Semismooth Newton method for frictional contact between pseudo-rigid bodies

Tomasz Koziara

Department of Civil Engineering
University of Glasgow, Glasgow G12 8LT, UK

Nenad Bićanić

Department of Civil Engineering
University of Glasgow, Glasgow G12 8LT, UK

Abstract

Recently developed semismooth Newton approach is adopted in the context of the frictional contact between three-dimensional pseudo-rigid bodies. The Signorini-Coulomb problem is formulated according to the formalism of the Contact Dynamics method. Hybrid linearisation, penalty scaling and line search techniques are combined as the global convergence enhancements of the Newton algorithm. Quasi-static simulations of dry masonry assemblies exemplify performance of the presented framework.

Key words: semismooth Newton, Signorini-Coulomb problem, frictional contact, multi-body contact, finite kinematics, pseudo-rigid body

1 Introduction

This paper deals with an adaptation of the semismooth Newton step developed by Hüeber *et al.* [1] within the formalism of the Contact Dynamics method [2,3]. The Newton method under consideration stems from a broader range of schemes for non-differentiable systems. General developments of this kind were discussed by Pang [4] and Qi and Sun [5]. In the context of the mixed frictional contact formulation Alart and Curnier [6] discuss the generalised Newton method (GNM),

Email addresses: t.koziara@civil.gla.ac.uk (Tomasz Koziara), n.bicanic@civil.gla.ac.uk (Nenad Bićanić).

which belongs to the same category. An observation made in [6], about the practical robustness of GNM applied to frictionless problems, was later confirmed under the umbrella of the primal-dual active set method [7]. The latter was shown to be equivalent to the semismooth Newton method by Hintermüller *et al.* [8]. In case of frictional problems, Christensen *et al.* [9] developed a linearisation along the lines of [4], and presented two-dimensional linearly elastic examples. In three dimensions, the non-smooth Newton scheme was recently applied by Jones and Papadopoulos [10] to solve anisotropic frictional problems. Large multi-body contact problems were not extensively studied within the context of Newton methods. Two-dimensional frictionless developments involving the primal-dual active set approach can be found in [11,12]. More general formulations, including friction and finite kinematics usually resort to methods avoiding formation of global tangents [3,13].

The Contact Dynamics method developed by Moreau [2] and Jean [3] is particularly convenient for kinematically modest finite deformation formulations, where the dual form of the contact problem can be inexpensively utilised (e.g. assembly of rigid bodies, for which an explicit inversion of the inertia operator is computationally feasible). The main features of the method comprise the use of a velocity level time stepping, the non-regularised treatment of frictional contact law, and the block Gauss-Seidel relaxation utilised stepwise in order to resolve unilateral constraints. The method was developed within the context of rigid and finite element discretised continua. Applications range from granular flow [14], through statics of masonry [15], to deep drawing simulations [16]. The Contact Dynamics method belongs to a broader category of schemes dealing with the non-smooth dynamical systems. For example developments by Wösle and Pfeiffer [17] and Pfeiffer *et al.* [18] utilise the same theoretical apparatus, although they differ in details (accelerations are involved, Newton scheme is used rather than the Gauss-Seidel).

The adopted kinematic model is the pseudo-rigid body, derived by Cohen and Muncaster [19] as a simplified counterpart of the finite elastodynamics. The model is straightforward to implement and facilitates a self-contained presentation. It is rich enough to serve as a rough post-processor of the macroscopic stress field, which is not without practical merit when large assemblies of nearly rigid bodies are to be considered. The main feature of the model is the spatial homogeneity of the deformation gradient. It was shown by Solberg and Papadopoulos [20] that the pseudo-rigid body is equivalent to the constant strain finite element. Due to simple nonlinear ordinary differential form of governing equations the model has been exploited as a basis for theoretical developments [21], as well as for numerical studies of chaotic dynamics [22].

The contribution of this paper is twofold. Firstly, the utility of the pseudo-rigid continuum in the context of large scale multi-body contact simulations is exemplified. This can be placed somewhere in between the rigid body and finite element based Contact Dynamics developments. Furthermore, the semismooth Newton technique

[1] is adopted and modified in order to effectively deal with the resultant frictional contact problem. In consequence, the Newton scheme is validated on a prototype of the finite kinematics three-dimensional multi-body contact formulation.

In the next two sections brief accounts of the kinematic model and the time stepping are given. Section 4 introduces the considered contact problem. The Newton step is detailed in Section 5. Numerical performance study is presented in Section 6, while the conclusions follow in the last section. Some of the technical details have been placed in the Appendix.

2 Pseudo-rigid bodies

The motion of a pseudo-rigid body is restricted to

$$\mathbf{x} = \mathbf{F}\left(\mathbf{X} - \bar{\mathbf{X}}\right) + \bar{\mathbf{x}} \tag{1}$$

where $\mathbf{x}$ is the current image of a referential point $\mathbf{X}$, $\mathbf{F} = \mathbf{F}(t)$ is a spatially homogeneous deformation gradient ($\det(\mathbf{F}) > 0$), $\bar{\mathbf{X}}$ is a selected referential point, and $\bar{\mathbf{x}} = \bar{\mathbf{x}}(t)$ is the current image of $\bar{\mathbf{X}}$. Assuming $\bar{\mathbf{X}}$ to be the mass centre, angular and linear momentum balance principles read

$$\ddot{\mathbf{F}}\mathbf{E}_0 + V_0\bar{\mathbf{P}} = \mathbf{S} \tag{2}$$

$$m\ddot{\bar{x}} = \mathbf{s} \tag{3}$$

where $\mathbf{E}_0$ is the referential Euler tensor, V_0 is the referential volume, $\bar{\mathbf{P}}$ is the mean first Piola-Kirchhoff stress tensor, $\mathbf{S}$ represents an angular momentum contribution of external forces, m is the mass, and $\mathbf{s}$ the resultant force vector. These are defined as follows

$$\mathbf{E}_0 = \int_{\mathscr{B}_0} \left(\mathbf{X} - \bar{\mathbf{X}}\right) \otimes \rho_0 \left(\mathbf{X} - \bar{\mathbf{X}}\right) dV \tag{4}$$

$$V_0 = \int_{\mathscr{B}_0} dV \tag{5}$$

$$\bar{\mathbf{P}} = \frac{1}{V_0} \int_{\mathscr{B}_0} \mathbf{P} dV \tag{6}$$

$$\mathbf{S} = \int_{\partial\mathscr{B}_0} \mathbf{p} \otimes \left(\mathbf{X} - \bar{\mathbf{X}}\right) dA + \int_{\mathscr{B}_0} \rho_0 \mathbf{b} \otimes \left(\mathbf{X} - \bar{\mathbf{X}}\right) dV \tag{7}$$

$$m = \int_{\mathscr{B}_0} \rho_0 dV \tag{8}$$

$$\mathbf{s} = \int_{\partial\mathscr{B}_0} \mathbf{p} dA + \int_{\mathscr{B}_0} \rho_0 \mathbf{b} dV \tag{9}$$

where ρ_0 is the mass density, $\mathbf{P}$ is the first Piola-Kirchhoff stress tensor, $\mathbf{p}$ and $\mathbf{b}$ are respectively surface and body forces. In order to close the set of equations (2), (3) a constitutive relation between $\bar{\mathbf{P}}$ and $\mathbf{F}$ is necessary. A hyperelastic pseudo-rigid continuum considered here admits the strain energy function Ψ such that

$$\bar{\mathbf{P}} = \partial_{\mathbf{F}} \Psi(\mathbf{F}) \tag{10}$$

$$\Psi = \frac{1}{4} \left[\mathbf{F}^T \mathbf{F} - \mathbf{I}\right] : \mathbf{C} : \left[\mathbf{F}^T \mathbf{F} - \mathbf{I}\right] \tag{11}$$

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \left[\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}\right] \tag{12}$$

where the Saint Venant - Kirchhoff material was adopted. In the above λ and μ are Lamé constants, while δ_{ij} is the Kronecker delta ($I_{ij} = \delta_{ij}$). Equations (2), (3), (10) with prescribed initial conditions and kinematic constraints form an initial value problem for a single body.

3 Time stepping

Quasi-static simulation of a multi-body assembly calls for a regularisation, as the tangent stiffness operator is singular. While this regularisation is expected to provide a meaningful representation of rigid modes, it seems most natural to adopt the dynamic formulation for that purpose. In the context of the Contact Dynamics method Acary and Jean [23] discuss such an adaptation. A straightforward relaxation technique results from assuming velocities to be zero at the beginning of each iteration. This leads to the modified implicit Euler scheme (22), (23).

Considering a set of N bodies $\{\mathscr{B}_i\}$, the balance principles (2) and (3) can be integrated over a time interval $[t, t+h]$ delivering

$$\mathbf{M}_i \left(\mathbf{u}_i^{t+h} - \mathbf{u}_i^t\right) = \int_t^{t+h} \mathbf{f}_i(t, \mathbf{q})\, dt + \sum_\alpha h \mathbf{r}_\alpha^{t+h} \tag{13}$$

where $i \in \{1, ..., N\}$ and α enumerates the set of contact points attached to $\mathscr{B}_i$. A column-wise vector and related matrix notation is adopted, thus omitting the i-indexing

$$\mathbf{u} = \left[\dot{F}_{11}\ \dot{F}_{21}\ \ldots\ \dot{\bar{x}}_1\ \dot{\bar{x}}_2\ \dot{\bar{x}}_3 \right]^T \tag{14}$$

$$\mathbf{f}(t,\mathbf{q}) = \left[\mathbf{S}(t)\ \mathbf{s}(t) \right]^T - V_0 \left[\bar{\mathbf{P}}(\mathbf{q})\ \mathbf{0} \right]^T \tag{15}$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{E}_0 \odot \mathbf{I} & \\ & m\mathbf{I} \end{bmatrix} \tag{16}$$

where $\mathbf{E}_0 \odot \mathbf{I} = E_{jl}\delta_{ik}$ is a 9×9 sub-matrix and $m\mathbf{I} = m\delta_{ij}$ is a 3×3 matrix. The configuration can be readily computed once the velocity is known up to some time $t+h$

$$\mathbf{q}^{t+h} = \mathbf{q}^t + \int_t^{t+h} \mathbf{u}(t)\, dt \tag{17}$$

The average generalised reaction at the contact point $\mathbf{x}_\alpha$ is defined as

$$\mathbf{r}_\alpha^{t+h} = \mathbf{H}^T \mathbf{R}_\alpha^{t+h} \tag{18}$$

where $\mathbf{H}$ acts between the twelve-component space of generalised velocities $\mathbf{u}$ and the three-component space of local velocities $\mathbf{U}$ (power conjugacy of velocities and forces results in the transpose)

$$\mathbf{U}(\mathbf{X}, \mathbf{e}_k) = \mathbf{H}(\mathbf{X}, \mathbf{e}_k)\, \mathbf{u} \tag{19}$$

The operator $\mathbf{H}$ depends upon the choice of a referential point $\mathbf{X}$ and the local base $\mathbf{e}_k$. It transforms the generalised velocity into the velocity of $\mathbf{X}$, expressed with respect to $\mathbf{e}_k$. For the pseudo-rigid body $\mathbf{H}$ reads

$$\mathbf{H}(\mathbf{X}, \mathbf{e}_k) = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \mathbf{e}_3^T \end{bmatrix} \left[(X_1 - \bar{X}_1)\mathbf{I} \quad (X_2 - \bar{X}_2)\mathbf{I} \quad (X_3 - \bar{X}_3)\mathbf{I} \quad \mathbf{I} \right] \tag{20}$$

where $\bar{\mathbf{X}}$ is the mass centre, and $\mathbf{I}$ is the 3×3 identity matrix. The average local reaction is defined as

$$\mathbf{R}_\alpha^{t+h} = \frac{1}{h} \int_t^{t+h} d\mathbf{R}_\alpha \tag{21}$$

where an abstract measure $d\mathbf{R}_\alpha$ incorporates contributions of persistent contacts and percussions. The modified backward Euler step follows

$$\mathbf{u}_i^{t+h} = \mathbf{A}_i^{-1} h \mathbf{f}_i \left(t+h, \mathbf{q}^t \right) + \mathbf{A}_i^{-1} \sum_\alpha \mathbf{H}_{i\alpha}^T h \mathbf{R}_\alpha^{t+h} \tag{22}$$

$$\mathbf{q}_i^{t+h} = \mathbf{q}_i^t + h\mathbf{u}_i^{t+h} \tag{23}$$

where

$$\mathbf{A}_i = \mathbf{M}_i + h^2 \left. \frac{\partial^2 \Psi_i}{\partial \mathbf{q}_i \partial \mathbf{q}_i} \right|_{\mathbf{q}^t} \tag{24}$$

$$\mathbf{H}_{i\alpha} = \mathbf{H}\left(\mathbf{X}_\alpha, \mathbf{e}_{\alpha k}\right)|_{\bar{\mathbf{X}}(\mathscr{B}_i)} \tag{25}$$

and $\mathbf{X}_\alpha$ is the referential image of the contact point $\mathbf{x}_\alpha$, while $\mathbf{e}_{\alpha k}$ is the associated local basis. The equation (22) holds true with the strong assumption of the contact geometry remaining unchanged over the time interval $[t, t+h]$. This simplifies the implementation, as the linearisation with respect to $\mathbf{H}$ is avoided.

4 Contact problem

Combining (19) and (22), the following saddle point problem results for the set of all bodies

$$\begin{bmatrix} \mathbf{A} & -\mathbf{H}^T h \\ \mathbf{H} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{R} \end{bmatrix} = \begin{bmatrix} h\mathbf{f} \\ \mathbf{U} \end{bmatrix} \tag{26}$$

where $\mathbf{A}$ is the block diagonal matrix composed of $\mathbf{A}_i$ blocks, $\mathbf{H}$ comprises $\mathbf{H}_{i\alpha}$ blocks, and the remaining vectors gather the relevant entities in a column-wise manner. Traditionally, in the Contact Dynamics method, the Schur complement system is formed

$$\mathbf{H}\mathbf{u} = \mathbf{H}\mathbf{A}^{-1}\mathbf{H}^T h\mathbf{R} + \mathbf{H}\mathbf{A}^{-1}h\mathbf{f} \quad \text{or} \quad \mathbf{U} = \mathbf{W}\mathbf{R} + \mathbf{B} \tag{27}$$

and the following form of the Signorini-Coulomb problem is adopted

$$\begin{cases} \mathbf{U}_\alpha = \mathbf{B}_\alpha + \sum_\beta \mathbf{W}_{\alpha\beta}\mathbf{R}_\beta \\ R_{\alpha N} = \mathrm{proj}_{R_+}\left(R_{\alpha N} - \rho \bar{U}_{\alpha N}\right) \\ \mathbf{R}_{\alpha T} = \mathrm{proj}_{D(\mu R_{\alpha N})}\left(\mathbf{R}_{\alpha T} - \rho \mathbf{U}_{\alpha T}\right) \end{cases} \tag{28}$$

In general, the above system stems from the discretised form of any dynamic frictional contact problem. The first equation in (28) depends upon the underlying kinematics. $\mathbf{W}_{\alpha\beta}$ are the 3×3 dense sub-blocks of the otherwise sparse representation of $\mathbf{W}$. The first equation states the relation between the average contact reactions $\mathbf{R}_\beta$ and the relative contact velocities $\mathbf{U}_\alpha$. The relation is expressed at time $t+h$, and the average related to the contact reactions corresponds to the interval $[t, t+h]$.

The index set of α and β enumerates all contacts. Thus for zero reactions the relative velocities become equal to $\mathbf{B}_\alpha$, which are therefore referred to as the free velocities. The details on derivation of the first equation in (28) can be found in the Appendix.

The second and the third equation in (28) describe the nonlinear constraints related to the Signorini contact condition and the Coulomb friction law. The discretised Signorini condition was adopted in the form of the *quasi-inelastic shock law* [3]

$$\bar{U}_N^{t+h} \geq 0 \; R_N^{t+h} \geq 0 \; \bar{U}_N^{t+h} R_N^{t+h} = 0 \tag{29}$$

where

$$\bar{U}_N^{t+h} = \frac{\max(0, g^t)}{h} + U_N^{t+h} \tag{30}$$

and $g(t)$ is the gap function defined in the Appendix. As the assumed normal direction is consistent with the positive gap velocity, (29) states that any violation of the non-penetration results in a reactive force or velocity driving at the penetration-free configuration. The relation (29) corresponds to the gap-force complementarity for a not established contact, and to the velocity-force complementarity for an established contact (cf. Jean [3]).

The discretised Coulomb friction law can be expressed in the form of the maximal dissipation principle

$$\mathbf{R}_T^{t+h} \in D\left(\mu R_N^{t+h}\right) \; \forall \mathbf{S} \in D\left(\mu R_N^{t+h}\right) \; \left\langle \mathbf{U}_T^{t+h}, \mathbf{S} - \mathbf{R}_T(t) \right\rangle \geq 0 \tag{31}$$

where μ is the friction coefficient, $D\left(\mu R_N^{t+h}\right)$ is a two-dimensional $\mathbf{0}$-centred disc of radius μR_N^{t+h}, and $\langle \cdot, \cdot \rangle$ stands for the scalar product in an Euclidean 2-space. Thus the friction force smaller than μR_N^{t+h} implies sticking, while sliding occurs for the tangential force of value μR_N^{t+h}, and with the direction opposite to the slip velocity.

Standard rewriting of the inequality constraints (29), (31) as equalities leads to the projection formulae in (28), where the regularisation parameter ρ is an arbitrary positive number, and R_+ stands for the semi-positive real half-space. A more comprehensive discussion about the related issues can be found in [17].

4.1 Sources of difficulty

A possible singularity of $\mathbf{W}$, the inherent non-smoothness of contact constraints, and the lack of normality of the Coulomb law are among the main reasons rendering (28) hard to solve. These are to be commented on in some detail here.

4.1.1 *Singularity of* **W**

The operator **W** derives from a formula

$$\mathbf{W} = \mathbf{H}\mathbf{A}^{-1}\mathbf{H}^T h \tag{32}$$

where **A** is the $n \times n$ generalised tangent, **H** is the $m \times n$ transformation matrix, and h is a positive scalar. The tangent **A** is symmetric and positive definite. It follows that **W** remains symmetric and positive definite, provided rows of **H** are linearly independent. Thus $m \leq n$. This puts a constraint on the number of contact points M. For an assembly of N pseudo-rigid bodies $3M \leq 12N$, as twelve degrees of freedom describe kinematics of a single body in three dimensions. At the same time, not only the number of contacts influences the rank of **H**. It is also down to the realisation of contact constraints. The pseudo-rigid body has a linear distribution of the instantaneous velocity over an arbitrary flat surface. Thus, the relative velocity between two bodies in a flat contact is fully parametrised by three points and any larger number of contacts results in the singularity of **W**. So does their colinearity. One can then generally speak about the global and local over-restraining of the system. In practice though, **W** becomes singular for some geometrical configurations of contact points. This happens regardless of the fact that the full rank of **H** is seemingly ensured by the algorithmic realisation of constraints. The phenomenon corresponds to static indeterminacy, naturally occurring in any complex network of forces. Numerical observations in that respect will be provided in Section 6.

4.1.2 *Non-smoothness and solution structure*

Newton method requires calculation of derivatives. The projections in (28) are nonsmooth functions, thus calculation of derivatives requires a generalisation. To picture that, let us consider a one dimensional simplification of the contact problem

$$\begin{cases} u = b + wr \\ r = \mathrm{proj}_{R_+}(r - u) \end{cases} \tag{33}$$

where the friction constraint was discarded and $\rho = 1$ was assumed. The above system can be rewritten as

$$c(r) = r - \max(0, r(1-w) - b) = 0 \tag{34}$$

which is a nonlinear equation the root of which is searched for. One can see that for $r < b/(1-w)$ the root is $r = 0$ while in the remaining case $r = -b/w$. In the former case $u = b \geq 0$, as $b < 0$ suppresses the root $r = 0$. In the latter case $u = 0$ and $r \geq 0$, which recovers the Signorini condition. The multi-dimensional version of (34) reads

$$c_\alpha(\mathbf{r}) = r_\alpha - \max\left(0, r_\alpha(1-w_{\alpha\alpha}) - b_\alpha - \sum_{\beta\neq\alpha} w_{\alpha\beta} r_\beta\right) = 0 \qquad (35)$$

A series of plots of $c(r)$ for various values of b corresponds to a series of sections of $c_\alpha(\mathbf{r})$ for some fixed $r_{\beta\neq\alpha}$. This can be observed for the two-dimensional case in Figure 1. What is also visible is the non-smoothness of the constraint graphs. Each surface plot splits into the part where $c_\alpha(\mathbf{r}) = r_\alpha$, and into another one where $c_\alpha(\mathbf{r})$ is an arbitrarily inclined half-plane. Both parts are connected in a continuous manner along the line $r_\alpha - \rho u_\alpha = 0$. This is where the non-smoothness occurs. In order to solve (35) by means of the Newton method, one has to sort out the differentiation along $r_\alpha - \rho u_\alpha = 0$. In fact the only troublesome part concerns the max function. Without resorting to details, it is sufficient to say that the max function is Lipschitz continuous, therefore the analysis presented in [4,5] is feasible. This allows to derive a locally convergent Newton scheme.

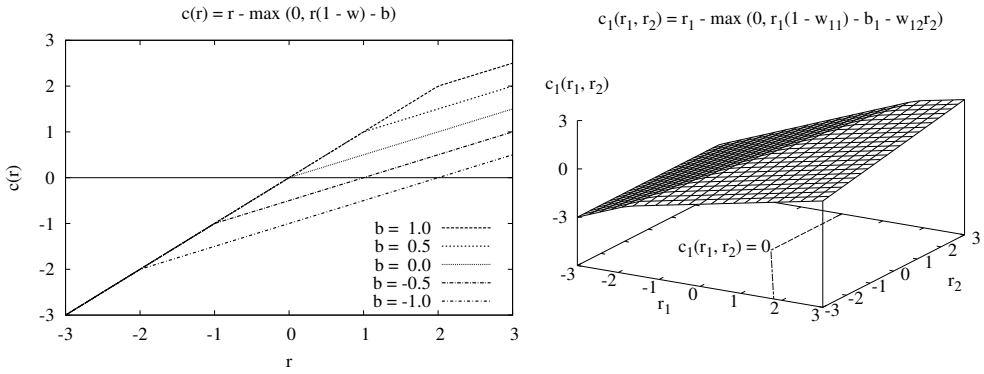

Figure 1. Plots of the Signorini constraints (34) and (35) for $w = w_{11} = 0.5$, $w_{12} = 0.3$, $b_1 = 0$.

It is relevant to ask whether the solution for (35) actually exists. As suggested by the contour line $c_1(r_1, r_2) = 0$ in Figure 1, each constraint contributes a curve composed of two straight half-lines. If all such curves have a nonempty intersection, the solution exists. If the intersection happens to be a point, the solution is unique. The half-line components are either $r_\alpha = 0$ or $b_\alpha + w_{\alpha\beta} r_\beta = 0$, where the summation over β is assumed. The matrix $\mathbf{w}$ is the principal sub-matrix of $\mathbf{W}$, obtained by removing all rows and columns involved in the tangential response $w_{\alpha\beta} = W_{NN\alpha\beta}$. Constructed this way, $\mathbf{w}$ remains symmetric and positive definite [24]. In other words, $\mathbf{W}$ and $\mathbf{w}$ are P-matrices. It is readily seen that this property assures the existence of a unique solution. For the two extreme cases one has either all $r_\alpha = 0$ (no contact) or the system of equations $b_\alpha + w_{\alpha\beta} r_\beta = 0$ is uniquely solved (all points in contact). In the remaining cases some $r_{\gamma\in I} = 0$ and some emerge as a solution of $b_\alpha + w_{\alpha\beta} r_\beta = 0$, where $\alpha, \beta \notin I$. The latter system can always be uniquely solved due to the P-property of $\mathbf{w}$. This is in fact a classical result related to convex optimisation or solution of linear complementary problems [8].

Geometric interpretation of the solution set is more troublesome for the frictional case. The friction cone introduces additional nonlinearity and the projection onto its section contributes yet another source of non-smoothness. It is possible though, to quite clearly visualise the structure of a single contact point problem. Adopting after [1] the Coulomb law in the form

$$\begin{array}{c} \max\left(\mu\left(R_N-\rho U_N\right),\left\|\mathbf{R}_T-\rho\mathbf{U}_T\right\|\right)\mathbf{R}_T \\ -\mu\max\left(0,\left(R_N-\rho U_N\right)\right)\left(\mathbf{R}_T-\rho\mathbf{U}_T\right)=0 \end{array} \tag{36}$$

where $\bar{U}_N = U_N$ was assumed, and the first projection in (28) utilised, three implicit equations follow

$$\mathbf{C}_T\left(\mathbf{R}\right)=\max\left(\mu d_N\left(\mathbf{R}\right),\left\|\mathbf{d}_T\left(\mathbf{R}\right)\right\|\right)\mathbf{R}_T-\mu\max\left(0,d_N\left(\mathbf{R}\right)\right)\mathbf{d}_T\left(\mathbf{R}\right)=\mathbf{0} \tag{37}$$

$$C_N\left(\mathbf{R}\right)=R_N-\max\left(0,d_N\left(\mathbf{R}\right)\right)=0 \tag{38}$$

where $\mathbf{d}\left(\mathbf{R}\right)=\mathbf{R}-\rho\left(\mathbf{B}+\mathbf{W}\mathbf{R}\right)$. The above constraints are depicted in Figure 2. The inclined surface describes the normal constraint $C_N\left(\mathbf{R}\right)=0$. The two nearly vertical (aligned with the R_N axis) surfaces correspond to the components of the tangential constraint $\mathbf{C}_T\left(\mathbf{R}\right)=\mathbf{0}$. The solution rests at the intersection point of all three surfaces. Small friction coefficient in the left picture results in frictional slipping. Larger coefficient in the right picture allows to pronounce the transition from the slip to the stick state. The constraint surfaces are visibly curved, as the non-smooth transition bends them along vertical lines. The solution point in the right picture is in the state of sticking. Essentially, for problems with many contact points, one is interested in finding intersection points like those in Figure 2.

4.1.3 Lack of normality

The lack of normality of the Coulomb law contributes notably to the overall difficulty. This can be illustrated by investigating an operator composed of (37) and (38)

$$\mathbf{C}\left(\mathbf{R}\right)=\left[\begin{array}{c} \max\left(\mu d_N\left(\mathbf{R}\right),\left\|\mathbf{d}_T\left(\mathbf{R}\right)\right\|\right)\mathbf{R}_T-\mu\max\left(0,d_N\left(\mathbf{R}\right)\right)\mathbf{d}_T\left(\mathbf{R}\right) \\ R_N-\max\left(0,d_N\left(\mathbf{R}\right)\right) \end{array}\right] \tag{39}$$

where $\mathbf{d}\left(\mathbf{R}\right)=\mathbf{R}-\rho\left(\mathbf{W}\mathbf{R}+\mathbf{B}\right)$, $\bar{U}_N = U_N$ holds, and a single contact point problem is considered. The lack of normality refers to the property already expressed in formulae (29) and (31). For the adopted assumptions, in case of the frictional slip, the velocity reads $\mathbf{U}_T=-\lambda\mathbf{R}_T$, $U_N=0$ ($\lambda>0$). The velocity is then normal to the section of constraints surface, rather to the surface itself (friction cone).

One can notice, that the equation $\mathbf{U} = \mathbf{WR} + \mathbf{B}$ can be viewed as the gradient of the function $\mathscr{P}(\mathbf{R}) = \mathbf{R}^T\mathbf{WR} + \mathbf{R}^T\mathbf{B}$, thus the entire problem might be regarded as a minimisation of $\mathscr{P}(\mathbf{R})$, subject to the constraints (29), (31). The lack of normality implies, that the classical optimality conditions cannot be established, and the problem itself cannot be regarded as "optimisation". Alart and Curnier propose the term "quasi-optimisation" instead [6]. Regardless of the terminology, a root of $\mathbf{C}(\mathbf{R}) = \mathbf{0}$ is sought. The lack of normality articulates itself in the nonmonotonicity of $\mathbf{C}(\mathbf{R})$. This implies that there exist $\mathbf{R}_1$, $\mathbf{R}_2$ such that

$$\langle \mathbf{C}(\mathbf{R}_1) - \mathbf{C}(\mathbf{R}_2), \mathbf{R}_1 - \mathbf{R}_2 \rangle < 0 \tag{40}$$

A simple numerical experiment shows that for data from Figure 2, and force pairs generated randomly on a unit ball the above inequality holds true in 30% of cases ($\mu = 0.4$). Thus, a linear expansion of (39) can experience a negative definite Jacobian. This is not a desirable feature, as the Newton scheme applied to the root finding problem $\mathbf{C}(\mathbf{R}) = \mathbf{0}$, can be viewed as a sequence of minimisations of quadratic approximations of a general nonconvex function. In such case the scheme requires globalisation (line search) in order to avoid divergence or convergence to a local maximum. A technique of this sort will be examined later (Section 5.3).

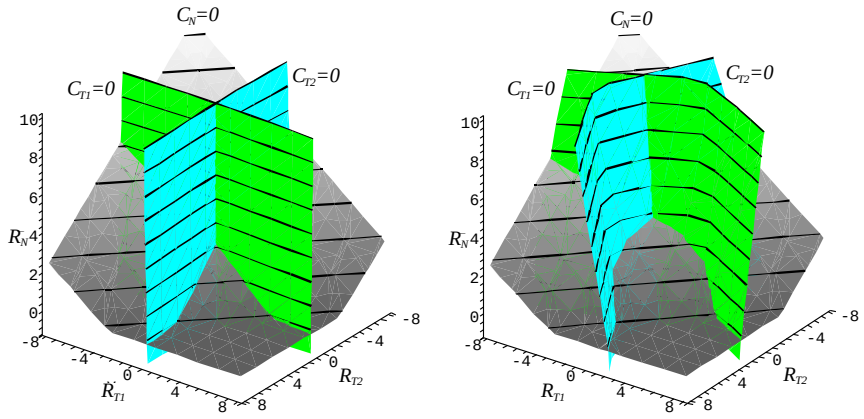

Figure 2. Plots of the Signorini-Coulomb constraints (37) and (38) for $\rho = 1$, $W_{ii} = 0.8$, $W_{i\neq j} = 0.4$, $B_{T1} = B_{T2} = 0$, $B_N = -2$. The coefficient of friction was $\mu = 0.1$ in the left picture, and $\mu = 0.9$ in the right one.

5 Newton method

The frictional contact problem (28) can be rewritten as

$$\begin{cases} \mathbf{U} = \mathbf{B} + \mathbf{WR} \\ \mathbf{C}(\mathbf{R}, \mathbf{U}) = \mathbf{0} \end{cases} \tag{41}$$

where

$$\mathbf{C}\left(\mathbf{R},\mathbf{U}\right)=\begin{bmatrix}\mathbf{C}_{\alpha T}\left(\mathbf{R}_{\alpha},\mathbf{U}_{\alpha}\right)\\ C_{\alpha N}\left(\mathbf{R}_{\alpha},\mathbf{U}_{\alpha}\right)\end{bmatrix} \tag{42}$$

$$\mathbf{C}_{\alpha T}\left(\mathbf{R}_{\alpha},\mathbf{U}_{\alpha}\right)=\max\left(\mu d_{\alpha N},\|\mathbf{d}_{\alpha T}\|\right)\mathbf{R}_{\alpha T}-\mu\max\left(0,d_{\alpha N}\right)\mathbf{d}_{\alpha T} \tag{43}$$

$$C_{\alpha N}\left(\mathbf{R}_{\alpha},\mathbf{U}_{\alpha}\right)=R_{\alpha N}-\max\left(0,d_{\alpha N}\right) \tag{44}$$

$$\mathbf{d}_{\alpha T}=\mathbf{R}_{\alpha T}-\rho\mathbf{U}_{\alpha T} \tag{45}$$

$$d_{\alpha N}=R_{\alpha N}-\rho\bar{U}_{\alpha N} \tag{46}$$

It is readily seen that constraints (43) and (44) parallel the projection formulae in (28). Similar formulation is a starting point of the development in [1]. There however, the finite element mortar discretisation provides the first relation in (41). Contrary to the current case, it is stated in the primal form, with displacements (velocities here) acting on the tangent stiffness operator. As stated earlier, the current formulation is usually more suitable for kinematic models with a moderate amount of freedom, as **W** and **B** can be inexpensively computed.

In order to approximately solve (41), the active set strategy and the frictional Newton step proposed in [1] will be adopted. For the class of problems like the unilateral contact alone, the primal-dual active set technique was shown to be equivalent to the semismooth Newton method [8], so that the overall development can be regarded as a variant of the Newton algorithm.

5.1 Unilateral contact

The frictionless case is briefly examined. Finding normal reactions reduces to a well behaved problem, the structure of which was already discussed in Section 4.1.2. According to the reasoning presented therein, once the index sets of zero and nonzero reactions are identified, the solution can be obtained in one step. The two possible index sets will be denoted as active $\mathscr{A}_N$ and inactive $\mathscr{I}_N$. Although their immediate identification is usually not possible, the predictive formula (46) and the normal constraint (44) suggest the following approximation

$$\mathscr{A}_N=\{\alpha : d_{\alpha N}\geq 0\}\qquad \mathscr{I}_N=\{\alpha\}\setminus\mathscr{A}_N \tag{47}$$

The primal-dual active set algorithm solves a series of reduced linear systems for successive approximations of the above sets. This can be summarised as follows

Algorithm UNIL

1 $k = 0$

2 $\mathbf{T}^k = \mathbf{W}\mathbf{R}^k + \mathbf{B} - \mathbf{U}^k$

3 $\mathscr{A}_N^k = \left\{\alpha : d_{\alpha N}^k \geq 0\right\} \quad \mathscr{I}_N^k = \{\alpha\} \setminus \mathscr{A}_N^k$

4 **if** $k > 0 \wedge \mathscr{A}_N^k = \mathscr{A}_N^{k-1}$ **stop**

5 $\mathscr{X} = \mathscr{A}_N^k N \quad \mathscr{Y} = \left\{\mathscr{A}_N^k T\right\} \cup \mathscr{I}_N^k$

6 $$\begin{bmatrix} \mathbf{W}_{\mathscr{X}\mathscr{X}} & \mathbf{W}_{\mathscr{X}\mathscr{Y}} \\ \mathbf{0} & \mathbf{I}_{\mathscr{Y}} \end{bmatrix} \begin{bmatrix} \delta\mathbf{R}_{\mathscr{X}} \\ \delta\mathbf{R}_{\mathscr{Y}} \end{bmatrix}^k = \begin{bmatrix} -\bar{\mathbf{U}}_{\mathscr{X}} - \mathbf{T}_{\mathscr{X}} \\ -\mathbf{R}_{\mathscr{Y}} \end{bmatrix}^k$$

7 $\mathbf{R}^{k+1} = \mathbf{R}^k + \delta\mathbf{R}^k$

8 $\mathbf{U}^{k+1} = \mathbf{U}^k + \mathbf{W}\delta\mathbf{R}^k + \mathbf{T}^k$

9 $k = k + 1$

10 **goto** 2

For the sake of consistency with the forthcoming frictional linearisation, the incremental formulation is utilised above. An update of the residual $\mathbf{T}^k$ in line 2 is followed by the prediction of the active and inactive sets in line 3. From the complementarity considerations, it is seen that once the correct sets were predicted, they are not changed in line 3. Thus, the termination criterion takes a quite specific form (line 4). In line 5 two index sets are created: $\mathscr{X}$ enumerating normal components in the active set, and $\mathscr{Y}$ enumerating all of the tangential components together with the inactive normal ones. The linear system in line 6 follows from

$$\mathbf{U}^k + \delta\mathbf{U}^k = \mathbf{W}\left(\mathbf{R}^k + \delta\mathbf{R}^k\right) + \mathbf{B} \tag{48}$$

when considered with $\bar{U}_{\alpha N}^k + \delta U_{\alpha N}^k = 0$ for $\alpha \in \mathscr{A}_N$ (resulting from (29), the quasi-inelastic shock law), $R_{\alpha N}^k + \delta R_N^k = 0$ for $\alpha \in \mathscr{I}_N$ and $\mathbf{R}_{\alpha T}^k + \delta\mathbf{R}_{\alpha T}^k = \mathbf{0}$ for all α. The last four lines conclude the algorithm in an obvious way. The above recipe can be optimised by eliminating all tangential components.

5.2 Frictional tangents

As explained by Hüeber *et al.* [1], a numerically robust linearisation of the frictional constraint (43) requires some heuristic modifications. The authors examine a number of such modifications, one of which proves to be the most effective. Here, a brief derivation of the relevant formulae is provided.

The basic technical step relies on the differentiation of the max function, as the non-smoothness of the Euclidean norm in (43) will not play any role (the term vanishes for sticking points and is nonzero otherwise). The max function is Lipschitz continuous and semismooth in the sense of [5]. Roughly speaking, this means that any two points on its graph can be connected by a finitely steep line. As it was already mentioned, the analysis provided by Pang [4] and Qi and Sun [5] allows to establish a locally superlinearly convergent Newton scheme for such a case. This result is of practical interest, when an incremental sequence of frictional contact problems is to be solved.

The generalised partial derivative of the function $f(x,y) = \max(x,y)$ can be written as $Gf_x = 1$ if $x \geq y$ and $Gf_x = 0$ if $x < y$. Gf_y is calculated analogously. As adopted in [5], at any point the generalised derivative belongs to the set-valued gradient defined by Clarke [25]. As a consequence, the partial derivatives at $x = y$ can be equal to any number in the range $[0,1]$. Thus when comparing x and y, the equality can be adopted on either side. From the algorithmic point of view this corresponds to a nuance in the definition of the active and inactive sets, utility of which will be commented on at a later point (Section 5.3). The active and inactive tangential sets are defined as follows

$$\mathscr{A}_T = \{\alpha \in \mathscr{A}_N : \|\mathbf{d}_{\alpha T}\| - \mu d_{\alpha N} \geq 0\} \qquad \mathscr{I}_T = \mathscr{A}_N \setminus \mathscr{A}_T \tag{49}$$

Let us focus on a contact point with index α, and temporarily neglect the α-indexing. The characteristic function $\chi_{\mathscr{S}} = 1$ if $\alpha \in \mathscr{S}$ and $\chi_{\mathscr{S}} = 0$ otherwise. According to the above definitions the differential of the tangential constraint reads

$$\begin{gathered} G_{\mathbf{C}_T}(\delta\mathbf{R}, \delta\mathbf{U}) = \chi_{\mathscr{A}_T} \frac{\mathbf{d}_T(\delta\mathbf{R}_T - \rho\delta\mathbf{U}_T)}{\|\mathbf{d}_T\|}\mathbf{R}_T \\ +\chi_{\mathscr{I}_T}\mu(\delta R_N - \rho\delta U_N)\mathbf{R}_T + \max(\mu d_N, \|\mathbf{d}_T\|)\delta\mathbf{R}_T \\ -\chi_{\mathscr{A}_N}\mu(\delta R_N - \rho\delta U_N)\mathbf{d}_T - \mu\max(0, d_N)(\delta\mathbf{R}_T - \rho\delta\mathbf{U}_T) \end{gathered} \tag{50}$$

and the tangential Newton step takes the form

$$G_{\mathbf{C}_T}\left(\delta\mathbf{R}^k, \delta\mathbf{U}^k\right) = -\mathbf{C}_T\left(\mathbf{R}^k, \mathbf{U}^k\right) \tag{51}$$

$$\left(\mathbf{R}^{k+1}, \mathbf{U}^{k+1}\right) = \left(\mathbf{R}^k, \mathbf{U}^k\right) + \left(\delta\mathbf{R}^k, \delta\mathbf{U}^k\right) \tag{52}$$

In case of frictional sticking ($\|\mathbf{d}_T\| < \mu d_N$), equation (51) simplifies to

$$\delta\mathbf{U}_T^k = -\frac{\mathbf{U}_T^k}{d_N^k}\delta R_N - \frac{\mathbf{U}_T^k}{d_N^k}\rho\bar{U}_N^k - \mathbf{U}_T^k \tag{53}$$

Condition $\bar{U}_N^k + \delta U_N^k = 0$ was utilised to derive the above (frictional linearisation is considered on the active normal set). Using (52), formula (53) can be rewritten as $\mathbf{U}_T^{k+1} = \left(1 - R_N^{k+1}/d_N^k\right)\mathbf{U}_T^k$, where it is seen that for a convergent sequence of iterates $\mathbf{U}_T^{k+1} \to \mathbf{0}$, as $\left|R_N^{k+1} - d_N^k\right| \to 0$. In the remaining case of frictional slipping ($\|\mathbf{d}_T\| \geq \mu d_N$), equation (51) takes the following form

$$\mathbf{R}_T^k + \left(\mathbf{I} - \mathbf{M}^k\right)\delta\mathbf{R}_T^k + \rho\mathbf{M}^k\delta\mathbf{U}_T^k = \mathbf{v}_T^k\mu\left(R_N^k + \delta R_N\right) \tag{54}$$

where $\mathbf{I}$ stands for the two-dimensional identity matrix, and

$$\mathbf{M}^k = e^k\left(\mathbf{I} - \mathbf{F}^k\right) \tag{55}$$

$$\mathbf{F}^k = \frac{\mathbf{R}_T^k \otimes \mathbf{d}_T^k}{\mu d_N^k \left\|\mathbf{d}_T^k\right\|} \tag{56}$$

$$e^k = \frac{\mu d_N^k}{\left\|\mathbf{d}_T^k\right\|} \tag{57}$$

$$\mathbf{v}_T^k = \frac{\mathbf{d}_T^k}{\left\|\mathbf{d}_T^k\right\|} \tag{58}$$

Equation (54) expresses a ray-wise Coulomb constraint along the predictor direction $\mathbf{v}_T^k$. Evidently, variations of the tangential reaction and velocity together contribute to the fulfilment of the linearised constraint. Thus, the iterates of the reaction $\mathbf{R}_T^{k+1}$ do not necessarily belong to the friction cone before the convergence tightens. The following modification

$$\tilde{\mathbf{F}}^k = \frac{\mathbf{R}_T^k \otimes \mathbf{d}_T^k}{\max\left(\mu d_N^k, \left\|\mathbf{R}_T^k\right\|\right)\left\|\mathbf{d}_T^k\right\|} \tag{59}$$

results in an approximate projection of $\mathbf{R}^{k+1}$ onto the tangent to the current section of the friction cone [1]. This results from the fact, that whenever $\mathbf{R}_T^k$ and $\mathbf{d}_T^k$ are nearly aligned, together with $\left\|\mathbf{R}_T^k\right\| \geq \mu d_N^k$, the matrix $\mathbf{I} - \tilde{\mathbf{F}}^k$ acts roughly as a projection on the direction perpendicular to $\mathbf{v}_T^k$. Therefore, the modified $\mathbf{M}^k$ filters out components parallel to $\mathbf{v}_T^k$. One can see, that when (59) is in power, $\mathbf{R}_T^k + \delta\mathbf{R}_T^k$ will approximately lay on the line perpendicular to $\mathbf{v}_T^k$ and tangent to the μR_N^{k+1} section

of the friction cone. This can be best observed in Figure 3. In practice then, the coefficients in relation (54) are computed with (56) replaced by (59). The modification results in faster and more robust convergence behaviour. This is because the iterates of $\mathbf{R}_T^{k+1}$ remain closer to the friction cone, thus less significantly interact trough the kinematic coupling in $\mathbf{W}$. This seems particularly helpful in the formulation admitting large rotations and therefore stronger normal-tangential coupling.

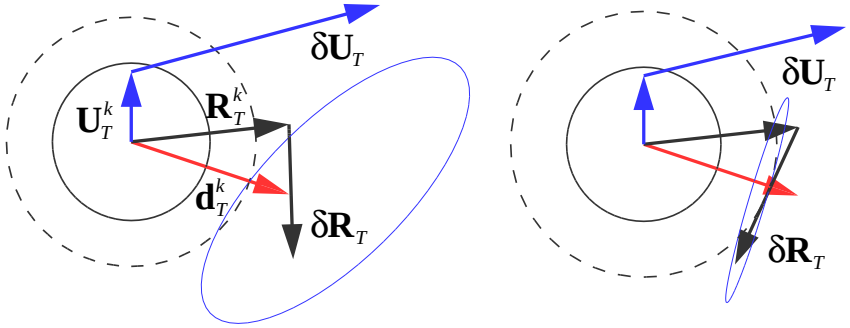

Figure 3. The Effect of the modification (59). The circle radii are μR_N^k (solid), and μR_N^{k+1} (dashed). On the left, the unmodified iteration of the Newton step (51), (52) is presented. The ellipsoid corresponds to the points of $\mathbf{R}_T^k + \left(\mathbf{I} - \mathbf{M}^k\right)\delta\mathbf{R}_T^k + \rho\mathbf{M}^k\mathbf{S}$, where $\|\mathbf{S}\| = \left\|\delta\mathbf{U}_T^k\right\|$. On the right, the same iteration with (59) enabled results in the narrowed ellipsoid, with one of its eigenvectors nearly orthogonal to $\mathbf{d}_T^k$. Thus, $\mathbf{R}_T^{i+1}$ is approximately placed on the tangent to the current section of the friction cone.

It is appropriate to mention another modification investigated in [1]. The authors regularise the operator $\mathbf{I} - \mathbf{M}^k$ in (54), so that it is always invertible and positive definite. This is not pursued here, as it proved not to be consistently beneficial in the numerical realisation. This might be due to the different way of eliminating variables in the current development.

5.3 Complete algorithm

The normal active set strategy from Section 5.1 can now be combined with the tangential linearisation, in order to deliver a complete Newton scheme for the frictional contact problem. As it was mentioned in Section 4.1.3, the nonmonotone character of the adopted contact law results in the need to globalise the Newton scheme. This is provided by means of the nonmonotone line search technique by Grippo *et al.* [26]. The choice seems to be more relevant to the nature of problem at hand. Nevertheless, the simple Armijo's [27] line search will also be investigated. This type of monotone globalisation was applied by Christensen *et al.* [9] in the context of two dimensional linearly elastic problems.

The sequence of iterates of contact reactions is generated according to

$$\mathbf{R}^{k+1} = \mathbf{R}^k + \alpha^k \delta \mathbf{R}^k \tag{60}$$

where $\delta\mathbf{R}^k$ is the search direction, and $\alpha^k \in (0,1]$ is the step size. The search direction results from the semismooth Newton step applied to the system (41). Three ways of calculating $\delta\mathbf{R}^k$ will be discussed. The first one results from the consistent linearisation of (41). The normal active set strategy and the tangential linearisation are combined as follows

Algorithm NEWT

1 $\mathbf{T}^k = \mathbf{W}\mathbf{R}^k + \mathbf{B} - \mathbf{U}^k$

2 $\mathscr{A}_N^k = \left\{\alpha : d_{\alpha N}^k \geq 0\right\}$ $\mathscr{I}_N^k = \{\alpha\} \setminus \mathscr{A}_N^k$

3 $\mathscr{A}_T^k = \left\{\alpha : \left\|\mathbf{d}_{\alpha T}^k\right\| - \mu d_{\alpha N}^k \geq 0 \wedge \alpha \in \mathscr{A}_N^k\right\}$ $\mathscr{I}_T^k = \mathscr{A}_N^k \setminus \mathscr{A}_T^k$

4 **solve** $\Omega^k \delta\mathbf{R}^k = \Pi^k$ **where**

5 **for** $\alpha \in \mathscr{I}_N^k$

$\Omega_{\alpha\alpha}^k = \mathbf{I}$ $\Omega_{\alpha\beta}^k = \mathbf{0}$ $\Pi^k = -\mathbf{R}^k$

6 **for** $\alpha \in \mathscr{A}_N^k$

$\Omega_{\alpha\beta N*}^k = \mathbf{W}_{\alpha\beta N*}$ $\Pi_{\alpha N}^k = -\bar{U}_{\alpha N}^k - T_{\alpha N}^k$

7 **for** $\alpha \in \mathscr{I}_T^k$

$\Omega_{\alpha\alpha T*}^k = \left[\, \mathbf{W}_{\alpha\alpha TT} \;\; \mathbf{W}_{\alpha\alpha NT} + \mathbf{U}_{\alpha T}^k / d_{\alpha N}^k \,\right]$

$\Omega_{\alpha\beta T*} = \mathbf{W}_{\alpha\beta T*}$ $\Pi_{\alpha T}^k = -\left(1 + \rho \bar{U}_{\alpha N}^k / d_{\alpha N}^k\right) \mathbf{U}_{\alpha T}^k - \mathbf{T}_{\alpha T}^k$

8 **for** $\alpha \in \mathscr{A}_T^k \wedge d_{\alpha N} = 0$

$\Omega_{\alpha\alpha T*}^k = \left[\, \mathbf{I} \;\; \mathbf{0} \,\right]$ $\Omega_{\alpha\beta T*}^k = \mathbf{0}$ $\Pi_{\alpha T}^k = -\mathbf{R}_{\alpha T}^k$

9 **for** $\alpha \in \mathscr{A}_T^k \wedge d_{\alpha N} > 0$

$\Omega_{\alpha\alpha T*}^k = \left[\, \left(\mathbf{I} - \mathbf{M}_\alpha^k\right) + \rho \mathbf{M}_\alpha^k \mathbf{W}_{\alpha\alpha TT} \;\; \rho \mathbf{M}_\alpha^k \mathbf{W}_{\alpha\alpha NT} - \mu \mathbf{v}_{\alpha T}^k \,\right]$

$\Omega_{\alpha\beta T*}^k = \rho \mathbf{M}_\alpha^k \mathbf{W}_{\alpha\beta T*}$ $\Pi_{\alpha T}^k = \mu \mathbf{v}_{\alpha T}^k R_{\alpha N}^k - \mathbf{R}_{\alpha T}^k - \rho \mathbf{M}_\alpha^k \mathbf{T}_{\alpha T}^k$

The residual update in the first line above is followed by the selection of active and inactive sets (lines 2, 3). The tangential sets are subsets of the normal active one, $\mathscr{A}_N^k = \mathscr{A}_T^k \cap \mathscr{I}_T^k$. The increment $\delta\mathbf{R}^k$ results from the solution of a linear system in line 4. The system matrix Ω is composed of dense 3×3 blocks $\Omega_{\alpha\beta}$, and has the

same block-sparsity pattern as $\mathbf{W}$. This is of use in the numerical realisation, as the symbolic factorisation of Ω can be computed only once. The structure of rows in line 5 results from the fact that $\mathbf{R}_\alpha^{k+1} = \mathbf{0}$ is assumed on $\mathscr{I}_N^k$ for the full Newton scheme ($\alpha^k = 1$). Thus $\delta\mathbf{R}_\alpha^k = -\mathbf{R}_\alpha^k$ on the inactive normal set. The normal row structure in line 6 [1] results from the reasoning already presented in Section 5.1. In short, it corresponds to the rows of

$$\mathbf{U}^k + \delta\mathbf{U}^k = \mathbf{W}\left(\mathbf{R}^k + \delta\mathbf{R}^k\right) + \mathbf{B} \tag{61}$$

written with the account of $\bar{U}_{\alpha N}^k + \delta U_{\alpha N} = 0$ for $\alpha \in \mathscr{A}_N$. In line 7, the coupling (53) is utilised in order to eliminate the velocity increment $\delta\mathbf{U}_{\alpha T}^k$ from the tangential rows of (61), with $\alpha \in \mathscr{I}_T^k$. As the active sets were defined with the equality inclusion $\geq$, one needs to deal with the active tangential case, corresponding to the zero friction bound (line 8). This is a pragmatic choice motivated by a faster communication during the solution process. In particular, considering a structure composed of contacting bodies with a force applied to only one of them, the above algorithm will assume the frictionless contact for all bodies not directly adjacent to the one with the nonzero force. The next iteration will then start from some nonzero reactions guess for all bodies connected in the contact graph (the nonzero block pattern graph of $\mathbf{W}$). If the sharp inequality $>$ was utilised to define the active sets, the nonzero contact forcing would have to gradually propagate according to the immediate adjacency in the contact graph. Coming back to the line 8, it is seen that the zero tangential response is enforced for $\alpha \in \mathscr{A}_T^k$ and $d_{\alpha N} = 0$. The remaining non-degenerate tangential case ($d_{\alpha N} > 0$) is considered in line 9. Here, the tangential rows of Ω are obtained by substituting $\delta\mathbf{U}_T^k$, calculated from (61) into the tangent relation (54). This way of eliminating variables is motivated by the intention of preserving the impact of modification (59).

In fact, the descent directions provided by the algorithm `NEWT` are not among the most effective, when calculated far from the solution. As a result of numerical experiments aimed at improvement of the global convergence properties, the following hybrid scheme has arisen

Algorithm HYB

1 $\mathbf{T}^k = \mathbf{W}\mathbf{R}^k + \mathbf{B} - \mathbf{U}^k$

2 $\mathscr{A}_N^k = \left\{\alpha : d_{\alpha N}^k \geq 0\right\}\ \ \mathscr{I}_N^k = \{\alpha\} \setminus \mathscr{A}_N^k$

3 $\mathscr{A}_T^k = \left\{\alpha : \left\|\mathbf{d}_{\alpha T}^k\right\| - \mu R_{\alpha N}^k \geq 0\right\}\ \ \mathscr{I}_T^k = \{\alpha\} \setminus \mathscr{A}_T^k$

4 **solve** $\Omega^k \delta\mathbf{R}^k = \Pi^k$ **where**

[1] In the algorithms presented in this section the asterisk subscript "$*$" replaces "all relevant indices", e.g. $\Omega_{N*} = \begin{bmatrix} \Omega_{NT} & \Omega_{NN} \end{bmatrix}$.

5 **for** $\alpha \in \mathscr{I}_N^k$

$\Omega_{\alpha\alpha}^k = \mathbf{I}$ $\Omega_{\alpha\beta}^k = \mathbf{0}$ $\Pi^k = -\mathbf{R}^k$

6 **for** $\alpha \in \mathscr{A}_N^k$

$\Omega_{\alpha\beta N*}^k = \mathbf{W}_{\alpha\beta N*}$ $\Pi_{\alpha N}^k = -\bar{U}_{\alpha N}^k - T_{\alpha N}^k$

7 **for** $\alpha \in \mathscr{I}_T^k$

$\Omega_{\alpha\beta T*} = \mathbf{W}_{\alpha\beta T*}$ $\Pi_{\alpha T}^k = -\mathbf{U}_{\alpha T}^k - \mathbf{T}_{\alpha T}^k$

8 **for** $\alpha \in \mathscr{A}_T^k \wedge d_{\alpha N} = 0$

$\Omega_{\alpha\alpha T*}^k = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}$ $\Omega_{\alpha\beta T*}^k = \mathbf{0}$ $\Pi_{\alpha T}^k = -\mathbf{R}_{\alpha T}^k$

9 **for** $\alpha \in \mathscr{A}_T^k \wedge d_{\alpha N} > 0$

$\Omega_{\alpha\alpha T*}^k = \begin{bmatrix} \left(\mathbf{I} - \mathbf{M}_\alpha^k\right) + \rho \mathbf{M}_\alpha^k \mathbf{W}_{\alpha\alpha TT} & \rho \mathbf{M}_\alpha^k \mathbf{W}_{\alpha\alpha NT} - \mu \mathbf{v}_{\alpha T}^k \end{bmatrix}$

$\Omega_{\alpha\beta T*}^k = \rho \mathbf{M}_\alpha^k \mathbf{W}_{\alpha\beta T*}$ $\Pi_{\alpha T}^k = \mu \mathbf{v}_{\alpha T}^k R_{\alpha N}^k - \mathbf{R}_{\alpha T}^k - \rho \mathbf{M}_\alpha^k \mathbf{T}_{\alpha T}^k$

Contrary to the previous case, the normal and tangential sets are independently set in lines 2 and 3. Also, the bound employed in line 3 is not any more based on the predictor $d_{\alpha N}^k$, but owes to the currently computed normal reaction $R_{\alpha N}^k$. The last difference with regard to the algorithm NEWT is the assumption of $\mathbf{U}_{\alpha T}^{k+1} = \mathbf{0}$ for $\alpha \in \mathscr{I}_T^k$, expressed in line 7. This means, that the strict linearisation of the inactive tangential case is not pursued. The above scheme can be linked to the fixed point technique, presented in the reference work [1]. The major inconsistency is in using the full Newton linearisation for the slip contact, in line 9. In the fixed point approach, the tangential slip relations are linearised according to the Tresca friction model. The linearisation can be obtained from (54), by discarding the term involving δR_N^k and employing a fixed normal bound b_N instead of d_N^k

$$\mathbf{R}_T^k + \left(\mathbf{I} - \mathbf{M}^k\right) \delta \mathbf{R}_T^k + \rho \mathbf{M}^k \delta \mathbf{U}_T^k = \mathbf{v}_T^k \mu b_N \tag{62}$$

where in the intermediate formulae (56), (57), and (59) the normal predictor d_N^k is replaced by the fixed bound b_N. The fixed point approach is summarised below

Algorithm FIX

1 $\mathbf{T}^k = \mathbf{W}\mathbf{R}^k + \mathbf{B} - \mathbf{U}^k$

2 $\mathscr{A}_N^k = \left\{\alpha : d_{\alpha N}^k \geq 0\right\}$ $\mathscr{I}_N^k = \{\alpha\} \setminus \mathscr{A}_N^k$

3 $\mathscr{A}_T^k = \left\{\alpha : \left\|\mathbf{d}_{\alpha T}^k\right\| - \mu b_{\alpha N} \geq 0\right\}$ $\mathscr{I}_T^k = \{\alpha\} \setminus \mathscr{A}_T^k$

4 **solve** $\Omega^k \delta \mathbf{R}^k = \Pi^k$ **where**

5 **for** $\alpha \in \mathscr{I}_N^k$

$\Omega^k_{\alpha\alpha} = \mathbf{I} \;\; \Omega^k_{\alpha\beta} = \mathbf{0} \;\; \Pi^k = -\mathbf{R}^k$

6 **for** $\alpha \in \mathscr{A}_N^k$

$\Omega^k_{\alpha\beta N*} = \mathbf{W}_{\alpha\beta N*} \;\; \Pi^k_{\alpha N} = -\bar{U}^k_{\alpha N} - T^k_{\alpha N}$

7 **for** $\alpha \in \mathscr{I}_T^k$

$\Omega_{\alpha\beta T*} = \mathbf{W}_{\alpha\beta T*} \;\; \Pi^k_{\alpha T} = -\mathbf{U}^k_{\alpha T} - \mathbf{T}^k_{\alpha T}$

8 **for** $\alpha \in \mathscr{A}_T^k \wedge d_{\alpha N} = 0$

$\Omega^k_{\alpha\alpha T*} = \left[\, \mathbf{I} \;\, \mathbf{0} \,\right] \;\; \Omega^k_{\alpha\beta T*} = \mathbf{0} \;\; \Pi^k_{\alpha T} = -\mathbf{R}^k_{\alpha T}$

9 **for** $\alpha \in \mathscr{A}_T^k \wedge d_{\alpha N} > 0$

$\Omega^k_{\alpha\alpha T*} = \left[\, \left(\mathbf{I} - \mathbf{M}^k_\alpha\right) + \rho \mathbf{M}^k_\alpha \mathbf{W}_{\alpha\alpha TT} \;\; \rho \mathbf{M}^k_\alpha \mathbf{W}_{\alpha\alpha NT} \,\right]$

$\Omega^k_{\alpha\beta T*} = \rho \mathbf{M}^k_\alpha \mathbf{W}_{\alpha\beta T*} \;\; \Pi^k_{\alpha T} = \mu \mathbf{v}^k_{\alpha T} b_{\alpha N} - \mathbf{R}^k_{\alpha T} - \rho \mathbf{M}^k_\alpha \mathbf{T}^k_{\alpha T}$

It is seen that the third line of the algorithm `HYB` is a special case of the corresponding line of `FIX`, with $b_{\alpha N} = R^k_{\alpha N}$. This corresponds to an update of the friction bound in every iteration of the fixed point scheme. In practice, at least in case of the current kinematic formulation, that frequent update of the friction bound prevents convergence of the fixed point approach. This happens because the tangential-normal coupling in line 9 results only from the problem kinematics. Nevertheless, the Tresca regularisation usually results in a good global convergence behaviour, provided the updates of $b_{\alpha N}$ are sparse enough. The hybrid approach provides then an intermediate case between the full Newton and the fixed point methods. The poor convergence of the fixed point approach with the most frequent update of normal bounds is remedied by the full linearisation of the tangential slip. This heuristic attempt of synergy between the Newton and fixed point approaches is to be numerically investigated in the next section.

One thing to be noted about all three algorithms is that they result in an unsymmetric systems to be solved for $\delta \mathbf{R}^k$. This can be to some extent remedied by multiplying the two tangential rows given in lines 9 by the operator $\left[\rho \mathbf{M}^k_\alpha\right]^{-1}$. This introduces more symmetry into Ω and seems to be particularly advantageous for the fixed point scheme, where the system matrix becomes gradually symmetric with progressing convergence. On the other hand, the modification (59) does not act any more on the entire row of Ω. The experience shows that (at least within the current formulation) this way of eliminating variables is not effective. Thus, it will

not be further investigated.

The above three schemes need to be embraced by some global criteria of advancing the iterations. In case of NEWT and HYB this will be provided by the mentioned line search technique. The fixed point scheme, being generally better behaved in terms of the global convergence, will only be wrapped into a suitable external loop updating the normal bounds $b_{\alpha N}$. All together, this can be stated as follows

Algorithm SCSOL(ALG, $\sigma, \gamma, \beta, J, \varepsilon, K, \phi, L$**)**

1 $k = 0$

2 **do**

3 $\delta \mathbf{R}^k =$ **ALG()**

4 $\alpha^k = 1$

5 **while** $\mathscr{M}\left(\mathbf{R}^k + \alpha^k \delta \mathbf{R}^k\right) > \max\limits_{0 \le j \le \min(k,J)} \mathscr{M}^{k-j} - 2\gamma\alpha^k \mathscr{M}^k$

$\wedge\, k > 0 \wedge \alpha^k > \beta \wedge$ **ALG** $\neq$ **FIX do** $\alpha^k = \sigma\alpha^k$

6 $\mathbf{R}^{k+1} = \mathbf{R}^k + \alpha^k \delta \mathbf{R}^k$

7 $\mathbf{U}^{k+1} = \mathbf{U}^k + \alpha^k \mathbf{W} \delta \mathbf{R}^k + \mathbf{T}^k$

8 $err = \left\|\delta \mathbf{R}^k\right\| / \left\|\mathbf{R}^{k+1}\right\|$

9 $k = k + 1$

10 **if ALG** $=$ **FIX** $\wedge\, err \le \varepsilon$ **then**

11 $err = \left\|\mathbf{R}_N^k - \mathbf{b}_N\right\| / \left\|\mathbf{R}_N^k\right\|$

12 $\mathbf{b}_N = \mathbf{R}_N^k$

13 **endif**

14 **if ALG** $\neq$ **FIX** $\wedge\, k > L$ **then SCALE(**$\{\rho_\alpha\}, \phi, L$**)**

15 **while** $err > \varepsilon \wedge k \le K$

The argument ALG can be NEWT, HYB or FIX. The next three arguments $\sigma, \gamma, \beta \in (0,1)$ correspond to the line search step. The ε describes numerical accuracy, K bounds the maximal number of iterations, and J is the length of memory buffer used by the line search. The remaining arguments $\phi, L \ge 1$ are used for the purpose of the penalty scaling and will be commented on later in this section. In the third line

above, the current increment of reactions is computed by ALG. The initial scaling parameter α^k is set equal to 1 in the following line. The loop in line 5 corresponds to the nonmonotone line search [26]. Note, that for $J = 0$ it is equivalent to the line search of Armijo's type [27]. Both approaches were originally designed for smooth problems. The analysis suitable for the nonsmooth setting was provided by Ferris and Lucidi [28]. The auxiliary merit function is defined as

$$\mathscr{M}(\mathbf{R}) = \frac{1}{2}\mathbf{C}^T(\mathbf{R},\mathbf{U})\,\mathbf{C}(\mathbf{R},\mathbf{U})\Big|_{\mathbf{U}=\mathbf{W}\mathbf{R}+\mathbf{B}} \tag{63}$$

where (41) was utilised. The symbol $\mathscr{M}^k$ refers to $\mathscr{M}\left(\mathbf{R}^k\right)$. If the minimisation in line 5 is successful (iterations end before $\alpha^k \leq \beta$), it is seen that a monotonic decrease of the merit function is enforced for $J = 0$, while this is not necessarily the case for $J > 0$. The acceptability criterion

$$\mathscr{M}\left(\mathbf{R}^k + \alpha^k \delta\mathbf{R}^k\right) \leq \max_{0\leq j\leq \min(k,J)} \mathscr{M}^{k-j} - 2\gamma\alpha^k\mathscr{M}^k \tag{64}$$

allows for the temporary growth of the merit function if only $J > 0$. At the same time the solution point remains inside of the nested level sets $\mathbf{R}^k \in \Lambda^k \subseteq \Lambda^{k-1}$

$$\Lambda^k = \left\{\mathbf{R} : \mathscr{M}(\mathbf{R}) \leq \max_{0\leq j\leq \min(k,J)} \mathscr{M}\left(\mathbf{R}^{k-j}\right)\right\} \tag{65}$$

The parameter J is then proportional to the allowed extent of the temporary growth of the merit function. Grippo *et al.* [26] prove that this relaxation does not hinder the global convergence, if only some conditions hold (roughly, the merit function must be bounded below, and $\delta\mathbf{R}$ must be a descend direction). At the same time, for the nonmonotone problems this may lead to a faster convergence, as a convergent sequence of iterates does not have to correspond to a monotonically decreasing sequence of function values. The threshold value β is used due to the finite precision of numerical computations (the line search loop exits after a finite number of steps). For this reason the line search cannot be fully robust in practice. In the above algorithm, the line search technique is applied for $k > 0$, which results in $\alpha^0 = 1$. This is a heuristic dictated by an observation, that usually it is more effective to start globalisation from the iterate obtained by the pure Newton step corresponding to an initial residual. In other words, it often happens that the subsequent alphas are "large", while if the line search was performed for $k = 0$, initial alphas often happen to be "small". Finally, it is seen that the line search is omitted for the fixed point scheme. The update of reactions and velocities follows in lines 6, 7. Line 7 corresponds to the Newton step

$$\delta\mathbf{U}^k = \mathbf{W}\delta\mathbf{R}^k + \mathbf{T}^k \tag{66}$$

$$\mathbf{T}^k = \mathbf{W}\mathbf{R}^k + \mathbf{B} - \mathbf{U}^k \tag{67}$$

thus the residual $\mathbf{T}^{k+1}$ is always zero

$$\mathbf{U}^{k+1} = \mathbf{U}^k + \alpha^k \mathbf{W}\delta\mathbf{R}^k + \mathbf{W}\mathbf{R}^k + \mathbf{B} - \mathbf{U}^k = \mathbf{W}\mathbf{R}^{k+1} + \mathbf{B} \tag{68}$$

It is possible to modify UNIL, NEWT, HYB and FIX by replacing the computation of $\mathbf{T}^k$ by the update of velocity $\mathbf{U}^k = \mathbf{W}\mathbf{R}^k + \mathbf{B}$. This, combined with the removal of line 7 from SCSOL still provides a feasible framework. Nevertheless, experience shows that processing the residual is advantageous and results in smaller numbers of iterations. From the lines 8-13 it is seen that in case of the fixed point method, after each convergent run with a fixed normal bound, the bound is updated and the relative error of this update replaces the error controlling the termination of the algorithm. Note also, that scaling is not applied in case of the fixed point scheme (line 14), as we are interested in testing the plainest possible version of this approach.

Algorithms like the one above, where the tangent operator results from a nonsmooth, and to some extent combinatorial structure, are prone to cycling. By this it is meant that for some parameter sets, the algorithm may get caught into a cycle (here corresponding to a sequence of contact states) preventing further convergence. In the context of frictional contact problems, this was mentioned by Alart and Curnier [6], or DeSaxcé and Feng [29]. In general, cycling is more frequent for stiff problems and its occurrence is related to the regularisation parameter ρ, used in predictive formulae (45) and (46). This issue is more thoroughly commented in [6]. In this work, initial values of ρ_α are independently set for each contact $\rho_\alpha = 1/\lambda_{max}(\mathbf{W}_{\alpha\alpha})$, where λ_{max} is the maximal eigenvalue of the diagonal block matrix $\mathbf{W}_{\alpha\alpha}$. This corresponds to convergence criteria of the projected gradient method applied to the set of diagonal sub-problems of (41) [13]. In the course of solution, each change of the tangential contact state (from stick to slip or vice versa) is further penalised by increasing ρ_α. As the maximal value of ρ_α ought to be bounded by $2/\lambda_{max}(\mathbf{W})$ [6], only a finite number (bounded by L) of such increases is performed, although an explicit estimate of $\lambda_{max}(\mathbf{W})$ is not accounted for. In order to avoid excessive number of heuristic parameters, L provides also the lower iterations bound after which the scaling is applied. This strategy seems to be sufficient in practice. The SCALE routine is summarised below (initial l_α are assumed equal zero)

Algorithm SCALE($\{\rho_\alpha\}, \phi, L$)

1 **for each** α **do**

2 **if** $l_\alpha < L \wedge (stick \rightarrow slip \vee slip \rightarrow stick)_\alpha$ **then**

3 $\rho_\alpha = \phi\rho_\alpha$

4 $l_\alpha = l_\alpha + 1$

5 **endif**

6 **endfor**

It is relevant to ask why does not the line search procedure suffice to avoid cycling. In Section 4.1.3 it was shown that the single contact point problem behaves in a non-monotone way. Using the unconstrained minimisation analogy, one could say that the merit function corresponding to this simplest case possesses a region of concavity. For problems with many contact points the corresponding merit functions possess regions of concavity not only juxtaposed from the single point problems, but also created through their kinematic interactions. In other words the problem becomes highly nonlinear. Theoretically, enforcing a monotone decrease of the merit function ($J = 0$) should guarantee a descent towards the local minimum. In practise though, the line search loop is forced to end after a relatively small number of steps (one is not interested in updating the solution with α^k close to the numerical zero). After an "unfinished" search the solution point may jump to a neighbouring hill. This process may continue in a cyclic manner. The nonmonotone line search ($J > 0$) only increases the probability of such scenario (nevertheless, it is potentially beneficial otherwise).

6 Performance study

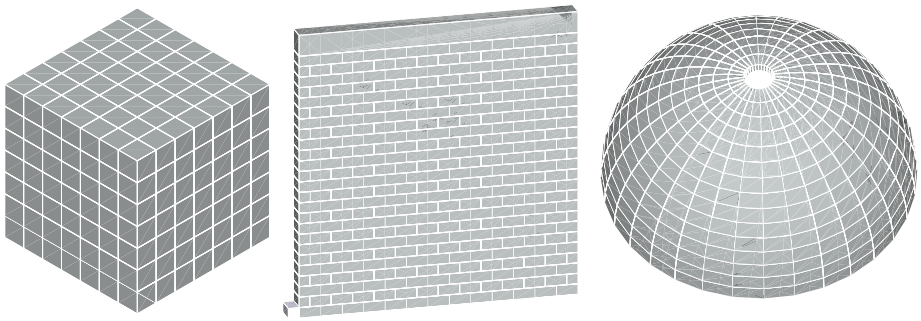

Figure 4. The cube, wall and dome assemblies, all resting on a rigid foundation. Cube units have dimensions $0.1 \times 0.1 \times 0.1\,m$, and are subjected to gravity of $(2,2,-10)\,m/s^2$. Wall units are of dimensions $0.2 \times 0.2 \times 0.1\,m$, and are subjected to gravity of $(0,0,-10)\,m/s^2$, along with the upper bar load of $(0,0,-30)\,kN$. The inner radius of dome is $10\,m$, and the thickness is $0.6\,m$. It deforms under the gravity of $(0,0,-10)\,m/s^2$. The lower left corner of the wall is restrained by a rigid cubic obstacle. Material properties are $15.5\,GPa$ for Young modulus, 0.2 for Poisson ratio, and $2200\,kg/m^3$ for the density.

Three examples are studied for three sizes of assemblies and a range of friction coefficients. The two-dimensional wall example corresponds to the experimental setup by Lourenco *et al.* [30]. The three-dimensional cube, and the dome examples have been selected to picture convergence for various geometrical placements of

contact points. The focus is on the numerical properties of SCSOL, rather then on the mechanical response of test examples. Assembly geometries, loading conditions and material properties are given in Figure 4.

The time stepping from Section 3 is employed. As a quasi-static response is considered, inertia properties were scaled in order to impose uniform numerical damping. For the implicit Euler scheme, a reasonable amount of damping can be obtained for $\lambda h \geq 4$, where λ is a selected eigenvalue of $\mathbf{M}^{-1}\mathbf{K}$, h is the time step, and $\mathbf{K}$ is the current stiffness tangent [31]. Here $h = 1$ was assumed, and inertia tensors $\mathbf{E}_0$ were scaled, so that $\lambda_{max}\left(\mathbf{M}^{-1}\mathbf{K}\right) = 4$ for all bodies.

Table 1
Numbers of bodies, contact points, and condition numbers of $\mathbf{W}$.

Example	Bodies	Contacts	$\mathbf{W}$ conditioning
CUBE1	27	63	2E+7
CUBE2	125	325	5E+7
CUBE3	343	931	9E+7
WALL1	56	147	3E+6
WALL2	162	451	4E+7
WALL3	338	963	2E+8
DOME1	60	120	7E+5
DOME2	220	440	2E+6
DOME3	480	960	8E+6

Table 1 summarises numbers of bodies, contact points, and condition numbers of respective $\mathbf{W}$ operators. Assemblies of variable size preserve geometrical features described in Figure 4. The condition numbers were obtained with dgscon routine of the sparse factorisation package SuperLU [32], which was also employed as the linear solver. The condition numbers are high, yet far from singular. Nevertheless, for the wall example the ill-conditioning of $\mathbf{W}$ significantly grows with the structure size. This corresponds to the discussion presented in Section 4.1.1. Conditioning of $\mathbf{W}$ does not directly correlate to that of Ω. In fact $\Omega = \mathbf{W}$ only if all contact points are in the frictional stick state. In most cases $\Omega \neq \mathbf{W}$ and Ω ought to be assembled with some care. As $\mathbf{W}$ corresponds to the inverse of a stiffness matrix, its entries are likely to be quite small ($O\left(10^{-8}\right)$ for example). The linearised constraints though, are usually of the order $O(1)$. For this reason, to prevent ill-conditioning, system rows corresponding to those constraints are scaled by the relevant diagonal entries of $\mathbf{W}$. For example a row $\ldots 0\,1\,0 \ldots \mathbf{R} = \Pi_i$ is replaced by $\ldots 0\,W_{ii}\,0 \ldots \mathbf{R} = W_{ii}\Pi_i$. Generally, scaling is applied to system rows defined in lines 5, 8, and 9 of NEWT, HYB, and FIX. As a result, the condition numbers of Ω are comparable to those of $\mathbf{W}$, provided the scaling of the regularisation parameter ρ_α is not excessive (routine SCALE).

Table 2
Parameters of SCSOL used in the performance study.

σ	γ	β	J	ε	K	ϕ	L
0.9	0.1	0.034	0 or 10	1E-10	1000	10	6

The input parameters of SCSOL are summarised in Table 2. Both the monotone ($J = 0$) and nonmonotone ($J = 10$) variants were investigated. The set of tested friction coefficients was $\mu \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}$. For each discretisation (Table 1), one hundred incremental runs of the time stepping were performed. In all test cases the zero initial guess was used for **R** and **U** for the first run of SCSOL. The consecutive runs started from the previous solution. To report averages of entities spanning several orders of magnitude, the following procedure was applied

$$average = \exp\left(\sum_{i=1}^{n} \log\left(value_i\right)/n\right) \tag{69}$$

where n is either the total number of system solutions (when reporting the conditioning of Ω) or the total number of convergent runs (when reporting the average final value of the merit function $\mathscr{M}$). In the sequel, instead of referring to SCSOL with a particular argument ALG, a direct reference to NEWT, HYB or FIX is sometimes made. The monotone (Armijo's type, $J = 0$) line search based algorithms are denoted by NEWT(A) and HYB(A), while the nonmonotone (Grippo's type, $J = 10$) line search based ones are denoted by NEWT(G), HYB(G).

Table 3
Results for the frictionless case, $\mu = 0$.

Maximal number of iterations	5
Average conditioning of Ω	2E+3
Average final value of $\mathscr{M}$	1E-23

For the frictionless problems SCSOL reduces to UNIL, regardless of the argument ALG. Results for this case are presented in Table 3. For all examples numbers of iterations are smaller than five. It is also seen that the system matrices are rather well behaved. This case can be tackled very efficiently.

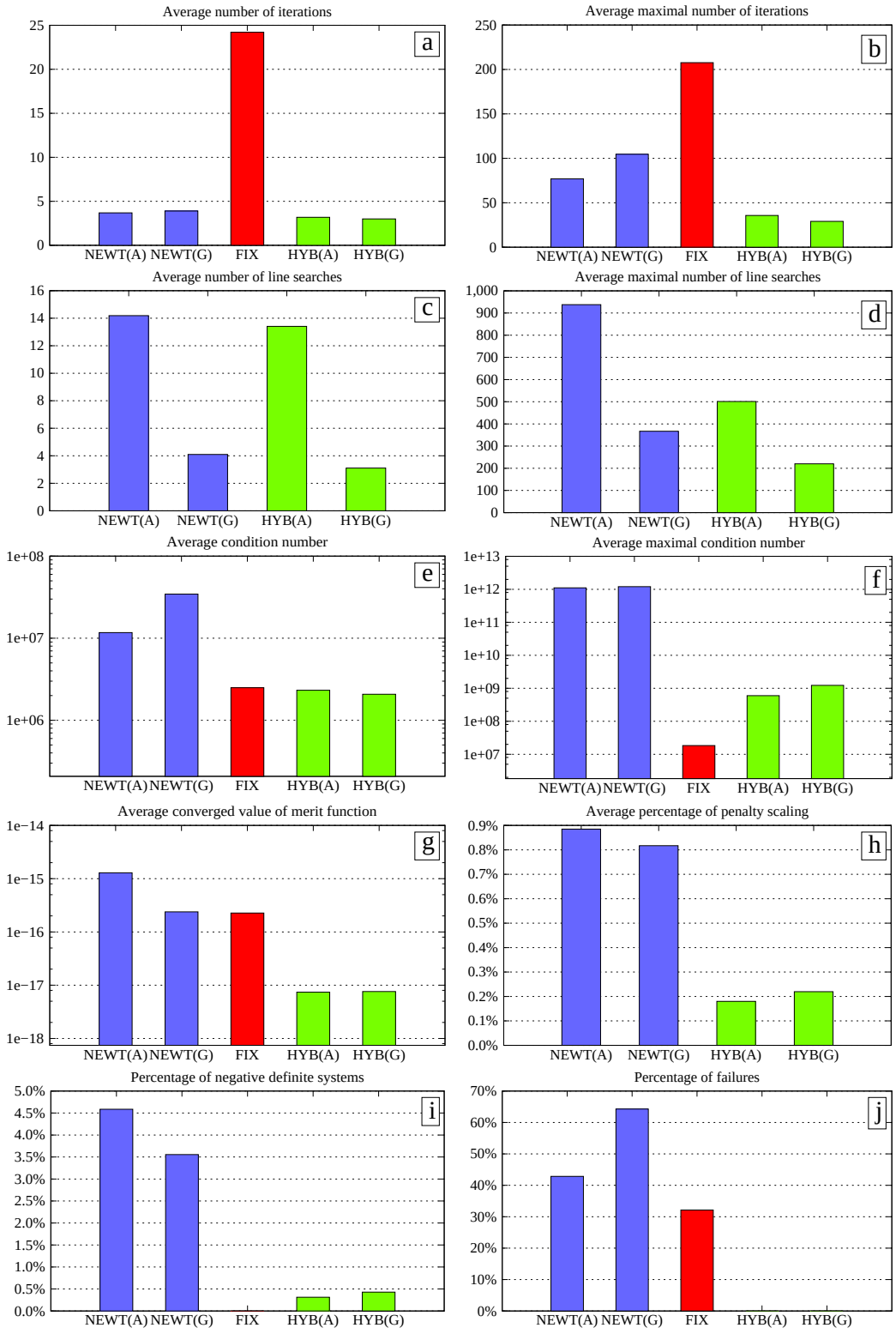

Figure 5. Aggregate statistics for `NEWT`, `FIX` and `HYB` (nine examples, three friction levels $\mu \in \left\{\frac{1}{3}, \frac{2}{3}, 1\right\}$, hundred increments). Note that statistics on line searches and scaling are not applicable to `FIX` and therefore omitted. `NEWT(A)` and `HYB(A)` correspond to the monotone (Armijo's) line search, while `NEWT(G)` and `HYB(G)` to the nonmonotone (Grippo's) line search.

Aggregate statistics for all frictional computations with $\mu \in \left\{\frac{1}{3}, \frac{2}{3}, 1\right\}$ have been summarised in Figure 5. The reported *average* values correspond to the 2700 runs of `SCSOL` (nine examples, three friction levels, hundred increments), while the reported *average maximal* values correspond to the averages of the 27 maxima (nine examples, three friction levels) taken over the hundred increments. As the maximal values usually correspond to the first run of `SCSOL` (starting from the zero initial guess), the average maxima give an estimate of the worst case performance.

In Figure 5 (j) it is seen, that while `NEWT` and `FIX` often failed to converge within the prescribed 1000 iterations, `HYB` is the only scheme which succeeded in all cases. It must be stated though, that while the failures of the full Newton approach correspond to the divergence (unbounded growth of the auxiliary merit function), those of the fixed point method correspond to the insufficient number of converging iterations. At the same time the full Newton method is more prone to divergence, when combined with the nonmonotone line search. This is because the minimisation along a given direction is not always successful ($\beta > 0$), and an unbounded growth of the auxiliary merit function (63) is thus possible. In the nonmonotone search case, a number of such failed minimisations can be stored and the maximal of them used as the reference value in the line search loop, resulting in a grater probability of divergence.

Comparison of the average iteration numbers in Figure 5 (a) shows that the hybrid approach inherits good local convergence properties of the full Newton scheme - the numbers of iterations are similar for both approaches (less then 5). At the same time the fixed point method needs considerably more iterations to converge (25 on average). In Figure 5 (b) it is seen that the average worst case performance of `HYB` compares favourably with the competitors. The nonmonotone version of the line search results in slightly smaller numbers of iterations for the hybrid approach, while it is quite on the contrary for the full Newton scheme (cf. comments in the previous paragraph). It should be noted, that the number of iterations for the fixed point scheme was found to be clearly related to the problem size (although it cannot be deduced from the presented figures).

In Figures 5 (c), (d) it can be seen that the nonmonotone line search consistently results in a smaller average numbers of line searches, when compared to the Armijo's type line search.

In terms of the system conditioning, it is seen in Figure 5 (e) that the hybrid linearisation inherits good properties of the fixed point scheme. The high worst case averages in Figure 5 (f) correspond to the nearly singular systems occurring towards the end of the first solver run. This issue does not represent a significant numerical difficulty, as SuperLU is capable of tackling ill-conditioned problems. The ill conditioning of systems produced by `HYB` is milder, compared to those resulting from `NEWT`.

Figure 5 (g) shows that the hybrid scheme on average results in the smallest final values of the auxiliary merit function. This is in relation with the amount of penalty scaling, presented in Figure 5 (h), which is smaller for the hybrid method (the penalty scaling percentage equals, for one solver run, to the percentage of regularisation parameters ρ_α affected by the routine `SCALE`). Similarly, the penalty scaling is related to the average worst case system conditioning presented in Figure 5 (f).

It is seen in Figure 5 (i), that the full Newton scheme generated roughly ten times more negative definite systems, compared with the hybrid method (which produced less then 0.5% of them). Using the unconstrained minimisation analogy, one could say that the full Newton method visits the tops of the hills more frequently then the hybrid scheme. This might to some extent explain its poor robustness.

In conclusion, the full Newton scheme (`NEWT`) appears to be unreliable in our setting, although it performs pretty well, whenever convergent. The fixed point method (`FIX`) performs robustly, and usually deals with well conditioned systems. Nevertheless it does fail to converge within a thousand iterations for relatively elementary test examples. The hybrid linearisation (`HYB`) nearly consistently delivers the best performance, especially when combined with the nonmonotone line search.

7 Closure

Based on the pseudo-rigid continuum model, a multi-body frictional contact problem has been formulated and applied to steady state problems involving some elementary masonry assemblies. The problem was stated as root finding of a nonsmooth vector function of contact forces, for which three solution methods have been investigated. In this context the consistent semismooth Newton scheme and the fixed point Tresca regularisation have proven relatively ineffective. At the same time, the heuristic hybrid linearisation combined with the nonmonotone line search appears to be the fastest and the most robust approach.

The Contact Dynamics method was originally developed for rigid bodies [2], and eventually extended to the finite element case [3]. It remains the matter of the future research to investigate whether the proposed hybrid scheme performs for these classical approaches as well as it does in the pseudo-rigid setting.

Acknowledgements

The authors would like to thank Georg Stadler for the helpful discussion and numerous hints on the semismooth Newton algorithm. We would also like to thank the anonymous referees for their contribution to the improvement of the paper.

Appendix (*contact formulation*)

Bodies are assumed to be convex polyhedrons. The contact point $\mathbf{x}_\alpha$ is located at the mass centre of an initial contact surface between two bodies. During subsequent steps of an incremental simulation, its referential image $\mathbf{X}_\alpha$ remains attached to a selected (master) body $\mathscr{M}_\alpha$ and its projection on the surface of the opposite (slave) body $\mathscr{S}_\alpha$ defines the gap function g_α and the local base $\mathbf{e}_{\alpha k}$. The symbol $\mathscr{C}_\alpha$ stands for totality of entities attached to the α-indexed contact point.

Considering evolution of the multi-body system over the interval $[t,t+h]$ and selecting $\mathscr{B}_i$ (slave) and $\mathscr{B}_j$ (master) in contact $\mathscr{C}_\alpha$, equation (22) can be written down for i and j, subtracted by sides, and multiplied by $\mathbf{H}_\alpha$ resulting in

$$\mathbf{U}_\alpha = \mathbf{B}_\alpha + \sum_\beta \mathbf{W}_{\alpha\beta}\mathbf{R}_\beta \tag{70}$$

where the upper $t+h$ index by $\mathbf{R}$ is omitted, $\mathbf{U}_\alpha$ is the local relative velocity at time $t+h$, $\mathbf{B}_\alpha$ is the local relative velocity at the point $\mathbf{x}_\alpha$ provided there are no contacts attached to $\mathscr{B}_i$ and $\mathscr{B}_j$ (free velocity), $\mathbf{W}_{\alpha\beta}$ is the local inverse of the generalised inertia matrix, and $\mathbf{R}_\beta$ are the average contact reactions over $[t,t+h]$. Definitions follow below

$$\mathbf{U}_\alpha = \mathbf{H}_{i\alpha}\mathbf{u}_i^{t+h} - \mathbf{H}_{j\alpha}\mathbf{u}_j^{t+h} \tag{71}$$

$$\mathbf{B}_\alpha = \mathbf{H}_{i\alpha}\mathbf{A}_i^{-1}h\mathbf{f}_i\left(t+h,\mathbf{q}^t\right) - \mathbf{H}_{j\alpha}\mathbf{A}_j^{-1}h\mathbf{f}_j\left(t+h,\mathbf{q}^t\right) \tag{72}$$

$$\mathbf{W}_{\alpha\beta}\big|_{\alpha\neq\beta} = s_{\alpha\beta}\mathbf{H}_{k_\beta\alpha}h\mathbf{A}_{k_\beta}^{-1}\mathbf{H}_{k_\beta\beta}^T \tag{73}$$

$$\mathbf{W}_{\alpha\alpha} = \mathbf{H}_{i\alpha}h\mathbf{A}_i^{-1}\mathbf{H}_{i\alpha}^T + \mathbf{H}_{j\alpha}h\mathbf{A}_j^{-1}\mathbf{H}_{j\alpha}^T \tag{74}$$

$$k_\beta = \begin{cases} i \text{ if } \mathscr{B}_i \in \mathscr{C}_\beta \\ j \text{ if } \mathscr{B}_j \in \mathscr{C}_\beta \end{cases} \tag{75}$$

$$s_{\alpha\beta} = \begin{cases} -1 \text{ if } \mathscr{B}_{k_\beta} \text{ is } \left(\mathscr{M}_\alpha \wedge \mathscr{S}_\beta\right) \vee \left(\mathscr{S}_\alpha \wedge \mathscr{M}_\beta\right) \\ \ \ 1 \text{ otherwise} \end{cases} \tag{76}$$

References

[1] S. Hüeber, G. Stadler, and B. I. Wohlmuth, "A primal-dual active set algorithm for three-dimensional contact problems with Coulomb friction," *SIAM Journal on Scientific Computing*, vol. 30, no. 2, pp. 572–596, 2007.

[2] J. J. Moreau, "Numerical aspects of the sweeping process," *Computer Methods in Applied Mechanics and Engineering*, vol. 177, no. 3-4, pp. 329–349, 1999.

[3] M. Jean, "The non-smooth contact dynamics method," *Computer Methods in Applied Mechanics and Engineering*, vol. 177, no. 3-4, pp. 235–257, 1999.

[4] J.-S. Pang, "Newton's method for β-differentiable equations.," *Mathematics of Operations Research*, vol. 15, pp. 311–341, 1990.

[5] L. Qi and J. Sun, "A nonsmooth version of Newton's method," *Mathematical Programming*, vol. V58, pp. 353–367, 1993.

[6] P. Alart and A. Curnier, "Mixed formulation for frictional contact problems prone to Newton like solution methods," *Computer Methods in Applied Mechanics and Engineering*, vol. 92, pp. 353–375, 1991.

[7] M. Hintermüller, V. Kovtunenko, and K. Kunisch, "The primal-dual active set method for a crack problem with non-penetration," *IMA J Appl Math*, vol. 69, no. 1, pp. 1–26, 2004.

[8] M. Hintermüller, K. Ito, and K. Kunisch, "The primal-dual active set strategy as a semismooth Newton method," *SIAM Journal on Optimization*, vol. 13, pp. 865–888, 2003.

[9] P. W. Christensen, A. Klarbring, J. S. Pang, and N. Strömberg, "Formulation and comparison of algorithms for frictional contact problems," *International Journal for Numerical Methods in Engineering*, vol. 42, pp. 145–173, 1998.

[10] R. E. Jones and P. Papadopoulos, "Simulating anisotropic frictional response using smoothly interpolated traction fields," *Computer Methods in Applied Mechanics and Engineering*, vol. 195, pp. 588–613, January 2006.

[11] S. Hüeber and B. I. Wohlmuth, "A primal-dual active set strategy for non-linear multibody contact problems," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, pp. 3147–3166, 2005.

[12] M. Ainsworth and L. A. Mihai, "A comparison of solvers for linear complementarity problems arising from large-scale masonry structures," *Appl. Math.*, vol. 51, no. 2, pp. 93–128, 2006.

[13] M. Renouf and P. Alart, "Conjugate gradient type algorithms for frictional multi-contact problems: Applications to granular materials," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, pp. 2019–2041, 2005.

[14] M. Renouf, F. Dubois, and P. Alart, "A parallel version of the non smooth contact dynamics algorithm applied to the simulation of granular media," *Journal of Computational and Applied Mathematics*, vol. 168, pp. 375–382, 2004.

[15] B. Chetouane, F. Dubois, M. Vinches, and C. Bohatier, "Nscd discrete element method for modelling masonry structures," *International Journal for Numerical Methods in Engineering*, vol. 64, pp. 65–94, 2005.

[16] F. Jourdan, P. Alart, and M. Jean, "A Gauss-Seidel like algorithm to solve frictional contact problems," *Computer Methods in Applied Mechanics and Engineering*, vol. 155, pp. 31–47, 1998.

[17] M. Wosle and F. Pfeiffer, "Dynamics of multibody systems containing dependent unilateral constraints with friction," *Journal of Vibration and Control*, vol. 2, pp. 161–192, 1996.

[18] F. Pfeiffer, M. Foerg, and H. Ulbrich, "Numerical aspects of non-smooth multibody dynamics," *Computer Methods in Applied Mechanics and Engineering*, vol. 195, pp. 6891–6908, October 2006.

[19] H. Cohen and R. G. Muncaster, *The Theory of Pseudo-rigid Bodies*. New York: Springer, 1988.

[20] J. M. Solberg and P. Papadopoulos, "Impact of an elastic pseudo-rigid body on a rigid foundation," *International Journal of Engineering Science*, vol. 38, pp. 589–603, 2000.

[21] T. R. Nordenholz and O. M. O'Reilly, "On steady motions of isotropic, elastic cosserat points," *IMA Journal of Applied Mathematics*, vol. 60, pp. 55–72, 1998.

[22] E. Kanso and P. Papadopoulos, "Dynamics of pseudo-rigid ball impact on rigid foundation," *International Journal of Non-Linear Mechanics*, vol. 39, pp. 299–309, 2004.

[23] V. Acary and M. Jean, "Numerical modeling of three dimensional divided structures by the non smooth contact dynamics method," in *The Fifth International Conference on Computational Structures Technology* (B. H. V. Topping, ed.), pp. 211–222, Civil-Comp Press, Edimburgh, 2000.

[24] G. Strang, *Linear algebra and its applications*. Brooks Cole, 1988.

[25] F. H. Clarke, *Optimization and Nonsmooth Analysis*. New York: John Wiley and Sons, 1983.

[26] L. Grippo, F. Lampariello, and S. Lucidi, "A nonmonotone line search technique for Newton's method," *SIAM J. Numer. Anal.*, vol. 23, no. 4, pp. 707–716, 1986.

[27] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

[28] M. Ferris and S. Lucidi, "Nonmonotone stabilization methods for nonlinear equations," *Journal of Optimization Theory and Applications*, vol. 81, pp. 53–71, 1994.

[29] G. De Saxcé and Z. Q. Feng, "The bipotential method: a constructive approach to design the complete contact law with friction and improved numerical algorithms," *Mathematical and Computer Modelling*, vol. 28, pp. 225–245, 1998.

[30] P. B. Lourenco, D. V. Oliveira, P. Roca, and A. Orduna, “Dry joint stone masonry walls subjected to in-plane combined loading,” *Journal of Structural Engineering*, vol. 131, pp. 1665–1673, 2005.

[31] T. J. R. Hughes, “Analysis of Transient Algorithms with Particular Reference to Stability Behavior,” in *Computational methods for transient analysis* (T. Belytschko and T. Hughes, eds.), pp. 67–155, Elsevier Science, 1983.

[32] J. W. Demmel, S. C. Eisenstat, J. R. Gilbert, X. S. Li, and J. W. H. Liu, “A supernodal approach to sparse partial pivoting,” *SIAM J. Matrix Analysis and Applications*, vol. 20, no. 3, pp. 720–755, 1999.