{aligned}&\sqrt{\tilde{\lambda}_1\tilde{\lambda}_1} = \sqrt{\lambda_1\lambda_2} = \sqrt{\det(\mathbf{x})\det(\mathbf{y})}\\ \leq&\frac{1}{2}(\tilde{\lambda}_1 + \tilde{\lambda}_2) = -\frac{1}{2}(\lambda_1 + \lambda_2) = -\frac{\mathbf{x}^T\mathbf{y}}{2}\\ =&\frac{1}{2}|\mathbf{x}^T\mathbf{y}|.\end{aligned}$$

(b) Without loss of generality, let $\mathbf{y} \in C$. If $\mathbf{y} \in -C$, we follow the same steps using $\tilde{\lambda}_1 = -\lambda_1$ and $\tilde{\lambda}_2 = -\lambda_2$ just as in the first part of the proof.

It holds $\sqrt{\det(\mathbf{x}) \cdot \det(\mathbf{y})} = \frac{\mathbf{x}^T\mathbf{y}}{2}$, if the arithmetic mean of $\lambda_1(\mathbf{v})$ and $\lambda_2(\mathbf{v})$ is equal to the geometric mean, i.e., if and only if

$$\begin{aligned}\lambda_1 &= \frac{1}{\sqrt{2}}\left(v_{\mathbf{n}} - \|\mathbf{v_t}\|\right)\\ =\lambda_2 &= \frac{1}{\sqrt{2}}\left(v_{\mathbf{n}} + \|\mathbf{v_t}\|\right).\end{aligned}$$

Therefore it needs to be shown, that

$$\mathbf{v_t} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{x} \circ \mathbf{y} = \alpha\mathbf{e} \quad \alpha \in \mathbb{R}.$$

Since $\mathbf{x}$ is invertible, $P(\mathbf{x}^{\frac{1}{2}}) = P(\mathbf{x})^{\frac{1}{2}}$ is invertible and using 3(h) we know that

$$\mathbf{x}^{-1} = P(\mathbf{x}^{-\frac{1}{2}})\mathbf{e}.$$

In addition, it holds $L(\mathbf{x}^{-1}) = L(\mathbf{x})^{-1}$ because of

$$\begin{aligned}&\mathbf{e} = L(\mathbf{x}^{-1})\mathbf{x}\\ \Leftrightarrow\quad &\mathbf{x} = L(\mathbf{x})\mathbf{e} = L(\mathbf{x})L(\mathbf{x}^{-1})\mathbf{x}\\ \Leftrightarrow\quad &L(\mathbf{x})L(\mathbf{x}^{-1}) = \mathrm{Id}\end{aligned}$$

for all invertible $\mathbf{x}$. Therefore,

$$\begin{aligned}\mathbf{v_t} = \mathbf{0} \quad &\Leftrightarrow \quad \mathbf{v} = P(\mathbf{x}^{\frac{1}{2}})\mathbf{y} = \alpha\mathbf{e}, \quad \alpha \in \mathbb{R}\\ &\Leftrightarrow \quad \mathbf{y} = \alpha P(\mathbf{x}^{-\frac{1}{2}})\mathbf{e}\\ &\Leftrightarrow \quad \mathbf{y} = \alpha\mathbf{x}^{-1} = \alpha L(\mathbf{x}^{-1})\mathbf{e}\\ &\Leftrightarrow \quad L(\mathbf{x})\mathbf{y} = \alpha\mathbf{e}\\ &\Leftrightarrow \quad \mathbf{x} \circ \mathbf{y} = \alpha\mathbf{e}.\end{aligned}$$

□

Theorem 1. *Let $(\mathbf{x}, \mathbf{y}) \in S_{++}$. Then it holds*

(a) $f_{cen}(\mathbf{x}, \mathbf{y}) \geq 0$,

(b) $f_{cen}(\mathbf{x}, \mathbf{y}) = 0 \quad \Leftrightarrow \quad \mathbf{x} \circ \mathbf{y} = \alpha\mathbf{e}$ *for some* $\alpha > 0$.

Proof.

(a) From Lemma 1(a) it follows

$$\mathbf{x}_i^T\mathbf{y}_i \geq 2\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)} \quad \text{for all } i = 1, ..., n. \tag{18}$$

Since the terms on the left and right hand side of the equation are both non–negative, we can take the product over all i on both sides and apply the n–th root to obtain

$$\left(\prod_{i=1}^{n}\mathbf{x}_i^T\mathbf{y}_i\right)^{\frac{1}{n}} \geq \left(\prod_{i=1}^{n}2\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)}\right)^{\frac{1}{n}}. \tag{19}$$

The result follows from the fact that the arithmetic mean of a set of positive numbers is always larger or equal to its geometric mean,

$$\frac{1}{n}\mathbf{x}^T\mathbf{y} \geq \left(\prod_{i=1}^{n}\mathbf{x}_i^T\mathbf{y}_i\right)^{\frac{1}{n}} \geq \left(\prod_{i=1}^{n}2\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)}\right)^{\frac{1}{n}}. \tag{20}$$

Therefore the term in the logarithm in (17) is larger than one and we have

$$f_{cen}(\mathbf{x}, \mathbf{y}) \geq 0.$$

(b) With Lemma 1(b) we know, that the inequalities in (18) and (19) hold with equality if and only if $\mathbf{x}_i \circ \mathbf{y}_i = \alpha_i\mathbf{e}$ for all $i = 1, ..., n$ and some $\alpha_i \geq 0$.

The first inequality in (20) holds with equality if and only if

$$0 \leq \mathbf{x}_i^T\mathbf{y}_i = \mathbf{x}_j^T\mathbf{y}_j \quad \text{for all } i, j \in \{1, ..., n\}.$$

Nominator and denominator in (17) are equal if and only if $\alpha_i = \alpha_j$ for all $i, j \in \{1, ..., n\}$ and therefore it must hold for some $\alpha \geq 0$

$$\mathbf{x}_i \circ \mathbf{y}_i = \alpha\mathbf{e} \quad \text{for all } i = 1, ..., n$$

and the claim is proved.

□

Because of Theorem 1(a) we know that

$$f(\mathbf{x}, \mathbf{y}) \geq \rho \log \mathbf{x}^T\mathbf{y}$$

for any $\rho \geq 0$, and we can regard each $(\mathbf{x}, \mathbf{y}) \in S_{++}$ with a small $f(\mathbf{x}, \mathbf{y})$ as an approximate solution to the CCP (15).

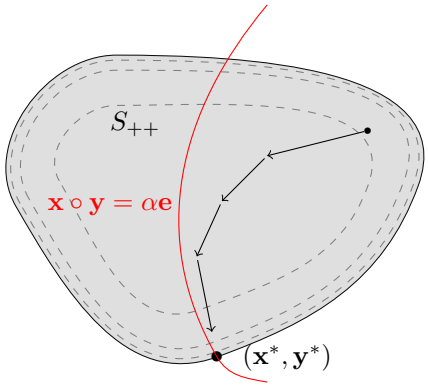


Figure 5: A sequence $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ approaching the optimum $(\mathbf{x}^*, \mathbf{y}^*)$ in a small neighborhood of the central path.

4.2. The Central Path

The logarithmic barrier f_{cen} penalizes values close to the boundary, since the denominator in the logarithm in equation (17) tends towards zero as $(\mathbf{x}, \mathbf{y}) \in S_{++}$ approaches the boundary of the feasible set. Let $\left(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}\right)$ be a sequence in S_{++} that approaches an optimum $(\mathbf{x}^*, \mathbf{y}^*)$ of (16). If $f_{cen}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) = 0$, the sequence approaches the boundary of S_{++} from the interior as fast as it decreases the cost function $\mathbf{x}^T\mathbf{y}/n$, and therefore, the sequence approaches the optimum strictly from within the feasible set, staying clear from the constraints.

The central path is defined as

$$\begin{aligned} S_{cen} &= \{\ (\mathbf{x}, \mathbf{y}) \in S_{++} \ \mid\ f_{cen}(\mathbf{x}, \mathbf{y}) = 0\,\} \\ &= \{\ (\mathbf{x}, \mathbf{y}) \in S_{++} \ \mid\ \mathbf{x} \circ \mathbf{y} = \alpha \mathbf{e},\ \ \alpha > 0\,\}\,. \end{aligned}$$

In practice, it is not possible to rigorously enforce $f_{cen}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) = 0$. Instead, one tries to find a sequence in a small neighborhood of the central path, see Figure 5.

A Newton step applied to the function

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{x} \circ \mathbf{y} - \alpha \mathbf{e} = \mathbf{0}$$

is a step towards the central path at α. Given $\left(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}\right) \in S_{++}$, the search direction of the Interior Point method is given by a solution to

$$\begin{bmatrix} \nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) & \nabla_{\mathbf{y}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) \\ \nabla F(\mathbf{x}^{(k)}) & -I \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}^{(k)} \\ \Delta\mathbf{y}^{(k)} \end{bmatrix} = \begin{bmatrix} \alpha\mathbf{e} - \mathbf{x}^{(k)} \circ \mathbf{y}^{(k)} \\ \mathbf{0} \end{bmatrix}$$

The matrices $\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) \in \mathbb{R}^{3n\times 3n}$ and $\nabla_{\mathbf{y}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) \in \mathbb{R}^{3n\times 3n}$ are block–diagonal with 3×3 blocks

$$\begin{aligned} \left(\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\right)_{i,i} &= L(\mathbf{y}_i^{(k)}) \\ \left(\nabla_{\mathbf{y}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\right)_{i,i} &= L(\mathbf{x}_i^{(k)}), \end{aligned}$$

for $i = 1, ..., n$. In the remainder of this paper we write

$$\begin{aligned} L(\mathbf{y}^{(k)}) &:= \nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) \\ L(\mathbf{x}^{(k)}) &:= \nabla_{\mathbf{y}}\mathbf{g}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) \end{aligned}$$

according to our comments in 3(k).

Finally, we choose a special value for α given by

$$\alpha = \beta \frac{\mathbf{x}^{(k)T}\mathbf{y}^{(k)}}{2n},$$

where $\beta \in (0, 1]$ is a parameter. A choice of $\beta = 1$ means we are looking for the point $(\mathbf{x}', \mathbf{y}')$ on the central path with minimal Euclidean distance $\|(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) - (\mathbf{x}', \mathbf{y}')\|$, while $\beta = 0$ would yield a Newton step aiming at optimality of (16) regardless of the path of centers. We choose the value for β at the beginning of each Newton step depending on the momentary value of $f_{cen}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$. If $f_{cen}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ is large, we are close to the boundary and we make a centralizing step towards the path of centers using a large value of $\beta \in (0, 1]$. If $f_{cen}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ is small, we are already fairly central and we can venture a minimizing step with a small $\beta \in (0, 1]$.

4.3. A feasible starting point

For the IPM to work, we need a feasible starting point $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}) \in S_{++}$. In other words we need to find $\mathbf{x}^{(0)} \in \operatorname{int}\mathcal{C}$ such that $F(\mathbf{x}^{(0)}) \in \operatorname{int}\mathcal{C}$. This is not a trivial task. The goal of this section is to generalize an approach given in (Kojima et al., 1991) to construct a feasible starting point to a similar CCP. The idea is to introduce an artificial variable and transform the original complementarity problem of size $3n$ into a complementarity problem of size $3n + 1$ with an obvious starting point.

We introduce the additional variable $s \in \mathbb{R}_+$ and a vector $\mathbf{d} \in \mathbb{R}^{3n}$ and consider the complementarity problem

$$\begin{aligned} \tilde{\mathcal{C}} \ni \tilde{\mathbf{x}} &= \begin{bmatrix} \mathbf{x} \\ s \end{bmatrix}, \\ \tilde{\mathcal{C}} \ni \tilde{\mathbf{y}} &= \begin{bmatrix} \bar{\mathbf{y}} \\ 1 \end{bmatrix} \\ &= \tilde{F}(\tilde{\mathbf{x}}) = \begin{bmatrix} F(\mathbf{x}) + s \cdot \mathbf{d} \\ 1 \end{bmatrix}, \qquad (21) \\ 0 &= \tilde{\mathbf{x}}^T\tilde{\mathbf{y}}, \end{aligned}$$

where $\tilde{\mathcal{C}} = \mathcal{C} \times \mathbb{R}_+$. Note that $\mathbb{R}_+$ is the cone of squares in the Jordan algebra $(\mathbb{R}, \cdot)$, where $\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the standard multiplication of two scalars. $(\mathbb{R}, \cdot)$ has the same structure as $(\mathbb{R}^3, \circ)$ and all results from Section 3 apply analogously. The central path of the new CCP is the zero set of

$$f_{cen}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = 2(n+1) \log \frac{\left(\mathbf{x}^T\bar{\mathbf{y}} + s\right)/(n+1)}{\prod_{i=1}^{n}\left[2s\sqrt{\det(\mathbf{x}_i)\det(\bar{\mathbf{y}}_i)}\right]^{1/n}}.$$

Assume, an initial feasible guess $\mathbf{x}^{(0)}$ for the original problem is given. For example, $\mathbf{x}_i^{(0)} = (1, 0, 0)^T$ obviously

lies in int C for all $i = 1, ..., n$. We now choose $s^{(0)}$ and $\mathbf{d}$, such that

$$\bar{\mathbf{y}}^{(0)} = \tilde{\mathbf{y}}^{(0)}_{\{1,...,3n\}} = F(\mathbf{x}^{(0)}) + s^{(0)}\mathbf{d} \in \operatorname{int} \mathcal{C}.$$

We can even choose these parameters in such a way that $(\tilde{\mathbf{x}}^{(0)}, \tilde{\mathbf{y}}^{(0)})$ lies on a point $\alpha > 0$ of our choice on the central path

$$\tilde{\mathbf{x}}^{(0)} \circ \tilde{\mathbf{y}}^{(0)} = \alpha \mathbf{e}$$

of the CCP (21). For all $i = 1, ..., n$ it must hold

$$\begin{bmatrix} \sqrt{2}\alpha \\ \mathbf{0} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} x^{(0)}_{in}\bar{y}^{(0)}_{in} + \mathbf{x}^{(0)T}_{it}\bar{\mathbf{y}}^{(0)}_{it} \\ x^{(0)}_{in}\bar{\mathbf{y}}^{(0)}_{it} + \bar{y}^{(0)}_{in}\mathbf{x}^{(0)}_{it} \end{bmatrix}$$

$$\Rightarrow \qquad \bar{y}^{(0)}_{in} = \frac{2\alpha}{x^{(0)}_{in} - \frac{\|\mathbf{x}^{(0)}_{it}\|^2}{x^{(0)}_{in}}}$$

$$\text{and} \quad \bar{\mathbf{y}}^{(0)}_{it} = -\left(\frac{\bar{y}^{(0)}_{in}}{x^{(0)}_{in}}\right)\mathbf{x}^{(0)}_{it}.$$

For the last complementarity condition we have to make sure that

$$s^{(0)} \cdot 1 = s^{(0)} = \mathbf{x}^{(0)T}_i \bar{\mathbf{y}}^{(0)}_i = 2\alpha.$$

This in turn means that

$$\mathbf{d} = \frac{1}{2\alpha}\left(\bar{\mathbf{y}}^{(0)} - F(\mathbf{x}^{(0)})\right).$$

Note that in theory, we could choose an initial guess arbitrarily close the the optimal solution of (21) by choosing a small α. But a small α implies a large $\mathbf{d}$, so that the system is badly scaled if $s^{(0)} = 2\alpha$ is too small.

A search direction for (21) is given by

$$\begin{bmatrix} L(\bar{\mathbf{y}}^{(k)}) & L(\mathbf{x}^{(k)}) & \mathbf{0} \\ \nabla F(\mathbf{x}^{(k)}) & -I & \mathbf{d} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}^{(k)} \\ \Delta\bar{\mathbf{y}}^{(k)} \\ \Delta s^{(k)} \end{bmatrix} = \begin{bmatrix} \alpha\mathbf{e} - L(\mathbf{x}^{(k)})\bar{\mathbf{y}}^{(k)} \\ \mathbf{0} \\ 2\alpha - s^{(k)} \end{bmatrix} \tag{22}$$

or equivalently by

$$\begin{aligned} \Delta s^{(k)} &= 2\alpha - s^{(k)}, \\ \tilde{A}\Delta\mathbf{x}^{(k)} &= \mathbf{b}, \\ \Delta\bar{\mathbf{y}}^{(k)} &= \nabla F(\mathbf{x}^{(k)})\Delta\mathbf{x}^{(k)} + \Delta s^{(k)}\mathbf{d}, \end{aligned} \tag{23}$$

where

$$\tilde{A} = \left[L(\mathbf{x}^{(k)})^{-1}L(\bar{\mathbf{y}}^{(k)}) + \nabla F(\mathbf{x}^{(k)})\right]$$

$$\mathbf{b} = \alpha\left(\mathbf{x}^{(k)}\right)^{-1} - \bar{\mathbf{y}}^{(k)} - \Delta s^{(k)}\mathbf{d}.$$

Since an optimal solution $(\tilde{\mathbf{x}}^*, \tilde{\mathbf{y}}^*)$ implies that $s^* = 0$, it follows that $(\mathbf{x}^*, \bar{\mathbf{y}}^*) = (\mathbf{x}^*, \mathbf{y}^*)$ is optimal for the original CCP. If $s^{(k)} \not\to 0$, the original CCP (15) is infeasible. During the iteration, $s^{(k)}$ decreases and eventually $\mathbf{y}^{(k)} = \bar{\mathbf{y}}^{(k)} - s^{(k)}\mathbf{d}$ is feasible. Then we can use the current iterate $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ as a starting point for a feasible Interior Point algorithm for the original CCP (15). In other words, we continue with the iteration, but replace $\bar{\mathbf{y}}^{(k)}$ with $\mathbf{y}^{(k)}$ and set $s^{(l)} = 0$ and $\Delta s^{(l)} = 0$ for all future iterations $l > k$.

4.4. Maximum Step Length

Assume without loss of generality that $(\mathbf{x}, \bar{\mathbf{y}}) \in S_{++}$ is feasible. When an inexact iterative solver is used to determine the Newton direction some care has to be taken not to leave the set S_{++}. Hence, the next task in the derivation of the Interior Point Method is to identify the supremum of step sizes $\theta > 0$, such that

$$\begin{aligned} &\mathbf{x} + \theta\Delta\mathbf{x} \in \operatorname{int} C \text{ and} \\ &\bar{\mathbf{y}} + \theta\Delta\bar{\mathbf{y}} \in \operatorname{int} C. \end{aligned}$$

Lemma 2. *Let $\mathbf{x} \in int\, C$ and $\Delta\mathbf{x} \in \mathbb{R}^3$. Then*

$$\mathbf{x} + \theta\Delta\mathbf{x} \in int\, C \quad \Leftrightarrow \quad \begin{cases} \Delta\mathbf{x} \in C \\ or \quad \theta < \theta_{\max} \end{cases}$$

where

$$\theta_{\max} := \frac{\det(\mathbf{x})}{\sqrt{\left[\frac{(\Delta\mathbf{x})^T J\mathbf{x}}{2}\right]^2 - \det(\Delta\mathbf{x})\det(\mathbf{x})} - \frac{(\Delta\mathbf{x})^T J\mathbf{x}}{2}}$$

Proof.
Let $\mathbf{v} = P(\mathbf{x}^{-\frac{1}{2}})\Delta\mathbf{x}$.

- It holds $P(x^{\frac{1}{2}})\mathbf{e} = \mathbf{x}$ and thus

$$\begin{aligned} \mathbf{x} + \theta\Delta\mathbf{x} \in \operatorname{int} C \quad &\Leftrightarrow \quad P(\mathbf{x}^{\frac{1}{2}})(\mathbf{e} + \theta\mathbf{v}) \in \operatorname{int} C \\ &\Leftrightarrow \quad \mathbf{e} + \theta\mathbf{v} \in \operatorname{int} C \\ &\Leftrightarrow \quad \left\{ \begin{array}{l} \lambda_1(\mathbf{e} + \theta\mathbf{v}) > 0 \\ \lambda_2(\mathbf{e} + \theta\mathbf{v}) > 0 \end{array} \right\} \\ &\Leftrightarrow \quad \left\{ \begin{array}{l} 1 + \theta\lambda_1(\mathbf{v}) > 0 \\ 1 + \theta\lambda_2(\mathbf{v}) > 0 \end{array} \right\}, \end{aligned}$$

  where the last equivalence holds because of

$$\begin{aligned} \lambda_i(\mathbf{e} + \theta\mathbf{v}) &= \frac{1}{\sqrt{2}}\left(\sqrt{2} + v_n \mp \|\theta\mathbf{v}_t\|\right) \\ &= 1 + \theta\lambda_i(\mathbf{v}), \quad i = 1, 2. \end{aligned}$$

  If both spectral values of $\mathbf{v}$ are non–negative, i.e. if $\mathbf{v} \in \mathbf{C} \Leftrightarrow \Delta\mathbf{x} \in C$, there is no restriction on θ. If $\lambda_1 = \min(\lambda_1(\mathbf{v}), \lambda_2(\mathbf{v}))$ is smaller than zero, we have

$$\theta < -\frac{1}{\lambda_1}. \tag{24}$$

- Using Section 3(i) we know that

$$\begin{aligned}\lambda_1 =& \frac{(\Delta\mathbf{x})^T\mathbf{x}^{-1}}{2} \\ &- \sqrt{\left[\frac{(\Delta\mathbf{x})^T\mathbf{x}^{-1}}{2}\right]^2 - \det(\mathbf{x}^{-1})\det(\Delta\mathbf{x})} \\ =& \frac{(\Delta\mathbf{x})^T J\mathbf{x}}{2\det(\mathbf{x})} - \sqrt{\left[\frac{(\Delta\mathbf{x})^T J\mathbf{x}}{2\det(\mathbf{x})}\right]^2 - \frac{\det(\Delta\mathbf{x})}{\det(\mathbf{x})}} \\ =& \frac{\frac{(\Delta\mathbf{x})^T J\mathbf{x}}{2} - \sqrt{\left[\frac{(\Delta\mathbf{x})^T J\mathbf{x}}{2}\right]^2 - \det(\mathbf{x})\det(\Delta\mathbf{x})}}{\det(\mathbf{x})}.\end{aligned} \tag{25}$$

  Observe that there are no imaginary solutions. If $\det(\Delta\mathbf{x}) < 0$, the term in the square root is positive. If $\det(\Delta\mathbf{x}) \geq 0$, the positivity of the term in the square root follows from Lemma 1.

- Finally we only need to check when λ_1 is non–negative. We distinguish two cases.

  (a) $(\Delta\mathbf{x})^T\mathbf{x}^{-1} < 0$.
  We can easily see from (25), that λ_1 is always negative.

  (b) $(\Delta\mathbf{x})^T\mathbf{x}^{-1} \geq 0$.
  From (25) we see that λ_1 is negative if $\det(\Delta\mathbf{x}) < 0$. With Lemma 1 follows that λ_1 is non–negative, if $\det(\Delta\mathbf{x})$ is non–negative. Because of $\mathbf{x}^{-1} \in C \;\Leftrightarrow\; \mathbf{x} \in C$ we have $(\Delta\mathbf{x})^T\mathbf{x}^{-1} \geq 0 \;\Leftrightarrow\; (\Delta\mathbf{x})^T\mathbf{x} \geq 0$ and therefore

$$\begin{aligned}\lambda_1 \geq 0 \quad &\Leftrightarrow \quad \left\{\begin{array}{c}(\Delta\mathbf{x})^T\mathbf{x} \geq 0 \\ \det(\Delta\mathbf{x}) \geq 0\end{array}\right\} \\ &\Leftrightarrow \quad \Delta\mathbf{x} \in C.\end{aligned} \tag{26}$$

  The upper bound for θ in the case $\Delta\mathbf{x} \notin C$ follows by inserting (25) into (24).

□

4.5. Nesterov–Todd–Scaling

As the dot–product of $\mathbf{x}_i$ and $\bar{\mathbf{y}}_i$ decreases, at least one of the two vectors approaches the boundary of its feasible set C. But the search directions $\Delta\mathbf{x}_i$ and $\Delta\bar{\mathbf{y}}_i$ cannot be chosen independently, and closeness to the boundary restricts the choice in feasible step sizes. The convergence of the Interior Point Method can be improved substantially by rescaling the space in which the cone $\mathcal{C}$ lives at the beginning of each iteration. This is done using the Nesterov–Todd scaling scheme, or NT–Scaling, see (Nesterov and Todd, 1997; Bai et al., 2004). Loosely speaking, the idea is to rescale the vectors $\mathbf{x}_i \mapsto \hat{\mathbf{x}}_i$ and $\bar{\mathbf{y}}_i \mapsto \hat{\mathbf{y}}_i$ such that $\hat{\mathbf{x}}_i = \hat{\mathbf{y}}_i$, and neither vector is closer to the boundary than the other.

Consider the block–diagonal automorphism $\hat{W} = \left[\begin{smallmatrix} W \\ & w \end{smallmatrix}\right]$ introduced in Section 3(j) satisfying $\hat{W}\tilde{\mathbf{x}}^{(k)} = \hat{W}^{-1}\tilde{\mathbf{y}}^{(k)}$. Define $\tilde{\mathbf{v}} = \left[\begin{smallmatrix}\mathbf{v} \\ v_s\end{smallmatrix}\right] = \frac{1}{\sqrt{\alpha}}\hat{W}\tilde{\mathbf{x}}^{(k)}$, i.e.

$$\begin{aligned}\mathbf{v} &= \frac{1}{\sqrt{\alpha}}W\mathbf{x}^{(k)} = \frac{1}{\sqrt{\alpha}}W^{-1}\bar{\mathbf{y}}^{(k)}, \\ v_s &= \frac{1}{\sqrt{\alpha}}ws^{(k)} = \frac{1}{\sqrt{\alpha}}w^{-1}\end{aligned}$$

with $w = \frac{1}{\sqrt{s^{(k)}}}$. Using

$$\begin{aligned}\mathbf{d}_x &:= \frac{1}{\sqrt{\alpha}}W\Delta\mathbf{x}^{(k)}, \\ \mathbf{d}_y &:= \frac{1}{\sqrt{\alpha}}W^{-1}\Delta\bar{\mathbf{y}}^{(k)}, \\ d_s &:= \frac{1}{\sqrt{\alpha}}\frac{1}{\sqrt{s^{(k)}}}\Delta s^{(k)},\end{aligned}$$

we can write (22) as

$$\begin{aligned}&\begin{bmatrix} L(W\mathbf{v})W^{-1} & L(W^{-1}\mathbf{v})W & \mathbf{0} \\ \nabla F(\mathbf{x}^{(k)})W^{-1} & -W & \sqrt{s^{(k)}}\mathbf{d} \\ \mathbf{0} & \mathbf{0} & v_s \end{bmatrix}\begin{bmatrix}\mathbf{d}_x \\ \mathbf{d}_y \\ d_s\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{e} - L(W^{-1}\mathbf{v})W\mathbf{v} \\ \mathbf{0} \\ 2 - v_s^2\end{bmatrix}.\end{aligned} \tag{27}$$

Next, we replace $L(W\mathbf{v})W^{-1}$ and $L(W^{-1}\mathbf{v})W$ by $L(\mathbf{v})$, premultiply the first equation by $L(\mathbf{v})^{-1} = L(\mathbf{v}^{-1})$ and the last equation by v_s^{-1}. This gives the system of equations

$$\begin{aligned}&\begin{bmatrix} I & I & \mathbf{0} \\ \nabla F(\mathbf{x}^{(k)})W^{-1} & -W & \sqrt{s^{(k)}}\mathbf{d} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}\begin{bmatrix}\mathbf{d}_x \\ \mathbf{d}_y \\ d_s\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v}^{-1} - \mathbf{v} \\ \mathbf{0} \\ 2v_s^{-1} - v_s\end{bmatrix}\end{aligned} \tag{28}$$

or equivalently

$$\begin{aligned}\Delta s^{(k)} &= 2\alpha - s^{(k)}, \\ A\Delta\mathbf{x}^{(k)} &= \mathbf{b}, \\ \Delta\bar{\mathbf{y}}^{(k)} &= \nabla F(\mathbf{x}^{(k)})\Delta\mathbf{x}^{(k)} + \Delta s^{(k)}\mathbf{d}.\end{aligned} \tag{29}$$

Here,

$$\begin{aligned}A &= \left[W^2 + \nabla F(\mathbf{x}^{(k)})\right] \\ &= \left[P(\mathbf{w}) + \nabla F(\mathbf{x}^{(k)})\right] \\ \mathbf{b} &= \sqrt{\alpha}W\mathbf{v}^{-1} - \bar{\mathbf{y}}^{(k)} - \Delta s^{(k)}\mathbf{d} \\ &= \sqrt{\alpha}\left(P(\mathbf{w}^{-\frac{1}{2}})\mathbf{v}\right)^{-1} - \bar{\mathbf{y}}^{(k)} - \Delta s^{(k)}\mathbf{d} \\ &= \sqrt{\alpha}\left(W^{-1}\mathbf{v}\right)^{-1} - \bar{\mathbf{y}}^{(k)} - \Delta s^{(k)}\mathbf{d} \\ &= \alpha\left(\mathbf{x}^{(k)}\right)^{-1} - \bar{\mathbf{y}}^{(k)} - \Delta s^{(k)}\mathbf{d},\end{aligned}$$

and $\mathbf{w} \in \operatorname{int} C$ is the scaling point for $\mathbf{x}^{(k)}$ and $\bar{\mathbf{y}}^{(k)}$.

The linear system (29) is an approximation of the linear system (23). By successively solving (29) and updating W accordingly in an inner iteration, $\mathbf{v}$ eventually satisfies $\mathbf{v}^{-1} = \mathbf{v}$ which is equivalent to $\mathbf{x} \circ \bar{\mathbf{y}} = \alpha \mathbf{e}$. Then the centralizing parameter α is updated and the process is repeated until the complementarity problem is solved (Bai et al., 2004). One could alternatively update α after each inner iteration to find a new search direction. For instance, centralizing steps that do not improve the value of the cost function must not be solved with a very high accuracy and one inner iteration might suffice to get closer to the central path. Hence, there must not be a strict distinction between inner and outer iterations.

Apart from improving the convergence rate of the method, the approximation (29) yields a symmetric linear system, if $\nabla F(\mathbf{x}^{(k)})$ is symmetric. This is not the case for (23), since $L(\mathbf{x}^{(k)})^{-1}$ and $L(\bar{\mathbf{y}}^{(k)})$ do not necessarily commute. One easily checks[4] that

$$\begin{aligned}\nabla F(\mathbf{x}^{(k)}) &= \mathcal{T}_y \nabla \bar{F}(\mathcal{T}_x^{-1}\mathbf{x}^{(k)})\mathcal{T}_x^{-1} \\ &\in \mathcal{T}_y \left(\bar{N} + \partial\tilde{\mathbf{r}}(\mathcal{T}_x^{-1}\mathbf{x}^{(k)})\right)\mathcal{T}_x^{-1}\end{aligned}$$

is only symmetric if the term $\tilde{\mathbf{r}}(\boldsymbol{\lambda})$ is ignored (see our comments at the end of Section 2.4) and if the friction coefficients satisfy $\mu_i = \mu_j$ for all $i, j \in \{1, ..., n\}$. For $\mu_i \neq \mu_j$ we can symmetrize the system via the transformation

$$\begin{aligned}A &\leftarrow \mathcal{T}_y^{-1} P(\mathbf{w}) T_x + \bar{N}, \\ \mathbf{b} &\leftarrow \mathcal{T}_y^{-1}\mathbf{b}, \\ A\Delta\boldsymbol{\lambda}^{(k)} &= \mathbf{b}, \\ \Delta\mathbf{x}^{(k)} &= \mathcal{T}_x \Delta\boldsymbol{\lambda}^{(k)}\end{aligned}$$

because of the block–diagonality of $P(\mathbf{w})$.

We conclude the section on NT–scaling with an explicit formula for the scaling point of a contact.

Lemma 3. *The scaling point* $\mathbf{w} \in \mathbb{R}^3$ *satisfying* $P(\mathbf{w}^{\frac{1}{2}})\mathbf{x} = P(\mathbf{w}^{-\frac{1}{2}})\mathbf{y}$ *for* $\mathbf{x}, \mathbf{y} \in int\, C$ *is*

$$\mathbf{w} = \frac{\mathbf{y} + \lambda J\mathbf{x}}{\sqrt{\mathbf{x}^T\mathbf{y} + 2\sqrt{\det(\mathbf{x})\det(\mathbf{y})}}},$$

where

$$\lambda := \det(\mathbf{w}) = \sqrt{\frac{\det(\mathbf{y})}{\det(\mathbf{x})}}.$$

Proof. This proof is based on the one from (Bai et al., 2004). We know from Section 3(j), that

$$\mathbf{w} = P(\mathbf{x}^{-\frac{1}{2}})\left(P(\mathbf{x}^{\frac{1}{2}})\mathbf{y}\right)^{\frac{1}{2}}.$$

With $\mathbf{u} := \mathbf{x}^{-1}$ and $\mathbf{v} := \left(P(\mathbf{x}^{\frac{1}{2}})\mathbf{y}\right)^{\frac{1}{2}}$ we can use Section 3(i) to get

$$\det(\mathbf{w}) = \det(\mathbf{u})\det(\mathbf{v}) = \frac{\det(\mathbf{v})}{\det(\mathbf{x})}.$$

Using Section 3(i) once more on $\mathbf{v}^2 = P(\mathbf{x}^{\frac{1}{2}})\mathbf{y}$ we obtain

$$\det(\mathbf{v})^2 = \det(\mathbf{v}^2) = \det(\mathbf{x})\det(\mathbf{y})$$

and thus

$$\lambda = \det(\mathbf{w}) = \sqrt{\frac{\det(\mathbf{y})}{\det(\mathbf{x})}}.$$

With $P(\mathbf{w}) = \mathbf{w}\mathbf{w}^T - \det(\mathbf{w})J$ we can write

$$\begin{aligned}&\mathbf{y} = P(\mathbf{w})\mathbf{x} = \mathbf{w}\mathbf{w}^T\mathbf{x} - \lambda J\mathbf{x} \\ \Rightarrow\quad &\mathbf{w} = \frac{1}{\mathbf{w}^T\mathbf{x}}\left(\mathbf{y} + \lambda J\mathbf{x}\right). \qquad (30)\end{aligned}$$

Taking the inner product of (30) with $J\mathbf{w}$ gives

$$\begin{aligned}2\lambda &= 2\det(\mathbf{w}) = \mathbf{w}^T J\mathbf{w} \\ &= \frac{1}{(\mathbf{w}^T\mathbf{x})^2}\left(\mathbf{y}^T J\mathbf{y} + 2\lambda\mathbf{x}^T JJ\mathbf{y} + \lambda^2\mathbf{x}^T JJJ\mathbf{x}\right).\end{aligned}$$

Making use of the fact that $JJ = I$ results in

$$\begin{aligned}\left(\mathbf{w}^T\mathbf{x}\right)^2 &= \frac{1}{\lambda}\det(\mathbf{y}) + \mathbf{x}^T\mathbf{y} + \lambda\det(\mathbf{x}). \\ &= \mathbf{x}^T\mathbf{y} + 2\sqrt{\det(\mathbf{x})\det(\mathbf{y})}. \qquad (31)\end{aligned}$$

Inserting (31) into (30) yields the desired result. □

4.6. Inexact Search Directions

Usually, in the context of soil simulations, the number of unknowns in every time step is quite high. Depending on the configuration, every soil particle is in contact with approximately 5 to 10 other particles and every contact has three unknowns – one for the normal reaction impulse and two for the tangential reaction impulses. In a simulation with one million particles one can expect to have around 15 to 30 million unknowns in every time step and a direct solver for the search direction is not an option.

Instead of directly solving the Newton step, we want to apply an iterative method to obtain inexact search directions in each IPM iteration. When using iterative methods such as Krylov subspace schemes, the convergence rate depends directly on the condition of the matrix A. Unfortunately, one intrinsic property of IPMs is that the condition of the system matrix A deteriorates close to the optimal solution. An additional difficulty is that $\nabla F(\mathbf{x})$ is highly rank–deficient in almost all cases and the CCP does not have a unique solution. Figure 6 shows a simple, static, two–dimensional example with merely ten resting disks subject to frictionless contact. The associated Linear Complementarity Problem allows infinitely many solutions for the reaction impulses, of which all lead to the same zero–velocity state.

[4] Simply verify, that for any symmetric $X \in \mathbb{R}^{3\times3}$ the matrix $T^y_{\mu_i} X (T^x_{\mu_j})^{-1}$ is equal to $T^y_{\mu_j} X (T^x_{\mu_i})^{-1}$ only if $\mu_i = \mu_j$.

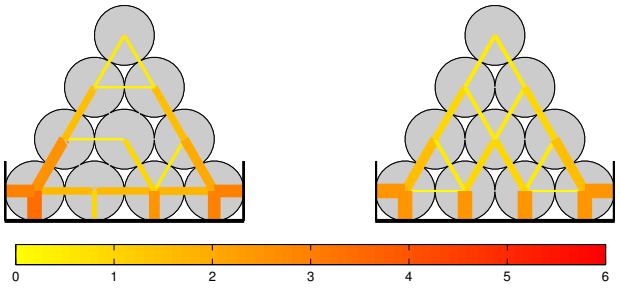


Figure 6: Two different valid impulse solutions to a simple example of a particle configuration allowing infinitely many solutions. The magnitude of the reaction impulses is visualized by the width and color of the lines connecting the two associated bodies.

Almost the entire computational effort of the Interior Point Method is spent on the solution of the Newton step. Therefore, some effort has be put into the regularization and preconditioning of the system. We follow the approach of (Gondzio, 2012) of regularizing the system first and then preconditioning it.

4.6.1. Regularization

One way of regularizing a rank–deficient linear system is by adding a diagonal matrix R with small values to it,

$$A \leftarrow A + R.$$

$R = \mathrm{diag}(\rho_1, ..., \rho_{3n})$ is called the Tikhonov matrix and solving $A\Delta\mathbf{x} = \mathbf{b}$ is equivalent to solving the minimization problem

$$\min_{\Delta\mathbf{x}} \|A\Delta\mathbf{x} - \mathbf{b}\|^2 + \|R\Delta\mathbf{x}\|^2.$$

The condition of the regularized system improves with the size of the regularization parameters ρ_i, but larger ρ_i give poorer approximations of the original problem.

Fortunately, in NSCD, the regularization parameters have a direct physical interpretation: They can be identified with finite normal and tangential stiffnesses in the contacts. Loosely speaking, regularization re–introduces smoothness into the non–smooth formulation without suffering from instabilities as they occur in the classical DEM, see (Lacoursire, 2007; Tasora et al., 2013). Instead, re–introducing finite stiffnesses improves the numerical properties of the linear systems. As perfectly rigid bodies are always just an approximation of the real world, the regularized system is in fact closer to reality than the original one, as long as the regularization parameters are sufficiently small. We follow the idea of (Tasora et al., 2013) and set $R = \mathcal{T}_y \bar{R} \mathcal{T}_x^{-1}$ and

$$\bar{R} = \mathrm{diag}\left([\rho_{1,\mathbf{n}}, \rho_{1,\mathbf{t}}, \rho_{1,\mathbf{t}}, ..., \rho_{n,\mathbf{n}}, \rho_{n,\mathbf{t}}, \rho_{n,\mathbf{t}}]\right). \tag{32}$$

The regularization parameters

$$\rho_{i,\mathbf{n}} = \frac{1}{\Delta t^2 k_{i,\mathbf{n}}} \quad \text{and} \quad \rho_{i,\mathbf{t}} = \frac{1}{\Delta t^2 k_{i,\mathbf{t}}}$$

are calculated from normal and tangential stiffnesses $k_{i,\mathbf{n}}$ and $k_{i,\mathbf{t}}$ respectively for all contacts $i = 1, ..., n$. The diagonal matrices $\mathcal{T}_y$ and $\mathcal{T}_x^{-1}$ are the transformations defined in the beginning of Section 4.1.

If the stiffness values are chosen carefully, regularization has more desirable properties. The authors of (Obermayr et al., 2011) calculate stiffnesses from the deformation of an elastic rod connecting the centers of mass the particles. If the contact with index i is associated with the bodies A and B, the stiffnesses are given by

$$k_{i,\mathbf{n}} = \frac{\pi}{4}\hat{E}(r_A + r_B) \quad \text{and} \quad \frac{k_{i,\mathbf{n}}}{k_{i,\mathbf{t}}} = \frac{(2-\nu)(1+\nu)}{2(1-\nu^2)} \tag{33}$$

where $\hat{E}$ is the Young's modulus of the material, ν the Poisson ratio, r_A and r_B are the radii of body A and B respectively. The stiffnesses in (33) yield a scale invariant model with respect to the particle size.

4.6.2. Eigenvalues of the linear systems

The matrix of the linear system that has to be solved in every time step is

$$A = W^2 + \nabla F(\mathbf{x}^{(k)}) + R$$

where R is a suitable regularization matrix, $W = P(\mathbf{w}^{\frac{1}{2}})$ and $\mathbf{w}$ is the scaling point for $\mathbf{x}^{(k)}$ and $\bar{\mathbf{y}}^{(k)}$.

Lemma 4. *Let $\mathbf{x} \circ \mathbf{y} = \alpha\mathbf{e}$ be a point on the central path for $\alpha \in (0, 1]$ and let $\mathbf{w}$ be the scaling point for $\mathbf{x}$ and $\mathbf{y}$. Then the eigenvalues ω_i, $i = 1, ..., 3n$ of the automorphism*

$$W = P(\mathbf{w}^{\frac{1}{2}}) \in \mathbb{R}^{3n\times 3n}$$

satisfy

$$\min \omega_i = \mathcal{O}(\sqrt{\alpha}) \quad \textit{and} \quad \max \omega_i = \mathcal{O}(\frac{1}{\sqrt{\alpha}}).$$

Proof.

Recall that $W = P(\mathbf{w}^{\frac{1}{2}}) \in \mathbb{R}^{3n\times 3n}$ is block diagonal with blocks $W_i = P(\mathbf{w}_i^{\frac{1}{2}}) \in \mathbb{R}^{3\times 3}$ associated to each contact.

We know from Section 3(h) that the eigenvalues of W_i are

$$\begin{aligned}
\omega_{i1}\left(P(\mathbf{w}_i^{\frac{1}{2}})\right) &= \lambda_1(\mathbf{w}_i^{\frac{1}{2}})^2 = \lambda_1(\mathbf{w}_i),\\
\omega_{i2}\left(P(\mathbf{w}_i^{\frac{1}{2}})\right) &= \lambda_2(\mathbf{w}_i^{\frac{1}{2}})^2 = \lambda_2(\mathbf{w}_i),\\
\omega_{i3}\left(P(\mathbf{w}_i^{\frac{1}{2}})\right) &= \det(\mathbf{w}_i^{\frac{1}{2}}) = \sqrt{\det(\mathbf{w}_i)}.
\end{aligned}$$

By inserting the explicit formula from Lemma 3 we obtain

$$\begin{aligned}
\omega_{i1} &= \frac{y_{i\mathbf{n}} + \lambda x_{i\mathbf{n}} - \|\mathbf{y}_{i\mathbf{t}} - \lambda\mathbf{x}_{i\mathbf{t}}\|}{\sqrt{2}\sqrt{\mathbf{x}_i^T\mathbf{y}_i + 2\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)}}},\\
\omega_{i2} &= \frac{y_{i\mathbf{n}} + \lambda x_{i\mathbf{n}} + \|\mathbf{y}_{i\mathbf{t}} - \lambda\mathbf{x}_{i\mathbf{t}}\|}{\sqrt{2}\sqrt{\mathbf{x}_i^T\mathbf{y}_i + 2\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)}}},\\
\omega_{i3} &= \sqrt{\lambda},
\end{aligned}$$

where $\lambda = \sqrt{\frac{\det(\mathbf{y}_i)}{\det(\mathbf{x}_i)}}$.

Because of $\mathbf{x}_i \circ \mathbf{y}_i = \alpha \mathbf{e}$ we know

- $\mathbf{x}_i = \alpha \mathbf{y}_i^{-1} = \frac{\alpha}{\det(\mathbf{y}_i)} J\mathbf{y}_i$,
- $2\alpha = \mathbf{x}_i^T \mathbf{y}_i$ and
- $2\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)} = \mathbf{x}_i^T \mathbf{y}_i$ *(with Lemma 1).*

Rewriting the equations for $\omega_{i1}, \omega_{i2}, \omega_{i3}$ using these relations yields

$$\begin{aligned}
\omega_{i,1/2} &= \frac{y_{i\mathbf{n}} + \lambda \frac{\alpha}{\det(\mathbf{y}_i)} y_{i\mathbf{n}} \mp \|\mathbf{y}_{i\mathbf{t}} + \lambda \frac{\alpha}{\det(\mathbf{y}_i)} \mathbf{y}_{it}\|}{\sqrt{2}\sqrt{2\alpha + 2\alpha}} \\
&= \left(1 + \alpha \sqrt{\frac{\det(\mathbf{y}_i)}{\det(\mathbf{x}_i)\det(\mathbf{y}_i)^2}}\right) \frac{y_{i\mathbf{n}} \mp \|\mathbf{y}_{i\mathbf{t}}\|}{2\sqrt{2}\sqrt{\alpha}} \\
&= (1 + \frac{\alpha}{\alpha}) \frac{y_{i\mathbf{n}} \mp \|\mathbf{y}_{i\mathbf{t}}\|}{2\sqrt{2}\sqrt{\alpha}} \\
&= \frac{1}{\sqrt{\alpha}} \frac{y_{i\mathbf{n}} \mp \|\mathbf{y}_{i\mathbf{t}}\|}{\sqrt{2}}, \\
\omega_{i3} &= \sqrt{\lambda} = \sqrt{\sqrt{\frac{\det(\mathbf{y}_i)}{\det(\mathbf{x}_i)}}} \\
&= \sqrt{\det(\mathbf{y}_i) \frac{1}{\sqrt{\det(\mathbf{x}_i)\det(\mathbf{y}_i)}}} \\
&= \sqrt{\frac{\det(\mathbf{y}_i)}{\alpha}}.
\end{aligned}$$

Analogously by using $\mathbf{y}_i = \alpha \mathbf{x}_i^{-1}$ we can calculate

$$\begin{aligned}
\omega_{i,1/2} &= \sqrt{\alpha} \frac{\sqrt{2}}{x_{i\mathbf{n}} \pm \|\mathbf{x}_{i\mathbf{t}}\|}, \\
\omega_{i3} &= \sqrt{\frac{\alpha}{\det(\mathbf{x}_i)}}.
\end{aligned}$$

Taking the minimum and maximum of ω_{i1}, ω_{i2} and ω_{i3} for all $i = 1, ..., n$ yields the result. □

Let the sorted eigenvalues of W be

$$0 < \omega_1 \leq \omega_2 \leq ... \leq \omega_{3n},$$

the sorted eigenvalues of $\nabla F(\mathbf{x}^k) + R$

$$0 < \nu_1 \leq \nu_2 \leq ... \leq \nu_{3n}$$

and the sorted eigenvalues of A

$$0 < a_1 \leq a_2 \leq ... \leq a_{3n}.$$

Using Weyl's inequality we can derive bounds for the eigenvalues of A

$$\begin{aligned}
0 < a_{\min} &:= \omega_1^2 + \nu_1 \leq a_1 \\
\text{and} \quad a_n \leq a_{\max} &:= \omega_n^2 + \nu_n. \qquad (34)
\end{aligned}$$

Assume now that the current iterate $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ is in a neighborhood of a point on the central path with a sufficiently small value $\alpha > 0$. Then

$$a_{\min} = \mathcal{O}(\rho_{\min}) \quad \text{and} \quad a_{\max} = \mathcal{O}(\alpha^{-1})$$

where $\rho_{\min}$ is the smallest regularization parameter. Thus the condition of the system matrix satisfies

$$\kappa_A = \mathcal{O}\left(\frac{1}{\rho_{\min} \cdot \alpha}\right).$$

Note, that without regularization, the lower bound on the eigenvalues of A would be given by $a_{\min} = \alpha$ because ν_1, as the smallest singular value of $\nabla F(\mathbf{x}^k)$, is zero in all interesting cases. Without regularization the condition of A would be $\mathcal{O}(\alpha^{-2})$.

Sufficiently large regularization bounds the smallest eigenvalues away from zero. Our expectation on a preconditioner is that the range of eigenvalues is reduced substantially.

We conclude the description of the Interior Point Method with a pseudo–code for every time step of a simulation in Algorithm 1.

5. Numerical Tests

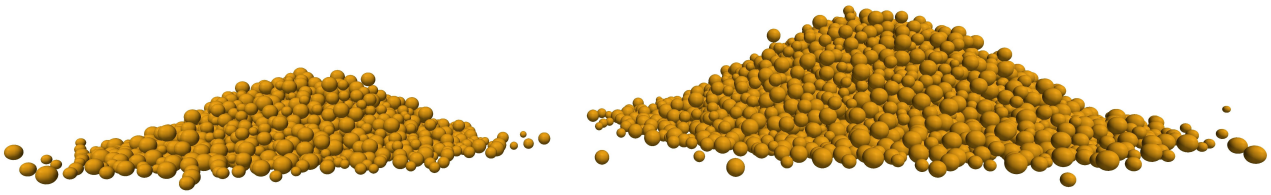
We implement the Interior Point Method in C++, making use of the PETSc library for Krylov solvers (Balay et al., 2013). In this section we compare different preconditioners and iterative linear solvers for the Newton step in each iteration. All system matrices are explicitly constructed in a standard sparse matrix structure and no parallelization is used so far. Once a choice for a solution technique is made, this specific method will be optimized in a future task.

5.1. Test Problems and Solver Setup

Our first benchmark is one time step from a simulation of a pile of 2048 non–rotational spheres, see Figure 7(a). The spheres have radii between 8 mm and 16 mm. With $n = 8378$ potential contacts between pairs of spheres there are 25135 unknowns to the Cone Complementarity Problem (21). It will be called Test Problem 1. Test Problem 2 is a similar pile, with 5040 non–rotational spheres, see Figure 7(b).There are $n = 21052$ potential contacts and thus 63157 unknowns to (21). The time step size is $\Delta t = 0.01$ s in all test cases.

For the analysis in this paper, we ignore the small term $\check{\mathbf{r}}(\boldsymbol{\lambda})$ from the end of Section 2.4 and we use the same frictional coefficient $\mu = \mu_i = 0.4$ for all contacts $i = 1, ..., n$. Therefore, $\nabla F(\mathbf{x}) = N$ is constant and symmetric. We use the same starting value $\mathbf{x}_i = (0.1, 0, 0)^T \in \operatorname{int} C$ for all contacts. Recall that we have freedom in the choice of the initial value for α. In our experience,

$$\alpha = \sum_{i=1,...,n} \frac{|\mathbf{x}_i^{(0)^T} \mathbf{y}_i^{(0)}|}{2n}$$



(a) Test Problem 1: A pile of 2048 spheres.

(b) Test Problem 2: A pile of 5040 spheres.

Figure 7: Two Test Problems.

is appropriate.

After every IPM iteration, the new value for α is chosen based on the value of f_{cen}. We do this in accordance with (Kojima et al., 1991) by writing

$$\alpha = \beta \frac{\mathbf{x}^T \bar{\mathbf{y}}}{2n}.$$

where

$$\beta = \begin{cases} \beta_{cen}, & \text{if } f_{cen}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \leq \alpha_{cen}, \\ \beta_1, & \text{if } \alpha_{cen} < f_{cen}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \leq \alpha_1, \\ \beta_{bd}, & \text{if } \alpha_1 < f_{cen}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}). \end{cases}$$

In all tests we use the bounds $\alpha_{cen} = 0.1$ and $\alpha_1 = 1$ for the centralizing function f_{cen}. We test an aggressive choice $\beta_{cen} = 0.01$, $\beta_1 = 0.1$ and $\beta_{bd} = 0.5$ and compare it to a more cautious strategy using $\beta_{cen} = 0.1$, $\beta_1 = 0.5$ and $\beta_{bd} = 1$. The aggressive choice follows the spirit of potential reduction algorithms, while the more cautious strategy tries to stay as central as possible, and hence it behaves like a path–following algorithm.

5.2. Test Studies

First, we analyze how regularization and preconditioning effects the condition of the Matrix A. Next, we study different Krylov methods to solve the linear subproblems. We shortly compare the previously stated aggressive choice of β to the more conservative one. Finally, we compare the presented IPM to the Projected Gauß–Jacobi Method and study its convergence behavior for different problem sizes.

5.2.1. Regularization and Preconditioning

In this section we examine numerically how regularization and preconditioning influence the linear systems. Incomplete Cholesky (IC(0)) and Incomplete LU (ILU(0)) factorizations are considered as preconditioners, see (Saad, 2003) for details.

Figure 8 shows approximations for the range of eigenvalues of the linear systems in every IPM iteration applied to Test Problem 1. The system matrices are calculated using Algorithm 1, where the linear systems are directly solved using an LU factorization.

As expected, in the non–regularized and non–preconditioned case, the condition grows quadratically as α tends towards zero. With regularization, the smallest eigenvalue of the regularized matrix $N+R$ stays bounded,

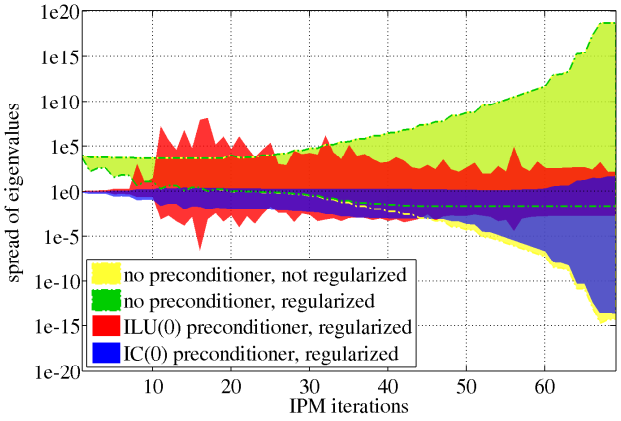


Figure 8: The spread of eigenvalues of the linear systems for Test Problem 1.

while the smallest eigenvalue of $W = P(\mathbf{w})$ approaches zero. Thus, for sufficiently small α, regularization bounds the smallest eigenvalue of $A = W + N + R$ away from zero.

The effect of the preconditioners is surprising, however. IC(0) very effectively reduces the range of eigenvalues in the first few iterations. Yet close to the exact solution, the preconditioned system is not much better than the regularized non–preconditioned system. ILU(0), on the other hand, actually worsens the condition of the systems between the 10th and the 40th iteration. Towards the exact solution, the eigenvalues are nicely clustered around one. In fact, all Krylov methods tested by us fail between the 10th and 16th iteration if the systems are preconditioned with ILU(0) or ILU(1): The preconditioners are indefinite and the inaccuracy of the solution of the linear system causes the maximum step size θ to converge to zero.

This result implies that, if a very high accuracy is required, it might make sense to switch to ILU(0) closer to the optimal solution after using IC(0) as an initial preconditioner. Yet, all our tests fail because ILU(0) is always indefinite after switching the preconditioner. In addition, there is no intuitive criterion to decide when to switch the preconditioners.

In conclusion, IC(0) is the more stable choice even though the decrease in the range of eigenvalues close to optimality is better with ILU(0).

Algorithm 1 time step $t \leftarrow t + \Delta t$

Require: positions $\mathbf{q}(t) \in \mathbb{R}^{6m}$, velocities $\mathbf{v}(t) \in \mathbb{R}^{6m}$, external forces $\mathbf{f}_{\text{ext}}(t) \in \mathbb{R}^{6m}$ and masses $M \in \mathbb{R}^{6m \times 6m}$ of m rigid bodies.

Require: $\alpha > 0$ and tolerances tol_1 and tol_2.

Perform collision detection to obtain contact normals and signed distances of potential contacts.
$feas$ =FALSE.
Construct $\bar{N}$ using (11) and $\bar{\mathbf{r}}$ using (10) for function evaluations of $F = \mathcal{T}_y \bar{F}(\mathcal{T}_x^{-1})$.
Calculate regularization matrix $\bar{R}$ as in (32).
$N = \mathcal{T}_y \bar{N} \mathcal{T}_x^{-1}$.
$R = \mathcal{T}_y \bar{R} \mathcal{T}_x^{-1}$.
$\mathbf{r} = \mathcal{T}_y \bar{\mathbf{r}}$.
Guess a feasible initial $\mathbf{x}^{(0)} = \mathcal{T}_x \boldsymbol{\lambda}^{(0)}$.
Construct $\mathbf{d}$ and $s^{(0)}$ as in Section 4.3.
$\bar{\mathbf{y}}^{(0)} = F(\mathbf{x}^{(0)}) + s^{(0)}\mathbf{d}$.
$k = 0$
while $s^{(k)}\|\mathbf{d}\| > \text{tol}_1$ **and** $\frac{\mathbf{x}^{(k)T}\mathbf{y}^{(k)}}{n} > \text{tol}_2$ **do**
    Set $\nabla F(\mathbf{x}^{(k)}) = N$ **or** evaluate $\nabla F(\mathbf{x}^{(k)})$ using (13).
    Calculate $P(\mathbf{w})$ using 3(h) and Lemma 3.
    if $feas$ =TRUE **then**
        $\mathbf{b} \leftarrow \alpha \left(\mathbf{x}^{(k)}\right)^{-1} - \bar{\mathbf{y}}^{(k)}$.
    else
        $\Delta s^{(k)} = 2\alpha - s^{(k)}$.
        $\mathbf{b} \leftarrow \alpha \left(\mathbf{x}^{(k)}\right)^{-1} - \bar{\mathbf{y}}^{(k)} - \Delta s^{(k)}\mathbf{d}$.
    end if
    $A \leftarrow P(\mathbf{w}) + \nabla F(\mathbf{x}^{(k)}) + R$.
    Calculate preconditioner P for $(A, \mathbf{b})$.
    Solve $P^{-1}A\Delta\mathbf{x} = P^{-1}\mathbf{b}$.
    Calculate stepping length θ with Lemma 2.
    $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \theta\Delta\mathbf{x}$.
    if $feas$ =TRUE **then**
        $\bar{\mathbf{y}}^{(k+1)} = \bar{\mathbf{y}}^{(k)} + \theta\nabla F(\mathbf{x}^{(k)})\Delta\mathbf{x}$.
    else
        $\bar{\mathbf{y}}^{(k+1)} = \bar{\mathbf{y}}^{(k)} + \theta\left(\nabla F(\mathbf{x}^{(k)})\Delta\mathbf{x} + \Delta s^{(k)}\mathbf{d}\right)$.
        $s^{(k+1)} = s^{(k)} + \theta\Delta s^{(k)}$.
    end if
    if $feas$ =FALSE **and** $\bar{\mathbf{y}}^{(k+1)} - s^{(k+1)}\mathbf{d} \in \text{int}\,\mathcal{C}$ **then**
        $\bar{\mathbf{y}}^{(k+1)} \leftarrow \bar{\mathbf{y}}^{(k+1)} - s^{(k+1)}\mathbf{d}$.
        $s^{(k+1)} \leftarrow 0$.
        $feas$ =TRUE.
    end if
    $k \leftarrow k + 1$.
    $\alpha \leftarrow \beta\frac{\mathbf{x}^{(k)T}\bar{\mathbf{y}}^{(k)}}{2n}$ for some $\beta \in (0, 1]$.
end while
$\boldsymbol{\lambda} = \mathcal{T}_x^{-1}\mathbf{x}^{(k)}$.
$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \Delta t M^{-1}\mathbf{f}_{\text{ext}}(t) + M^{-1}D\boldsymbol{\lambda}$.
$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \Delta t\mathbf{v}(t + \Delta t)$.
$t \leftarrow t + \Delta t$.

5.2.2. Krylov Solvers

Next, we test the performance of Krylov solvers applied to the preconditioned system. We use the cautious choice for β, i.e. the path–following strategy, for these tests. We consider the Generalized Minimum Residual Method (GMinres), the Conjugate Gradient (CG) method and the stabilized Bi–Conjugate Gradient (BiCGstab) method. IC(0) will serve as a preconditioner. The maximum number of Krylov iterations is set to 500 and we perform 100 outer IPM iterations. We only stop the IPM before the maximum number of outer iterations is reached, if the step size θ falls below a predefined threshold.

We compare the methods with regard to computing time and the value of the cost function

$$cost(\boldsymbol{\lambda}, \mathbf{u}) = |\boldsymbol{\lambda}^T\mathbf{u}|/n,$$

where $\boldsymbol{\lambda}$ and $\mathbf{u}$ are as defined in Section 2 and n is the number of contacts. Note, that for the optimum we have

$$cost(\boldsymbol{\lambda}^*, \mathbf{u}^*) = 0.$$

In addition, we test the feasibility $\boldsymbol{\lambda} \in \mathcal{K}_\mu$ and $\mathbf{u} \in \mathcal{K}_\mu^*$ using the functions

$$feas(\boldsymbol{\lambda}, \mathbf{u}) = \max_{i=1,\dots,n} \left(\max\left(f_u(\mathbf{u}_i), f_\lambda(\boldsymbol{\lambda}_i)\right)\right)$$

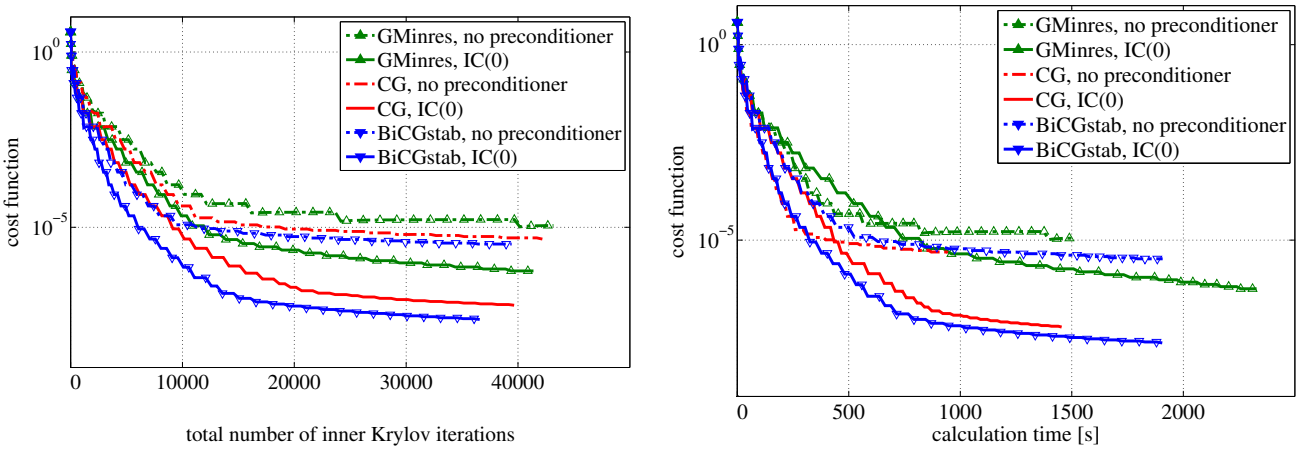
and

$$f_u(\mathbf{u}_i) = -\min\left(0, u_{i\mathbf{n}} - \mu_i\|\mathbf{u}_{i\mathbf{t}}\|\right) \text{ and}$$
$$f_\lambda(\boldsymbol{\lambda}_i) = -\min\left(0, \mu_i\lambda_{i\mathbf{n}} - \|\boldsymbol{\lambda}_{i\mathbf{t}}\|\right).$$

Figure 9 presents convergence plots for Test Problem 2. Figure 9(a) depicts the cost function value over the total number of inner Krylov iterations. While the linear solves from the first few iterations are always cheap, the maximum number of 500 Krylov iterations is reached during all later IPM iterations. In theory, IPMs converge superlinearly. Not surprisingly, using an inexact linear solver reduces the rate of convergence. This has two reasons. Firstly, the condition of the matrices deteriorates close to optimality. Secondly, as the error of the CCP decreases, the accuracy requirements for the linear solvers increase: The residual of the linear problem has to be of some orders of magnitude smaller than the current error of the CCP. Otherwise we cannot expect the error to decrease any further.

We learn from Figure 9(a) that BiCGstab delivers the smallest error in the fewest number of inner iterations. Figure 9(b) shows that, if the accuracy requirements are not too high, preconditioning can be skipped altogether when using CG. This fits to our findings from the previous section that IC(0) is not necessarily better than the not preconditioned regularized system close to optimality.

We observe for all calculations that, initially, the feasibility error decreases with a comparable rate as the cost function. Then, it drops suddenly from around $10^{-1.3}$ to zero by the 19th IPM iteration for all Krylov solvers and the algorithm switches to a feasible Interior Point Method.



(a) cost function value over number of inner Krylov iterations.

(b) cost function value over calculation time.

Figure 9: Comaprison of different Krylov Solvers.

5.2.3. Path–following vs. Potential Reduction

In this section we compare the path–following method ($\beta_{cen} = 0.1$, $\beta_1 = 0.5$, $\beta_{bd} = 1.0$) to the potential reduction technique ($\beta_{cen} = 0.01$, $\beta_1 = 0.1$, $\beta_{bd} = 0.5$) applied to Test Problem 2.

Figure 10 shows the convergence behavior of the two approaches. Figure 10(a) shows that the more aggressive potential reduction scheme reduces the error faster than the path–following scheme, regardless of whether a direct solver or a Krylov method is used. When we use BiCGstab as the linear solver, the step size θ converges to zero quickly and we are forced to stop after around 30 iterations. This is not the case for path–following, but the convergence rate decreases after 50 iterations. Figure 10(a) also shows that, if the linear systems are solved accurately, the error is reduced to machine precision at a constant rate, both for path–following and potential reduction.

Figure 10(b) compares the two techniques in terms of calculation time. Path–following allows larger step sizes θ and the linear solvers close to the central path seem to be better conditioned, so that in total less inner Krylov iterations are needed. Therefore, if the IPM is combined with an inexact linear solver, the more conservative path–following method converges faster than potential reduction.

5.2.4. Comparison to PGJ

As a final test, we compare the IPM to the Projected Gauß–Jacobi solver (PGJ) in terms of accuracy. We use the path–following strategy, IC(0) as a preconditioner and BiCGstab as a solver. The IPM decreases the cost function at about the same rate as it decreases the feasibility error. In our experience, PGJ reduces the cost function much faster than the feasibility error. To compare the methods more easily, we examine them in terms of the error

$$error(\boldsymbol{\lambda}, \mathbf{u}) = \max\left(cost(\boldsymbol{\lambda}, \mathbf{u}), feas(\boldsymbol{\lambda}, \mathbf{u})\right),$$

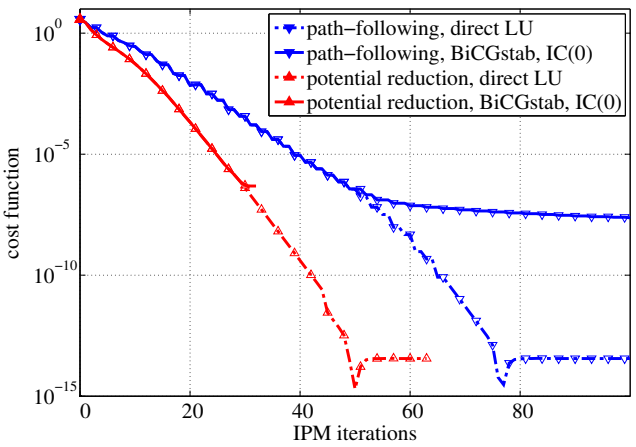
where the functions *cost* and $feas$ are the same as in Section 5.2.2.

Figure 11(a) shows the convergence of PGJ and IPM for both test cases. PGJ struggles specifically with larger problems, its convergence rate stalls almost completely before a reasonably small error is achieved. The convergence rate of IPM, on the other hand, decreases slowly with the problem size. For Test Problem 2, PGJ requires about 43000 iterations to reduce the error below 10^{-3}. The Interior Point Method requires a total of around 2545 Krylov iterations and a little more than a fifth of the calculation time for the same accuracy. From Figure 11(b) we learn that the decrease in the convergence rate of the IPM is almost entirely due to the inaccuracy of the linear solves. We see the error plotted over the number of outer IPM iterations when a direct solver is used. The number of outer IPM iterations needed to achieve a certain accuracy are nearly the same for both test problems.

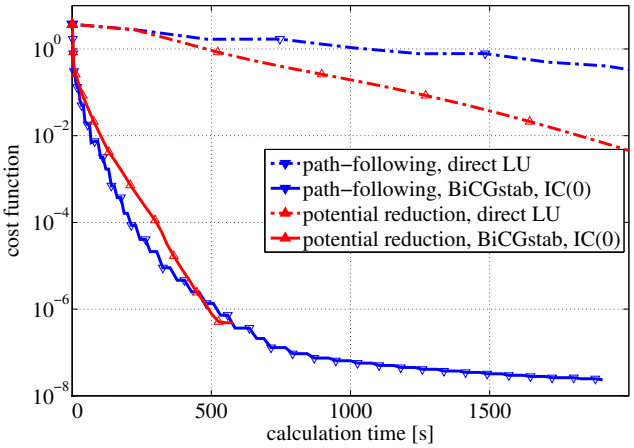
IPM scales better with the problem size than PGJ does and achieves a smaller error for both test cases. The scalability of the IPM with the problem size depends almost entirely on the scalability of the linear solver.

6. Conclusion

We conclude that the presented IPM is well–suited to solve the CCP to a higher accuracy than PGJ. With a direct solver, IPM converges with a constant rate until machine precision is reached. Since the IPM scales with the number of particles just as well as the linear solvers do, a direct solver cannot be applied in reasonable time for large scale problems. We tested three different Krylov solvers and two preconditioners and found that regularizing the subsystems, preconditioning them with Incomplete Cholesky factorizations and solving them with BiCGstab gives the fastest convergence. The inaccuracy of the linear solves decreases the overall convergence rate of the

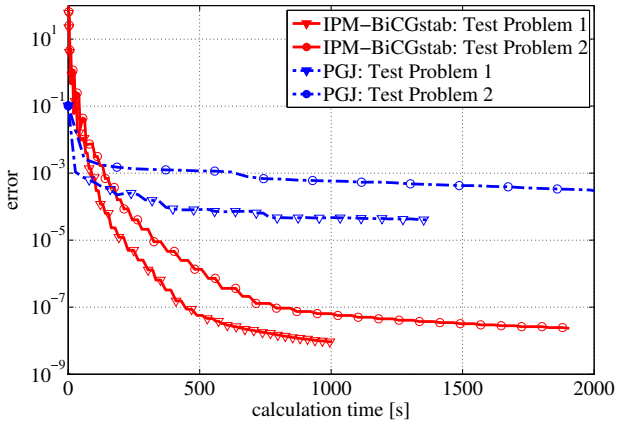


(a) cost function value over number of IPM iterations.

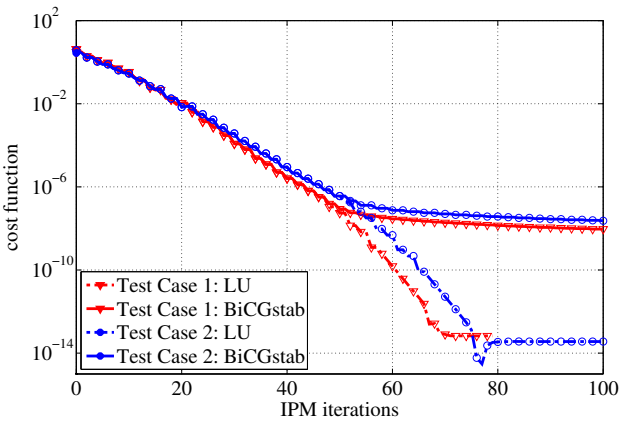


(b) cost function value over calculation time.

Figure 10: Path–Following vs. Potential Reduction.



(a) error value over calculation time.



(b) cost function value over IPM iterations.

Figure 11: The convergence of PGJ and IPM for different problem sizes.

IPM, but the inexact IPM still produces smaller errors than PGJ. IPMs superiority over PGJ becomes especially clear as the problem size increases.

We also found that, towards the exact solution, IC(0) does not decrease the range of eigenvalues much further than plain regularization does. The rate of convergence of the IPM and its calculation time depend highly on the efficiency of the preconditioner and the linear solver. While the method has shown its potential, IC(0) and BiCGstab still require a large number of iterations and the calculation time is higher than desirable.

In the future, the highest priority is to further analyze different preconditioners. Possibly there are problem–related ways to improve the numerical properties of the linear systems, either by taking advantage of the physical model or by manipulating the Interior Point Method itself. State–of–the–art linear solvers such as Algebraic Multigrid Methods could be applied if the accuracy requirement is very high. Once it is clear which solution technique fits best into the IPM framework, there are a lot of further optimizations to think of. One could start solving the CCP using PGJ and once the convergence rate decreases, the PGJ solution can be used to construct a good starting value for the IPM. Similarly, one could exploit the fact that the linear systems behave nicely in the first few IPM iterations. Hence, a simple and cheap linear solver suffices. As the error decreases it might make sense to switch to a more powerful solver. Ideally, the linear solvers should lend themselves for a matrix–free implementation and they should be parallelizable.

For the prediction of draft forces from the interaction of an excavator with soil, we have to demand very high standards from the simulation method. It has to be stable, accurate and it should be able to run a simulation in a reasonably short amount of time. So far, neither DEM nor NSCD are entirely satisfactory: One method is accurate, but instable and slow; while the other is reasonably fast and stable, but inaccurate. In this paper, we use results from abstract algebra, convex analysis, measure theory, optimization, nonlinear programming and numerics to im-

prove the accuracy of NSCD. So far, the presented Interior Point Method is little more than a prototype, but it is already substantially more accurate than PGJ and proved to be promising for further development.

References

Acary, V., Brogliato, B., 2008. Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics. Springer Verlag.

Alizadeh, F., Goldfarb, D., 2003. Second-order cone programming. Mathematical Programming 95, pp. 3 – 51.

Anitescu, M., 2005. Optimization-based simulation of nonsmooth rigid multibody dynamics. Mathematical Programming 105, pp. 113 – 143.

Anitescu, M., Cremer, J.F., Potra, F.A., 1995. Formulating 3D Contact Dynamics Problems. Technical Report. The University of Iowa.

Anitescu, M., Potra, F., 1997. Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. Nonlinear Dynamics , pp. 231 – 247.

Anitescu, M., Tasora, A., 2008. An iterative approach for cone complementarity problems for nonsmooth dynamics. Computational Optimization and Applications 47, pp. 207 – 235.

Bai, Y., Wang, G., Roos, C., 2004. A new primal-dual interior-point algorithm for second-order cone optimization. Optimization Online .

Balay, S., Brown, J., Buschelman, K., Gropp, W.D., Kaushik, D., Knepley, M.G., McInnes, L.C., Smith, B.F., Zhang, H., 2013. PETSc Web page. URL: http://www.mcs.anl.gov/petsc.

Balzer, M., Kleinert, J., Obermayr, M., 2013. Parallel implementation of the non-smooth contact dynamics method for large particle systems, in: Proceedings of the 3rd Conference on Particle-Based Methods, Stuttgart (PARTICLES).

Cundall, P.A., Strack, O.D.L., 1979. A discrete numerical model for granular assemblies. Géotechnique 29, pp. 47 – 65.

Daviet, G., Bertails-Descoubes, F., Boissieux, L., 2011. A hybrid iterative solver for robustly capturing coulomb friction in hair dynamics. ACM Trans. Graph. 139 30, pp. 1–12.

DeSaxcé, G., Feng, Z.Q., 1998. The bipotential method: A constructive approach to design the complete contact law with friction and improved numerical algorithms. Mathematical and Computer Modelling 28, pp. 225 – 245. Recent Advances in Contact Mechanics.

Faraut, J., Korányi, A., 1994. Analysis on symmetric cones. Oxford mathematical monographs, Clarendon Press, Oxford.

Faybusovich, L., 2002. A jordan-algebraic approach to potential-reduction algorithms. Mathematische Zeitschrift 239, pp. 117 – 129.

Fukushima, M., Luo, Z., Tseng, P., 2002. Smoothing functions for second-order-cone complementarity problems. SIAM Journal on Optimization 12, pp. 436 – 460.

Gondzio, J., 2012. Matrix-free interior point method. Computational Optimization and Applications 51, pp. 457 – 480.

Halmos, P.R., 1950. Measure Theory. Van Nostrand, New York, NY. July 1969 reprinting.

Hayashi, S., 2004. Studies on Second–Order Cone Complementarity Problems. Ph.D. thesis. Kyoto University.

Heyn, T., 2013. On the Modeling, Simulation, and Visualization of Many–Body Dynamics Problems with Friction and Contact. Ph.D. thesis. Uiversity of Wisconsin–Madison.

Heyn, T., Anitescu, M., Tasora, A., Negrut, D., 2013. Using krylov subspace and spectral methods for solving complementarity problems in many-body contact dynamics simulation. International Journal for Numerical Methods in Engineering 95, pp. 541–561.

Kleinert, J., Obermayr, M., Balzer, M., 2013. Modeling of large scale granular systems using the discrete element method and the non-smooth contact dynamics method: A comparison, in: Proceedings of the ECCOMAS Multibody Dynamics Conference, Zagreb.

Kojima, M., Megiddo, N., Noma, T., Yoshise, A., 1991. A Unified approach to interior point algorithms for linear complementarity problems. Springer-Verlag, Berlin, New York.

Krabbenhoft, K., Lyamin, A., Huang, J., da Silva, M.V., 2012. Granular contact dynamics using mathematical programming methods. Computers and Geotechnics 43, pp. 165 – 176.

Lacoursire, C., 2007. Regularized, stabilized, variational methods for multibodies, in: The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), pp. 40–48.

Moreau, J., Panagiotopoulos, P., 1990. Nonsmooth mechanics and applications. ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik 70, pp. 472 – 472.

Moreau, J.J., 1988. Bounded variation in time. Topics in nonsmooth mechanics, pp. 1-74 (1988).

Nesterov, Y.E., Todd, M.J., 1997. Self-scaled barriers and interior-point methods for convex programming 22, pp. 1 – 42.

Obermayr, M., Dressler, K., Vrettos, C., Eberhard, P., 2011. Prediction of draft forces in cohesionless soil with the discrete element method. Journal of Terramechanics 48, pp. 347 – 358.

Saad, Y., 2003. Iterative Methods for Sparse Linear Systems. 2nd ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Stewart, D.E., 2000. Rigid-body dynamics with friction and impact. SIAM Review 42, pp. 3 – 39.

Tasora, A., Anitescu, M., 2011. A matrix-free cone complementarity approach for solving large-scale, nonsmooth, rigid body dynamics. Computer Methods in Applied Mechanics and Engineering 200, pp. 439 – 453.

Tasora, A., Anitescu, M., Negrini, S., Negrut, D., 2013. A compliant visco-plastic particle contact model based on differential variational inequalities. International Journal of Non-Linear Mechanics 53, pp. 2 – 12. Multibody System Dynamics: A Selective Perspective.