SIAM J. SCI. COMPUT.
Vol. 41, No. 1, pp. B114–B138
© 2019 Society for Industrial and Applied Mathematics

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 for hyperelastic materials. The discretization uses the mortar method which is known to be more stable than node-to-segment approaches. The resulting nonconvex constrained minimization problems are solved using a filter–trust-region scheme, and we prove global convergence towards first-order optimal points. The constrained Newton problems are solved robustly and efficiently using a truncated nonsmooth Newton multigrid method with a monotone multigrid linear correction step. For this we introduce a cheap basis transformation that decouples the contact constraints. Numerical experiments confirm the stability and efficiency of our approach.

Key words. contact problem, finite-strain, filter, trust-region, multigrid

AMS subject classifications. 65N55, 65K15

DOI. 10.1137/17M1142338

1. Introduction.

Although large deformation contact problems arise in many important applications, only very few methods today can solve them fast and robustly. All of these methods have their advantages and disadvantages.

Discretization of such problems leads to constrained nonconvex minimization problems. The prevailing methods for these problems are primal–dual active set strategies [10, 11, 17] and penalty methods [18]. For both methods only local convergence can be expected [12]. Furthermore, the resulting linearized Newton problems can be indefinite due to the nonconvexity of the strain energy.

In this work we construct a filter–trust-region method [7] for the constrained nonconvex minimization problem. The filter technique ensures asymptotic fulfillment of the nonlinear nonpenetration constraints by rejecting iterates that are neither improving the energy nor the infeasibility compared to all previous iterates. The trust-region method provides a natural way to handle indefiniteness of the linearized problems. We show that the modifications we need to make to the method to apply it to contact problems stay within the realm of the general filter–trust-region convergence theory, and hence we obtain global convergence of the method to first-order stationary points.

A priori, the Newton problems of a filter–trust-region method are quadratic minimization problems with convex inequality constraints. Such problems are generally expensive to solve. We extend an efficient multigrid strategy originally introduced for contact problems in small strain elasticity [13] to the case of large strains. This requires rewriting the inequality constraints as sets of bound constraints. The inequality constraints consist of two parts: the trust-region constraint and the linearized contact condition. We define the trust-region in terms of the max-norm. With this choice, the trust-region constraints form a set of bound constraints by construction. To decouple

*Submitted to the journal's Computational Methods in Science and Engineering section August 7, 2017; accepted for publication (in revised form) November 21, 2018; published electronically February 14, 2019.

http://www.siam.org/journals/sisc/41-1/M114233.html

Funding: This work was supported by the DFG Matheon project CH1 funded by ECMath.

†Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin (youett@math.fu-berlin.de, kornhuber@math.fu-berlin.de).

‡Institute of Numerical Mathematics, Technische Universität Dresden, Zellescher Weg 12–14, 01069 Dresden (oliver.sander@tu-dresden.de).

the contact constraints we extend the technique used in [13] to the finite strain case. The idea there is to construct a basis transformation that replaces the nodal basis at the contact boundaries by a system of relative movements. The construction of this transformation requires the solution of a linear system involving a contact surface mass matrix (the nonmortar matrix) during each Newton-type iteration. The additional computational cost of this is negligible because the size of the mass matrix is much smaller than the overall problem size and grows with a lower order. The transformation also leads to a slight modification of the Newton matrix, but we show that this modification does not influence the convergence behavior of the overall method.

In previous work, the truncated nonsmooth Newton multigrid (TNNMG) method has been shown to be very fast and effective for quadratic minimization problems with bound constraints such as small-strain contact problems and obstacle problems [8, 9]. Since we have found a way to uncouple the contact constraints for finite-strain contact problems, we can also harness the performance of TNNMG for the Newton problems of a finite-strain contact problem. Unfortunately, this only works if the quadratic models are convex. For the nonconvex case, we extend the TNNMG method by combining it with a monotone multigrid (MMG) method for the linear correction step. The MMG method will handle the trust-region constraints (which have a comparatively simple structure) while the more complicated contact constraints will be left to the TNNMG step. The resulting scheme is globally convergent even for indefinite trust-region problems. At the same time, we observe multigrid-type convergence rates in numerical experiments.

This paper is organized as follows. In section 2 the static large deformation contact problem is described and its weak formulation is derived. In section 3 we summarize the mortar discretization of the problem, which we use because it avoids most instabilities and unphysical oscillations of the node-to-segment approaches [18]. As a stepping stone, we then construct a locally convergent, efficient solver based on sequential quadratic programming in section 4. We introduce the TNNMG multigrid algorithm and the constraints decoupling strategy needed to solve the quadratic constrained Newton problems. To globalize the local SQP solver, in section 5 we then introduce the filter–trust-region algorithm and the combined TNNMG/MMG scheme for the solution of the linearized problems. We show global convergence of both methods. The final section 6 is dedicated to numerical experiments. While our algorithm exhibited unconditional global convergence with multigrid convergence speed in numerical computations, a primal–dual active-set method in the spirit of [17] failed to converge, unless the loading increments were selected to be sufficiently small.

2. Static large deformation contact problems.

In this section we will briefly summarize the equations of equilibrium of two nonlinear hyperelastic bodies subject to mutual contact. A more detailed introduction can be found, e.g., in [14].

2.1. Strong formulation.

Let $\Omega^i \subset \mathbb{R}^d$, $i = 1, 2$, $d = 2, 3$ denote the disjoint reference configurations of two deformable objects. Assume that the boundaries of the Ω_i are such that the outer unit normal fields $\mathbf{n}_R^i : \partial\Omega^i \to \mathbb{R}^d$ exist everywhere. Let the boundaries be decomposed into disjoint relatively open sets $\partial\Omega^i = \overline{\Gamma}_D^i \cup \overline{\Gamma}_N^i \cup \overline{\Gamma}_C^i$ corresponding to Dirichlet, Neumann, and contact boundaries. 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 objects. For example $\Omega = \Omega^1 \cup \Omega^2$ 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 $\boldsymbol{\varphi} := (\boldsymbol{\varphi}^1, \boldsymbol{\varphi}^2) : \Omega \to \mathbb{R}^d$

$$\begin{aligned} \operatorname{div} \mathbf{P}(\boldsymbol{\varphi}) + \mathbf{f} &= 0 \qquad \text{in } \Omega, \\ \mathbf{P}(\boldsymbol{\varphi})\mathbf{n}_R &= \mathbf{t} \qquad \text{on } \Gamma_N, \\ \boldsymbol{\varphi} &= \boldsymbol{\varphi}_D \quad \text{on } \Gamma_D. \end{aligned} \tag{1}$$

Here, $\mathbf{P} : \Omega \to \mathrm{Mat}^+(d)$ is the first Piola–Kirchhoff stress field, and $\mathrm{Mat}^+(d)$ is the set of $d \times d$ matrices with positive determinant. The functions $\mathbf{f} \in \mathbf{L}^2(\Omega)$ and $\mathbf{t} \in \mathbf{L}^2(\Gamma_N)$ are prescribed external volume and traction force densities, which are assumed to be independent of the deformation. The function $\boldsymbol{\varphi}_D \in C(\Gamma_D)^d$ specifies the Dirichlet boundary conditions. We will only consider hyperelastic continua, i.e., materials for which there exists a stored energy functional $\mathcal{W} : \Omega \times \mathrm{Mat}^+(d) \to \mathbb{R}$, $(x, F) \mapsto \mathcal{W}(x, F)$, that links the stresses to the deformation via

$$\frac{\partial \mathcal{W}}{\partial F}(\cdot, \nabla \boldsymbol{\varphi}) = \mathbf{P}(\boldsymbol{\varphi}). \tag{2}$$

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

$$\det \nabla \boldsymbol{\varphi}(x) > 0 \quad \forall\, x \in \Omega, \tag{3}$$

in the sense that

$$\mathcal{W}(x, \nabla \boldsymbol{\varphi}) \to \infty \quad \text{if } \det \nabla \boldsymbol{\varphi}(x) \searrow 0.$$

As a consequence, we will not explicitly enforce (3) 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. For $i = 1, 2$, let $\mathbf{n}^i$ denote the outer unit normal field on the deformed contact boundary $\gamma_C^i := \boldsymbol{\varphi}^i(\Gamma_C^i)$. Modeling of nonpenetration can be done in several ways, depending on which projection is chosen to identify the contact surfaces with each other. Earlier papers used the closest-point projection from γ_C^1 to γ_C^2 [14, 15, 25]. Recently, using the projection along $\mathbf{n}^1$ has become more popular [11, 17, 18, 20]. In the following we only consider the closest-point projection approach, but others can be used equally well. The deformed contact boundaries are identified with each other through the projection $\Phi : \gamma_C^1 \to \gamma_C^2$

$$\Phi(s) := \operatorname*{arg\,min}_{r \in \gamma_C^2} \|s - r\|.$$

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

$$g(s) := \mathbf{n}^2(\Phi(s)) \cdot (s - \Phi(s)), \tag{4}$$

where we have used the fact that $s - \Phi(s)$ is orthogonal to γ_C^2 at $\Phi(s)$. With these definitions, nonpenetration of the bodies is enforced by requiring

$$g(s) \geq 0 \quad \forall\, s \in \gamma_C^1; \tag{5}$$

compare Figure 1.

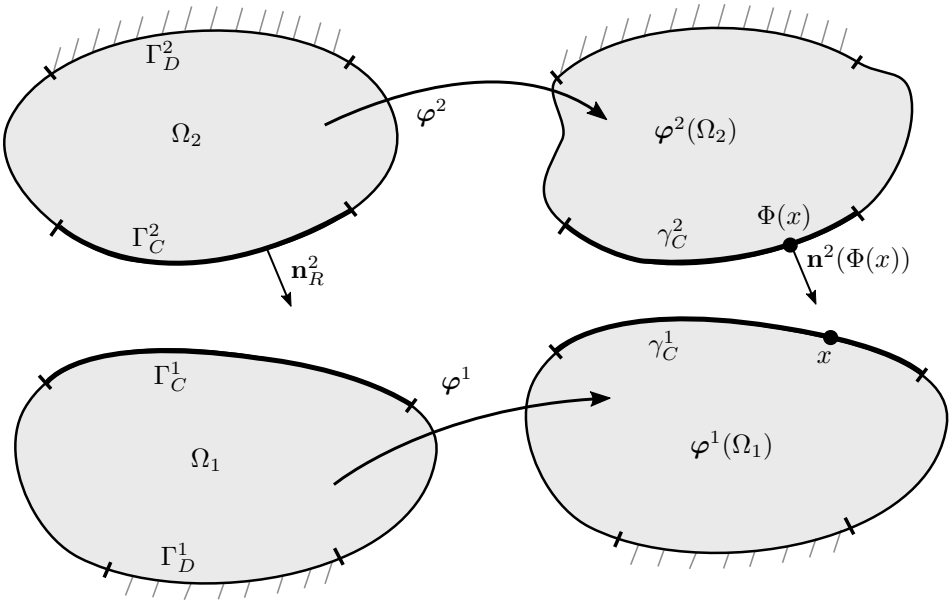

FIG. 1. *Reference and deformed configuration of the two bodies.*

So far the nonpenetration constraint was derived only from a kinematical point of view. To investigate the effect of these constraints on the elastic system we examine the resulting contact forces. 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 boundary traction*

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

then represents the contact forces on γ_C^1. It can be decomposed into normal and tangential parts t_N and $\mathbf{t}_T$ with respect to $\mathbf{n}^1$. We consider frictionless contact only so the tangential traction $\mathbf{t}_T$ at the contact boundary vanishes. The contact normal stresses fulfil the *Karush–Kuhn–Tucker (KKT) conditions*

$$t_N \le 0, \quad g \ge 0, \quad g \cdot t_N = 0 \quad \text{on } \gamma_C^1,$$

where the first one states that traction is a pressure, the second one is (5), and the last one is the *complementary condition* [14].

2.2. Weak formulation. The equilibrium configurations of hyperelastic continua are characterized as stationary points of the energy functional

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

where $\mathcal{W}$ is the hyperelastic energy density (2), and $\mathcal{F}$ and $\mathcal{G}$ are potentials of the external forces. Stable configurations are the local minimizers of this energy [3, Theorem 4.1-2]. Existence of minimizers has been shown for the case of a polyconvex and coercive strain energies [3, Theorem 7.7-1]. The corresponding first-order optimality condition is the weak form of the elasticity problem (1)

We now add the contact constraints. In a Sobolev space setting, the nonpenetration constraint (5) takes the form

$$g(s) \geq 0 \quad \text{for almost all } s \in \gamma_C^1, \tag{6}$$

and similarly for the other two KKT conditions. In anticipation of the mortar discretization we rewrite this condition in a variationally consistent form. Let $\mathbf{H}^1_D(\Omega)$ denote the Sobolev space of d-valued weakly differentiable functions fulfilling the Dirichlet boundary conditions in the sense of traces. 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

$$\begin{aligned} W^+ &:= \{v \in W : v \geq 0 \text{ a.e.}\}, \\ M^+ &:= \left\{\mu \in M : \langle \mu, v\rangle \geq 0 \ \ \forall\, v \in W^+\right\}, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the dual paring of M and W. Now, the resulting weak formulation of the nonpenetration constraint (6) is given by

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

The equivalence of this to (6) is shown in [24]. 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 nonconvex set of feasible deformations. The weak formulation of the large deformation contact problem now reads:

$$\text{Find a local minimiser } \boldsymbol{\varphi} \text{ of } \mathcal{J} \text{ in } \mathcal{K}. \tag{7}$$

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

3. Discretization. In this section we will describe the discretization of the minimization problem (7) using first-order Lagrangian finite elements and mortar elements for the contact constraints. Let $\mathcal{T}_h$ be a shape-regular grid of the bodies $\Omega := \Omega^1 \cup \Omega^2$, and $\mathcal{N}(\mathcal{T}_h)$ the set of vertices. The space of d-valued first-order finite elements is $\mathbf{S}_h = (S_h)^d$, and for each node $p \in \mathcal{N}(\mathcal{T}_h)$ the scalar nodal basis function corresponding to p is denoted by $\psi_p \in S_h$. We discretize the hyperelasticity problem (7) 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 discretization of the contact constraints. We use dual mortar functions [23] to discretize the mortar cone M^+, but Lagrange elements can be used equally well. For a given discrete deformation $\boldsymbol{\varphi}_h \in \mathbf{S}_{D,h}$, let γ_h^i be the grid of the deformed contact boundary obtained by restricting $\mathcal{T}_h$ to the reference contact boundary Γ_C^i, and then deforming this restriction using $\boldsymbol{\varphi}_h$. We denote the basis of the Lagrange multiplier space by

$$\Theta_h := \left\{\theta_p : p \in \mathcal{N}(\gamma_h^1)\right\}. \tag{8}$$

Note that due to the biorthogonality of the dual basis θ_p and the nodal basis functions ψ_q, i.e.,

$$\int_{\gamma_h^1} \theta_p \psi_q \, ds = \delta_p^q \int_{\gamma_h^1} \psi_q \, ds,$$

the functions θ_p are in general depending on the deformation. The discrete mortar cone $M_h^+ \not\subset M^+$ is then given by

$$M_h^+ := \Big\{ \mu_h \in \operatorname{span} \Theta_h : \int_{\gamma_h^1} \mu_h(s)\, v_h(s)\, ds \geq 0 \quad \forall\, v_h \in S_h(\gamma_h^1),\, v_h \geq 0 \Big\}.$$

This leads to the weak nonpenetration constraint

$$\int_{\gamma_h^1} g(s)\, \mu_h(s)\, ds \geq 0 \quad \forall\, \mu_h \in M_h^+,$$

which, considering the definition (4) of the gap function g, is

$$\int_{\gamma_h^1} \mathbf{n}^2(\Phi(s)) \cdot (s - \Phi(s))\, \mu_h(s)\, ds \geq 0 \quad \forall\, \mu_h \in M_h^+. \tag{9}$$

As the normal field of a piecewise polynomial surface, $\mathbf{n}^2$ is not continuous on γ_C^2. We therefore replace it by a smoothed normal field $\mathbf{n}_h$. Define vertex normals by averaging the adjacent face normals, i.e., for each vertex $p \in \mathcal{N}(\gamma_C^2)$ with neighboring faces $\mathcal{E}(p)$ on the contact boundary we set

$$\mathbf{n}_p := \frac{\sum_{e \in \mathcal{E}(p)} \mathbf{n}_e}{\big\| \sum_{e \in \mathcal{E}(p)} \mathbf{n}_e \big\|},$$

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

$$\mathbf{n}_h := \sum_{p \in \mathcal{N}(\gamma_h^2)} \psi_p \mathbf{n}_p, \tag{10}$$

and we replace $\mathbf{n}^2$ in (9) with $\mathbf{n}_h$. This continuous approximation yields a smoother behavior when sliding occurs compared to using discontinuous element normals; compare [18]. The resulting discrete nonpenetration constraint with

$$g_h(s) := \mathbf{n}_h(\Phi(s)) \cdot (s - \Phi(s)) \tag{11}$$

reads

$$\int_{\gamma_h^1} g_h(s)\, \mu_h(s)\, ds \geq 0, \quad \mu_h \in M_h^+. \tag{12}$$

We denote the corresponding discrete feasible set by

$$\mathcal{K}_h := \Big\{ \boldsymbol{\varphi}_h \in \mathbf{S}_{D,h} : \int_{\gamma_h^1} g_h(s)\, \mu_h(s)\, ds \geq 0 \quad \forall \mu_h \in M_h^+ \Big\}.$$

Summarizing, the discrete problem is given by:

$$\text{Find a local minimiser } \boldsymbol{\varphi}_h \text{ of } \mathcal{J} \text{ in } \mathcal{K}_h. \tag{13}$$

As for the nondiscrete case (7), the existence of solutions of (13) appears to be an open question.

3.2. Algebraic contact problem. For the rest of this paper we will denote the pth component of a (block-)vector v by v_p, the pth row of a matrix A by A_p, and the (p,q)th entry of a (block-)matrix A by A_{pq}. The algebraic representation of the finite-strain contact problem is derived using the canonical isomorphism $I : \mathbb{R}^{dn} \to \mathbf{S}_h$

(where $n := |\mathcal{N}(\mathcal{T}_h)|$) that identifies finite element functions with their coefficient (block-)vectors. The algebraic energy is then given by

$$J : \mathbb{R}^{dn} \to \mathbb{R}, \qquad J(z) := \int_\Omega \mathcal{W}(I(z))\,dx - b^T z,$$

with $b \in \mathbb{R}^{dn}$ given componentwise by

$$(b_p)_i := \int_\Omega \mathbf{f}\, e^i \psi_p\,dx + \int_{\Gamma_N} \mathbf{t}\, e^i \psi_p\,ds, \quad 1 \le p \le n,\ 0 \le i < d,$$

where e^i denotes the ith Euclidean basis vector. The nonpenetration constraint (12) is represented algebraically by a function $c : \mathbb{R}^{dn} \to \mathbb{R}^{m_1}$, with $m_1 := |\mathcal{N}(\gamma_h^1)|$, defined by testing the weak constraint (12) with the mortar basis functions (8)

$$c_q(z) := \int_{\gamma_h^1} g_h(s)\theta_q(s)\,ds, \quad 1 \le q \le m_1. \tag{14}$$

Note that the dependency of c_q on the algebraic deformation z is hidden in the deformed coordinates $s = s(z)$ and in the dual basis functions $\theta_p = \theta_p(z, s)$. Details on how to assemble the algebraic constraints can be found in [17]. Summarizing, the nonconvex algebraic contact problem reads:

$$\text{Find a local minimiser } z \text{ of } J \text{ in } K, \tag{15}$$

where

$$K := \left\{ z \in \mathbb{R}^{dn} :\ c_q(z) \ge 0, \quad 1 \le q \le m_1 \right\}.$$

4. Inexact SQP multigrid methods for contact problems. In this section we show how (15) can be solved locally using SQP. We propose a basis transformation that decouples the linearized constraints for each quadratic subproblem. The transformed problems can then be solved robustly and efficiently using a TNNMG method. The transformation involves minor modifications to the tangent stiffness matrices that do not harm the overall convergence properties.

4.1. SQP. Consider the constrained optimisation problem (15). The first-order optimality conditions are given by the following theorem.

THEOREM 4.1 (see [16, Theorem 12.1]). *Let z^* be a local minimizer of (15). If the rows of the active constraint Jacobian, i.e., those rows p of $\nabla c(z^*)$ for which $c_p(z^*) = 0$, are linearly independent, then there exists a unique Lagrange multiplier $\lambda \in \mathbb{R}^{m_1}$ such that*

$$\begin{aligned} \nabla J(z^*) + \lambda^T \nabla c(z^*) &= 0, \\ c(z^*) &\ge 0, \end{aligned} \tag{16}$$

and

$$\lambda_p \le 0, \quad \lambda_p c_p(z^*) = 0, \quad 1 \le p \le m_1.$$

The SQP method is derived by applying a Josephy–Newton method to the first-order optimality system (16) and eliminating the Lagrange multiplier. In the following, let upper indices $k \in \mathbb{N}$ be the iteration number of the Josephy–Newton method, and introduce a quadratic model energy by

$$m^k(u) := \nabla J(z^k)^T u + \frac{1}{2} u^T H^k u, \tag{17}$$

with $H^k \in \mathbb{R}^{dn \times dn}$ symmetric. The SQP constraints are derived by replacing the constraint $c(z) \ge 0$ with its linearization at z^k

$$\nabla c(z^k)u + c(z^k) \geq 0, \tag{18}$$

where $u \in \mathbb{R}^{dn}$ is the argument of m^k. Then, the Josephy–Newton problems can be reformulated as quadratic minimization problems for the correction $u^k \in \mathbb{R}^{dn}$

$$\min_{u \in \mathbb{R}^{dn}} m^k(u), \quad \nabla c(z^k)u + c(z^k) \geq 0. \tag{QP}$$

Local linear convergence of this scheme can be proven if H^k is a symmetric positive definite approximation of the Hessian of the Lagrangian

$$H^k \approx \nabla^2 J(z^k) + (\lambda^k)^T \nabla^2 c(z^k); \tag{19}$$

see [16, Theorem 18.7].

4.2. Multigrid methods for bound-constrained quadratic minimization problems. In an SQP method, solving the constraint quadratic problems (QP) is by far the most costly part. We will solve these problems with multigrid efficiency using the TNNMG method [9]. To illustrate the method we assume for the rest of this section that the local models m^k are strictly convex.

Consider the quadratic functional (17) with $H^k \in \mathbb{R}^{dn \times dn}$ symmetric and positive definite. For simplicity we drop the superscript k for this section. Also, for this section only, we assume that the linear constraints decouple into bound constraints. Hence, we want to find the unique minimizer of m subject to

$$a_i \leq u_i \leq b_i, \quad 1 \leq i \leq dn, \tag{20}$$

where the a_i may be $-\infty$ and the b_i may be $+\infty$. One iteration step of TNNMG for this problem can be separated into the following four substeps: let $u^\nu \in \mathbb{R}^{dn}$ be a given iterate.

1. *Projected Gauss–Seidel step*
Set $w_0 = u^\nu$; then for $p = 1, \ldots, dn$, set

$$\begin{aligned} \alpha^p &= \operatorname*{arg\,min}_{a_p \leq \alpha + u_p^\nu \leq b_p} m(w^{p-1} + \alpha e^p), \\ w^p &= w^{p-1} + \alpha^p e^p, \end{aligned} \tag{21}$$

where e^p is the pth Euclidean basis vector.
Denote by $u^{\nu+\frac{1}{2}} := w_{dn}$ the resulting presmoothed iterate.

2. *Truncated linear correction*
To accelerate the convergence, the smoothing is followed by a linear correction step for the defect problem

$$\min_{v \in \mathbb{R}^{dn}} \frac{1}{2} v^T H v - r^T v,$$

where the residual is given by

$$r := \nabla J(z) - H u^{\nu+\frac{1}{2}}.$$

For this step the active components

$$\mathcal{A}(u) := \left\{ p \in \{1, \ldots, dn\} : u_p = a_p \text{ or } u_p = b_p \right\}$$

are truncated [8], i.e., temporarily frozen. This is achieved by multiplying the defect problem with the truncation matrix

$$Q^\nu \in \mathbb{R}^{dn\times dn}: \quad Q^\nu_{pq} := \begin{cases} 1, & p = q \text{ and } p \notin \mathcal{A}(u^{\nu+\frac{1}{2}}), \\ 0, & \text{else.} \end{cases} \tag{22}$$

The linear truncated defect problem therefore reads

$$v^\nu := \operatorname*{arg\,min}_{v\in\mathbb{R}^{dn}} \frac{1}{2} v^T Q^\nu H Q^\nu v - (r^\nu Q^\nu)^T v, \tag{23}$$

where all the entries of v^ν corresponding to truncated components are set to zero. Note that the defect problem (23) is unconstrained. For the approximate solution of this problem on the space spanned by the inactive components one (or a few) geometric or algebraic linear multigrid step(s) is used.

3. *Projection*
The resulting correction v^ν may violate the defect constraints. To ensure feasibility it is projected back onto the defect obstacles in the l^2-sense, i.e., we define $\hat{v}^\nu$ by

$$\hat{v}^\nu_i := \begin{cases} b_i - u_i^{\nu+\frac{1}{2}} & \text{if } v_i^\nu > b_i - u_i^{\nu+\frac{1}{2}}, \\ a_i - u_i^{\nu+\frac{1}{2}} & \text{if } v_i^\nu < a_i - u_i^{\nu+\frac{1}{2}}, \\ v_i^\nu, & \text{else.} \end{cases}$$

4. *Line search*
The projection in step 3 can lead to an increase of model energy. To ensure monotonicity of the algorithm a line search is performed

$$\alpha^\nu = \operatorname*{arg\,min}_{\alpha\in\mathbb{R}} m^k(u^{\nu+\frac{1}{2}} + \alpha\hat{v}^\nu) \quad \text{such that (s.t.) } u^{\nu+\frac{1}{2}} + \alpha\hat{v}^\nu \text{ admissible.} \tag{24}$$

This one-dimensional constrained quadratic problem can be solved analytically. As a result we obtain $u^{\nu+1} := u^{\nu+\frac{1}{2}} + \alpha^\nu \hat{v}^\nu$ with

$$m(u^{\nu+1}) \le m(u^{\nu+\frac{1}{2}}) \le m(u^\nu).$$

Global convergence of the algorithm follows immediately from the convergence of the presmoothing Gauss–Seidel step and the monotonicity.

THEOREM 4.2 (see [8, Theorem. 6.4]). *Suppose that $H \in \mathbb{R}^{dn\times dn}$ is symmetric positive definite, and the constraints have the form* (20). *Then the TNNMG method converges globally to a minimizer of* (17) *subject to* (20).

4.3. Decoupling the constraints. In this section we will construct a basis transformation of $\mathbb{R}^{dn}$ that decouples the linearized contact constraints (18). This generalizes an idea from [13], which did the same in the infinitesimal strain framework. We start by considering $\nabla c(z)$ in more detail: the linearization

$$\delta c_p(z)u := \lim_{t\searrow 0} \frac{c_p(z+tu) - c_p(z)}{t}, \quad u \in \mathbb{R}^{dn},$$

of the pth component of the algebraic contact constraint (14) in the direction of $u \in \mathbb{R}^{dn}$ can be divided into three parts

$$\begin{aligned} \delta c_p(z) = &\int_{\gamma_h^1} \delta \mathbf{n}_h(\Phi(s)) \cdot (s - \Phi(s))\, \theta_p(s)\, ds \\ &+ \int_{\gamma_h^1} \mathbf{n}_h(\Phi(s)) \cdot \delta\Big[(s - \Phi(s))\, \theta_p(s)\Big]\, ds \\ &+ \int_{\gamma_h^1} \mathbf{n}_h(\Phi(s)) \cdot (s - \Phi(s))\, \theta_p(s)\, \delta ds. \end{aligned} \tag{25}$$

The first part is the linearization of the nodally averaged normal field (10). In the continuous case this term vanishes due to the colinearity of the normal $\mathbf{n}^2(\Phi(s))$ with the closest point projection $s - \Phi(s)$; see [14]. The second part is the linearization of the discretized gap function (11) and the mortar basis function, and the third summand is the linearization of the deformation dependent integral domain, which we denote by δds.

Let $m_2 := |\mathcal{N}(\Gamma_C^2)|$ be the number of vertices on the contact boundary Γ_C^2, and as before $m_1 := |\mathcal{N}(\Gamma_C^1)|$. In the following we assume for simplicity that the coefficient vectors $u \in \mathbb{R}^{dn}$ are ordered such that $u = (u_C^1, u_C^2, u^I)$, where $u_C^1 \in \mathbb{R}^{dm_1}$ and $u_C^2 \in \mathbb{R}^{dm_2}$ are the degrees of freedom on the contact boundaries Γ_C^1 and Γ_C^2, respectively, and u^I denotes all other degrees of freedom. Then, the algebraic form (25) of the constraint Jacobian $\nabla c(z)$ can be split into a *nonmortar* and a *mortar* part, corresponding to the linearizations with respect to z_C^1 and z_C^2, respectively,

$$\nabla c(z) = \begin{pmatrix} D(z) & M(z) & 0 \end{pmatrix},$$

where $D(z) := \frac{\partial c(z)}{\partial z_C^1} \in \mathbb{R}^{m_1 \times dm_1}$, $M(z) := \frac{\partial c(z)}{\partial z_C^2} \in \mathbb{R}^{m_1 \times dm_2}$ are sparse block-matrices given by

$$\begin{aligned} D(z)_{pq} &= \int_{\gamma_h^1} \mathbf{n}_h(\Phi(s)) \frac{\partial}{\partial z_q^1} \Big[(s - \Phi(s))\, \theta_p(s) \Big]\, ds \\ &\quad + \int_{\gamma_h^1} \mathbf{n}_h(\Phi(s)) \cdot (s - \Phi(s))\, \theta_p(s)\, \frac{\partial}{\partial z_q^1} ds \in \mathbb{R}^{1 \times d}, \\ M(z)_{pq} &= \int_{\gamma_h^1} \frac{\partial}{\partial z_q^2} \big[\mathbf{n}_h(\Phi(s)) \big] (s - \Phi(s))\, \theta_p(s) - \mathbf{n}_h(\Phi(s)) \frac{\partial}{\partial z_q^2} \Big[\Phi(s) \Big] \theta_p(s)\, ds \\ &\quad + \int_{\gamma_h^1} \mathbf{n}_h(\Phi(s)) \cdot (s - \Phi(s))\, \theta_p(s)\, \frac{\partial}{\partial z_q^2} ds \in \mathbb{R}^{1 \times d}, \end{aligned}$$

and 0 denotes a $m_1 \times d(n - m_1 - m_2)$ zero matrix. The algebraic linearized constraints (18) then take the form

$$D(z) u_C^1 + M(z) u_C^2 \geq -c(z). \tag{26}$$

In our aim to decouple these constraints we first separate the normal from the tangential components. Let $O(z) \in \mathbb{R}^{dm_1 \times dm_1}$ be the block-diagonal matrix consisting of Householder transformations $O_{11}, \ldots, O_{m_1 m_1}$ such that $O_{pp}(z) \in \mathbb{R}^{d \times d}$ rotates the first Euclidean basis vector $e^1 \in \mathbb{R}^d$ onto the normal $\mathbf{n}_h$ at the projected vertex $\Phi(p) \in \gamma_h^2$ $\forall p \in \mathcal{N}(\gamma_h^1)$. We use $O(z)$ to transform the nonmortar matrix by

$$\big(D(z)O(z)\big)_{pq} =: \big(\overbrace{D_N(z)_{pq}}^{\in \mathbb{R}} \quad \overbrace{D_T(z)_{pq}}^{\in \mathbb{R}^{d-1}}\big). \tag{27}$$

In the normal part $D_N(z) \in \mathbb{R}^{m_1 \times m_1}$, the first component of each $(1 \times d)$-block of $D(z)O(z)$ is collected. Analogously, the $d - 1$ tangential components are collected in $D_T(z) \in \mathbb{R}^{m_1 \times (d-1)m_1}$. The crucial insight of [13] was to see that the contact constraints can be decoupled by inverting D_N. For small-strain contact problems this could be trivially achieved, because the biorthogonality of the dual mortar basis lead to a diagonal matrix D_N. In the finite-strain setting, D_N is sparse but no longer diagonal. We suppose that the matrix remains invertible for all relevant configurations z. For

the sake of the argument we use its inverse D_N^{-1} now and comment later on how to compute it efficiently.

Consider the following deformation-dependent transformation $T(z) \in \mathbb{R}^{dn \times dn}$

$$T(z) := \begin{pmatrix} O(z)K(z) & -O(z)L(z) & 0 \\ 0 & \mathrm{Id} & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix}, \tag{28}$$

where the $(d \times d)$-block-matrices $K(z), L(z)$ are given component-wise by

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

The inverse of T is sparse and has the form

$$T(z)^{-1} = \begin{pmatrix} U(z)O(z) & -V(z) & 0 \\ 0 & \mathrm{Id} & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix},$$

where

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

LEMMA 4.3. *In the transformed coordinates*

$$\bar{u} = T^{-1}(z) u,$$

the linearised contact constraints (26) *take the form*

$$\bar{u}^1_{C,0} \le c(z), \tag{29}$$

where $\bar{u}^1_{C,0}$ *is the vector that contains the first of each block of* d *degrees of freedom on* γ_h^1.

Proof. We omit the dependencies on z for simplicity. The linearized constraints (26) transform according to

$$\begin{aligned} (\nabla c)\, T &= \begin{pmatrix} D & M & 0 \end{pmatrix} T \\ &= \begin{pmatrix} DOK & \big[M - DOL\big] & 0 \end{pmatrix}. \end{aligned}$$

The first column of this is an $(m_1 \times dm_1)$-matrix. It can be simplified by noting that for any $p, q = 1, \dots, m_1$

$$\begin{aligned} (DOK)_{pq} &= \sum_{j=1}^{m} \begin{pmatrix} (D_N)_{pj} & (D_T)_{pj} \end{pmatrix} \begin{pmatrix} -(D_N^{-1})_{jq} & -(D_N^{-1} D_T)_{jq} \\ 0 & \delta_j^q \mathrm{Id} \end{pmatrix} \\ &= \begin{pmatrix} -\delta_p^q & \big[-(D_T)_{pq} + (D_T)_{pq}\big] \end{pmatrix} = \begin{pmatrix} -\delta_p^q & 0^{1\times(d-1)} \end{pmatrix}. \end{aligned}$$

The second column vanishes since for $p \in \{1, \dots, m_1\}, q \in \{1, \dots, m_2\}$ we have

$$(DOL)_{pq} = \sum_{j=1}^{m_1} \begin{pmatrix} (D_N)_{pj} & (D_T)_{pj} \end{pmatrix} \begin{pmatrix} (D_N^{-1} M)_{jq} \\ 0 \end{pmatrix} = M_{pq},$$

where we have used the relationship (27). Therefore, if u is a vector such that (18) holds, we obtain that (29) and vice versa. ∎

In transformed coordinates, subproblem (QP) turns into

$$\text{(TQP)} \qquad \min_{\bar{u}\in\mathbb{R}^{dn}} m_T^k(\bar{u}), \quad (\bar{u}^1_{C,0})_p \le c_p(z^k), \quad p = 1,\dots,m_1,$$

with transformed quadratic energy

$$m_T^k(\bar{u}) := f_T^k \bar{u} + \frac{1}{2}\bar{u}^T H_T^k \bar{u},$$
$$f_T^k := \nabla J(z^k)^T T(z^k), \quad H_T^k := T(z^k)^T H^k T(z^k).$$

If the transformed model energy is strictly convex, this quadratic minimization problem with bound constraints can be solved by the TNNMG method of the previous section. To avoid having to assemble (28) on all grid levels, only the presmoothing Gauss–Seidel step is applied to the decoupled formulation (TQP). The truncated defect problem and coarse grid correction are computed in Euclidean coordinates.

4.4. Avoiding the inverse nonmortar matrix. The decoupling strategy of the previous section uses the explicit inverse of the sparse matrix $D_N(z)$, whose size corresponds to the number of degrees of freedom on the nonmortar contact boundary γ_h^1. While the inverse itself can be computed in reasonable time using a direct sparse solver, it leads to a considerable increase of density of the tangent stiffness matrix H_T^k compared to the untransformed matrices H^k. This severely slows down the multigrid solver. In the following we show how the matrix inversion and the resulting density increase can be avoided while conserving the convergence of the SQP method and the filter method presented in the next section. To this end, we first consider a lumped approximation of the nonmortar matrix D_N

$$(\widetilde{D}_N)_{pq} := \begin{cases} \sum_{j=1}^{m_1} (D_N)_{pj}, & p = q, \\ 0, & \text{else.} \end{cases}$$

We then define a new transformation $\widetilde{T}(z^k)$ by formula (28), but using the diagonal matrix $\widetilde{D}_N^{-1}$ instead of D_N^{-1}. Then, we apply this new transformation to the tangent stiffness matrix H^k only, but we keep the exact transformation $T(z^k)$ for the gradient f_T^k. The resulting approximate SQP problem reads

$$\text{(IQP)} \qquad \min_{\bar{u}\in\mathbb{R}^{dn}} \widetilde{m}^k(\bar{u}), \quad (\bar{u}^1_{C,0})_p \le c_p(z^k), \quad p = 1,\dots,m_1,$$

with

$$\widetilde{m}^k(\bar{u}) := f_T^k \bar{u} + \frac{1}{2}\bar{u}^T \widetilde{H}_T^k \bar{u}, \quad \widetilde{H}_T^k := \widetilde{T}(z^k)^T \widetilde{H}^k \widetilde{T}(z^k).$$

In other words, we still compute the subproblem in the transformed coordinates of subsection 4.3, but we have replaced the tangent matrix by a sparser approximation. Note that we retain the first-order consistency of the SQP model (QP), because the linear term f_T^k is still transformed according to the exact mapping T^k. This guarantees the convergence of the SQP method and of the filter–trust-region method presented in the next section. Further, this transformation can be done without explicitly computing D_N^{-1} by solving the small linear system

$$(D_N(z^k))^T \bar{f} = f^1_{C,0},$$

where $f^1_{C,0}$ consists of the first components of the entries of $\nabla \mathcal{J}(z^k)$ that correspond to degrees of freedom on the nonmortar contact boundary γ^1_h. The transformed gradient f^k_T can then be directly computed from $\bar{f}$ by multiplication with $O(z^k)$ and $(D_T(z^k))^T$, respectively, $(M(z^k))^T$ compare (28). Similarly, the transformation back to Euclidean coordinates

$$u = T(z^k)\bar{u},$$

can be computed without the explicit inverse D_N^{-1} by solving the small linear system

$$D_N(z^k)\hat{u}^1_{C,0} = -(\bar{u}^1_C + D_T(z^k)\bar{u}^1_{C,T} + M(z^k)\bar{u}^2_C),$$

and rotating the block vector $\mathbb{R}^{dm_1} \ni w := \begin{pmatrix} \hat{u}^1_{C,0} & \bar{u}^1_{C,T} \end{pmatrix}$

$$u^1_C = Ow, \quad u^2_C = \bar{u}^2_C, \quad u^I = \bar{u}^I,$$

where $\bar{u}^1_{C,T} \in \mathbb{R}^{(d-1)m_1}$ denotes the block-vector corresponding to the tangential nonmortar degrees of freedom.

Remark 4.4. Approximating the algebraic problem is often done in large deformation contact problems to simplify the unknown contact forces that show up explicitly in the weak formulation when applying an active-set method [10, 17]. This allows us to eliminate the Lagrange multipliers at the cost of losing angular momentum conservation. In contrast, by conserving the first-order consistency of the subproblems (QP), the approximation of the Hessian suggested here is only affecting the convergence rate of the SQP method and preserves the angular momentum.

5. Globalization by filter–trust-region methods. We globalize the SQP method of the previous chapter by extending it to a filter–trust-region method. In contrast to the active-set strategies widely used in contact mechanics [10, 11, 17], this method can be shown to converge globally even for rather general nonconvex strain energy functionals.

5.1. Filter–trust-region methods. The SQP method of the previous chapter converges only locally. Furthermore, away from local minimisers of J, the exact Hessian (19) does not have to be positive definite. Hence, approximating it by a positive definite matrix may result in poor performance of the SQP method [16]. In the following we will use the popular approximation of (19) by the Hessian of the energy

$$H^k = \nabla^2 J(z^k),$$

which avoids the need to compute the Lagrange multipliers during the SQP iteration. To handle the possible unboundedness from below of the local problems (QP) with this definition of H^k, the *trust-region globalization* adds a norm constraint on the correction

$$\|\bar{u}\| \le \Delta^k, \quad k = 0, 1, \dots \tag{30}$$

We choose the infinity norm as then (30) is equivalent to a set of bound constraints, which fits naturally with the nonsmooth multigrid solver of subsection 4.2.

The constraint is adjusted dynamically according to how well the local model approximates the nonlinear functional. We measure the approximation quality by the scalar quantity

$$\rho^k := \frac{J(z^k) - J(z^k + u^k)}{\widetilde{m}^k(0) - \widetilde{m}^k(\bar{u}^k)}, \tag{31}$$

where $\bar{u}^k$ is the solution of the SQP subproblem (IQP) in transformed coordinates and $u^k = T^k \bar{u}^k$.

Incorporating (30) into (IQP) yields the constrained quadratic optimization problems

$$\begin{aligned} &\min_{\bar{u} \in \mathbb{R}^{dn}} \widetilde{m}^k(\bar{u}), \\ &-\Delta^k \le \bar{u}_p \le c_p^{\Delta^k}, \quad 1 \le p \le dn, \end{aligned} \tag{TRQP}$$

with

$$c_p^{\Delta^k} := \begin{cases} \min\left\{c_p(z^k), \Delta^k\right\}, & p \text{ first component of degree of freedom on } \gamma_h^1, \\ \Delta^k, & \text{else.} \end{cases}$$

These problems always have at least one solution, even if $\widetilde{m}^k$ is nonconvex.

To arrive at a globally convergent scheme one also has to control the possible infeasibility of the intermediate iterates z^k, which results from replacing the nonlinear contact constraint from (15) by a linearized one. We measure the infeasibility of an iterate using the nonsmooth function

$$\vartheta(z) := \max_{p=1,\dots,m_1} \{0, -c_p(z)\}.$$

A filter method creates tentative new iterates by solving (TRQP) and accepting or rejecting them based on a set of criteria. In the following we use the abbreviations $J^k := J(z^k)$ and $\vartheta^k := \vartheta(z^k)$ to denote the energy and infeasibility of the kth iterate. Let z^{k+1} be a potential new iterate, i.e., $z^{k+1} = z^k + u^k$, with u^k an approximate solution of (TRQP). If $J^{k+1} \le J^i$ and $\vartheta^{k+1} \le \vartheta^i$ for all previous iterates i, then the step can be accepted. If there is a previous iterate $z^i, i \le k$, such that

$$J^i \le J^{k+1} \quad \text{and} \quad \vartheta^i \le \vartheta^{k+1},$$

then the candidate z^{k+1} should be rejected. The critical question is what to do if

$$J^{k+1} < J^i, \quad \text{but} \quad \vartheta^{k+1} > \vartheta^i,$$

or vice versa, for all previous iterates. To overcome this difficulty Fletcher and Leyffer introduced the notion of a *filter* [7].

DEFINITION 5.1. *Let $0 < \xi < 1$. A pair (J^k, ϑ^k) ξ-dominates (J^i, ϑ^i) if*

$$J^k < J^i - \xi\vartheta^k \quad \text{and} \quad \vartheta^k < (1-\xi)\vartheta^i.$$

For a fixed constant $0 < \xi < 1$, a set of tuples (J^i, ϑ^i) is called a filter $\mathcal{F}_\xi$ if no tuple ξ-dominates any other tuple in $\mathcal{F}_\xi$ (Figure 2).

A filter defines a region of acceptable new iterates.

DEFINITION 5.2. *An iterate z^{k+1} is acceptable to the filter $\mathcal{F}_\xi$ if*

$$J(z^{k+1}) < J^i - \xi\vartheta(z^{k+1}) \quad \text{or} \quad \vartheta(z^{k+1}) < (1-\xi)\vartheta^i \quad \forall\, (J^i, \vartheta^i) \in \mathcal{F}_\xi.$$

Certain acceptable iterates are added to the filter during the filter iteration, and all pairs that are dominated by the new iterate are removed.

Remark 5.3. This criterion guarantees the convergence towards the feasible set $\mathcal{K}_h$ of every acceptable sequence of iterates that is subsequently added to the filter if $\xi > 0$; see [5, Lemma 15.5.2].

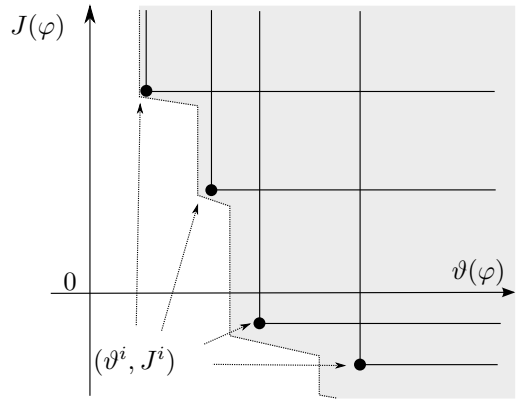

FIG. 2. *Illustration of a filter with four points. The grey area corresponds to points that are not acceptable.*

The filter–trust-region algorithm is given by the following steps.

1. *Computing a candidate*
Compute a new candidate $z^k + u^k$ by approximately solving (TRQP), and evaluate the corresponding energy $J(z^k + u^k)$ and infeasibility $\vartheta(z^k + u^k)$.

2. *Acceptance tests*
If the candidate is not acceptable to the filter then the trust-region is decreased. Further, if the approximation quality of the model is poor, i.e., $\rho^k < \eta_1 < 1$, for some fixed constant $0 < \eta_1 < 1$, the candidate is also rejected whenever the current infeasibility is small. This is estimated by checking if

$$\widetilde{m}^k(0) - \widetilde{m}^k(\bar{u}^k) \geq \kappa_\vartheta (\vartheta^k)^2 \tag{32}$$

for some fixed $0 < \kappa_\vartheta < 1$. In the affirmative case the trust-region is also decreased $\Delta^{k+1} < \Delta^k$.

3. *ϑ-type iteration*
If the feasibility check (32) fails, the previous iterate z^k is added to the filter and the candidate is accepted by the filter method. Hence (32) enables the method to accept candidates that improve the infeasibility $\vartheta^{k+1} < \vartheta^k$ while possibly increasing the energy. This is called a ϑ-type iteration.

4. *J-type iteration*
If (32) is fulfilled and the approximation quality of the model is high, i.e., $\rho^k \geq \eta_2$, with $0 < \eta_1 \leq \eta_2 < 1$, then additionally the trust-region radius can be increased. This potentially allows to achieve a larger energy reduction in the following iteration.

5. *Ensuring admissibility*
The combination of the trust-region constraints with the linearized nonpenetration constraints can lead to local problems (TRQP) that do not have a solution. This happens when the infeasibility is too large while the trust-region is very small

$$c_p(z^k) < -\Delta^k \quad \text{for some } p \in \{1, \dots, m_1\}\,.$$

This case is treated by the filter method as follows: first, the tuple (J^k, ϑ^k) of the previous iterate is added to the filter; it is always acceptable by construction. Then, the algorithm enters the so-called *feasibility restoration phase*. In this phase a new

iterate z^{k+1} and trust-region radius Δ^{k+1} are computed such that z^{k+1} is acceptable to the filter and the local problem (TRQP) is admissible again. This is done by minimizing the infeasibility directly

$$\text{(FRP)} \qquad \min_{z \in \mathbb{R}^{dn}} \vartheta(z),$$

e.g., by using a semismooth trust-region method [5]. To ensure that a point which is acceptable to the filter can be computed, it is crucial that only infeasible points are included in the filter

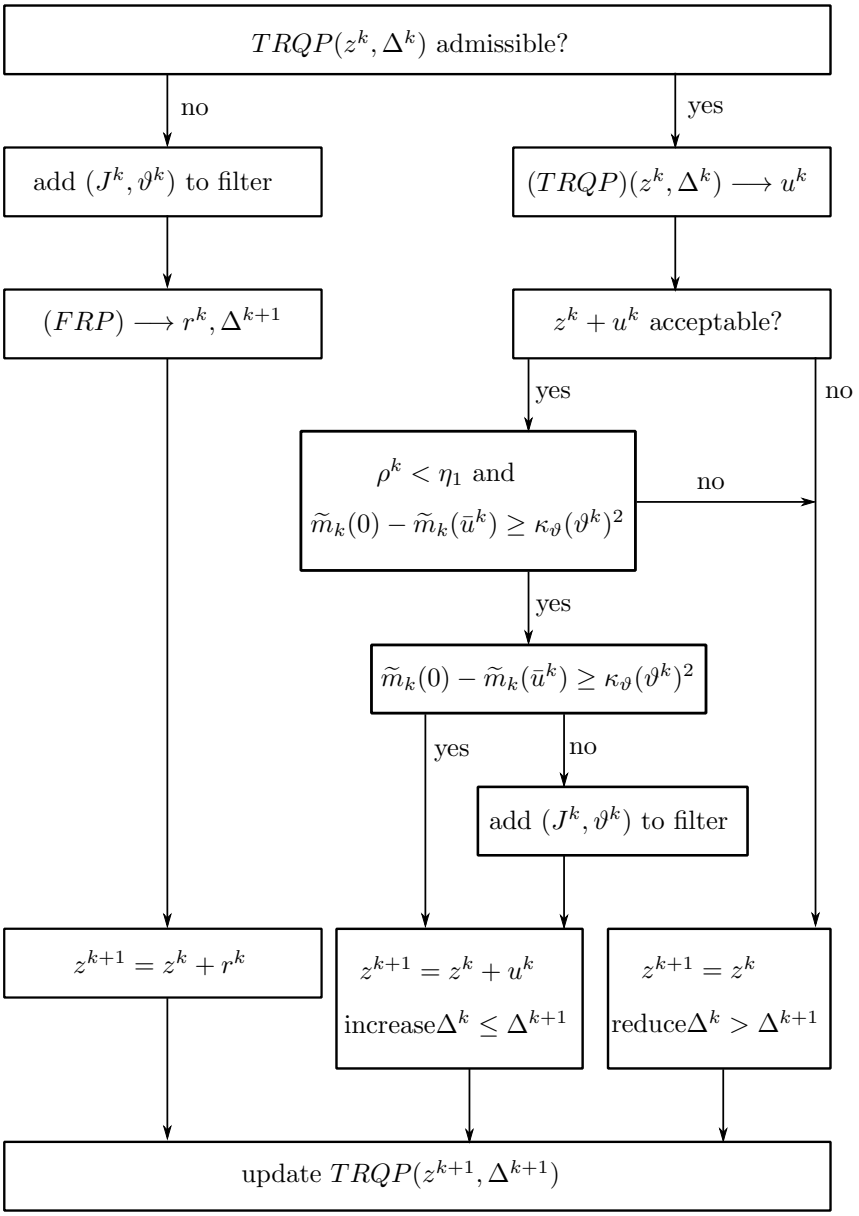

FIG. 3. *Illustration of the filter–trust-region method.*

$$(J^i, \vartheta^i) \in \mathcal{F} \Rightarrow \vartheta^i \neq 0.$$

This is achieved by only adding the tuple (J^k, ϑ^k) to the filter if (32) fails. A flowchart of the method can be found in Figure 3.

5.2. Global convergence of the filter–trust-region method. The general filter–trust-region theory shows global convergence of the method to first-order optimal points under mild assumptions on the problem [6]. We state these assumptions here for the case of the finite-strain contact problem and then formally state the convergence result.

Assumption 1. The iterates z^k generated by the filter method stay in a compact set $\mathcal{L}$.

Unfortunately, this does not immediately follow from coercivity of the hyperelastic energy functional, as the filter–trust-region algorithm is not a monotone descent method.

Assumption 2. The contact constraint $c : \mathbb{R}^{dn} \to \mathbb{R}^{m_1}$ and the energy are are both twice continuously differentiable on $\mathcal{L}$.

The contact constraint is smooth enough if the contact boundary is and if the occurring deformations are not too extreme.

Assumption 3. The normal nonmortar matrix $D_N(z)$ is regular and has a bounded inverse on $\mathcal{L}$.

This assumption again only rules out a few extreme deformations. As D_N is a mass matrix, the assumption is mainly about the grid quality of the deformed configurations.

The smoothness of c and the boundedness of D_N^{-1} imply the boundedness of the exact and lumped transformations that decouple the linearized contact constraints

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

with a constant $\kappa_T > 0$ independent of k. This in turn implies the boundedness of the transformed Hessians $\left\|\widetilde{H}_T^k\right\|$ and gradients $\left\|f_T^k\right\|$

$$\left\|\widetilde{H}_T^k\right\| \leq \max_{z\in\mathcal{L}} \left\|\nabla^2 J(z)\right\| \kappa_T^2, \quad \left\|f_T^k\right\| \leq \max_{z\in\mathcal{L}} \left\|\nabla J(z)\right\| \kappa_T.$$

This last boundedness is the assumption that typically appears in general filter–trust-region results.

The final assumption is to ensure that the model energy $\widetilde{m}^k$ is reduced sufficiently during each filter iteration. Therefore, let $\chi : \mathbb{R}^{dn} \to \mathbb{R}$ be an optimality measure of the subproblem (TRQP), i.e., a nonnegative, continuous function that vanishes if and only if $\bar{u}$ is a stationary point of the inexact SQP subproblem (TRQP).

Assumption 4. The numerical solution $\bar{u}^k$ of (TRQP) fulfills the *sufficient Cauchy decrease* condition

$$\widetilde{m}^k(0) - \widetilde{m}^k(\bar{u}^k) \geq \kappa_{\mathrm{scd}} \chi(z^k) \min\left\{ \frac{\chi(z^k)}{\left\|\widetilde{H}_T^k\right\|}, \Delta^k \right\}$$

for some constant $\kappa_{\mathrm{scd}} > 0$.

Assumption 4 means that at least a fixed fraction of the decrease that is generated by following the projected gradient has to be achieved [4]. This assumption is fulfilled eventually when the globally convergent TNNMG method is applied to (TRQP). Numerical tests indicate that already one iteration is enough to exceed the desired decrease.

From these assumptions the general filter–trust-region theory [5, 6] deduces the following global convergence result.

THEOREM 5.4 (see [5, Theorem 15.5.13]). *Let Assumptions 1 to 4 hold and $(z^k)_{k\in\mathbb{N}}$ be a sequence generated by the filter–trust-region method. Then, either the feasibility restoration phase terminates unsuccessfully by converging to a critical point of* (FRP) *or there exists a subsequence $(z^{k_l})_{l\in\mathbb{N}} \subseteq (z^k)_{k\in\mathbb{N}}$ such that*

$$\lim_{l\to\infty} z^{k_l} = z_*,$$

where z_* is a first-order critical point of the nonlinear problem (15).

We summarize that under reasonable assumptions in the context of finite-strain contact problems, the filter–trust-region multigrid method converges globally.

5.3. Multigrid solution of the trust-region subproblems. The TNNMG method of subsection 4.2 cannot be used to solve the trust-region subproblems (TRQP). They are still quadratic minimization problems with bound constraints; however, the functionals may now be nonconvex. Remember that TNNMG includes an unconstrained minimization step for the truncated energy (23). This minimization does not have a solution if the quadratic energy is not convex. We circumvent this problem by adding an additional set of bound constraints to (23), which can be interpreted as applying a trust-region method to compute the coarse grid correction. The resulting obstacle problem can be solved using the classical MMG from [8, Algorithm 5.10]. In contrast to the TNNMG, the MMG method does not neglect the obstacles on the coarser grids. In principle, MMG could be used directly to solve (TRQP). However, this would require the transformations (28) on each level of the grid hierarchy, which complicates the implementation [13, 19].

In the following we revisit the substeps of a TNNMG iteration and describe the necessary modifications. We use ν to denote the TNNMG iteration number, but for brevity we omit the SQP iteration index k. Let $u^\nu \in \mathbb{R}^{dn}$ be the current iterate.

1. *Projected Gauss–Seidel step*
The nonlinear smoother remains unchanged, noting that the one-dimensional minimization problems (21) always have at least one solution, because we minimize over a compact set now. If the global minimizer (21) is not unique, we pick the one with a larger α^p. Let $\bar{u}^{\nu+\frac{1}{2}}$ denote the resulting presmoothed iterate.

2. *Truncated linear correction*
We then set up the truncated defect problem (23). In transformed coordinates it reads

$$v^\nu := \operatorname*{arg\,min}_{\bar{v}\in\mathbb{R}^{dn}} \frac{1}{2}\bar{v}^T Q^\nu \widetilde{H}_T Q^\nu \bar{v} - (r_T^\nu Q^\nu)^T \bar{v},$$

where

$$r_T^\nu := f_T - \widetilde{H}_T \bar{u}^{\nu+\frac{1}{2}},$$

and $Q^\nu := Q(\bar{u}^{\nu+\frac{1}{2}})$ is the truncation matrix (22). Next, one transforms the defect problem back into Euclidean coordinates

$$\min_{v\in\mathbb{R}^{dn}} \frac{1}{2} v^T \widetilde{H}^\nu v - (r^\nu)^T v, \tag{33}$$

with

$$r^\nu := r_T^\nu Q^\nu T^{-1}, \quad \widetilde{H}^\nu := T^{-T} Q^\nu \widetilde{H}_T^\nu Q^\nu T^{-1}$$

to avoid multigrid prolongation operators in transformed coordinates. To handle the possible unboundedness of the defect problem (33) we additionally prescribe a set of finite bound constraints

$$a_i \le v_i \le b_i \quad i = 1, \dots, dn, \tag{34}$$

which lead to a minimization problem on a compact set.

The constraints are constructed such that a correction v^ν in untransformed coordinates that complies with (34) will not violate the trust-region constraints of (TRQP) when converted to transformed coordinates.

LEMMA 5.5. *Let w_i denote the number of nonzero entries in the ith row of the sparse transformation matrix T^{-1}, and let*

$$R_j := \left\{1 \le i \le dn \ : \ T_{ij}^{-1} \ne 0\right\}$$

for all $j = 1, \dots, dn$. Let

$$a_j := \max_{i\in R_j}\left\{\frac{-\operatorname{sign}(T_{ij}^{-1})\Delta^k - \bar{u}_i^{\nu+\frac{1}{2}}}{w_i T_{ij}^{-1}}\right\}, \quad b_j := \min_{i\in R_j}\left\{\frac{\operatorname{sign}(T_{ij}^{-1})\Delta^k - \bar{u}_i^{\nu+\frac{1}{2}}}{w_i T_{ij}^{-1}}\right\}. \tag{35}$$

Then if $v \in \mathbb{R}^{dn}$ is such that (34) holds, we get

$$\|\bar{u}^{\nu+\frac{1}{2}} + \bar{v}\|_\infty \le \Delta^k. \tag{36}$$

Proof. Insert (35) into (34) to obtain

$$\max_{i\in R_j}\left\{\frac{-\operatorname{sign}(T_{ij}^{-1})\Delta_k - \bar{u}^{\nu+\frac{1}{2}}}{w_i T_{ij}^{-1}}\right\} \le v_j \le \min_{i\in R_j}\left\{\frac{\operatorname{sign}(T_{ij}^{-1})\Delta_k - \bar{u}^{\nu+\frac{1}{2}}}{w_i T_{ij}^{-1}}\right\}.$$

Now consider the pth constraint in decoupling coordinates

$$\begin{aligned}\bar{u}_p^{\nu+\frac{1}{2}} + \bar{v}_p &= \bar{u}_p^{\nu+\frac{1}{2}} + \sum_{j=1}^{dn} T_{pj}^{-1} v_j \\ &\le \bar{u}_p^{\nu+\frac{1}{2}} + \sum_{\substack{j=1\\ T_{pj}^{-1}\ne 0}}^{dn} T_{pj}^{-1}\frac{\Delta_k - \bar{u}_p^{\nu+\frac{1}{2}}}{w_p T_{pj}^{-1}} = \Delta_k,\end{aligned}$$

which is the upper bound of (36). The lower bound is shown in the same way. ∎

We construct the defect problem constraints by replacing the feasible weights (35) by an averaged version

$$\begin{aligned} a_j &:= \frac{1}{|R_j|}\sum_{i\in R_j}\left\{\frac{-\operatorname{sign}(T_{ij}^{-1})\Delta^k - \bar{u}_i^{\nu+\frac{1}{2}}}{w_i T_{ij}^{-1}}\right\}, \\ b_j &:= \frac{1}{|R_j|}\sum_{i\in R_j}\left\{\frac{\operatorname{sign}(T_{ij}^{-1})\Delta^k - \bar{u}_i^{\nu+\frac{1}{2}}}{w_i T_{ij}^{-1}}\right\}, \end{aligned} \tag{37}$$

which is less restrictive than (35) while capturing the scaling of the decoupling transformation T.

For the approximate solution of the defect problem (33) with constraints (34) and (37), a standard MMG method is applied [8]. Like in the presmoothing step, the Gauss–Seidel smoothers of that method have to take into account the possible nonconvexity of the local one-dimensional problems.

3. *Projection*
The resulting correction v^ν is transformed back into the coordinates in which the linearized nonpenetration constraints decouple

$$\bar{v}^\nu := T^{-1} v^\nu.$$

This transformed correction is then projected onto the defect obstacles of (TRQP), i.e., we define $\hat{v}^\nu$ by

$$\hat{v}_p^\nu := \begin{cases} c_p^{\Delta^k} - \bar{u}_p^{\nu+\frac{1}{2}} & \text{if} \quad \bar{v}_p^\nu > c_p^{\Delta^k} - \bar{u}_p^{\nu+\frac{1}{2}}, \\ \hat{v}_p^\nu, & \text{else.} \end{cases}$$

4. *Line search*
The tentative new iterate $\bar{u}^{\nu+\frac{1}{2}} + \hat{v}^\nu$ is feasible, but it may violate the monotonicity of the TNNMG method. We ensure energy decrease by performing an exact line search in the direction of $\hat{v}^\nu$, as in (24). This is a scalar, quadratic, possibly nonconvex minimization problem. Since it is posed on a closed interval it is guaranteed to have a solution, which can be computed explicitly.

The modified TNNMG algorithm converges globally towards first-order optimal points of the constrained quadratic minimization problem (TRQP).

A convergence proof for the variant without truncation of active components is given in [26].

THEOREM 5.6. *The TNNMG method with an MMG correction described in subsection* 5.3 *either stops at a first-order optimal point of* (TRQP) *or the limit of every convergent subsequence is first-order optimal.*

6. Numerical examples. In this section we illustrate the robustness and global convergence of the filter–trust-region method. The numerical simulations were done using the DUNE framework [1, 2]. A detailed description of the implementation of the linearized nonpenetration constraint (25) can be found in [17, 18].

6.1. Ironing. The ironing problem is often used to test the robustness of the mortar discretization and the applied algebraic solver [18]. In this example a rectangular block is placed under a half-pipe (Figure 4).

The block is fixed at the bottom with homogeneous Dirichlet conditions. For the half-pipe, nonhomogeneous Dirichlet conditions are prescribed on the top boundary: first, the half-pipe is pressed vertically into the block with a prescribed total displacement of 1.4 units (Phase 1). Then, in a second phase, it is swiped over the block horizontally for 2.1 units; see Figure 5.

This benchmark problem is usually solved in small loading steps to stabilize the widely used active-set and penalty methods, which only converge locally [10, 11, 17, 18]. For comparison we implemented a primal–dual active-set method as described in [17]. However, we implemented the correct (c.f. [11]) contact forces generated by the constraint derivative instead of the simplification utilized in [17]. The evaluation

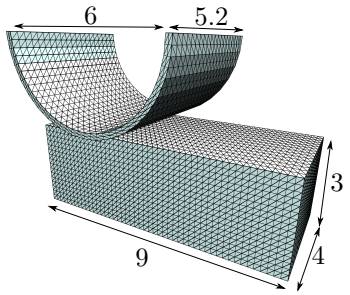

FIG. 4. *The initial configuration of the refined grids.*

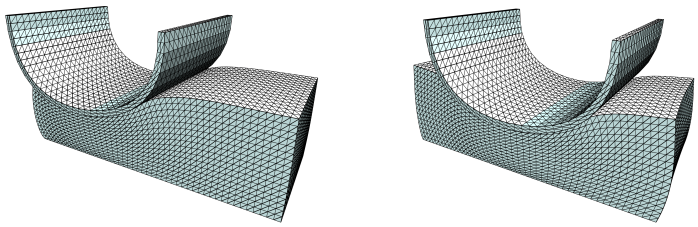

FIG. 5. *Left: Deformed grids after the vertical displacement. Right: Deformed grids after the horizontal displacement.*

of the Hessian of the Lagrangian is performed by a finite difference approximation of the Hessian of the constraint $\nabla^2 c(z)$ and a least-square estimate for the initial Lagrange multipliers. To depict the superior robustness of the proposed method, we solve the ironing problem in only two steps, one for each of the two phases. This lead to divergence of the primal–dual active set method during the second phase and illustrates a crucial advantage of our method: global convergence implies that it is not necessary to detect loading increments such that the initial iterate lies in the region of local convergence. In both phases we choose the block to be the nonmortar body. The bodies are modeled by the nonlinear homogeneous Neo–Hookean material law

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

where $E(\nabla\boldsymbol{\varphi}) := \frac{1}{2}(\nabla\boldsymbol{\varphi}^T\nabla\boldsymbol{\varphi} - \mathrm{Id})$ denotes the Green–Lagrange strain tensor, and we choose the Lamé parameters as

$$\lambda_{\mathrm{pipe}} = 450, \qquad \mu_{\mathrm{pipe}} = 225,$$
$$\lambda_{\mathrm{block}} = \frac{3}{4}, \qquad \mu_{\mathrm{block}} = \frac{3}{8}.$$

We use tetrahedral grids with $42,483$ and $6,993$ degrees of freedom, respectively, obtained by four steps and one step, respectively, of uniform refinement, of corresponding coarser grids.

The two problems are solved by the filter–trust-region method until the H^1-norm of the relative correction is less than 10^{-7}. For the solution of the subproblems (TRQP), we apply the extended TNNMG method from subsection 5.3 until the H^1-norm of the relative correction falls below a tolerance of 10^{-4}, and we use the IPOPT interior point algorithm [22] to solve the problem on the coarsest grid level.

In the filter–trust-region method we used the following constants suggested in [5]: to measure the approximation quality we set $\eta_1 = 0.1$ and $\eta_2 = 0.9$, and in the ϑ-type criterion (32) we use $\kappa_\vartheta = 10^{-4}$. When the trust region radius needs to be decreased we use

$$\Delta^{k+1} = 0.25 \min\left\{ \left\| \bar{u}^k \right\|_\infty, \Delta^k \right\},$$

and we skip increasing it during J-type iterations $\Delta^{k+1} = \Delta^k$. As initial trust-region we chose $\Delta^0 = 0.5$ for both phases. To monitor the convergence of the method towards first-order optimal points, we consider the optimality measure

$$\chi(z^k) := \left| \min_{\substack{\bar{d}^1_{C,0} \le c(z^k) \\ \|\bar{d}\|_\infty \le 1}} \left\langle \nabla \widetilde{m}^k(0), \bar{d} \right\rangle \right|, \tag{38}$$

which vanishes for first-order optimal points of J when additionally $\vartheta(z^k) \to 0$ (see [5, Theorem 12.1.6 and Theorem 15.5.13]). Its evaluation involves a linear minimization problem with bound constraints, which can be solved easily.

Figure 6 shows a comparison of convergence of the filter–trust-region method with sparse inexactly transformed Hessians (IQP) and the respective method using exactly transformed dense Hessians (TQP), which in the following will be denoted by inexact and exact Hessians. This is slightly ambiguous, as both methods use the Hessian of the objective instead of the Lagrangian. For the latter we constructed the decoupling coordinate transformation (28) by computing an LU-decomposition of D_N using UMFPack, which leads to dense blocks in the exactly transformed stiffness matrices H^k_T. For both problems the total iteration numbers are comparable, but due to the sparsity of the inexact Hessian $\widetilde{H}^k_T$, the total wall time required by the inexact version is over 80% smaller than for the filter method with exact Hessians; see Table 1. For the exact (sparse) Hessian, 19 (14) iterates were rejected by the filter, and 1 (15) because of insufficient model approximation (31). In total 10 steps in each phase had to be recomputed for the exact Hessian method and 13, respectively, 16 steps were repeated in the inexact Hessian case. The feasibility restoration phase of the filter method never occurred.

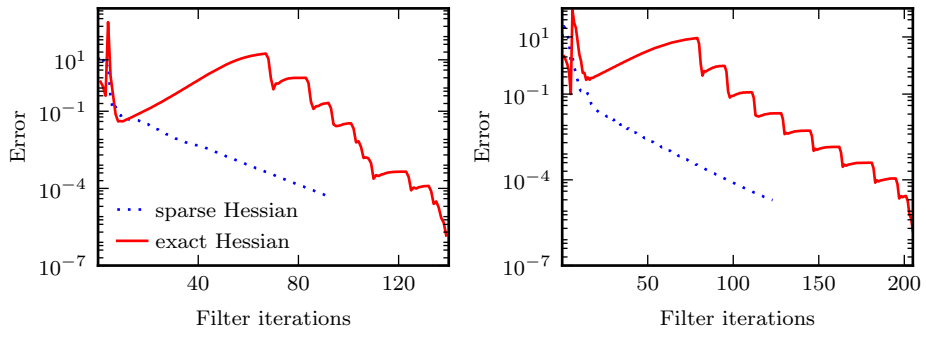

FIG. 6. *Convergence of the optimality measure χ for the filter–trust-region method with inexact sparse Hessians compared to the filter method with exact Hessians. Left: Vertical phase. Right: Horizontal phase.*

TABLE 1
Averaged CPU wall times for the exact and inexact filter for the vertical phase.

	Averaged wall time H_T^k	Averaged wall time $\widetilde{H}_T^k$
Assembly of (TRQP)	41.31s	9.8s
TNNMG solution	73.6s	15.5s
Total time to solution	17 227s	2 917s

In Figure 7 the trust-region radius Δ^k and the infeasibility of both variants are shown during the vertical phase. Once the approximation quality of the subproblems becomes too bad, the step is rejected and the trust-region is decreased to achieve a better approximation of the nonlinear energy J. Surprisingly, in the case of the inexactly transformed Hessians, the growing instability is detected much earlier than in the case of exactly transformed Hessians, leading to a faster convergence in this test problem.

In the left panel of Figure 8 the average number of TNNMG iterations is shown, that are required to solve the local problems (TRQP) for different grid refinement numbers while fixing a given error tolerance bound. The iteration numbers appear

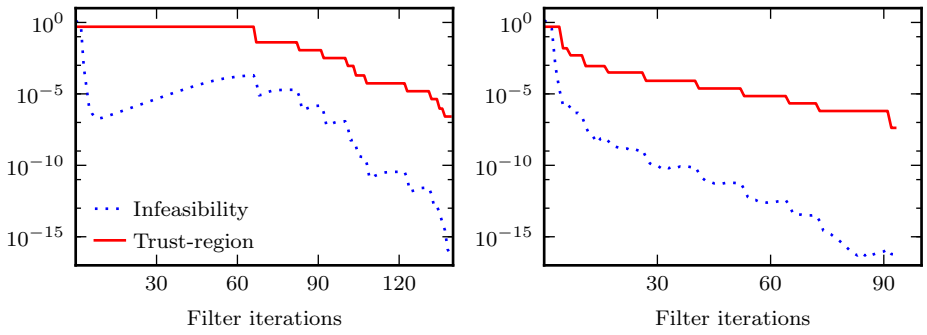

FIG. 7. *The trust-region radius Δ^k and infeasibility ϑ^k of the filter method. Left: Vertical phase using the exactly transformed dense Hessians. Right: Vertical phase using the sparse approximation.*

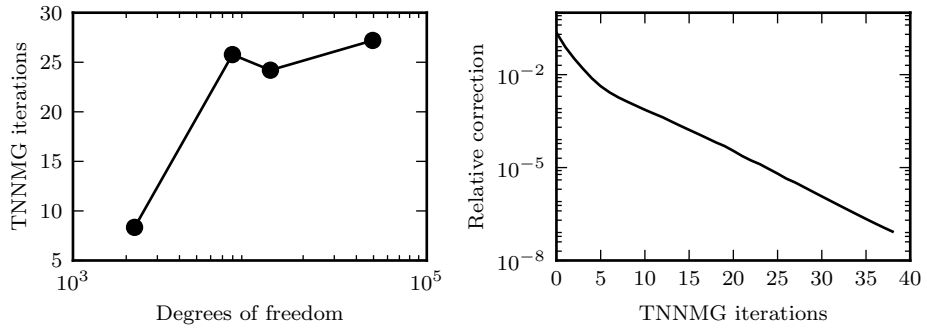

FIG. 8. *Left: TNNMG iterations steps averaged over the total filter–trust-region iteration for an increasing number of degrees of freedom. Right: The H^1-norm of the relative correction for the TNNMG method applied to single problem* (TRQP).

to be bounded, which indicates the mesh independent convergence often observed for multigrid methods [8]. In the right of Figure 8 the fast convergence of the TNNMG method applied to a single (TRQP) is plotted for an error tolerance of 10^{-7}.

7. Conclusion. In this paper we presented a globally convergent solver for large deformation contact problems. The solver is using a decoupling of the linearized contact constraints that allows to apply a fast and efficient multigrid method for the solution of the quadratic constrained subproblems. The method stands out due to its superior robustness over locally convergent methods, enabling to solve problems without applying incremental loading steps. To improve the convergence speed, second-order consistent SQP models could be used to achieve locally super-linear convergence [21], which is part of future work.

REFERENCES

[1] 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 (2008), pp. 121–138, https://doi.org/10.1007/s00607-008-0004-9.

[2] M. Blatt, A. Burchardt, A. Dedner, C. Engwer, J. Fahlke, B. Flemisch, C. Gersbacher, C. Gräser, F. Gruber, C. Grüninger, and O. Sander, *The distributed and unified numerics environment, version* 2.4, Arch. Numer. Softw., 4 (2016), pp. 13–29.

[3] P. Ciarlet, *Mathematical Elasticity, Volume* I*: Three-Dimensional Elasticity*, Elsevier, New York, 1988.

[4] 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 J. Optim., 3 (1993), pp. 164–221.

[5] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000, https://doi.org/10.1137/1.9780898719857.

[6] 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 J. Optim., 13 (2002), pp. 635–659, https://doi.org/10.1137/S1052623499357258.

[7] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function.*, Math. Program., 91 (2002), pp. 239–269, https://doi.org/10.1007/s101070100244.

[8] C. Gräser and R. Kornhuber, *Multigrid methods for obstacle problems*, J. Comput. Math., 27 (2009), pp. 1–44.

[9] C. Gräser, U. Sack, and O. Sander, *Truncated nonsmooth Newton multigrid methods for convex minimization problems*, in Domain Decomposition Methods in Science and Engineering XVIII, vol. 70 of Lecture Notes in Computational Science and Engineering, Springer, New York, 2009, pp. 129–136, https://doi.org/10.1007/978-3-642-02677-5.

[10] 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 (2008), pp. 245–258, https://doi.org/10.1016/j.finel.2007.11.018.

[11] C. Hesch and P. Betsch, *A mortar method for energy–momentum conserving schemes in frictionless dynamic contact problems*, Int. J. Numer. Meth. Engrg., 77 (2009), pp. 1468–1500, https://doi.org/10.1002/nme.2466.

[12] S. Hüeber and B. Wohlmuth, *A primal–dual active set strategy for non-linear multibody contact problems*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 3147–3166.

[13] R. Krause and B. Wohlmuth, *Monotone methods on nonmatching grids for nonlinear contact problems*, SIAM J. Sci. Comput., 25 (2003), pp. 324–347, https://doi.org/10.1137/S1064827502405318.

[14] T. A. Laursen, *Computational Contact and Impact Mechanics*, Springer, New York, 2003, https://doi.org/10.1007/978-3-211-77298-0.

[15] 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. Numer. Meth. Engrg., 36 (1993), pp. 3451–3485, https://doi.org/10.1002/nme.1620362005.

[16] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2006, https://doi.org/10.1007/978-0-387-40065-5.

[17] A. Popp, M. W. Gee, and W. Wall, *Finite deformation contact based on* 3*d dual mortar and semi–smooth Newton approach*, in Trends in Computational Contact Mechanics, Springer, New York, 2010, pp. 57–77.
[18] M. Puso and T. Laursen, *A mortar segment-to-segment contact method for large deformation solid mechanics*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 601–629, https://doi.org/10.1016/j.cma.2003.10.010.
[19] O. Sander, *Multidimensional Coupling in a Human Knee Model*, Ph.D. thesis, Freie Universität Berlin, 2008.
[20] M. Tur, F. J. Fuenmayor, and P. Wriggers, *A mortar-based frictional contact formulation for large deformations using Lagrange multipliers*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 2860–2873.
[21] S. Ulbrich, *On the superlinear local convergence of a filter-SQP method*, Math. Program., 100 (2004), pp. 217–245.
[22] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57.
[23] B. Wohlmuth, *Discretization Methods and Iterative Solvers based on Domain Decomposition*, LNCSE vol. 17, Springer Verlag, New York, 2001, https://doi.org/10.1007/978-3-642-56767-4.
[24] B. Wohlmuth, *Variationally consistent discretization schemes and numerical algorithms for contact problems*, Acta Numerica, 20 (2011), pp. 569–734, https://doi.org/10.1017/S0962492911000079.
[25] P. Wriggers, *Computational Contact Mechanics*, Springer, New York, 2006.
[26] J. Youett, *Dynamic large deformation contact problems and applications in virtual medicine*, Ph.D. thesis, Freie Universität Berlin, 2016, http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000102281.