A GLOBALLY CONVERGENT FILTER–TRUST-REGION METHOD FOR LARGE DEFORMATION CONTACT PROBLEMS

JONATHAN YOUETT, OLIVER SANDER, AND RALF KORNHUBER

Abstract. We present a globally convergent method for the solution of frictionless large deformation contact problems involving hyperelastic materials. For the discretisation we apply the dual mortar method which is known to be more stable than node-to-segment approaches. The resulting non-convex constrained minimisation problems are solved using a filter–trust-region scheme. This method combines several techniques from non-linear optimisation to achieve global convergence towards first-order optimal points. To speed up the method inexact linearisations of the non-penetration constraint are used whenever the current iterate is far away from a critical point. A monotone multigrid method is applied for the fast solution of the constrained Newton problems.

1. Introduction

Although large deformation contact problems arise in many important industrial applications, only very few methods exist for their fast and robust solution. All of them have their advantages and disadvantages. In the last decade node-to-segment methods have been very popular where non-penetration is only enforced at the Lagrange points [1, 21, 15, 16]. Even for linear problems these approaches show unphysical oscillations [29]. When sliding occurs these methods generate artificial jumps in the contact forces, which lead to instabilities and possible divergence [23].

Alternatively, the use of the variationally consistent mortar discretisation, where non-penetration is enforced in a weak sense, has become more and more popular [19, 23]. The idea is to discretise a variational formulation of the constraints, which in the continuous case is equivalent to enforcing the point wise inequality in the L^2-sense. This approach has the advantage that locking is avoided and forces are exactly transferred during contact, i.e., the patch test is fulfilled [23].

The prevailing method for the solution of the resulting algebraic non-linear system are primal–dual active set strategies [14, 17, 25] and penalty methods [23]. For these methods only local convergence can be expected [18]. Furthermore, the resulting linearised Newton problems can be indefinite due to the possible non-convexity of the strain energy. In this work we apply a filter–trust-region method that can be shown to converge globally to first-order optimal points [9]. The filter technique ensures asymptotic fulfilment of the non-linear non-penetration constraints by rejecting iterates that are neither improving the energy nor the infeasibility of all previous iterations. The trust-region method provides a natural way to handle indefiniteness of the linearised problems by introducing an automatic damping strategy.

A priori, the Newton problems of such a method are quadratic minimisation problems with convex inequality constraints. Such problems are in general expensive to solve. We extend an efficient multigrid strategy originally introduced for contact problems in small strain elasticity. In the case of large deformation contact problems, the constraints consist of two parts: the trust-region constraint and the linearised

This work has been done within a DFG Matheon project funded by ECMath.

contact condition. The trust-region constraint is a set of individual bound constraints if the trust-region is defined in terms of the infinity norm. To decouple the contact constraints we extend the technique used in [19] for the small strain case to the setting of large deformations. This decoupling of the non-penetration constraints requires the computation of a LU decomposition of the non-mortar matrix during each iteration. The additional computational cost of the decomposition is negligible but due to fill-in the transformed quadratic problems have a dense blocks corresponding to the degrees of freedom on the contact boundaries. Several different non-smooth multigrid algorithms have been proposed in the literature. The monotone multigrid method can handle problems of the given type, but the implementation for contact problems is difficult [19]. Instead we propose to combine it with a TNNMG method, which cannot handle indefinite problems, but which is much easier to implement [12]. The resulting scheme can be shown to be globally convergent even for indefinite problems. To accelerate the filter method we use an inexact constraint linearisation that by using the bi-orthogonality of the dual mortar functions leads to a diagonal structure avoiding the LU decomposition. By controlling the additionally introduced error the global convergence can be preserved.

This paper is organised as follows: In Section 2 the static large deformation contact problem is described and the weak formulation is derived. In Section 3 we shortly summarise the dual mortar discretisation of the problem. In Section 4 we describe basic sequential quadratic programming and introduce the inexact constraint linearisation together with the decoupling strategy. In Section 5 we then depict the filter–trust-region algorithm. We prove its global convergence and show how inexact linearisations can be used to speed up the scheme. The final Section 6 is dedicated to a numerical example.

2. Static large deformation contact problems

In this section we will briefly summarise the equations of equilibrium of two non-linear hyperelastic bodies subject to mutual contact. A more detailed introduction can be found, e.g., in [20].

2.1. **Strong formulation.**

Let $\Omega^i \subset \mathbb{R}^d$, $d = 2, 3$ denote the reference configurations of two bodies. Assume that the boundaries of the domains are smooth enough for the outer unit normal fields $\mathbf{n}_R^i : \partial\Omega^i \longrightarrow \mathbb{R}^d$ to exist everywhere. Let the boundaries be decomposed into disjoint relatively open sets $\partial\Omega^i = \bar{\Gamma}_D^i \cup \bar{\Gamma}_N^i \cup \bar{\Gamma}_C^i$ corresponding to Dirichlet, Neumann and possible contact parts. We assume that Γ_D^i has positive $(d-1)$-dimensional measure for $i = 1, 2$ and that Γ_D^1 is compactly embedded in $\partial\Omega^1 \setminus \overline{\Gamma_C^1}$.

In the following, unindexed variables are used to denote quantities defined over both bodies. For example $\Omega = \left(\Omega^1, \Omega^2\right)$ denotes the reference configuration of both bodies together. Neglecting the inertia terms the balance of linear momentum yields the following system of partial differential equations in reference coordinates for the deformation function $\boldsymbol{\varphi} = (\boldsymbol{\varphi}^1, \boldsymbol{\varphi}^2) : \Omega \longrightarrow \mathbb{R}^d$

$$\begin{aligned} \operatorname{div} \mathbf{P}(\boldsymbol{\varphi}) + \mathbf{f} &= 0 && \text{in } \Omega, \\ \mathbf{P}(\boldsymbol{\varphi})\mathbf{n}_R &= \mathbf{t} && \text{on } \Gamma_N, \\ \boldsymbol{\varphi} &= \boldsymbol{\varphi}_D && \text{on } \Gamma_D. \end{aligned}$$

The functions $\mathbf{f} \in \mathbf{L}^2(\Omega)$ and $\mathbf{t} \in L^2(\Gamma_N)^d$ are prescribed external volume and traction force densities, which are assumed to be dead loads, i.e., independent of the deformation. The function $\boldsymbol{\varphi}_D \in C(\Gamma_D)^d$ specifies the Dirichlet boundary conditions.

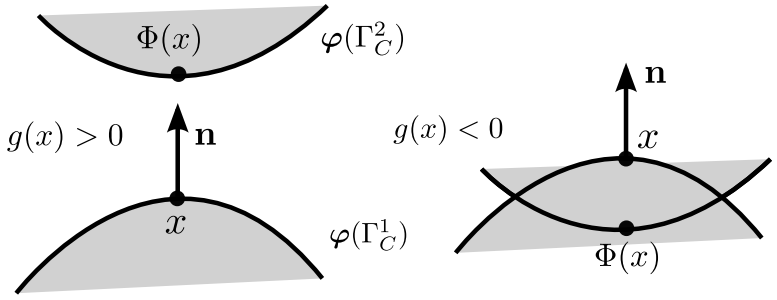

FIGURE 1. *Left: Feasible configuration. Right: Unfeasible configuration.*

Let $\mathrm{Mat}^+(d)$ be the set of second-order tensors with positive determinant. The first Piola–Kirchhoff stress tensor field $\mathbf{P} : \Omega \longrightarrow \mathrm{Mat}^+(d)$ represents the internal stress that arises due to external loading and boundary conditions. In the following, we will only consider hyperelastic continua, i.e., materials for which there exists a stored energy functional $W(x, F) : \Omega \times \mathrm{Mat}^+(d) \longrightarrow \mathbb{R}$ that links the stresses to the deformation via

$$\frac{\partial W}{\partial F}(x, \nabla\boldsymbol{\varphi}(x)) = \mathbf{P}(x, \nabla\boldsymbol{\varphi}(x)).$$

We assume that the hyperelastic energy is penalising any violation of the orientation-preserving condition

$$\det \nabla\boldsymbol{\varphi}(x) > 0 \quad \forall\, x \in \Omega, \tag{1}$$

in the sense that

$$W(x, \nabla\boldsymbol{\varphi}(x)) = \infty \quad \text{if } \det \nabla\boldsymbol{\varphi}(x) \leq 0.$$

As a consequence we will not explicitly enforce (1) as a hard constraint.

The subsets Γ_C^i denote the parts of the boundaries where contact may occur. Contact constraints are naturally formulated on the deformed domain. Let $\mathbf{n}^i$ denote the outer unit normal fields on the deformed contact boundaries $\gamma_C^i := \boldsymbol{\varphi}^i(\Gamma_C^i)$. Modelling of non-penetration can be done in several ways, depending on which projection is chosen to identify the contact surfaces with each other. In earlier papers the closest-point projection, minimising the Euclidean distance, was used [22, 20, 30]. Recently using the normal projection along $\mathbf{n}^1$ has become more popular [23, 25, 17, 26]. In the following we only consider the normal projection approach, although also the closest-point projection can be used in the discretisation and algorithm developed hereafter. The deformed contact boundaries are identified with each other through the projection $\Phi : \gamma_C^1 \longrightarrow \gamma_C^2$

$$\Phi(x) = \min_{\mu \in \mathbb{R}} (x + \mu \mathbf{n}^1(x)) \quad \text{s.t. } x + \mu \mathbf{n}^1(x) \in \gamma_C^2. \tag{2}$$

The resulting normal distance function or *gap function* $g : \gamma_C^1 \longrightarrow \mathbb{R}$ is given by

$$g(x) := \mathbf{n}^1(x) \cdot (\Phi(x) - x), \tag{3}$$

and non-penetration of the bodies is enforced by requiring

$$g(x) \geq 0 \quad \forall\, x \in \gamma_C^1, \tag{4}$$

which is illustrated in Figure 1.

So far the non-penetration constraint was derived only from a kinematical point of view. To investigate the effect of these constraints on the elastic system we examine resulting contact forces in more detail. Consider the *Cauchy stress tensor*

$\boldsymbol{\sigma}(\boldsymbol{\varphi}) := \det(\nabla\boldsymbol{\varphi})^{-1}\mathbf{P}(\boldsymbol{\varphi})\nabla\boldsymbol{\varphi}^T$ which expresses the stress relative to the deformed configuration $\boldsymbol{\varphi}(\Omega)$. The *Cauchy traction*

$$\mathbf{t}_C := \boldsymbol{\sigma}(\boldsymbol{\varphi}^1)\mathbf{n}^1,$$

then represents the contact forces on Γ_C^1. First, the Cauchy traction can be decomposed into a normal and a tangential part

$$\mathbf{t}_C = t_N\mathbf{n}^1 + \mathbf{t}_T,$$

where

$$t_N := \mathbf{n}^1 \cdot (\boldsymbol{\sigma}(\boldsymbol{\varphi}^1)\mathbf{n}^1),$$
$$\mathbf{t}_T := \mathbf{t}_C - t_N\mathbf{n}^1.$$

We consider frictionless contact only so the tangential traction $\mathbf{t}_T$ at the contact boundary vanishes

$$\mathbf{t}_C = t_N\mathbf{n}^1.$$

The contact normal stresses fulfil the following *Karush–Kuhn–Tucker conditions*

$$t_N \le 0, \quad g \ge 0, \quad g \cdot t_N = 0,$$

where the first one states that traction is a pressure and the last one is the *complementary condition*.

2.2. **Weak formulation.** The equilibrium configurations of hyperelastic continua are characterised as local minimisers of the energy functional

$$\mathcal{J}(\boldsymbol{\varphi}) := \int_\Omega W(\boldsymbol{\varphi}) - \mathcal{F}(\boldsymbol{\varphi})\,dx - \int_{\Gamma_N} \mathcal{G}(\boldsymbol{\varphi})\,ds, \tag{5}$$

where $\mathcal{F}$ and $\mathcal{G}$ are potentials of the external forces. Let $\mathbf{H}^1_D(\Omega) := \left(H^1_D(\Omega)\right)^d$ denote the d-valued Sobolev space of weakly differentiable functions fulfilling the Dirichlet boundary conditions in the sense of traces. Then the weak formulation of static hyperelasticity is given by:

$$\boldsymbol{\varphi} \in \mathbf{H}^1_D(\Omega): \quad \mathcal{J}(\boldsymbol{\varphi}) \le \mathcal{J}(\mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}^1_D(\Omega), \tag{6}$$

cf. [5, Theorem 4.1-2]. Existence of minimisers has been shown for the case of a poly-convex and coercive strain energies [5, Theorem 7.7-1]. The corresponding first-order optimality condition is the principle of virtual work

$$\int_\Omega \mathbf{P}(\boldsymbol{\varphi}) : \nabla\mathbf{v} - \mathbf{f}\mathbf{v}\,dx - \int_{\Gamma_N} \mathbf{t}\mathbf{v}\,ds = 0.$$

In a Sobolev space setting, the non-penetration constraint (4) takes the form

$$g(x) \ge 0 \quad \forall\, x \in \gamma_C^1 \quad \text{almost everywhere.} \tag{7}$$

In anticipation of the mortar discretisation we rewrite this condition in a variationally consistent form. We assume that the gap function g is smooth enough such that

$$\boldsymbol{\varphi} \mapsto g(\boldsymbol{\varphi}),$$

maps every $\mathbf{H}^1_D(\Omega)$ function to a function in $W := H^{\frac{1}{2}}(\gamma_C^1)$. We denote the dual trace space by

$$M := H^{\frac{1}{2}}(\gamma_C^1)',$$

and the cones of positive functions and dual functionals by

$$W^+ := \{v \in W : v \ge 0 \text{ a.e.}\},$$
$$M^+ := \left\{\mu \in M : \langle \mu, v\rangle_{M\times W} \ge 0,\, \forall\, v \in W^+\right\},$$

where $\langle \cdot, \cdot \rangle_{M \times W}$ denotes the dual paring. We will just write $\langle \cdot, \cdot \rangle$ if the spaces are clear from the context. Now, the resulting weak formulation of the non-penetration constraint (7) is given by

$$\langle \mu, g(\boldsymbol{\varphi}) \rangle \geq 0 \quad \forall\, \mu \in M^+. \tag{8}$$

Proposition 2.1. *The weak non-penetration constraint* (8) *is equivalent to* (7)

Proof. [31, Proposition 1.3.4] □

We denote by

$$\mathcal{K} := \left\{ \boldsymbol{\varphi} \in \mathbf{H}^1_D(\Omega) : \langle \mu, g(\boldsymbol{\varphi}) \rangle \geq 0 \quad \forall\, \mu \in M^+ \right\},$$

the closed non-convex set of feasible deformations. The weak formulation of the large deformation contact problem now reads:

$$\boldsymbol{\varphi} \in \mathcal{K} : \quad \mathcal{J}(\boldsymbol{\varphi}) \leq \mathcal{J}(\mathbf{v}) \quad \forall\, \mathbf{v} \in \mathcal{K}. \tag{9}$$

To our knowledge the existence of solutions is still an open question.

3. Discretisation

In this section we will describe the discretisation of the minimisation problem (9) using first-order Lagrangian finite elements, and dual mortar elements for the contact constraints. Let $\mathcal{T}_h$ be a shape-regular grid of the bodies $\Omega = \Omega^1 \cap \Omega^2$, and $\mathcal{N}(\mathcal{T}_h)$ the set of vertices. The space of d-valued first-order finite elements is $\mathbf{S}_h$ and for each node $p \in \mathcal{N}(\mathcal{T}_h)$ the scalar nodal basis functions corresponding to p is denoted by $\psi_p \in \mathbf{S}_h$. We discretise the hyperelasticity problem (6) by replacing the solution space $\mathbf{H}^1_D(\Omega)$ by the finite dimensional subspace $\mathbf{S}_{D,h} := \mathbf{S}_h \cap \mathbf{H}^1_D(\Omega)$.

3.1. Dual mortar discretisation of the contact constraints.

We discretise the mortar cone M^+ by dual functions. This technique has already been applied for small displacement contact problems [19]. In the large deformation case it is used to condensate the Lagrange multipliers in active-set approaches [25, 14]. In the large deformation setting they lose their ability to decouple the obstacle constraints, unless appropriate inexact linearisations are used (Section 4.1). The dual functions are constructed such that they fulfil the following bi-orthogonality condition: Let θ_q denote the dual basis function corresponding to the node $q \in \mathcal{N}(\gamma_C^1)$. Then,

$$\int_{\gamma_h} \psi_p \, \theta_q \, ds = \delta_q^p \int_{\gamma_h} \psi_p \, ds, \quad \forall\, p, q \in \mathcal{N}(\gamma_C^1), \tag{10}$$

where γ_h denotes the discretisation of the deformed contact boundary γ_C^1 and δ_p^q is the Kronecker symbol. The dual functions are discontinuous. In general they are deformation-dependent because the orthogonality condition (10) is posed on the deformed contact boundary. They are constructed as linear combination of Lagrange basis functions by solving a respective linear system (10) element wise [13]. For simplicial grids this can be done independently of the deformation as the determinant of the deformation gradient is constant in this case and cancels out.

$$\theta_{p|T} = (d\psi_p - \sum_{\substack{q \in T \\ q \neq p}} \psi_q)_{|T} \quad \text{for triangle } T \in \mathcal{T}_h.$$

We denote the dual basis by

$$\Theta_h^{\varphi} := \{ \theta_p : p \in \mathcal{N}(\gamma_h) \}. \tag{11}$$

The discrete mortar cone $M_h^+ \not\subset M^+$ is then given by

$$M_h^+ := \left\{ \mu_h \in \operatorname{span} \Theta_h^{\varphi} : \int_{\gamma_h} \mu_h \, v_h \, ds \geq 0 \quad \forall\, v_h \in S_h(\gamma_h), v_h \geq 0 \right\}.$$

This leads to the following weak non-penetration constraint

$$\int_{\gamma_h} g(s)\,\mu_h\,ds \geq 0 \quad \forall\,\mu_h \in M_h^+.$$

With Definition 3 for g we obtain

$$\int_{\gamma_h} \mathbf{n}^1(s)\cdot(\Phi(s)-s)\,\mu_h\,ds \geq 0 \quad \forall\,\mu_h \in M_h^+.$$

For the discretisation of the normal field $\mathbf{n}^1$ we first define vertex normals by averaging the adjacent face normals, i.e., for each vertex $p \in \mathcal{N}(\gamma_h)$ with neighbouring faces $\mathcal{E}(p)$

$$\mathbf{n}_{h,p} = \frac{\sum_{e\in\mathcal{E}(p)} \mathbf{n}_e}{\left\|\sum_{e\in\mathcal{E}(p)} \mathbf{n}_e\right\|}, \tag{12}$$

where $\mathbf{n}_e$ is the face normal of e at the corner p. The discretised normal field is then defined as the finite element function

$$\mathbf{n}_h := \sum_{p\in\mathcal{N}(\gamma_h)} \psi_p \mathbf{n}_{h,p}. \tag{13}$$

This continuous approximation yields a smoother behaviour when sliding occurs compared to using discontinuous element normals, cf. [23]. The resulting discrete non-penetration constraint with

$$g_h(s) := \mathbf{n}_h(s)\cdot(\Phi(s)-s) \tag{14}$$

reads

$$\int_{\gamma_h} g_h(s)\,\mu_h \geq 0, \quad \mu_h \in M_h^+. \tag{15}$$

We denote the corresponding discrete feasible set

$$\mathcal{K}_h = \left\{ \boldsymbol{\varphi}_h \in \mathbf{S}_{D,h} : \int_{\gamma_h} g_h(s)\,\mu_h\,ds \geq 0, \quad \forall \mu_h \in M_h^+ \right\}.$$

Summarising, the fully discrete problems is given by

$$\boldsymbol{\varphi}_h \in \mathcal{K}_h : \mathcal{J}(\boldsymbol{\varphi}_h) \leq \mathcal{J}(\mathbf{v}_h) \quad \forall\,\mathbf{v}_h \in \mathcal{K}_h.$$

3.2. **Algebraic contact problem.** The algebraic representation is derived using the canonical isomorphism I to identify finite element functions with their coefficient (block-)vectors

$$\mathbb{R}^{dn} \ni (\varphi_p)_p = \varphi \mapsto I(\varphi) = \sum_{i=1}^{n} \psi_p \varphi_p \in \mathbf{S}_h,$$

where $n := |\mathcal{N}(\mathcal{T}_h)|$ and $\varphi_p \in \mathbb{R}^d$ for $1 \leq p \leq n$. The algebraic energy is then defined by

$$J(\varphi) := \int_\Omega W(I(\varphi))\,dx - b^T\varphi, \tag{16}$$

where $b = (b_p)$ is given component-wise by

$$b_p := \int_\Omega \mathbf{f}\,\psi_p\,dx - \int_{\Gamma_N} \mathbf{t}\,\psi_p\,dx.$$

The algebraic non-penetration constraint $c : \mathbb{R}^{dn} \longrightarrow \mathbb{R}^m$, with $m := |\mathcal{N}(\gamma_h)|$, is derived by testing (15) with the dual basis functions (11)

$$c_q(\varphi) := \int_{\gamma_h} g_h(s)\theta_q\,ds \quad 1 \leq q \leq m, \tag{17}$$

where c_p denotes the p-th component of c. The non-convex algebraic contact problem then reads

$$\min_{\varphi\in\mathbb{R}^{dn}} J(\varphi), \quad c(\varphi) \geq 0, \tag{18}$$

4. Inexact SQP multigrid methods for contact problems

In this section we describe classical *sequential quadratic programming* (SQP), which is an extension of Newton's method to constrained optimisation. SQP requires the subsequent solution of quadratic sub-problems with linearised constraints. For convex quadratic problems with box-constraints fast and robust non-linear multigrid methods are available [11, 12, 19]. In Section 4.1 we introduce an inexact linearisation of the non-penetration constraint (17). By exploiting the bi-orthogonality (10) these constraints can be decoupled into box-constraints. Next the transformation that is used for the decoupling is then extended to the exact constraints Section 4.2. In contrast to the inexact linearisation this requires the inversion of a $(m \times m)$-matrix using a LU-decomposition. To avoid the resulting fill-in in the stiffness matrix, the inexact linearisations are used whenever the current iterate is far from a critical point. The incorporation of the inexact linearisations and a non-linear multigrid method for the solution of the decoupled inexact and exact SQP sub-problems is described in the next Section 5.
In the following consider the constrained optimisation problem (18). The first-order optimality conditions for this system are given by the following theorem.

Theorem 4.1. *Let φ^* be a local minimiser of* (18)*. If the rows of the active constraint Jacobian, i.e., $\nabla c_p(\varphi^*)$ for which $c_p(\varphi^*) = 0$, are linearly independent, then there exists a Lagrange multiplier $\lambda \in \mathbb{R}^m$ such that*

$$\begin{aligned} \nabla J(\varphi) + \lambda^T \nabla c(\varphi) &= 0, \\ c(\varphi) &\geq 0, \end{aligned} \tag{19}$$

and

$$\lambda \leq 0, \quad c(\varphi^*) \geq 0, \quad \lambda_p c_p(\varphi^*) = 0, \quad 1 \leq p \leq m$$

Proof. [24, Theorem 12.1]. □

The SQP method is derived by applying Newton's method to the first-order optimality system (19) and eliminating the Lagrange multiplier. In the following let $k \in \mathbb{N}$ be the iteration counter in the Newton method. Denote the linearised energy by

$$m_k(u) := f_k u + \frac{1}{2} u^T H_k u,$$

with $f_k := \nabla J(\varphi_k)^T$ and $H_k \in \mathbb{R}^{dn\times dn}$ symmetric. Then, the linearised Newton problems can be reformulated as quadratic minimisation problems for the correction $u_k \in \mathbb{R}^{dn}$

$$\min_{u\in\mathbb{R}^{dn}} m_k(u), \quad C_k u + c_k \geq 0, \tag{QP}$$

where $C_k := \nabla c(\varphi_k)$ and $c_k := c(\varphi_k)$. Quadratic local convergence of this scheme can be proven if H_k is the Hessian of the Lagrangian

$$H_k = \nabla^2 J(\varphi_k) + \lambda_k^T \nabla^2 c(\varphi_k), \tag{20}$$

and if λ_k is a suitable approximation of the Lagrange multiplier (QP), e.g, a least-squares approximation [7, Theorem 15.2.2]. Furthermore, linear convergence still holds when an approximate H_k like the Hessian of the energy

$$H_k = \nabla^2 J(\varphi_k), \tag{21}$$

is chosen. This avoids the need to compute the Lagrange multipliers in (20).

4.1. **Inexact constraint linearisations.** The solution of the SQP sub-problems (QP) is in general expensive due to the linear inequality constraints $\nabla c(\varphi_k)$. In this section we derive an approximation $\widetilde{C}(\varphi_k) \approx \nabla c(\varphi_k)$ that allows to apply fast multigrid methods for the solution of the respective inexact sub-problems. Therefore, we first consider $\nabla c(\varphi)$ in more detail: The linearisation

$$\delta c(\varphi)u := \lim_{t \longrightarrow 0} c(\varphi + tu)$$

of each component $c(\varphi)_p$ of the algebraic contact constraint (17) can be divided into three parts

$$\begin{aligned} \delta c(\varphi)_p = &\int_{\gamma_h} \delta \mathbf{n}_h \cdot (\Phi(s) - s)\, \theta_p \, ds + \int_{\gamma_h} \mathbf{n}_h \cdot \delta \left[(\Phi(s) - s)\, \theta_p \right] ds \\ &+ \int_{\gamma_h} \mathbf{n}_h \cdot (\Phi(s) - s)\, \theta_p \, \delta ds. \end{aligned} \tag{22}$$

The first part involves the linearisation of the nodally averaged normal (13). In the continuous case this term vanishes due to the colinearity with the normal projection $\Phi(s) - s$. The second part is the linearisation of the discretised gap function (14) and the third summand labels the linearisation of the deformation-dependent integral domain, which we denote by δds.

The approximate linearisation that we consider is motivated by the infinitesimal strain framework. In that case the deformation is assumed to be small and hence no distinction is made between deformed and undeformed coordinates $\gamma_h = \Gamma_C^1$. Therefore, the corresponding linearised non-penetration constraint consists only of parts of the second term in (22). In algebraic form this approximation reads

$$\widetilde{C}(\varphi)u := N\widetilde{M}(\varphi)u_C^2 - N\widetilde{D}(\varphi)u_C^1, \tag{23}$$

where the inexact non-mortar resp. mortar matrix, $\widetilde{D}$ resp. $\widetilde{M}$, are given by

$$\begin{aligned} \widetilde{D}(\varphi)_{pq} &:= \mathrm{Id}^{d\times d} \int_{\gamma_h} \psi_q^1 \theta_p \, ds \qquad p, q \in \mathcal{N}(\gamma_C^1), \\ \widetilde{M}(\varphi)_{pq} &:= \mathrm{Id}^{d\times d} \int_{\gamma_h} \psi_q^2 \theta_p \, ds \qquad p \in \mathcal{N}(\gamma_C^1), q \in \mathcal{N}(\gamma_C^2), \end{aligned}$$

the block-diagonal matrix $N \in \mathbb{R}^{m \times dm}$ consists of the averaged normals

$$N_{pp} = \mathbf{n}_{h,p}^T,$$

and u_C^i denote the degrees of freedom corresponding to the vertices on the contact boundaries γ_C^i. Note that this approximation does not exactly coincide with the second term of (22) as we replaced the discretised normal field (13) by the nodally averaged normals (12). Further, the linearisation of the normal projection $\Phi(s)$ was neglected. Using the bi-orthogonality condition (10) one can see that the matrix $\widetilde{D}(\varphi)$ is block-diagonal

$$\widetilde{D}(\varphi)_{pk} = \int_{\gamma_h} \psi_k^1 \theta_p \, ds = \delta_p^k \int_{\gamma_h} \psi_k^1 \, ds.$$

This structure is used in [19] for linearised contact problems to construct a non-standard finite element basis in which the non-penetration constraints decouple: Let O be the block-diagonal matrix consisting of Householder transformations that rotate the first Euclidean basis vector onto the averaged normal $\mathbf{n}_{h,p}$. For simplicity we assume that the coefficient vector is ordered such that $u = (u_C^1, u_C^2, u^I)$ where

u^I denote all interior degrees of freedom. The basis transformation constructed in [19] is given by

$$\widetilde{T}(\varphi) := \begin{pmatrix} O(\varphi)\widetilde{D}(\varphi)^{-1} & \widetilde{D}(\varphi)^{-1}\widetilde{M}(\varphi) & 0 \\ 0 & \mathrm{Id} & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix}. \tag{24}$$

In this transformed basis

$$\{\overline{\psi}\} = \widetilde{T}(\varphi)^T \{\psi\}, \tag{25}$$

the constraints decouple and are given by simple bound constraints

$$\overline{u}^1_{C,0} \leq c(\varphi), \tag{26}$$

where $\overline{u}^1_{C,0} \in \mathbb{R}^m$ is the scalar non-mortar coefficient vector containing for each vertex on the contact boundary only the first component, c.f. [19]. Exchanging the exact constraint linearisation by $\widetilde{C}_k := \widetilde{C}(\varphi_k)$ and using (25) yields the inexact SQP problem

$$\begin{aligned} \min_{u\in\mathbb{R}^{dn}} \ \widetilde{m_k}(u) :=& u^T f_{\widetilde{T},k} + \frac{1}{2} u^T H_{\widetilde{T},k} u, \\ & u^1_{C,0} \leq c_k, \end{aligned} \tag{IQP}$$

with $\widetilde{T}_k = \widetilde{T}(\varphi_k)$ and

$$f_{\widetilde{T},k} := \widetilde{T}_k^T f_k, \qquad H_{\widetilde{T},k} := \widetilde{T}_k^T H_k \widetilde{T}_k.$$

For the case of a convex model $\widetilde{m_k}(u)$ fast multigrid methods are available[11, 12] for the solution (IQP). An extension of this method to the non-convex problem at hand is described in Section 5.1.

Remark 4.2. *A similar approximation was used in* [14, 25] *to simplify the unknown contact forces that show up explicitly in their weak formulation. This allowed the authors to condense the algebraic system by eliminating the Lagrange multipliers at the cost of losing angular momentum conservation. In contrast, we are using the approximation merely to speed-up our method and eventually use exact linearisations.*

4.2. **Decoupling the exact constraints.** The basis transformation (24) relies on the diagonal structure of the non-mortar matrix $\widetilde{D}$. For the exact constraint linearisation (22) in the framework of large deformations this structure is lost. In the following we construct a transformation that extends the technique of the previous section to decouple the constraints in the case of exact linearisations. Therefore, we first split $\nabla c(\varphi)$

$$\nabla c(\varphi) u = D(\varphi) u^1_C + M(\varphi) u^2_C,$$

where

$$D(\varphi) = \frac{\partial c(\varphi)}{\partial \varphi^1_C}, \quad M(\varphi) = \frac{\partial c(\varphi)}{\partial \varphi^2_C}.$$

Note that the non-mortar and mortar matrices are block-matrices with $1 \times d$ blocks. In the inexact case the transformation is constructed by decomposing the respective matrices into block matrices with $d \times d$ blocks and the normal matrix N (23). This is possible due to the nodal approximation of the discretised normal field $\mathbf{n}_h$. The same technique cannot be applied here as the normal field is coupling through the first term in (22). Therefore, we eliminate the normal field by first rotating the non-mortar degrees of freedom into normal and tangential coordinates using Householder reflections O. We denote the nodal and tangential coordinates of u^1_C by

$$O u^1_C =: \begin{pmatrix} u^1_{C,N} & u^1_{C,T} \end{pmatrix},$$

and the corresponding non-mortar matrix by

$$D(\varphi)O(\varphi) =: \begin{pmatrix} D_N(\varphi) & D_T(\varphi) \end{pmatrix}, \tag{27}$$

The normal part $D_N(\varphi) \in \mathbb{R}^{m\times m}$ is a square matrix with scalar entries. In contrast to the inexact non-mortar matrix (23) it is not diagonal any more. To compute the inverse we first construct a LU-decomposition of $D_N(\varphi)$ and then use backward substitution to get $D_N^{-1}(\varphi)$. We now define an analogous transformation to (24) for the exact linearisation. Therefore, let again the coefficient vector be ordered according to $u = (u_C^1, u_C^2, u^I)$. Then the transformation is given by

$$T(\varphi) := \begin{pmatrix} O(\varphi)K(\varphi) & -O(\varphi)L(\varphi) & 0 \\ 0 & \mathrm{Id} & 0 \\ 0 & 0 & \mathrm{Id}, \end{pmatrix}.$$

where, omitting the dependencies on φ, the block-matrices K, L with $d \times d$-blocks are given by

$$K_{pq} := \begin{pmatrix} -\overbrace{(D_N^{-1})_{pq}}^{\in\mathbb{R}} & \overbrace{-(D_N^{-1}D_T)_{pq}}^{\in\mathbb{R}^{1\times d-1}} \\ 0 & \delta_p^q \mathrm{Id} \end{pmatrix}, \quad L_{pq} := \begin{pmatrix} \overbrace{(D_N^{-1}M)_{pq}}^{\in\mathbb{R}^{1\times d}} \\ 0 \end{pmatrix},$$

and δ_p^q denotes the Kronecker delta.

Lemma 4.3. *In the transformed basis*

$$\left\{\overline{\psi}\right\} = T(\varphi)^T \left\{\psi\right\},$$

the linearised non-penetration constraints (22) *decouple*

$$\overline{u}_{C,0}^1 \le c(\varphi), \tag{28}$$

where $\overline{u}_{C,0}^1 \in \mathbb{R}^m$ *is the scalar non-mortar coefficient vector containing for each vertex on the contact boundary only the first component.*

Proof. The linearised constraints (22) transform according to

$$\frac{\partial c(\varphi)}{\overline{\varphi}} = C(\varphi)T(\varphi) = \begin{pmatrix} D(\varphi) & M(\varphi) & 0 \end{pmatrix} T(\varphi).$$

The result follows immediately by (27) and using the fact that the blocks of K and the non-diagonal entries of L are zero except for the first row. □

Now in accordance with (IQP) an equivalent formulation of the SQP problem (QP) is given by

$$\begin{aligned} \min_{u\in\mathbb{R}^{dn}} \; m_{T,k}(u) :=& u^T f_{T,k} + \frac{1}{2} u^T H_{T,k} u, \\ & u_{C,0}^1 \le c_k, \end{aligned} \tag{TQP}$$

with $T_k = T(\varphi_k)$, $c_k = c(\varphi_k)$ and

$$f_{T,k} := T_k^T f_k, \qquad H_{T,k} := T_k^T H_k T_k$$

Note that, compared to the exact SQP problems (TQP), the inexact problems are cheap to assemble as the non-mortar matrix $\widetilde{D}$ is diagonal and hence can be inverted easily. Further, in the exact case dense blocks corresponding to contact degrees of freedom appear in the transformed model due to fill-in of D_N^{-1}. Thus, performing an iteration step for the SQP problem (TQP) will in general take more CPU time compared to (IQP).

Remark 4.4. *If the size of the non-mortar matrix $D(\varphi)$ exceeds the applicability range of direct solvers for the computation of the inverse, one can always go back to solving the original formulation of the Newton problem involving linear inequality constraints* (QP)*. This can be done for example using an interior point method* [27]*, but is in general slower than the multigrid method described in Section 5.1 for large scale systems.*

In the next Section 5 we will now describe how these inexact steps can be incorporated into the SQP method and how global convergence is achieved. To this end, we define the approximation error that one makes by using inexact constraint linearisations as

$$e_k(u) = ||(C_k - \widetilde{C}_k)u||.$$

5. Filter–Trust-Region Methods

In this section we introduce a novel inexact version of the filter–trust-region method. The filter method was first introduced by Fletcher and Leyffer in [9], and it combines several techniques from non-linear optimisation. In contrast to the active-set strategies, widely used in contact mechanics [17, 25, 14], this method can be shown to converge globally even for rather general non-convex strain energy functionals (5).

The SQP method is derived by applying Newton's method to the first-order optimality system (19). As a consequence, only local convergence of the scheme can be expected. Furthermore, away from local solutions of (18) the quadratic models $m_{T,k}(\cdot)$ may not be bounded from below for general non-convex strain energy functionals (18). To handle this possible ill-posedness of the local problems the *trust-region globalisation* adds additional constraints on the step size of the correction

$$\|u\| \leq \Delta_k, \quad k = 0, 1, \dots \tag{29}$$

This step size is adjusted dynamically according to how well the local model $m_{T,k}$ approximates the non-linear functional J. The approximation quality is measured by the scalar quantity

$$\rho_k := \frac{m_k(u_k) - m_k(0)}{J(\varphi_k + u_k) - J(\varphi_k)}$$

Hence, the trust-region Δ_k acts like an automatic damping strategy.

The norm in (29) can be chosen arbitrarily. In this paper we choose the infinity norm as it fits naturally to the decoupled constraints (28) and (26). Incorporating (29) into (TQP) yields

$$\begin{aligned} &\min_{u \in \mathbb{R}^{dn}} m_{T,k}(u), \\ &-\Delta_k \leq u_p \leq c^{\Delta}_{k,p}, \quad 1 \leq p \leq n, \end{aligned} \tag{TRQP}$$

with

$$c^{\Delta}_{k,p} := \begin{cases} \min\{c_{k,p}, \Delta_k\}, & p \in \mathcal{N}_0(\gamma_h) \\ \Delta_k, & \text{else} \end{cases},$$

and

$$\mathcal{N}_0(\gamma_h) := \Big\{ p \in \{1, \dots, n\} : p \text{ first component of a node } q \in \mathcal{N}(\gamma_h) \Big\}.$$

Likewise, incorporating (29) into the inexact problem (IQP) results in

$$\begin{aligned} &\min_{u \in \mathbb{R}^{dn}} \widetilde{m_k}(u), \\ &-\Delta_k \leq u_p \leq c^{\Delta}_{k,p}, \quad 1 \leq p \leq n. \end{aligned} \tag{ITRQP}$$

Finally, to arrive at a globally convergent scheme one has to control the possible infeasibility of the intermediate iterates φ_k, which results from replacing the nonlinear contact constraint (18) by a linearised one. The infeasibility of an iterate can be measured using the non-smooth function

$$\vartheta(\varphi) := \max_{p=1,\dots,m} \{0, -c(\varphi)_p\}\,.$$

In the following we use the abbreviations $J_k := J(\varphi_k)$ and $\vartheta_k := \vartheta(\varphi_k)$ to denote the energy and infeasibility of the k-th iterate.
Let φ_{k+1} be a potential new iterate, i.e., a solution of (TRQP). We say that candidate φ_{k+1} is *dominated* by a previous iterate φ_j, $j < k$ if

$$J_j \le J_{k+1} \quad \text{and} \quad \vartheta_j \le \vartheta_{k+1}, \tag{30}$$

If there is such a φ_j then the candidate should be rejected. The critical case is to decide for the case where

$$J_{k+1} < J_j, \quad \text{but} \quad \vartheta_{k+1} > \vartheta_j, \quad j = 1, \dots, k$$

whether φ_{k+1} should be accepted or not. To overcome this difficulty Fletcher and Leyffer introduced the notion of a *filter* [9].

Definition 5.1. A pair (J_k, ϑ_k) *strongly dominates* (J_i, ϑ_i) if for $\gamma > 0$

$$\vartheta(\varphi_k) < (1 - \gamma_\vartheta)\vartheta_i \text{ and } J(\varphi_k) < J_i - \gamma_\vartheta \vartheta(\varphi_k) \tag{31}$$

A set of tuples (ϑ_i, J_i) is called a *filter* $\mathcal{F}$, if no tuple strongly dominates any other tuple in $\mathcal{F}$ for some fixed constant $0 < \gamma_\vartheta \le 1$, cf. Figure 2

A filter defines a region of acceptable new iterates.

Definition 5.2. An iterate φ_k is *acceptable to the filter* $\mathcal{F}$, if

$$\vartheta(\varphi_k) < (1 - \gamma_\vartheta)\vartheta_i \text{ or } J(\varphi_k) < J_i - \gamma_\vartheta \vartheta(\varphi_k) \quad \forall\, (\vartheta_i, J_i) \in \mathcal{F},$$

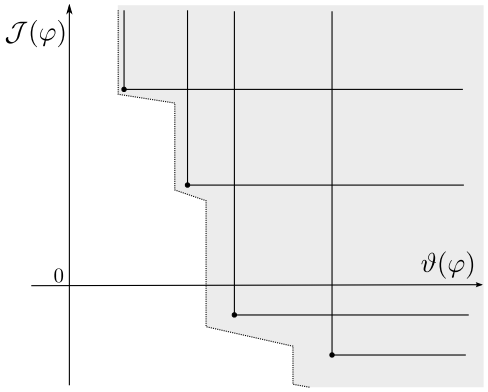

FIGURE 2. *Illustration of a filter with four points. The grey area corresponds to points that are not acceptable.*

Whenever an iterate is added to the filter, all filter elements that are strongly dominated by the new iterate are removed. From this acceptance criterion convergence of the iterates towards the feasible set $\mathcal{K}_h$ can be shown

Lemma 5.3. [7, Lemma 15.5.2] *Let $(\varphi_k)_{k\in\mathbb{N}}$ be a bounded sequence which is added to the filter, then*

$$\vartheta_k \longrightarrow 0.$$

Before stating the filter–trust-region method we define a measure for the optimality of an iterate. This is used on a theoretical side within the global convergence proof and can also serve as a stopping criterion. Define

$$\chi_k = \chi(\varphi_k) := \left| \min_{\substack{u^1_{C,0} \le c(\varphi_k) \\ \|u\| \le 1}} \langle \nabla m_{T,k}(0), u \rangle \right|, \tag{32}$$

and a respective inexact variant

$$\widetilde{\chi}_k = \widetilde{\chi}(\varphi_k) := \left| \min_{\substack{u^1_{C,0} \le c(\varphi_k) \\ \|u\| \le 1}} \langle \nabla \widetilde{m_k}(0), u \rangle \right|. \tag{33}$$

The evaluation of χ_k and $\widetilde{\chi}_k$ involves the solution of a simple linear problem.

Lemma 5.4 (Optimality measures)**.** *The functions χ_k and $\widetilde{\chi}_k$ are optimality measures for* (TQP) *and* (IQP)*, i.e., they are non-negative, continuous and vanish if and only if $u = 0$ is a local optimal point.*

Proof. [7, Theorem 12.1.6] □

We are now ready to formulate the inexact variant of the filter–trust-region method:

Algorithm FTR (Inexact Filter–Trust-Region Step)

Choose Tol, $\kappa_{\mathrm{ac}}, \kappa_\vartheta > 0$, $0 < \gamma_\vartheta \le 1$ \\ *filter parameter*
$0 < \eta_1 < \eta_2 \le 1$, $0 < \gamma_1 < 1 < \gamma_2$ \\ *trust−region parameter*

// Close to optimal points, solve the exact problem
1. If $\widetilde{\chi}_k <$ Tol
Solve $TRQP$*, go to 4.*

//Check if $ITRQP$ *is feasible*
2. If $ITRQP$ *not feasible* {
add φ_k *to the filter,*
call restoration phase $\to (\varphi_{k+1}, \Delta_{k+1})$*, go to 1.*
}

// Check approximation error of inexact constraints
3. Compute u_k *from* $ITRQP$*. If*

$$e_k(u_k) > \kappa_{\mathrm{ac}} \Delta_k^2, \tag{34}$$

then recompute u_k *from* $TRQP$

// Acceptance of iterate
4. Evaluate $\vartheta(\varphi_k + u_k)$*,* $J(\varphi_k + u_k)$

// Reject iterate and reduce trust-region
If $\varphi_k + u_k$ *not acceptable to* $\mathcal{F}$ *or if* $\rho_k < \eta_1$ *and*

$$m_k(0) - m_k(u_k) \ge \kappa_\vartheta \vartheta_k^2, \tag{35}$$

then $\varphi_{k+1} = \varphi_k$, $\Delta_{k+1} = \gamma_1 \Delta_k$

// Accept trial point
$\varphi_{k+1} = \varphi_k + u_k$

// $\vartheta-$type iteration to decrease infeasibility.
If (35) *does not hold*
    add φ_k *to filter*

// Very Successful Iteration, increase trust$-$region
If (35) *and* $\rho_k \geq \eta_2$
    $\Delta_{k+1} = \gamma_2 \Delta_k$

Incorporation of the trust-region constraint can lead to local problems $TRQP$ or $ITRQP$ that are not feasible. This happens when the infeasibility is too large while the trust-region is very small

$$c_p(\varphi_k) > \Delta_k \quad \text{for some } 1 \leq p \leq m.$$

If this happens the filter method enters the so-called *feasibility restoration phase* (FRP). In this phase a new iterate φ_{k+1} and trust-region radius $\Delta_{k+1} > 0$ is computed such that the new φ_{k+1} is acceptable to the filter and the local problem is feasible again. This is done by minimising the infeasibility directly

$$\min_{\varphi \in \mathbb{R}^{dn}} \vartheta(\varphi), \tag{36}$$

e.g., using a semi-smooth trust-region method [7]. To enable the restoration phase to find an acceptable new iterate it is thus crucial that the filter does not contain any feasible points, i.e.,

$$(\vartheta_j, J_j) \in \mathcal{F} \Longrightarrow \vartheta_j \neq 0.$$

This is ensured by condition (35) first proposed in [8]. If (35) fails, the filter iteration is merely trying to reduce the infeasibility, which is called an *ϑ-type* iteration. If the infeasibility is not dominating then the algorithm is using the trust-region functionality to achieve a sufficient reduction of the energy, which is called *J-type* iteration.
The condition (34) ensures that the error that we make by using approximate constraint Jacobians is controlled. If (34) fails the exact Newton problem TRQP has to be solved to ensure global convergence of the method.
Further, whenever the inexact criticality measure $\widetilde{\chi}_k$ is indicating that the iterate is close to a critical point, the exact local problem TRQP has to be solved. This ensures that for a sequence $(\varphi_k)_{k \in \mathcal{N}}$ generated by Algorithm FTR with

$$\lim_{k \longrightarrow \infty} \widetilde{\chi}_k = 0,$$

it also holds that

$$\lim_{k \longrightarrow \infty} \chi_k = 0.$$

5.1. **Solution of the linearised problems.** In this section we describe the multigrid method that we use for the iterative solution of the exact and inexact SQP subproblems (TRQP) and (ITRQP). The goal is to extend the truncated non-smooth Newton multigrid method (TNNMG), developed for convex energies composed of a smooth and a non-smooth block-separable part [12], to the non-convex problems at hand. This is achieved by combining the TNNMG method with the classical monotone multigrid method [11]. In the following we describe the single steps of the resulting algorithm in more detail. For brevity we only use the notation of the

exact problem (TRQP). Let $u_k^\nu \in \mathbb{R}^{dn}$ denote the initial iterate.

1. Non-linear pre-smoothing step
First a projected Gauss–Seidel step is applied to the transformed quadratic obstacle problem TRQP:
Set $w_0 = u_k^\nu$, then for $p = 1, \dots, dn$

$$\begin{aligned} &\alpha_p \in \operatorname*{arg\,min}_{\alpha\in\mathbb{R}} m_k(w_{p-1} + \alpha e_p), \quad \text{s.t. } -\Delta_k \le \alpha + u_{k,p}^\nu \le c_{k,p}^\Delta, \\ &w_p = w_{p-1} + \alpha_p e_p, \end{aligned} \tag{37}$$

where e_p denotes the p'th Euclidean basis vector. The 1-dimensional minimisation problems appearing in (37) can be solved analytically as $m_{T,k}$ is quadratic. In the case of a non-convex component an exact solution is found at one of the boundaries $\alpha_p \in \left\{-\Delta_k, c_{k,p}^\Delta\right\} - u_{k,p}^\nu$. Note that this solution does not have to be unique. The convergence proof for the algorithm does not depend on the particular choice of the solution. Hence, in the case of multiple solutions we always chose the solution on the right boundary of the domain $c_{k,p} - u_{k,p}^\nu$ if it decreases the energy or we stay at the initial position. We denote by $u_k^{\nu+\frac{1}{2}} = u_k^\nu + w_{dn}$ the pre-smoothed iterate.
2. Truncated defect problem
To accelerate the convergence of the multigrid method the active components

$$\mathcal{A}(u) := \left\{p \in \{1, \dots, dn\} : u_p = -\Delta_k \text{ or } u_p = c_{k,p}^\Delta\right\},$$

are truncated after pre-smoothing [11]. The resulting truncation matrix $Q^\nu \in \mathbb{R}^{dn\times dn}$ is given by

$$Q_{pq}^\nu := \begin{cases} 1 & p = q \text{ and } p \notin \mathcal{A}(u_k^{\nu+\frac{1}{2}}), \\ 0 & \text{else.} \end{cases}$$

At this point in the TNNMG method all constraints are dropped and a linear problem is solved. To regain feasibility the resulting correction is projected back onto the feasible set. We extend this idea to the problem at hand by neglecting only the constraints that correspond to the linearised non-penetration constraint. Note again, that the trust-region constraints are necessary to control possible unboundedness of the quadratic model and hence they should be kept. The truncated defect problem in nodal coordinates reads

$$\begin{aligned} &\min_{d\in\mathbb{R}^{dn}} d^T r_k^\nu + \frac{1}{2} d^T H_k^\nu d, \\ &-\Delta_k \le d_p + u_{k,p}^{\nu+\frac{1}{2}} \le \Delta_k, \quad 1 \le p \le dn, \end{aligned} \tag{38}$$

where

$$r_k^\nu := T_k^{-T} Q^\nu (f_{T,k} - H_{T,k} u_k^{\nu+\frac{1}{2}}), \quad H_k^\nu := T_k^{-T} Q^\nu H_{T,k} Q^\nu T_k^{-1}.$$

The inverse transformations can be computed explicitly by

$$T^{-1} = \begin{pmatrix} UO & -V & 0 \\ 0 & \mathrm{Id} & 0 \\ 0 & 0 & \mathrm{Id}, \end{pmatrix}. \tag{39}$$

where

$$U_{pq} := \begin{pmatrix} -\overbrace{(D_N)_{pq}}^{\in\mathbb{R}} & -\overbrace{(D_T)_{pq}}^{\in\mathbb{R}^{1\times d-1}} \\ 0 & \delta_p^q \mathrm{Id} \end{pmatrix}, \quad V_{pq} := \begin{pmatrix} \overbrace{M_{pq}}^{\in\mathbb{R}^{1\times d}} \\ 0 \end{pmatrix},$$

Likewise in the inexact case it is given by

$$\widetilde{T}^{-1} = \begin{pmatrix} O & -O\widetilde{D}^{-1}\widetilde{M} & 0 \\ 0 & \mathrm{Id} & 0 \\ 0 & 0 & \mathrm{Id}, \end{pmatrix}. \tag{40}$$

Note that (39) and (40) are sparse block matrices.

3. Coarse grid correction
For the approximate solution of the defect problem (38) a standard monotone multigrid method can be used which is described in detail in [11, Algorithm 5.10]. Only the Gauss–Seidel smoother have be adjusted to the non-convexity according to (37). A monotone multigrid for finite strain contact problems is described in [19], where also the coarse grid correction is computed in transformed coordinates. Compared to our approach this has the disadvantage that the mortar matrices are required on each grid level. Further, numerical experiments in the small strain framework suggest that the fast asymptotic convergence phase of the multigrid is reached faster if the contact constraints are ignored in the defect problem [12].
In the following we denote the coarse correction computed by performing one (or a few) monotone multigrid iterations to (38) by d^ν.
The correction is first transformed back into the coordinates in which the linearised non-penetration constraints decouple

$$\bar{d}^\nu := T_k^{-1} d^\nu.$$

To regain feasibility the correction is projected onto the defect obstacles of (TRQP), i.e., $\bar{d}^\nu$ such that

$$-\Delta_k \le u_k^{\nu+\frac{1}{2}} + \bar{d}^\nu \le c_k^\Delta. \tag{41}$$

4. Line search
The projection of the correction onto the feasible set can lead to an increase in energy. Therefore, to ensure energy decrease of the iteration step a line search is performed

$$\alpha^\nu \in \operatorname*{arg\,min}_{\alpha\in\mathbb{R}} m_k(u_k^{\nu+\frac{1}{2}} + \alpha\bar{d}^\nu), \quad \text{s.t. } -\Delta_k \le u_k^{\nu+\frac{1}{2}} + \alpha\bar{d}^\nu \le c_k^\Delta. \tag{42}$$

This one-dimensional problem can again be solved analytically similarly to (37). As a result it holds for $u_k^{\nu+1} := u_k^{\nu+\frac{1}{2}} + \alpha^\nu d^\nu$

$$m_{T,k}(u_k^{\nu+1}) \le m_{T,k}(u_k^{\nu+\frac{1}{2}}) \le m_{T,k}(u_k^\nu).$$

Note again, that in the case of a non-convex energy the local Gauss–Seidel problems (37) must not have a unique solution. Now depending on the choice of intermediate solutions the algorithm might lead to different iterates. To this end let $\mathcal{M} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be the set-valued mapping which maps the initial iterates onto the set of all possible next iterates. Global convergence of the method towards first-order optimal points can be shown using the following result for closed set-valued mappings:

Theorem 5.5. *Consider the continuous energy functional $J : \mathbb{R}^{dn} \longrightarrow \mathbb{R}$. Let an algorithm be given by the set-valued mapping $\mathcal{M} : \mathbb{R}^{dn} \rightrightarrows \mathbb{R}^{dn}$ and a procedure to specify the sequence of iterates $w^k \in \mathcal{M}(w^{k-1})$. Let*

$$\mathcal{K}^{\mathrm{opt}} := \big\{ u \in \mathbb{R}^{dn} : u \text{ first-order optimal point of } J \big\}$$

and assume that:

- *If $w \notin \mathcal{K}^{\mathrm{opt}}$, then*

$$J(v) < J(w) \quad \forall\, v \in \mathcal{M}(w). \tag{43}$$

- *If $w \in \mathcal{K}^{\mathrm{opt}}$, then either the algorithm stops or*

$$J(v) \leq J(w) \quad \forall\, v \in \mathcal{M}(w). \tag{44}$$

Further, assume that the range of $\mathcal{M}$ is contained in a compact set $\mathcal{H} \subset \mathbb{R}^{dn}$ and that $\mathcal{M}$ is closed at all $w \notin \mathcal{K}^{\mathrm{opt}}$, i.e., for every convergent sequence $w^k \longrightarrow \bar{w}$ and any sequence $v^k \in \mathcal{M}(w^k)$ with $v^k \longrightarrow \bar{v}$, it holds that $\bar{v} \in \mathcal{M}(\bar{w})$.
Then, either the algorithm stops at a $w \in \mathcal{K}^{\mathrm{opt}}$ or the limit of any convergent subsequence is in $\mathcal{K}^{\mathrm{opt}}$.

Proof. [32, Theorem A] □

Properties (43) and (44) clearly hold for the leading Gauss–Seidel step (37) and hence for the whole algorithm. Furthermore, the range of $\mathcal{M}$ is contained in the compact set defined by the trust-region constraints TRQP. The closedness of $\mathcal{M}$ can be seen by first observing that exact directional minimisation is closed. Therefore, define

$$\mathcal{M}_e(w) := \left\{ w + \alpha e : \alpha \in \operatorname*{arg\,min}_{\beta \in [a_e(w), b_e(w)]} J(w + \beta e) \right\}, \tag{45}$$

where e denotes the search direction and $a_e(\cdot), b_e(\cdot)$ represent position-dependant defect obstacles in that direction. In the proposed algorithm such directional minimisations are performed in the non-linear pre-smoothing step (37), the line-search (42) and the computation of the coarse grid projection using a monotone multigrid. In all cases the dependency of a_e, b_e on w is always continuous by construction of the defect obstacles.

Proposition 5.6. *The mapping $\mathcal{M}_e$ is closed.*

Proof. Let the sequences $(w^k), (v^k) \subset \mathcal{H}$ be given with $v^k \in \mathcal{M}_e(w^k)$, and let $w^k \longrightarrow \bar{w}$, $v^k \longrightarrow \bar{v}$. By definition each $v^k = w^k + \alpha^k e$ for some $\alpha^k \in [a_e(w^k), b_e(w^k)] := h(e, w^k)$ and further $\bar{v} = \bar{w} + \bar{\alpha} e$ with $\alpha^k \longrightarrow \bar{\alpha}$.
We now have to show that $\bar{v} \in \mathcal{M}_e(\bar{w})$ which by (45) is equivalent to

$$J(\bar{v}) = J(\bar{w} + \bar{\alpha} e) \leq J(\bar{w} + \alpha e) \quad \forall\, \alpha \in h(e, \bar{w}).$$

The continuity of $h(e, \cdot)$ yields that $a_e(w^k) \longrightarrow a_e(\bar{w})$ and $b_e(w^k) \longrightarrow b_e(\bar{w})$.
Thus, for all $\alpha \in (a_e(\bar{w}), b_e(\bar{w}))$ and k large enough we obtain $\alpha \in h(e, w^k)$ and

$$J(w^k + \alpha^k e) \leq J(w^k + \alpha e).$$

Let $k \longrightarrow \infty$, then from the continuity of J we deduce

$$J(\bar{v}) \leq J(\bar{w} + \alpha e) \quad \forall\, \alpha \in (a_e(\bar{w}), b_e(\bar{w})).$$

The case where $\alpha \in \{a_e(\bar{w}), b_e(\bar{w})\}$ follows similarly. □

As a result we deduce that performing a line search (42) is a closed mapping. The closedness of the leading Gauss–Seidel step and the monotone multigrid method follows by observing that the composition of closed maps is a closed mapping again.

Lemma 5.7. *Let $\mathcal{H}$ be compact set and $S, T : \mathcal{H} \rightrightarrows \mathcal{H}$. If T is closed at w and S on $T(w)$, then also the composition $S \circ T$ is closed at w.*

Proof. [32, Corollary 4.2.1] □

Setting up the truncated defect problem (38) and projecting the coarse grid correction d^ν in (41) are continuous and hence closed operations. We conclude

Corollary 5.8. *The combination of TNNMG and monotone multigrid described in Section 5.1 either stops at a first-order optimal point of* (TRQP) *or the limit of any convergent subsequence is first-order optimal.*

5.2. **Global Convergence of the filter SQP-method.** The global convergence of Algorithm FTR towards first-order optimal points can be shown by modifying the proof for the filter method with only exact SQP problems [8]. In this section we will only state the necessary standard assumptions and prove the main result. The detailed steps leading to the main theorem are described in Appendix A.

Assumption 1. *The iterates φ_k generated by Algorithm FTR stay in a compact set $\mathcal{C} \supseteq \mathcal{K}_h$.*

Assumption 2. *The discrete energy functional J is twice differentiable w.r.t the deformation.*

From these two assumptions the boundedness of $\|H_k\|, \|\nabla J\|$ and $\left\|\nabla^2 c\right\|$ follows.

$$\begin{aligned}\|H_k\| \leq \kappa_H &:= \max_{\varphi\in\mathcal{C}} \left\|\nabla^2 J(\varphi)\right\|,\\ \|f_k\| \leq \kappa_f &:= \max_{\varphi\in\mathcal{C}} \|\nabla J(\varphi)\|,\\ \kappa_g &:= \max_{\varphi\in\mathcal{C}} \left\|\nabla^2 g(\varphi)\right\|.\end{aligned}$$

Further, the inverse mortar transformations need to be bounded

Assumption 3. *The exact and inexact inverse mortar transformation* (40) *resp.* (40) *are bounded*

$$\left\|T(\varphi_k)^{-1}\right\| \leq \kappa_T \quad \left\|\widetilde{T}(\varphi_k)^{-1}\right\| \leq \kappa_T,$$

independent of k.

The inverse transformations (40) and (39) only involve the (in-)exact (non-)mortar matrices and the Householder reflection. As a result this assumption hold true as long as the faces on contact boundary γ_h don't degenerate which is enforced implicitly through the material law (1). The next assumption ensures that the energy is reduced sufficiently.

Assumption 4. *The approximate solutions u_k, $\widetilde{u}_k$ of TRQP and ITRQP respectively, fulfil the* ***sufficient Cauchy decrease*** *condition:*

$$\begin{aligned}m_k(u_k) - m_k(0) &\geq \kappa_{\mathrm{scd}}\chi_k \min\left\{\frac{\chi_k}{\|H_k\|}, \Delta_k\right\},\\ \widetilde{m_k}(\widetilde{u}_k) - \widetilde{m_k}(0) &\geq \kappa_{\mathrm{scd}}\widetilde{\chi}_k \min\left\{\frac{\widetilde{\chi}_k}{\|H_k\|}, \Delta_k\right\},\end{aligned}$$

for some constant $\kappa > 0$.

The sufficient Cauchy decrease holds for the first local minimum on the projected gradient path [6]. Hence, Assumption 4 can be interpreted as achieving at least a fraction of the decrease that is generated by following the negative projected gradient.
In our case this assumption holds when performing suitably many iteration steps of the globally convergent multigrid method 5.1 for the solution of the local problems TRQP and ITRQP. Numerical tests strongly indicate that already one iteration is enough to exceed the desired decrease.
From these assumptions the following convergence result can be shown:

Theorem 5.9. *Let Assumptions 1 to 4 hold and $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR. Then, either the feasibility restoration phase terminates unsuccessfully by converging to a critical point of* (36) *or there exists a subsequence $(\varphi_{k_l})_{l\in\mathbb{N}} \subseteq (\varphi_k)_{k\in\mathbb{N}}$ such that*

$$\lim_{l\longrightarrow\infty} \varphi_{k_l} = \varphi_*,$$

where φ_* is a first-order critical point of the non-linear problem (18).

Proof. Assume that (FRP) always terminates successfully. Then by Lemmata A.7, A.8 and A.10 there exists a subsequence $(\varphi_{k_l})_{l\in\mathbb{N}}$ with

$$\lim_{l\longrightarrow\infty} \vartheta_{k_l} = \lim_{l\longrightarrow\infty} \widetilde{\chi}_{k_l} = 0.$$

For k large enough we get $\widetilde{\chi}_{k_l} \leq \varepsilon$ and thus eventually only the exact local problems are solved. We conclude that also

$$\lim_{l\longrightarrow\infty} \chi_{k_l} = 0.$$

By the compactness Assumption 1 there exists a convergent subsequence which we denote again by $(\varphi_{k_l})_{l\in\mathbb{N}}$ such that

$$\lim_{l\longrightarrow\infty} \varphi_{k_l} = \varphi_*.$$

From the continuity of $\chi(\cdot)$ and $\vartheta(\cdot)$ it follows that also $\chi(\varphi_*) = \vartheta(\varphi_*) = 0$. Hence, φ_* is a first-order critical point of (18) which can be seen as follows:
Let d_* be the solution of the optimality measure, i.e.

$$d_* \in \mathbb{R}^n : \min_{\substack{C(\varphi_*)u+c(\varphi_*)\geq 0 \\ \|u\|\leq 1}} \langle \nabla J(\varphi_*), u\rangle .$$

Then the first-order optimality conditions (19) ensure the existence of a Lagrange multiplier $\lambda_* \in \mathbb{R}^m, \lambda \geq 0$ s.t.

$$\begin{aligned} \nabla J(\varphi_*) + \lambda_*^T C(\varphi_*) &= 0, \\ \lambda_*(C(\varphi_*)d_* + c(\varphi_*)) &= 0. \end{aligned} \tag{46}$$

Multiplying the first equation of (46) with d_* and using $\langle \nabla J(\varphi_*), d_*\rangle = 0$ leads to

$$\lambda_*^T C(\varphi_*)d_* = 0,$$

Inserting this into the second equation of (46) yields $\lambda_* c(\varphi_*)$. From the feasibility of φ_* we conclude that λ_* is also a Lagrange multiplier for the first-order optimality system of the non-linear problem (18). □

6. Numerical Examples

In this section we show a numerical examples to illustrate the performance of the filter–trust-region method Algorithm FTR. The implementation was done within the Dune environment which is a free c++ toolbox for the solution of partial differential equations [2, 3]. The implementation of the discrete normal projection 6.1 can be found in the dune-grid-glue module [4] and is shortly described in the next section 6.1. For the solution of the coarse grid problems within the multigrid step we apply the CoinIpOpt library which uses an optimised interior point algorithm [28].

6.1. Discretisation of the normal projection Φ.

For the discretisation of the normal projection (2) each pair of faces $(T^1, T^2) \in (\mathcal{T}_h(\gamma_C^1), \mathcal{T}_h(\gamma_C^2))$, needs to be checked for intersection. To avoid the quadratic effort of testing all pairs of faces, we use the optimal advancing front algorithm by Gander and Japhet [10]. In this scheme, after an initial intersecting pair is found, only neighbouring information is used resulting in linear complexity.
Now let $\mathbf{p}_i^1$ and $\mathbf{p}_i^2$ denote the corners of T^1 and T^2 respectively and let $\mathbf{n}_{h,i}^2$ denote the nodally averaged normals (12) associated to the mortar vertices $\mathbf{p}_i^1$. Then, computation of the intersection polygon for the pair of faces is done as follows:

1. Project each $\mathbf{p}_i^2$ along $\mathbf{n}_{h,i}^2$ onto the T^1-plane
2. Check which $\mathbf{p}_i^2$ are contained in T^1.

3. Check which $\mathbf{p}_i^1$ are contained in the projected $\widetilde{T}^2$
4. Compute edge intersections and determine which neighbours intersect
5. Triangulate the convex polygon

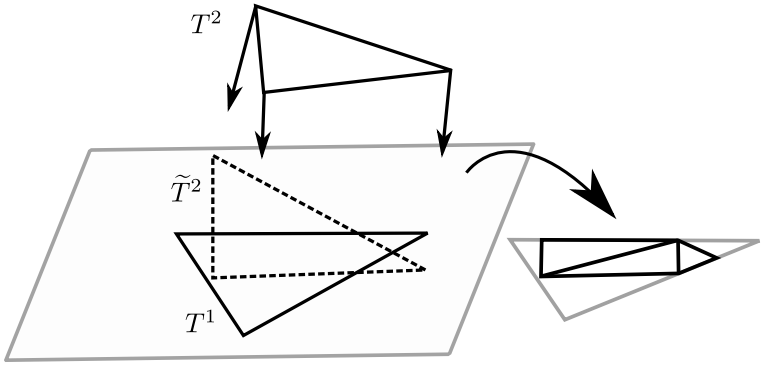

FIGURE 3. *Computation and triangulation of an intersection polygon.*

Remark 6.1. *While most other approaches* [25, 23] *use an element normal for the projection of the vertices* $\mathbf{p}_i^2$*, we use nodally averaged normals* $\mathbf{n}_{h,i}^2$ *which results in a globally continuous merged contact surface* $\gamma_h \cup \Phi^{-1}(\mathcal{T}_h(\gamma_C^2))$*. This is due to the fact that the inverse projection of the vertices* $\mathbf{p}_i^2$ *is independent of the face normal* $\mathbf{n}_{T^1}^1$*.*

6.1.1. *Ironing example.* In this section we illustrate the convergence of the proposed filter–trust-region method at the example of the ironing problem. This problem is used to test the robustness of the mortar discretisation and the applied algebraic solver due to its difficulty [23, 25]. We adopt the initial configuration and loading set-up from [25], see Figure 4. In this example a block is placed under a half-spherical shell. The block is fixed at the bottom with homogeneous Dirichlet conditions. For

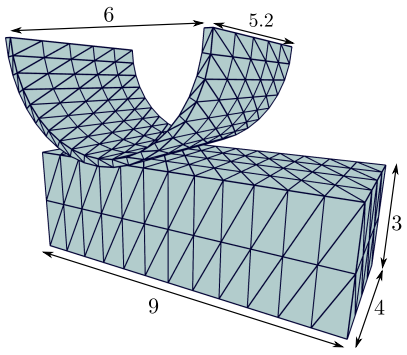

FIGURE 4. *The initial configuration of the coarse ironing grids.*

the shell non-homogeneous Dirichlet conditions are prescribed on the top boundary that are enforced in loading steps: First, the shell is pressed vertically into the block with a prescribed total displacement of 1.4 units. Then, in the second phase it is swiped over the block until a horizontal displacement of 2.1 is reached, see Figure 5.

As done in [25] we apply an equidistant loading increment of 0.14. Both bodies are modelled by the Neo–Hookean material law

$$W(\varphi) = \frac{\lambda}{4}(\det(\nabla\boldsymbol{\varphi})^2 - 1) - (\frac{\lambda}{2} + \mu)\log(\det(\nabla\boldsymbol{\varphi})) + \mu \operatorname{tr} E(\boldsymbol{\varphi}),$$

with

$$\lambda_{\text{shell}} = 750, \quad \mu_{\text{shell}} = 375,$$
$$\lambda_{\text{block}} = \frac{3}{4}, \quad \mu_{\text{block}} = \frac{3}{8}.$$

In this example the stiffness ratio of the two bodies is very high with $1:1000$. In the case of a softer contact $(1:100)$ similar convergence results were obtained. We

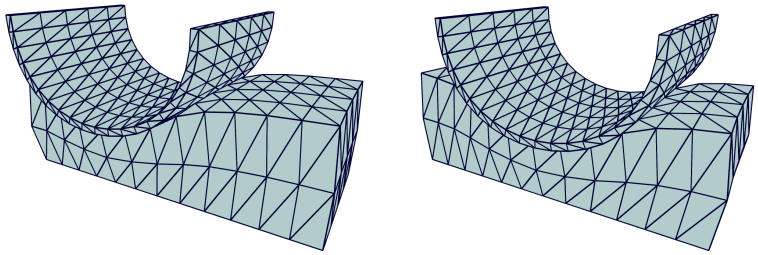

FIGURE 5. *Left:Ironing problem after the horizontal displacement. Right: Ironing problem after the vertical displacement*

use the Hessian of the strain energy within the SQP method, c.f., (21). The loading problems are solved with the inexact filter–trust-region method until the optimality measure χ, see (32), falls below 10^{-9} or the relative error in the H^1-norm is less than 10^{-10}. The constant in the control of the approximation error (34) is chosen $\kappa_{\text{ac}} = 1$. We switch from inexact to exact local problems when the approximate optimality $\widetilde{\chi}$, cf. (33), falls below 10^{-5}.

On the left side of Figure 6 the convergence of the inexact filter method and corresponding filter with exact Jacobians are shown for an exemplary step during the swiping phase at the increment 0.28. For this computation the block grid has been refined uniformly three times and the shell grid once. As can be seen in the picture, the decrease of the exact optimality is flattening when only the "cheap" inexact problems are solved. This is why it is important for the performance of the inexact filter method to choose the switching point, i.e. the tolerance for the inexact criticality measure, carefully or adjust it dynamically.

Already in this rather small problem with $26,550$ degrees of freedom, the computation time for the solution of the inexact local problems with the multigrid method is about ten times smaller than the time needed by the optimised solver for the exact problem ($1.5s$ resp. $15s$). In this exemplary loading step 12 inexact iterations were performed before switching to exact constraint Jacobians, which corresponds to a decrease of computation time of 15%. In total the computation time needed for this example with a three times refined block and once refined shell was reduced about 14%, see Figure 6 and Table 1

Only linear convergence can be expected due to the approximate Hessians (21) that are used within the filter method. Extending this approach to second-order consistent models will be done in future work.

In the right of Figure 7 the total inexact filter iterations are plotted for different refinement levels, where the grids are refined alternatingly. During the first phase the iteration numbers seem to remain constant for decreasing mesh size; in the

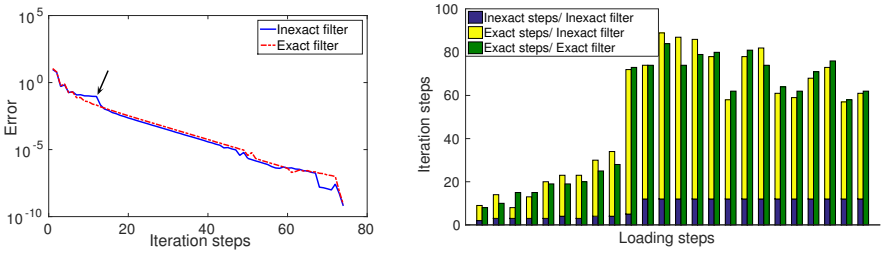

FIGURE 6. *Left: Reduction of the criticality measure χ of the exact and inexact filter method during the 12^{th} loading step. The arrow marks the spot where the approximate criticality measure reached its tolerance. The flattening in that area shows that exact steps are necessary for the convergence. Further, the switch between exact and inexact steps has to be chosen appropriately to guarantee a high reduction of computation time. Right: Comparsion of the inexact and exact steps that were computed during all loading steps. The total iteration numbers of the inexact and exact filter are comparable. As a result, the use of cheap inexact steps results in a decrease computation time.*

	exact	inexact	difference
comp. time	18,495s	16,027s	2,648s

TABLE 1. The total averaged computation times for the exact and inexact filter. Using inexact Jacobians results in a reduction of 14%.

second phase this independence is lost, although the increase is still in an acceptable range.

In Figure 7 we compare the solutions generated by our filter method to the results that are computed by only solving the inexact local problems until $\widetilde{\chi} < 10^{-10}$. We observe that, although this error is small, convergence to the true minimisers cannot be expected in general.

APPENDIX A. DETAILED CONVERGENCE PROOF

To a great extend the proof goes along the lines of the proof for the filter method with exact constraint Jacobians [7]. In this section will we shortly repeat these parts of the proof and adjust them to account for the inexactness of the constraint linearisation. First we show that the old and the new infeasibility can be bounded by the trust-region radius.

Lemma A.1. *Assume that TRQP is compatible. Then, there exists a constant $\kappa_{\mathrm{ui}} > 0$ such that*

$$\vartheta_k \leq \kappa_{\mathrm{ui}} \Delta_k,$$
$$\vartheta(\varphi_k + u_k) \leq \kappa_{\mathrm{ui}} \Delta_k^2.$$

Proof. Let $\overline{u}_k = T_k^{-1} u_k$ be the solution of (TRQP) in transformed coordinates. Then from the compatibility we immediately deduce

$$\vartheta_k \leq \|\overline{u}_k\|_\infty \leq \|\overline{u}_k\| = \left\|T_k^{-1} u_k\right\| \leq \left\|T_k^{-1}\right\| \|u_k\| \leq \kappa_T \|u_k\|.$$

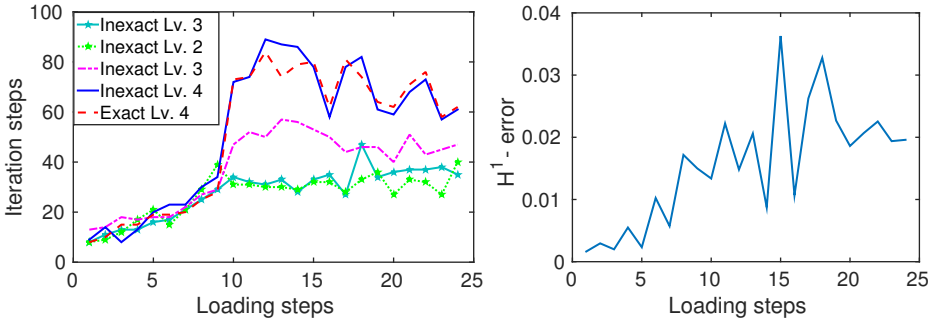

FIGURE 7. *Left: A comparsion of the filter iteration numbers for different refinement levels and the exact filter method on the finest grid. During the swiping phase (steps* $10-24$*) a mild mesh-dependence can be observed. Right: The* H^1-*error of the true minimizers to approximate ones that are computed from the sole solution of inexact problems. While the error seems to be constant and small, convergence of a purely inexact solver cannot be expected.*

To derive a bound for the new infeasibility we have to distinguish between the cases where the exact and inexact linearisations are used. We will only consider the inexact case, the exact one follows similarly. Performing a first-order Taylor expansion of the p'th constraint at φ_k yields

$$-c(\varphi_k + u_k)_p = -c_{k,p} - C_{k,p}u_k - \frac{1}{2}u_k^T \nabla^2 c(\xi_k)_p u_k,$$

where $\nabla^2 c(\xi_k)_p$ denotes the remainder term. Adding the zero $\widetilde{C}_{k,p} - \widetilde{C}_{k,p}$ and using the approximation error bound (34) we can deduce

$$\begin{aligned} -c(\varphi_k + u_k)_p &= -c_{k,p} - \widetilde{C}_{k,p}u_k + \left(\widetilde{C}_{k,p} - C_{k,p}\right) u_k - \frac{1}{2}u_k^T \nabla^2 c(\xi_k)_p u_k \\ &\leq 0 + \kappa_{\mathrm{ac}}\Delta_k^2 + \frac{1}{2}\max_{1\leq i\leq m}\max_{\varphi\in\mathcal{C}} \left\|\nabla^2 c(\varphi)_i\right\| \|u_k\|^2 \\ &\leq (\kappa_{\mathrm{ac}} + \frac{1}{2}\kappa_g)\Delta_k^2. \end{aligned}$$

The first term vanishes because u_k is a solution of the local problem ITRQPand the second term can be bounded because of the approximation error bound (34). The claim then follows with

$$\kappa_{\mathrm{ui}} := \max\left\{\kappa_{\mathrm{ac}} + \frac{1}{2}\kappa_g, \kappa_T\right\}.$$

□

The following results hold independent of which local problem is solved, the inexact (ITRQP) or the exact one (TRQP). For brevity, we will only state them for the exact problems.
One can see that the model decrease can be bounded from below by the trust-region if the radius is sufficiently small and the critically measure is not already zero.

Lemma A.2. *Assume that TRQP is compatible, that*

$$\chi_k \geq \varepsilon > 0, \tag{47}$$

and that

$$\Delta_k \leq \frac{\varepsilon}{\kappa_H} =: \delta_m.$$

Then

$$m_k(0) - m_k(u_k) \geq \frac{1}{2}\kappa_{\mathrm{scd}}\varepsilon\Delta_k.$$

Proof. By Assumption 4 it holds

$$m_k(u_k) - m_k(0) \geq \kappa_{\mathrm{scd}}\chi_k(u)\min\left\{\frac{\chi_k(u_k)}{\|H_k\|},\, \Delta_k\right\} \geq \kappa_{\mathrm{scd}}\varepsilon\Delta_k$$

□

By restricting the trust-region further the local model becomes a better approximation of the non-linear functional and very successful iterations are guaranteed

Lemma A.3. *Assume that TRQP is compatible, that* (47) *and*

$$\Delta_k \leq \min\left\{\delta_m, \frac{(1-\eta_2)\kappa_{\mathrm{scd}}\varepsilon}{\kappa_H}\right\} =: \delta_\rho.$$

Then

$$\rho_k \leq \eta_2.$$

Proof. Use Lemma A.2 and the second-order consistency of the model m_k

$$|J(\varphi_k + u_k) - m_k(u_k)| \leq \kappa_H\Delta_k^2,$$

[7, Lemma 15.5.5]. □

Next it can be shown that for a small trust-region the *ϑ-type* condition holds

Lemma A.4. *Assume that TRQP is compatible, that* (47) *and*

$$\Delta_k \leq \min\left\{\delta_m, \left(\frac{\kappa_{\mathrm{scd}}\varepsilon}{\kappa_\vartheta\kappa_{\mathrm{ui}}^2}\right)\right\} =: \delta_f.$$

Then

$$m_k(0) - m_k(u_k) \geq \kappa_\vartheta\vartheta_k^2.$$

Proof. Follows from Lemmata A.1 and A.2:

$$\kappa_\vartheta\vartheta_k^2 \leq \kappa_\vartheta\kappa_{\mathrm{ui}}^2\Delta_k^2 \leq \kappa_{\mathrm{scd}}\varepsilon\Delta_k \leq m_k(0) - m_k(u_k)$$

□

If the infeasibility is small enough one can deduce that enough energy decrease is generated so that the iterate is acceptable to the filter.

Lemma A.5. *Assume that TRQP is compatible, that* (47) *and*

$$\Delta_k \leq \min\left\{\frac{\eta_2\kappa_{\mathrm{scd}}\varepsilon}{\gamma_\vartheta\kappa_{\mathrm{ui}}}, \delta_\rho\right\} =: \delta_\vartheta.$$

Then

$$J(\varphi_k + u_k) \leq J_k - \gamma_\vartheta\vartheta_k.$$

Proof. Lemmata A.1 to A.3, see [7, Lemma 15.5.7]. □

The local problem TRQP is compatible if the infeasibility is bounded by the trust-region radius. To conclude that this holds when both the radius and the infeasibility are small is the result of the next lemma.

Lemma A.6. *Assume that* (47) *and*

$$\Delta_k \leq \min\left\{\gamma_1\delta_\rho, \frac{\gamma_1^2(1-\gamma_\vartheta)}{\kappa_{\mathrm{ui}}}\right\} =: \delta_R.$$

Then TRQPis compatible.

Proof. Assume that no compatibility holds. Then, the algorithm ensures that $k-1$ was compatible. Now assume that the previous iteration $k-1$ was unsuccessful, that means $\Delta_k = \gamma_1 \Delta_{k-1}$ and $\vartheta_k = \vartheta_{k-1}$. As φ_{k-1} is acceptable to filter, by Lemmata A.3 and A.5 the unsuccessfulness must hold because

$$\vartheta(\varphi_{k-1} + u_{k-1}) > (1-\gamma_\vartheta)\vartheta_{k-1} = (1-\gamma_\vartheta)\vartheta_k$$

From Lemma A.1 we deduce that

$$(1-\gamma_\vartheta)\vartheta_k \leq \kappa_{\mathrm{ui}} \Delta_{k-1}^2 \leq \frac{\kappa_{\mathrm{ui}}}{\gamma_1^2} \Delta_k^2.$$

As the iteration k is not compatible it holds $\Delta_k < \vartheta_k$. Combining this with the previous estimate leads to a contradiction of $\Delta_k \leq \delta_R$. Hence, the iteration $k-1$ was successful and $\vartheta_k = \vartheta(\varphi_k + u_{k-1})$. Using again that k is not compatible and Lemma A.1 we arrive at

$$\Delta_k < \vartheta_k \leq \kappa_{\mathrm{ui}} \Delta_{k-1}^2 \leq \frac{\kappa_{\mathrm{ui}}}{\gamma_1^2} \Delta_k^2,$$

which again contradicts $\Delta_k \leq \delta_R$ as $(1-\gamma_\vartheta) < 1$. Therefore, the initial assumption of k being not compatible must be wrong. □

With these lemmata at hand it can be concluded that whenever during the solution an infinite number of iterates is added to the filter, then there exist a convergent subsequence which is first-order optimal. To this end we define

$$\mathcal{Z} := \{k \in \mathbb{N} |\, \varphi_k \text{ is added to the filter}\},$$
$$\mathcal{R} := \{k \in \mathbb{N} |\, k \text{ is incompatible}\}.$$

Lemma A.7. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_k$ be a sequence generated Algorithm FTR such that $|\mathcal{Z}| = \infty$. Then there exists a subsequence $(k_l)_l \subseteq \mathcal{Z}$ such that*

$$\lim_{l \longrightarrow \infty} \vartheta_{k_l} = 0,$$

and

$$\lim_{l \longrightarrow \infty} \chi_{k_l} = 0.$$

Proof. Let $(\varphi_{k_l})_{l \in \mathbb{N}}$ be a subsequence with $k_l \in \mathcal{Z}$. Then by Lemma 5.3 it holds

$$\vartheta_{k_l} \longrightarrow 0. \tag{48}$$

Now assume that

$$\chi_{k_l} \geq \varepsilon_1 > 0, \quad \forall l \geq l_0.$$

Further assume that

$$\Delta_{k_m} \geq \varepsilon_2 > 0, \quad \forall m \geq m_0 \geq i_0. \tag{49}$$

By the sufficient Cauchy decrease 4 it then holds

$$m_{k_m}(0) - m_{k_m}(u_{k_m}) \geq \kappa_{\mathrm{scd}} \varepsilon_1 \min\left\{\frac{\varepsilon_1}{\kappa_H}, \varepsilon_2\right\} =: \delta > 0,$$

and therefore in the limit

$$\lim_{m \longrightarrow \infty} [m_{k_m}(0) - m_{k_m}(u_{k_m})] \geq \delta.$$

As φ_{k_m} was added to the filter, it must either be because $k_m \in \mathcal{R}$ or a ϑ-type took place, i.e. (35) fails:

$$m_{k_m}(0) - m_{k_m}(u_{k_m}) < \kappa_\vartheta \vartheta_{k_m}^2.$$

In the first case

$$\vartheta_{k_m} > \Delta_{k_m},$$

which contradicts (49) due to (48). The second case contradicts (48) as

$$0 < \delta \leq m_{k_m}(0) - m_{k_m}(u_{k_m}) < \kappa_\vartheta \vartheta_{k_m}^2.$$

Therefore the assumption (49) is wrong and $\Delta_{k_l} \longrightarrow 0$. Hence by Lemma A.6 eventually $k_l \notin \mathcal{R}$ for all k_l large enough. Now again as above these iterates are added to the filter and

$$m_{k_l}(0) - m_{k_l}(u_{k_l}) < \kappa_\vartheta \vartheta_{k_l}^2,$$

which cannot hold for arbitrary small Δ_{k_l} due to Lemma A.4. We conclude that also the first assumption is wrong from which the claim follows. □

Finally we prove Lemma A.7 for the case when only finitely many iterates were added to the filter.

Lemma A.8. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR such that $|\mathcal{Z}| < \infty$, then*

$$\lim_{k\longrightarrow\infty} \vartheta_k = 0.$$

Proof. From $|\mathcal{Z}| < \infty$ it follows that for k large enough, the ϑ-type condition (35)

$$m_k(0) - m_k(u_k) \geq \kappa_\vartheta \vartheta_k^2 \geq 0,$$

never fails. For successful iterations k it holds

$$J(\varphi_k) - J(\varphi_{k+1}) \geq \eta_1 m_k(0) - m_k(u_k) \geq \kappa_\vartheta \vartheta_k^2 \geq 0.$$

From Assumptions 1 and 2 one can deduce that $(J(\varphi_k) - J(\varphi_{k+1})) \longrightarrow 0$ and hence that the claim follows, as for all unsuccessful iterations j it holds $\vartheta_{j+1} = \vartheta_j$. □

With this lemma at hand one can deduce that for non-critical points the trust-region radius cannot become arbitrarily small.

Lemma A.9. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR such that $|\mathcal{Z}| < \infty$, and*

$$\chi_k \geq \varepsilon > 0,$$

for all k large enough. Then there exists $\Delta_{\min} > 0$ such that

$$\Delta_k \geq \Delta_{\min} \quad \forall\, k \in \mathbb{N}.$$

Proof. Proof by contradiction using the mechanism of the algorithm, see [7, Lemma 15.5.11]. □

This then yields the desired result for the case of a finite filter.

Lemma A.10. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR such that $|\mathcal{Z}| < \infty$. Then*

$$\liminf_{k\longrightarrow\infty} \chi_k = 0.$$

Proof. Assume that

$$\chi_k \geq \varepsilon > 0.$$

By Lemma A.8 it holds $\vartheta_k \longrightarrow 0$ and as in the proof of Lemma A.8 we can deduce that for a successful subsequence k_l it holds $(J(\varphi_{k_l-1}) - J(\varphi_{k_l})) \longrightarrow 0$.
Further, by Assumption 4 and Lemma A.9

$$J(\varphi_{k_l-1}) - J(\varphi_{k_l}) \geq \eta_1 (m_{k_l}(0) - m_{k_l}(u_{k_l})) \geq \kappa_{\mathrm{scd}} \varepsilon \min\left\{ \frac{\varepsilon}{\kappa_H}, \Delta_{\min} \right\} > 0$$

which is a contradiction. □

References

[1] F. Armero and E. Petőcz. Formulation and analysis of conserving algorithms for frictionless dynamic contact/impact problems. *Comp. Meth. Appl. Mech. and Eng.*, 158(3):269–300, 1998.

[2] P. Bastian, M. Blatt, A. Dedner, C. Engwer, R. Klöfkorn, R. Kornhuber, M. Ohlberger, and O. Sander. A Generic Grid Interface for Parallel and Adaptive Scientific Computing. Part II: Implementation and Tests in DUNE. *Computing*, 82(2–3):121–138, 2008.

[3] P. Bastian, M. Blatt, A. Dedner, C. Engwer, R. Klöfkorn, M. Ohlberger, and O. Sander. A Generic Grid Interface for Parallel and Adaptive Scientific Computing. Part I: Abstract Framework. *Computing*, 82(2–3):103–119, 2008.

[4] P. Bastian, G. Buse, and O. Sander. Infrastructure for the coupling of dune grids. In *Numerical Mathematics and Advanced Applications 2009*, pages 107–114. Springer, 2010.

[5] P. Ciarlet. *Mathematial Elasticity: Volume I Three-Dimensional Elasticity*. Elsevier, 1988.

[6] A. R. Conn, N. Gould, A. Sartenaer, and P. L. Toint. Global convergence of a class of trust region algorithms for optimization using inexact projections on convex constraints. *SIAM Journal on Optimization*, 3(1):164–221, 1993.

[7] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. Siam, 2000.

[8] R. Fletcher, N. Gould, S. Leyffer, P. Toint, and A. Wächter. Global convergence of a trust-region sqp-filter algorithm for general nonlinear programming. *SIAM Journal on Optimization*, 13(3):635–659, 2002.

[9] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Math. Programming*, 91:239–269, 2002.

[10] M. Gander and C. Japhet. An algorithm for non-matching grid projections with linear complexity. In *Domain Decomposition Methods in Science and Engineering XVIII*, pages 185–192. Springer, 2009.

[11] C. Gräser and R. Kornhuber. Multigrid methods for obstacle problems. *Journal of Computational Mathematics*, 27(1):1–44, 2009.

[12] C. Gräser, U. Sack, and O. Sander. Truncated nonsmooth Newton multigrid methods for convex minimization problems. In *Proc. of DD18*, pages 129–136, 2009.

[13] S. Hartmann, S. Brunssen, E. Ramm, and B. Wohlmuth. Unilateral non-linear dynamic contact of thin-walled structures using a primal-dual active set strategy. *International journal for numerical methods in engineering*, 70(8):883–912, 2007.

[14] S. Hartmann and E. Ramm. A mortar based contact formulation for non-linear dynamics using dual lagrange multipliers. *Finite Elements in Analysis and Design*, 44:245–258, 2008.

[15] P. Hauret and P. L. Tallec. Energy-controlling time integration methods for nonlinear elastodynamics and low-velocity impact. *Computer methods in applied mechanics and engineering*, 195(37):4890–4916, 2006.

[16] C. Hesch. *Mechanische Integratoren fr Kontaktvorgänge deformierbarer Körper unter großen Verzerrungen*. PhD thesis, Universität Siegen, 2007.

[17] C. Hesch and P. Betsch. A mortar method for energy–momentum conserving schemes in frictionless dynamic contact problems. *Int. J. Numer. Methods Eng.*, 77:1468–1500, 2009.

[18] S. Hüeber and B. Wohlmuth. A primal–dual active set strategy for non-linear multibody contact problems. *Computer Methods in Applied Mechanics and Engineering*, 194(27):3147–3166, 2005.

[19] R. Krause and B. Wohlmuth. Monotone methods on nonmatching grids for nonlinear contact problems. *SIAM J. Sci. Comput.*, 25(1):324–347, 2003.

[20] T. A. Laursen. *Computational Contact and Impact Mechanics*. Springer, 2003.

[21] T. A. Laursen and G. R. Love. Improved implicit integrators for transient impact problems–geometric admissibility within the conserving framework. *Int. J. Numer. Methods Eng.*, 53:245–274, 2002.

[22] T. A. Laursen and J. C. Simo. A continuum-based finite element formulation for the implicit solution of multibody, large deformation frictional contact problems. *Int. J. Num. Meth. in Eng.*, 36(20):3451–3485, 1993.

[23] T. A. L. M. A. Puso. A mortar segment-to-segment contact method for large deformation solid mechanics. *Comput. Methods Appl. Mech. Engrg.*, 193:601–629, 2004.

[24] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[25] A. Popp, M. W. Gee, and W. Wall. Finite deformation contact based on 3d dual mortar and semi–smooth newton approach. In *Trends in Computational Contact Mechanics*, pages 57–77. Springer, 2010.

[26] M. Tur, F. J. Fuenmayor, and P. Wriggers. A mortar-based frictional contact formulation for large deformations using lagrange multipliers. *Computer Methods in Applied Mechanics and Engineering*, 198(37):2860–2873, 2009.

[27] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.
[28] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.
[29] B. Wohlmuth. Variationally consistent discretization schemes and numerical algorithms for contact problems. *Acta Numerica*, 20:569–734, 2011.
[30] P. Wriggers and T. A. Laursen. *Computational contact mechanics*, volume 30167. Springer, 2006.
[31] J. Youett. *Dynamic large deformation contact problems and applications in virtual medicine*. PhD thesis, Freie Universität Berlin, 2016.
[32] W. Zangwill. *Nonlinear programming : a unified approach*. Prentice-Hall Englewood Cliffs, NJ, 1969.

(Jonathan Youett) Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin
E-mail address: youett@math.fu-berlin.de

(Oliver Sander) Institute of Numerical Mathematics, Technische Universität Dresden, Zellescher Weg 12–14, 01069 Dresden
E-mail address: oliver.sander@tu-dresden.de

(Ralf Kornhuber) Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin
E-mail address: kornhube@math.fu-berlin.de