A Multigrid Semismooth Newton Method for Semilinear Contact Problems

Michael Ulbrich · Stefan Ulbrich · Daniela Bratzke

Abstract This paper develops and analyzes multigrid semismooth Newton methods for a class of inequality-constrained optimization problems in function space which are motivated by and include linear elastic contact problems of Signorini type. We show that after a suitable Moreau-Yosida type regularization of the problem superlinear local convergence is obtained for a class of semismooth Newton methods. In addition, estimates for the order of the error introduced by the regularization are derived. The main part of the paper is devoted to the analysis of a multilevel preconditioner for the semismooth Newton system. We prove a rigorous bound for the contraction rate of the multigrid cycle which is robust with respect to sufficiently small regularization parameters and the number of grid levels. Moreover, it applies to adaptively refined grids. The paper concludes with numerical results.

1 Introduction

In this paper, a class of multigrid semismooth Newton methods for constrained optimization problems is developed and systematically analyzed. The considered problem class is motivated by linear elastic contact problems, which are included as a special case. We work with the same nonpenetration constraints as in linear contact, but compared to linear elasticity, we cover more general cost functions than the quadratic elastic energy. More precisely, the problems have the following form:

$$\min_{u \in \mathbf{U}} J(u) \quad \text{subject to} \quad \tau_C^n(u) \le \psi \quad \text{on } \Gamma_C. \tag{1}$$

Here, $\mathbf{U} = \{u \in H^1(\Omega)^d : \tau_D u = 0\}$, $\Omega \subset \mathbb{R}^d$ is a bounded open domain, and Γ_C and Γ_D are disjoint subsets of the boundary $\partial\Omega$ of Ω. Further, $\tau_D : H^1(\Omega)^d \to H^{1/2}(\Gamma_D)^d$ is the trace operator on Γ_D, i.e., $\tau_D(u)(x) = u(x)$ for all $x \in \Gamma_D$ if u is continuous on $\bar{\Omega}$;

M. Ulbrich
Technische Universität München, Department of Mathematics, Chair of Mathematical Optimization, Boltzmannstr. 3, D-85747 Garching b. München, Germany, E-mail: mulbrich@ma.tum.de

S. Ulbrich
TU Darmstadt, Fachbereich Mathematik, Chair of Nonlinear Optimization, Dolivostr. 15, D-64293 Darmstadt, Germany, E-mail: ulbrich@mathematik.tu-darmstadt.de. This author was supported by Deutsche Forschungsgemeinschaft within SFB 666 and SFB 805.

D. Bratzke
TU Darmstadt, Fachbereich Mathematik, Chair of Nonlinear Optimization, Dolivostr. 15, D-64293 Darmstadt, Germany, E-mail: bratzke@mathematik.tu-darmstadt.de. This author was supported by Deutsche Forschungsgemeinschaft within SFB 666.

$\tau_C^n = n^T \tau_C : \mathbf{U} \to V := H^{1/2}(\Gamma_C)$ is the normal trace operator on Γ_C, with n denoting the outer unit normal; further, $\psi : \Gamma_C \to \mathbb{R}$ is a given, sufficiently smooth function. Targeting for Newton-type methods, the objective function $J : \mathbf{U} \to \mathbb{R}$ is assumed to be twice continuously differentiable. Some additonal requirements on Ω, Γ_C, Γ_D, and J will be given below.

Problem of the form (1) arise, e.g., in elastic contact problems, where Ω is the reference configuration of an elastic body, $u(x)$ denotes the displacement of the reference point $x \in \Omega$, and J is the total energy. The constraint then expresses that the normal displacement on Γ_C shall not exceed ψ, which can be interpreted as the normal distance to a rigid obstacle.

Our approach uses a Moreau-Yosida (MY) regularization to obtain a nonsmooth approximation of the first order optimality conditions that is suitable for applying a semismooth Newton method. We develop a superlinear convergence theory in function space as well as error estimates for the MY-regularized solutions. A particular focus of the paper is put on a multigrid method for preconditioning or solving the linear semismooth Newton systems.

The MY-regularization is required since the problem contains a pointwise inequality constraint that is posed in a Sobolev space $V := H^{1/2}(\Gamma_C)$. The natural space for the Lagrange multiplier is then the dual space V' and thus the complementarity condition cannot be written in a pointwise almost everywhere form. For sufficiently smooth data, regularity results for the solution can be used to infer that the multiplier is an L^q-function. Then, a nonsmooth pointwise reformulation of the complementarity condition would in fact be possible. However, in a primal-dual formulation of the optimality system, replacing the multiplier space V' by L^q does not provide a framework where the linear operator in the Newton system is boundedly invertible. Thus, a dual regularization would be required to fix this [32,34]; as we will see, such a regularization is equivalent to the Moreau-Yosida approach, see also [29, Sec. 2.1] and [34, Sec. 8.2.4 and 9.2]. A different alternative, chosen, e.g., in [4,17,24,36], is to consider the problem after discretization and relying on the fact that then all norms are equivalent. However, this comes at the cost of dimension-dependent condition numbers and norm equivalence constants. This regularization by discretization (or well-posedness through discretization) strategy requires to combine it with a nested iteration from coarse to fine grids to compensate for the lack of mesh-independence since a function space counterpart of the discrete algorithm is then missing. We therefore prefer to work with the MY-regularization, which by our error estimates can be balanced with the discretization error, to have a well-posed algorithm also in function space.

Extending results in [29,32,34], we show that a regularization with parameter $\alpha > 0$ results in a solution that deviates at most by $o(\alpha^{1/2})$ (as $\alpha \to 0^+$) from the true solution if the true Lagrange multiplier is in $L^2(\Gamma_C)$. Further, if the Lagrange multiplier is in $H^s(\Gamma_C)$, $0 < s \le 1/2$ and the derivative of J is κ-Hölder continuous near the solution (which holds globally with $\kappa = 1$ for linear elasticity), we show that the convergence rate is $O(\alpha^{\frac{1+2s}{2+4s(1-\kappa)}})$. Multiplier regularity can be ensured under suitable assumptions by invoking regularity results for elastic obstacle problems [19,20,26]. We then introduce a finite element discretization and, based on this, a discrete counterpart of the semismooth Newton's method.

The main part of the paper is devoted to the analysis of a multigrid cycle that can be used stand alone or as a preconditioner to solve the semismooth Newton system to the desired accuracy. Due to the regularization, multigrid methods for the semismooth Newton system require special care. The regularization introduces an algebraic (i.e. non-differential) operator acting on Γ_C that is strongly weighted. This requires to develop a special multigrid iteration. Building on a general framework of multilevel convergence theory [38], we prove a guaranteed contraction rate that is independent of the number of grid levels and uniform for all regularization parameters $\alpha \in (0, \alpha_h^+]$, where the upper bound $\alpha_h^+ > 0$ depends on the mesh size h of the finest grid in the contact region, but is larger than required to balance the regularization and discretization error. This robustness with respect to $\alpha \to 0^+$ means that in the limit it is also applicable to the regularization via discretization approach. A direct application of the latter strategy would result in systems with Dirichlet boundary conditions on the estimated active set, which is usually not resolvable on coarser grids (in this context, see also [35]). In the MY-approach, instead of these Dirichlet conditions a penalized version of them occurs.

Several different multigrid approaches for contact problems and for related classes of variational inequalities have been proposed in the literature, see the surveys [9,36]. On the one hand, there are methods that target the variational inequality directly: the monotone multigrid method [21–23,37], projected subspace decomposition [1,25] as well as subset decomposition [30]. Furthermore, a class of optimal quadratic programming algorithms is systematically investigated in the book [7]. On the other hand, as in our approach, multigrid methods can be applied to linear subproblems arising in Newton-type methods. Semismooth Newton methods [8,11,18,28,32–34] for elastic contact problems have been investigated in, e.g., [4,12,17,27, 29,36]. Similar as in our approach, the methods in [12,29] use a regularized formulation for the semismooth Newton method in order to achieve an appropriate function space framework. Most other approaches build on discretized settings that do not have a suitable counterpart in function space. This, in particular, applies to the available literature on multigrid semismooth Newton methods [4,17,24,36]. In contrast, we consider multigrid methods for semismooth Newton systems that arise for regularized problems, which are derived from the function space theory of semismooth Newton methods. We also mention that it is possible to use semismooth approaches also in the case of frictional contact problems [5,15,29].

This paper is organized as follows: In section 2 we introduce the class of constrained optimization problems in function space addressed in this paper and relate them to elastic contact problems. Optimality conditions and the regularized problem as well as estimates for the order of the error introduced by the regularization are developed in section 3. A semismooth Newton method for the regularized contact problem is developed and analyzed in section 4. Section 5 forms the main part of the paper. Using a suitable discretization, it provides a detailed development and analysis of a multigrid preconditioner for solving the discretized semismooth Newton systems. The efficiency of the approach is demonstrated by numerical tests in section 6.

2 Problem setting

Let the bounded open set $\Omega \subset \mathbb{R}^d$ ($d = 2$ or $d = 3$) have a sufficiently smooth boundary.

The boundary Γ of Ω contains two *disjoint* subsets Γ_D and Γ_C with $\Gamma = \bar{\Gamma}_D \cup \bar{\Gamma}_N \cup \bar{\Gamma}_C$. On Γ_D we pose homogeneous Dirichlet conditions. To avoid technical difficulties, we assume throughout this section:

- The boundary Γ is sufficiently smooth (for $C^{1,1}$-boundary, the normal trace operator maps from $H^1(\Omega)^d$ to $H^{1/2}(\Gamma)$) [19, sec. 5.3])
- Γ_D is either empty or has positive surface measure.
- Γ_C has positive surface measure.
- The sets Γ_D and Γ_C have positive distance from each other.

We consider the problem (1) as described in section 1.

Introducing the closed convex cone $\mathcal{K} = \{v \in V : v \geq 0 \text{ on } \Gamma_C\}$, where $V = H^{1/2}(\Gamma_C)$, and the bounded linear operator $B : \mathbf{U} \to V$, $Bu = \tau_C^n(u)$, we can write (1) as a cone constrained optimization problem:

$$\min_{u \in \mathbf{U}} J(u) \quad \text{subject to} \quad \psi - Bu \in \mathcal{K}. \tag{2}$$

Example 1 (Linear elasticity) As an example, we discuss linear elasticity. Let the subset $\Gamma_N \subset \Gamma$ be disjoint to $\Gamma_C \cup \Gamma_D$ with sufficiently smooth boundary. The objective function

(energy)

$$J(u) = \tfrac{1}{2}a(u,u) - f(u)$$

is convex and quadratic, with a bilinear form $a : H^1(\Omega)^d \times H^1(\Omega)^d \to \mathbb{R}$ and a linear form $f \in (H^1(\Omega)^d)'$. The latter is defined as

$$f(u) = \int_\Omega f_V^T u \, dx + \int_{\Gamma_N} f_S^T u \, dS(x), \tag{3}$$

where $f_V \in L^2(\Omega)^d$ and $f_S \in L^2(\Gamma_N)^d$ denote the volume force in Ω and the surface traction on Γ_N, respectively. In the case of Lamé material, the bilinear form is given by

$$a(u,v) = 2\mu \int_\Omega \varepsilon(u){:}\varepsilon(v) \, dx + \lambda \int_\Omega \mathrm{div}(u)\mathrm{div}(v) \, dx, \tag{4}$$

with the Lamé parameters $\lambda, \mu > 0$ and the strain rate tensor $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$. There holds with a constant $M_a > 0$:

$$|a(u,v)| \le M_a |u|_{H^1(\Omega)^d} \, |v|_{H^1(\Omega)^d} \quad \forall\, u, v \in H^1(\Omega)^d.$$

Furthermore, if Γ_D has a positive surface measure, then Korn's inequality [13] yields that there exists $C_A = C_A(\Gamma_D, \Omega) > 0$ such that

$$\|u\|^2_{H^1(\Omega)^d} \le C_A a(u,u) \quad \forall\, u \in \mathbf{U}. \tag{5}$$

Hence, the objective function is quadratic and strongly convex. □

3 Optimality conditions and regularization

If the boundary $\partial\Omega$ (and thus n) is sufficiently smooth, then the normal trace operator $B : \mathbf{U} \to V$ is onto. This is a constraint qualification for (1). Therefore, we can state the following optimality conditions. To this end, we define the Lagrange function

$$L : \mathbf{U} \times V' \to \mathbb{R}, \quad L(u,y) = J(u) + \langle y, Bu - \psi \rangle_{V',V}. \tag{6}$$

Proposition 1 *Let $\bar u \in \mathbf{U}$ be a local solution of* (2). *Then there exists a unique Lagrange multiplier $\bar y \in V'$ such that the following Karush-Kuhn-Tucker (KKT) conditions hold:*

$$L_u(\bar u, \bar y) = J'(\bar u) + B^* \bar y = 0, \tag{7}$$

$$\bar y \in \mathcal{K}^*, \quad L_y(\bar u, \bar y) = B\bar u - \psi \le 0, \quad \langle \bar y, L_y(\bar u, \bar y) \rangle_{V',V} = 0. \tag{8}$$

Here, $\mathcal{K}^* = \{y \in V' : \langle y, v \rangle_{V',V} \ge 0 \; \forall\, v \in \mathcal{K}\}$ denotes the dual cone of the closed convex cone $\mathcal{K}$.

We target at applying a semismooth Newton's method to a suitable nonsmooth equation reformulation of the optimality system. A by now well established way to express a complementarity condition

$$w_1 \ge 0, \quad w_2 \ge 0, \quad w_1 w_2 = 0 \quad \text{a.e. in } \Omega \tag{9}$$

between Lebesgue-functions w_1, w_2 is to write them as

$$w_1 - [w_1 - \theta w_2]_+ = 0 \tag{10}$$

where $[\cdot]_+ := \max(0, \cdot)$ is applied pointwise and $\theta > 0$ is fixed. In our situation, however, unless we can invoke regularity results, we have a complementarity between V and V', and the latter is a space of distributions, not of pointwise a.e. defined functions. There thus are two options, either discretizing first and then applying the reformulation in finite dimensions, or regularizing the KKT conditions to recover an L^q setting. In the discretization approach, there then would be no clear and well-posed function space equivalent of the discrete formulation. Hence, we prefer a regularization and adopt the Moreau-Yosida approach in the following. It can be developed in at least two equivalent ways, either by primal penalization or by dual regularization. The primal approach approximates (1) by the unconstrained problem

$$\min_u \ J_\alpha(u) := J(u) + \frac{1}{2\alpha}\|[\alpha\hat{y} + Bu - \psi]_+\|^2_{L^2(\Gamma_C)}. \tag{11}$$

with $\alpha > 0$ and $\hat{y} \in L^2(\Gamma_C)_+ = \{y \in L^2(\Gamma_C) : y \geq 0 \text{ a.e. on } \Gamma_C\}$.

This setting is such that J_α is differentiable and J'_α is semismooth. In fact, the space V is continuously embedded in $L^q(\Gamma_C)$ for all $q \in [2,\infty]$ if $d = 1$, $q \in [2,\infty)$ if $d = 2$, and $q \in [2, 2(d-1)/(d-2)] \supsetneq \{2\}$ if $d \geq 3$. Since $t \mapsto \frac{1}{2}[t]_+^2$ is differentiable with Lipschitz continuous derivative $[t]_+$, the Nemytskii operator $v \mapsto \frac{1}{2}[v]_+^2$ is continuously differentiable from $L^q(\Gamma_C)$, $q \geq 2$, to $L^{q/2}(\Gamma_C) \subset L^1(\Gamma_C)$ with derivative $d \mapsto [v]_+ d$. Hence, $v \mapsto \frac{1}{2}\|[v]_+\|^2_{L^2} = \frac{1}{2}\int_{\Gamma_C}[v(x)]_+^2\, dS(x)$ is continuously differentiable from $L^q(\Gamma_C)$, $q \geq 2$, to $\mathbb{R}$ with derivative $[v]_+ \in L^{q'}(\Gamma_C)$, $1/q + 1/q' = 1$. Since the argument of $[\cdot]_+$ in (11) maps affine linearly to at least $L^2(\Gamma_C)$, we see that the objective function $J_\alpha(u)$ of (11) is continuously differentiable with derivative

$$J'_\alpha(u) = J'(u) + B^*[\hat{y} + \alpha^{-1}(Bu - \psi)]_+ \in \mathbf{U}'.$$

At a solution u_α of (11), there thus holds

$$J'(u_\alpha) + B^*[\hat{y} + \alpha^{-1}(Bu_\alpha - \psi)]_+ = 0. \tag{12}$$

The semismoothness of this nonsmooth equation will be shown in Theorem 4.

Comparing with (7), the quantity $y_\alpha := [\hat{y} + \alpha^{-1}(Bu_\alpha - \psi)]_+$ can be viewed as an approximation of $\bar{y}$. Thus, an extended way of writing (12) is

$$J'(u_\alpha) + B^* y_\alpha = 0, \tag{13}$$

$$y_\alpha - [\hat{y} + \alpha^{-1}(Bu_\alpha - \psi)]_+ = 0. \tag{14}$$

We briefly show that there is also a dual way of deriving (13), (14). For this, we add a dual regularization to the Lagrangian, which results in $L_\alpha : \mathbf{U} \times L^2(\Gamma_C) \to \mathbb{R}$,

$$L_\alpha(u,y) = J(u) + (y, Bu - \psi)_{\Gamma_C} - \frac{\alpha}{2}\|y - \hat{y}\|^2_{L^2(\Gamma_C)}. \tag{15}$$

We note that (7), (8) is a first order condition for $(\bar{u},\bar{y})$ to be a saddle point of L on $\mathbf{U} \times \mathcal{K}^*$. If we formulate the corresponding first order saddle point conditions for L_α at (u_α, y_α) on $\mathbf{U} \times L^2(\Gamma_C)_+$, we obtain

$$(L_\alpha)_u(u_\alpha, y_\alpha) = 0, \tag{16}$$

$$y_\alpha \geq 0, \ \ (L_\alpha)_y(u_\alpha, y_\alpha) \leq 0, \ \ \langle y_\alpha, (L_\alpha)_y(u_\alpha, y_\alpha)\rangle_{\Gamma_C} = 0. \tag{17}$$

Now (16) is exactly (13). Further, by the equivalence of (9) and (10), the condition (17) can be written as

$$y_\alpha - [y_\alpha + \theta(L_\alpha)_y(u_\alpha, y_\alpha)]_+ = 0. \tag{18}$$

Since $[L_\alpha]_y(u,y) = Bu - \psi - \alpha(y - \hat{y})$, the choice $\theta = \alpha^{-1}$ shows that (17), (18), and (14) are all equivalent.

We now will investigate the follwing aspects:

- error estimates for $u_\alpha - \bar{u}$ and $y_\alpha - \bar{y}$.
- a semismooth Newton's method for solving the regularized problem.
- a multigrid preconditioner for the semismooth Newton system.

3.1 Error Estimates

We now derive error etimates for the MY-regularized solution and the corresponding multiplier. We will work under the following assumption:

Assumption (E)
Let $\bar{u}$ solve (1) and denote by $\bar{y} \in V'$ the corresponding Lagrange multiplier. Let $\delta, \sigma > 0$ be such that $J : \mathbf{U} \to \mathbb{R}$ is continuously differentiable in a neighborhood of $\bar{\mathcal{U}}_\delta = \{u \in \mathbf{U} : \|u - \bar{u}\|_{\mathbf{U}} \le \delta\}$ and there holds:

$$J(u+d) - J(u) \ge \langle J'(u), d\rangle_{\mathbf{U}',\mathbf{U}} + \frac{\sigma}{2}\|d\|_{\mathbf{U}}^2 \quad \forall\, u, u+d \in \bar{\mathcal{U}}_\delta. \tag{19}$$

The condition (19) is equivalent to the following local strict monotonicity of J':

$$\langle J'(u+d) - J'(u), d\rangle_{\mathbf{U}',\mathbf{U}} \ge \sigma\|d\|_{\mathbf{U}}^2 \quad \forall\, u, u+d \in \bar{\mathcal{U}}_\delta.$$

In the case of linear elasticity (see Example 1), this assumption is satisfied if Korn's inequality is applicable to ensure (5).

For the analysis, we restrict the minimization of J_α to the set $\bar{\mathcal{U}}_\delta$:

$$\min_u \; J_\alpha(u) := J(u) + \frac{1}{2\alpha}\|[\alpha\hat{y} + Bu - \psi]_+\|_{L^2(\Gamma_C)}^2 \quad \text{s.t} \quad u \in \bar{\mathcal{U}}_\delta. \tag{20}$$

The constraint $u \in \bar{\mathcal{U}}_\delta$ ensures existence of a unique solution, since $\bar{\mathcal{U}}_\delta$ is closed, convex, and bounded in $\mathbf{U}$ and J is convex, thus weakly lower semicontinuous, on $\bar{\mathcal{U}}_\delta$. We will prove convergence $u_\alpha \to \bar{u}$ in $\mathbf{U}$ ($\alpha \to 0^+$) and also will give convergence rates under an additional assumption. The convergence in $\mathbf{U}$ implies that for $\alpha > 0$ sufficiently small the constraint $u \in \bar{\mathcal{U}}_\delta$ is inactive at u_α.

In the convergence analysis, we use the following abbreviations:

$$d_\alpha = u_\alpha - \bar{u}, \;\; r_\alpha = \alpha\hat{y} + Bu_\alpha - \psi, \;\; v_\alpha = [r_\alpha]_+, \;\; y_\alpha = \alpha^{-1}v_\alpha. \tag{21}$$

Further, we write $\langle\cdot,\cdot\rangle$ for the dual pairing of compatible $\mathbb{R}^n$-valued (generalized) functions on Ω, thus abbreviating $\langle\cdot,\cdot\rangle_{\mathbf{U}',\mathbf{U}}$, $\langle\cdot,\cdot\rangle_{L^{q'}(\Omega)^n, L^q(\Omega)^n}$, etc.. We use the notation $\langle\cdot,\cdot\rangle_{\Gamma_C}$ to abbreviate $\langle\cdot,\cdot\rangle_{V',V}$, etc., and we will use $\|\cdot\|_{L^2}$, $\|\cdot\|_{H^s}$, etc., to denote the respective function space norms on Γ_C.

We need the optimality conditions of (20), which are given by

$$u_\alpha \in \bar{\mathcal{U}}_\delta, \quad \langle J'(u_\alpha) + \alpha^{-1}B^*[\alpha\hat{y} + Bu_\alpha - \psi]_+, u - u_\alpha\rangle \ge 0 \quad \forall\, u \in \bar{\mathcal{U}}_\delta. \tag{22}$$

Theorem 1 *Under Assumption (E) and with $\hat{y} \in L^2(\Gamma_C)_+$ the following holds for $\alpha \to 0^+$:*

a) $u_\alpha \to \bar{u}$ *in* $\mathbf{U}$,

b) $\|[Bu_\alpha - \psi]_+\|_{L^2} \le \|[\alpha\hat{y} + Bu_\alpha - \psi]_+\|_{L^2} = o(\sqrt{\alpha})$,

c) $\alpha^{-1}[\alpha\hat{y} + Bu_\alpha - \psi]_+ \to \bar{y}$ *in* V'.

Proof From (22) with $u = \bar{u}$, we obtain with (19):

$$\begin{aligned} 0 &\le \langle J'(u_\alpha), -d_\alpha\rangle + \langle y_\alpha, -Bd_\alpha\rangle_{\Gamma_C} \\ &\le J(\bar{u}) - J(u_\alpha) - \frac{\sigma}{2}\|d_\alpha\|_{\mathbf{U}}^2 - \langle y_\alpha, Bd_\alpha\rangle_{\Gamma_C}. \end{aligned} \tag{23}$$

Further, using $y_\alpha = \alpha^{-1}v_\alpha = \alpha^{-1}[\alpha\hat{y} + Bu_\alpha - \psi]_+ \geq 0$ yields:

$$\begin{aligned} &- \langle y_\alpha, Bd_\alpha\rangle_{\Gamma_C} = \langle y_\alpha, B\bar{u} - \psi\rangle_{\Gamma_C} - \langle y_\alpha, Bu_\alpha - \psi\rangle_{\Gamma_C} \\ &\leq -\langle y_\alpha, Bu_\alpha - \psi\rangle_{\Gamma_C} = \langle v_\alpha, \hat{y}\rangle_{\Gamma_C} - \alpha^{-1}\langle v_\alpha, r_\alpha\rangle_{\Gamma_C} \\ &= \langle v_\alpha, \hat{y}\rangle_{\Gamma_C} - \alpha^{-1}\|v_\alpha\|_{L^2}^2 \leq \alpha\|\hat{y}\|_{L^2}^2 - \frac{3}{4\alpha}\|v_\alpha\|_{L^2}^2. \end{aligned} \tag{24}$$

Thus, $\frac{\sigma}{2}\|d_\alpha\|_{\mathbf{U}}^2 + \frac{3}{4\alpha}\|v_\alpha\|_{L^2}^2 \leq J(\bar{u}) - J(u_\alpha) + \alpha\|\hat{y}\|_{L^2}^2$.

Since Assumption (E) implies that J is bounded below on $\bar{\mathcal{U}}_\delta$, this shows that $(\alpha^{-1}\|v_\alpha\|_{L^2}^2)$ is bounded. In particular, $v_\alpha \to 0$ in $L^2(\Gamma_C)$.

Let (u_{α_l}), $\alpha_l \to 0^+$, be such that $\|d_{\alpha_l}\|_{\mathbf{U}}$ converges to $\limsup_{\alpha\to 0^+}\|d_\alpha\|_{\mathbf{U}}$. Due to boundedness, we can select this sequence in such a way that (u_{α_l}) converges weakly in $\mathbf{U}$ to a limit $\tilde{u} \in \mathbf{U}$. Then $Bu_{\alpha_l} \to B\tilde{u}$ weakly in V and thus, by compactness of $V \subset L^2(\Gamma_C)$, strongly in $L^2(\Gamma_C)$. This yields $[B\tilde{u} - \psi]_+ = \lim_{l\to\infty} v_{\alpha_l} = 0$ in $L^2(\Gamma_C)$ and, hence, $B\tilde{u} \leq \psi$ a.e. on Γ_C. Further, we obtain with (7), (23), and (24):

$$\begin{aligned} &\sigma\|d_\alpha\|_{\mathbf{U}}^2 \leq \langle J'(u_\alpha) - J'(\bar{u}), d_\alpha\rangle \leq -\langle y_\alpha, Bd_\alpha\rangle_{\Gamma_C} + \langle \bar{y}, Bd_\alpha\rangle_{\Gamma_C} \\ &\leq \alpha\|\hat{y}\|_{L^2}^2 - \tfrac{3}{4\alpha}\|v_\alpha\|_{L^2}^2 + \langle \bar{y}, Bu_\alpha - \psi\rangle_{\Gamma_C} - \langle \bar{y}, B\bar{u} - \psi\rangle_{\Gamma_C} \\ &= \alpha\|\hat{y}\|_{L^2}^2 - \tfrac{3}{4\alpha}\|v_\alpha\|_{L^2}^2 + \langle \bar{y}, Bu_\alpha - \psi\rangle_{\Gamma_C} \leq \alpha\|\hat{y}\|_{L^2}^2 + \langle \bar{y}, Bu_\alpha - \psi\rangle_{\Gamma_C}. \end{aligned} \tag{25}$$

Thus, by the choice of (u_{α_l}) and using $u_{\alpha_l} \to \tilde{u}$ weakly in $\mathbf{U}$:

$$\limsup_{\alpha\to 0^+}\|d_\alpha\|_{\mathbf{U}}^2 \leq \lim_{l\to\infty} \alpha_l\|\hat{y}\|_{L^2}^2 + \langle \bar{y}, Bu_{\alpha_l} - \psi\rangle_{\Gamma_C} = \langle \bar{y}, B\tilde{u} - \psi\rangle_{\Gamma_C} \leq 0,$$

hence $u_\alpha \to \bar{u}$ in $\mathbf{U}$ and $\tilde{u} = \bar{u}$. Further, we obtain from (25):

$$0 \leq \frac{3}{4\alpha}\|v_\alpha\|_{L^2}^2 \leq \alpha\|\hat{y}\|_{L^2}^2 + \langle \bar{y}, Bu_\alpha - \psi\rangle_{\Gamma_C} \to \langle \bar{y}, B\bar{u} - \psi\rangle_{\Gamma_C} \leq 0,$$

hence $\alpha^{-1}\|v_\alpha\|_{L^2}^2 \to 0$.

To show c), we note that for $\alpha > 0$ sufficiently small, a) implies that u_α lies in the interior of $\bar{\mathcal{U}}_\delta$. Hence, for $\alpha > 0$ small, (22) becomes (12). Subtracting (7) yields

$$B^*(y_\alpha - \bar{y}) = J'(\bar{u}) - J'(u_\alpha) \to 0 \quad \text{in } \mathbf{U}' \quad (\alpha \to 0^+).$$

Since B is surjective, the open mapping theorem yields $y_\alpha - \bar{y} \to 0$ in V'.

Our next results provide rates of convergence. To this end, we need additional regularity of the optimal Lagrange multiplier $\bar{y}$. It can be obtained from regularity results for the system (7), (8).

Example 2 (Linear elasticity) As an example for higher multiplier regularity, we consider the case of linear elasticity, see Example 1. The following can be shown, see ([19, Thm. 6.5]):

Let $\Omega \subset \mathbb{R}^3$ have a $C^{1,1}$-boundary and let f be given by (3) *with $f_V \in L^2(\Omega)^3$ and $f_S \in L^2(\Gamma \setminus \Gamma_D)^3 \cap H^1(\Gamma_C)^3$. Suppose that the part Γ_C of the boundary is smooth enough and that there exists a function u^1 such that $n^T u^1 = \psi$, $u^1 \in H^3(\Omega^*)$ for every compact domain $\Omega^* \subset \Omega \cup \Gamma_C$. Then, for a compact set Ω^0 in $\Omega \cup \Gamma_C$ whose relative neighborhood to $\bar{\Omega}$ belongs to Ω^*, the solution of* (1) *satisfies $u \in H^2(\Omega^0)^3$.*

We also refer to [20,26]. Now, if $\mathcal{O}$ is an open set such that $\hat{\Omega} := \Omega \cap \mathcal{O} \neq \emptyset$ is sufficiently smooth and the solution satisfies $\bar{u} \in H^2(\hat{\Omega})^d$ then an integration by parts of $a(\bar{u}, v) - F(v) + \int_{\Gamma_C} \bar{y}\, n^T v\, dS = 0$, $v \in \mathbf{U}$, $\operatorname{supp} v \subset \mathcal{O}$ yields:

$$-\operatorname{div}\sigma(\bar{u}) = f_V \;\text{ in } \hat{\Omega}, \;\; \sigma(\bar{u})n = f_S \;\text{ in } \hat{\Gamma}_N, \;\; \bar{u} = 0 \;\text{ in } \hat{\Gamma}_D, \;\; \sigma(\bar{u})n + \bar{y}n = 0 \;\text{ in } \hat{\Gamma}_C,$$

where $\hat{\Gamma}_C = \Gamma_C \cap \mathcal{O}$, etc., and $\sigma(u) = 2\mu\varepsilon(u) + \lambda\operatorname{div}(u)\, I$ is the stress tensor. Now $\bar{u} \in H^2(\hat{\Omega})^d$ implies $\bar{y} = -n^T\sigma(\bar{u})n \in H^{1/2}(\hat{\Gamma}_C) \subset L^2(\hat{\Gamma}_C)$.

The example thus shows that the multiplier regularity assumptions of the following theorem are reasonable.

Theorem 2 *Let the Assumption (E) and $\hat{y} \in L^2(\Gamma_C)_+$ hold.*

a) *Assume that the Lagrange multiplier $\bar{y}$ in* (7) *satisfies $\bar{y} \in L^2(\Gamma_C)$. Then there holds*

$$\|[Bu_\alpha - \psi]_+\|_{L^2} \le \|[\alpha\hat{y} + Bu_\alpha - \psi]_+\|_{L^2} \le \sqrt{2}\,\alpha(\|\bar{y}\|_{L^2} + \|\bar{y} - \hat{y}\|_{L^2}),$$
$$\|u_\alpha - \bar{u}\|_{\mathbf{U}} \le \sqrt{\frac{\alpha}{2\sigma}}\|\bar{y} - \hat{y}\|_{L^2}.$$

b) *Assume that for $s \in (0, 1/2]$ there holds $\bar{y} - \hat{y} \in H^s(\Gamma_C)$. Further, let there exist $\alpha_0 > 0$ such that* (12) *holds for all $0 < \alpha \le \alpha_0$ and that there exist $C_L > 0$ and $\kappa \in (0, 1]$ with*

$$\|J'(u_\alpha) - J'(\bar{u})\|_{\mathbf{U}} \le C_L\|u_\alpha - \bar{u}\|_{\mathbf{U}}^\kappa \quad \forall\, \alpha \in (0, \alpha_0]. \tag{26}$$

Then there holds for $\alpha \to 0^+$:

$$\frac{\sigma}{2}\|u_\alpha - \bar{u}\|_{\mathbf{U}}^2 + \frac{\alpha}{2}\|\bar{y} - y_\alpha\|_{L^2}^2 \le O((1 + \|\bar{y} - \hat{y}\|_{H^s})\alpha^{\frac{1+2s}{1+2s(1-\kappa)}}).$$

Remark 1 By Theorem 1, the assumption that (12) holds for sufficiently small $\alpha > 0$ is always satisfied. □

Remark 2 For the case of linear elasticity, see Example 1, (26) is satisfied with $\kappa = 1$. □

Proof For both assertions a) and b), we start as in (25) and use (24):

$$\begin{aligned}\sigma\|d_\alpha\|_{\mathbf{U}}^2 &\le \langle J'(u_\alpha) - J'(\bar{u}), d_\alpha\rangle \le -\langle y_\alpha, Bd_\alpha\rangle_{\Gamma_C} + \langle \bar{y}, Bd_\alpha\rangle_{\Gamma_C} \\ &\le \langle v_\alpha, \hat{y}\rangle_{\Gamma_C} - \alpha^{-1}\|v_\alpha\|_{L^2}^2 + \langle \bar{y}, Bd_\alpha\rangle_{\Gamma_C}.\end{aligned}$$

There holds, using $r_\alpha \le v_\alpha$, $\bar{y} \ge 0$ and $\langle \bar{y}, B\bar{u} - \psi\rangle_{\Gamma_C} = 0$:

$$\langle \bar{y}, Bd_\alpha\rangle_{\Gamma_C} = \langle \bar{y}, (Bu_\alpha - \psi) - (B\bar{u} - \psi)\rangle_{\Gamma_C} = \langle \bar{y}, r_\alpha - \alpha\hat{y}\rangle_{\Gamma_C} \le \langle \bar{y}, v_\alpha - \alpha\hat{y}\rangle_{\Gamma_C}.$$

Thus, we obtain, using $y_\alpha = \alpha^{-1}v_\alpha$:

$$\begin{aligned}\sigma\|d_\alpha\|_{\mathbf{U}}^2 &\le \langle v_\alpha, \hat{y}\rangle_{\Gamma_C} - \alpha^{-1}\|v_\alpha\|_{L^2}^2 + \langle \bar{y}, v_\alpha - \alpha\hat{y}\rangle_{\Gamma_C} = \langle \bar{y} - y_\alpha, v_\alpha - \alpha\hat{y}\rangle_{\Gamma_C} \\ &= \langle \bar{y} - y_\alpha, v_\alpha - \alpha\bar{y}\rangle_{\Gamma_C} + \alpha\langle \bar{y} - y_\alpha, \bar{y} - \hat{y}\rangle_{\Gamma_C} \\ &= -\alpha\|\bar{y} - y_\alpha\|_{L^2}^2 + \alpha\langle \bar{y} - y_\alpha, \bar{y} - \hat{y}\rangle_{\Gamma_C}.\end{aligned} \tag{27}$$

a) We now address the case $\bar{y} \in L^2(\Gamma_C)$. Then by Young's inequality:

$$\sigma\|d_\alpha\|_{\mathbf{U}}^2 \le -\alpha\|\bar{y} - y_\alpha\|_{L^2}^2 + \alpha\langle \bar{y} - y_\alpha, \bar{y} - \hat{y}\rangle_{\Gamma_C} \le -\frac{\alpha}{2}\|\bar{y} - y_\alpha\|_{L^2}^2 + \frac{\alpha}{2}\|\bar{y} - \hat{y}\|_{L^2}^2.$$

This shows $\|d_\alpha\|_{\mathbf{U}} \le \sqrt{\frac{\alpha}{2\sigma}}\|\bar{y} - \hat{y}\|_{L^2}$. Further,

$$-\|\bar{y} - y_\alpha\|_{L^2}^2 = -\|\bar{y}\|_{L^2}^2 - \|y_\alpha\|_{L^2}^2 + 2\langle \bar{y}, y_\alpha\rangle_{\Gamma_C} \le \|\bar{y}\|_{L^2}^2 - \frac{1}{2}\|y_\alpha\|_{L^2}^2.$$

Hence

$$\sigma\alpha^{-1}\|d_\alpha\|_{\mathbf{U}}^2 + \frac{1}{4}\|y_\alpha\|_{L^2}^2 \le \frac{1}{2}\|\bar{y}\|_{L^2}^2 + \frac{1}{2}\|\bar{y} - \hat{y}\|_{L^2}^2.$$

We use $\hat{y} \ge 0$, which implies $0 \le [Bu_\alpha - \psi]_+ \le v_\alpha$, and conclude

$$\|[Bu_\alpha - \psi]_+\|_{L^2} \le \|v_\alpha\|_{L^2} = \alpha\|y_\alpha\|_{L^2} \le \sqrt{2}\alpha(\|\bar{y}\|_{L^2} + \|\bar{y} - \hat{y}\|_{L^2}).$$

b) Next, we address the case $\bar{y} - \hat{y} \in H^s$ under the assumption stated in b). We can approximate $h := \bar{y} - \hat{y}$ by $h_\delta \in H^1(\Gamma_C)$ (e.g., by applying a mollifier; then the following inequalities can be shown by considering the spaces L^2 and H^1 and then using complex interpolation) such that

$$\|h_\delta\|_{H^{\frac{1}{2}}} \le C_1 \delta^{s-1/2} \|h\|_{H^s}, \quad \|h - h_\delta\|_{L^2} \le C_2 \delta^s \|h\|_{H^s}.$$

Further, there exists a bounded extension operator $E : H^{1/2}(\Gamma_C) \to \mathbf{U}$ with $BEy = y$ and $\|Ey\|_{\mathbf{U}} \le C_e \|y\|_{H^{1/2}}$ for all y. Thus, for $0 < \alpha \le \alpha_0$ we can use (12) to obtain:

$$\begin{aligned}
&\langle \bar{y} - y_\alpha, \bar{y} - \hat{y}\rangle_{\Gamma_C} = \langle \bar{y} - y_\alpha, h_\delta\rangle_{\Gamma_C} + \langle \bar{y} - y_\alpha, h - h_\delta\rangle_{\Gamma_C},\\
&\langle \bar{y} - y_\alpha, h_\delta\rangle_{\Gamma_C} = \langle \bar{y} - y_\alpha, BEh_\delta\rangle_{\Gamma_C} = \langle B^*(\bar{y} - y_\alpha), Eh_\delta\rangle\\
&= \langle J'(u_\alpha) - J'(\bar{u}), Eh_\delta\rangle \le \|J'(u_\alpha) - J'(\bar{u})\|_{\mathbf{U}'} \|Eh_\delta\|_{\mathbf{U}}\\
&\le C_E C_L \|d_\alpha\|_{\mathbf{U}}^\kappa \|h_\delta\|_{H^{\frac{1}{2}}} \le C_E C_L C_1 \delta^{s-1/2} \|h\|_{H^s} \|d_\alpha\|_{\mathbf{U}}^\kappa,\\
&\langle \bar{y} - y_\alpha, h - h_\delta\rangle_{\Gamma_C} \le \|\bar{y} - y_\alpha\|_{L^2} \|h - h_\delta\|_{L^2} \le C_2 \delta^s \|\bar{y} - y_\alpha\|_{L^2} \|h\|_{H^s}.
\end{aligned}$$

Inserting these estimates into (27) and using Young's inequality, we obtain:

$$\begin{aligned}
\sigma\|d_\alpha\|_{\mathbf{U}}^2 &\le -\alpha\|\bar{y} - y_\alpha\|_{L^2}^2 + \alpha\langle \bar{y} - y_\alpha, \bar{y} - \hat{y}\rangle_{\Gamma_C}\\
&\le -\alpha\|\bar{y} - y_\alpha\|_{L^2}^2 + \alpha C_E C_L C_1 \delta^{s-1/2} \|h\|_{H^s} \|d_\alpha\|_{\mathbf{U}}^\kappa\\
&+ \alpha C_2 \delta^s \|\bar{y} - y_\alpha\|_{L^2} \|h\|_{H^s}\\
&\le -\frac{\alpha}{2}\|\bar{y} - y_\alpha\|_{L^2}^2 + \frac{\alpha}{2} C_2^2 \delta^{2s} \|h\|_{H^s}^2 + \frac{\sigma}{2} \|d_\alpha\|_{\mathbf{U}}^2\\
&+ \frac{(2-\kappa)}{2\sigma^{\frac{\kappa}{2-\kappa}}} \kappa^{\frac{\kappa}{2-\kappa}} (\alpha C_E C_L C_1 \delta^{s-1/2} \|h\|_{H^s})^{\frac{2}{2-\kappa}}.
\end{aligned}$$

Making the ansatz $\delta = \alpha^\nu$ and balancing the exponents of α in the second and fourth term yields

$$1 + 2\nu s = (1 + \nu(s - 1/2))\frac{2}{2-\kappa}, \quad \text{thus} \quad \nu = \frac{\kappa}{1 + 2s(1-\kappa)}.$$

The exponent thus is $\frac{1+2s}{1+2s(1-\kappa)}$. We arrive at

$$\frac{\sigma}{2}\|u_\alpha - \bar{u}\|_{\mathbf{U}}^2 + \frac{\alpha}{2}\|\bar{y} - y_\alpha\|_{L^2}^2 \le O((1 + \|\bar{y} - \hat{y}\|_{H^s}^2)\alpha^{\frac{1+2s}{1+2s(1-\kappa)}}).$$

Remark 3 If $\kappa = 1$ we get $\nu = 1$ and the α-exponent $1 + 2s$.

In the case $\bar{y} \in L^2(\Gamma_C)$ we can improve the rate from $O(\alpha^{\frac{1}{2}})$ to $o(\alpha^{\frac{1}{2}})$:

Theorem 3 *Let Assumption (E) as well as $\hat{y} \in L^2(\Gamma_C)_+$ hold and assume $\bar{y} \in L^2(\Gamma_C)$. Then, for $\alpha \to 0^2$, there holds:*

$$\|u_\alpha - \bar{u}\|_{\mathbf{U}} = o(\alpha^{\frac{1}{2}}), \tag{28}$$

$$\|y_\alpha - \bar{y}\|_{V'} = O(\|J'(u_\alpha) - J'(\bar{u})\|_{\mathbf{U}}. \tag{29}$$

Proof See appendix A.1.

4 Semismooth Newton method

We consider now the efficient solution of the optimality system

$$H(u) := J'(u) + B^*[\hat{y} + \alpha^{-1}(Bu - \psi)]_+ = 0, \tag{30}$$

where we take the particular difficulties mentioned above into account.

We denote the nonsmooth part by

$$G(u) := B^*[\hat{y} + \alpha^{-1}(Bu - \psi)]_+. \tag{31}$$

The derivation of a Newton-type method for (30) requires the analysis of the nonsmooth operator $w \mapsto [w]_+$. In fact, we can apply the following result.

Lemma 1 *Consider, for $1 \le r < p \le \infty$ and bounded Γ_C the superposition operator $S : w \in L^p(\Gamma_C) \mapsto [w]_+ \in L^r(\Gamma_C)$. Define the set-valued generalized differential $\partial S : L^p(\Gamma_C) \rightrightarrows \mathcal{L}(L^p(\Gamma_C), L^r(\Gamma_C))$ consisting of all $M : v \mapsto m\, v$, where*

$$m \in L^\infty(\Gamma_C), \quad m|_{w<0} = 0, \quad m|_{w>0} = 1, \quad m|_{w=0} \in [0,1].$$

Then, S is Lipschitz continuous and it is semismooth in the following sense

$$\sup_{M \in \partial S(w+s)} \|S(w+s) - S(w) - Ms\|_{L^r(\Gamma_C)} = o(\|s\|_{L^p(\Gamma_C)}).$$

Proof This is a special case of results in [11,33]. □

We now use the continuous embedding $V \subset L^p(\Gamma_C)$, with $p > 2$ appropriately chosen depending on d, and assume that $\hat{y}$ is chosen such that $\hat{y} \in L^p(\Gamma_C)_+$. Then we have

$$\|B\|_{\mathbf{U}, L^p(\Gamma_C)} \le C_B \tag{32}$$

with a constant $C_B > 0$. Now we define for G in (31) a generalized differential $\partial G : \mathbf{U} \rightrightarrows \mathcal{L}(\mathbf{U}, \mathbf{U}')$ as follows:

$$\partial G(u) = \left\{ \alpha^{-1} B^* M B,\ M \in \partial S(\hat{y} + \alpha^{-1}(Bu - \psi)) \right\}. \tag{33}$$

Using the semismoothness of the max operator $S(w) = [w]_+$, see Lemma 1, and the embedding $V \subset L^p(\Gamma_C)$ with suitable $p > 2$, it is now straightforward to derive the semismoothness of G:

Theorem 4 *Consider the operator $G : \mathbf{U} \to \mathbf{U}'$ defined in* (31) *with differential ∂G given by* (33)*. Then G is Lipschitz continuous and moreover semismooth in the sense that, at every $u \in \mathbf{U}$,*

$$\sup_{Z \in \partial G(u+s)} \|G(u+s) - G(u) - Zs\|_{\mathbf{U}} = o(\|s\|_{\mathbf{U}}) \quad \textit{as} \ \|s\|_{\mathbf{U}} \to 0.$$

Proof Combining Lemma 1, (32), and the chain rule for semismooth operators [34, Prop. 3.8] shows that $G : \mathbf{U} \to \mathbf{U}'$ defined in (31) is semismooth with the differential given by (33). □

For the convergence analysis of a semismooth Newton method for the nonsmooth system (30) we make the following assumption that is in accordance with Assumption (E) and will be verified later for particular applications.

Assumption (S)

1. $u \in \mathbf{U} \mapsto J(u)$ is twice continuously F-differentiable.

2. (12) has a solution $u_\alpha \in \mathbf{U}$ at which $J''(u_\alpha) \in \mathcal{L}(\mathbf{U}, \mathbf{U}')$ is $\mathbf{U}$-coercive, i.e. there exists a constant $C_A > 0$ such that for $\rho > 0$ small enough

$$C_A\,(J''(u)v, v)_{\mathbf{U}} \geq \|v\|_{\mathbf{U}}^2 \quad \forall\, v \in \mathbf{U} \;\; \forall\, u \in \mathcal{U}_\rho := \{u \in \mathbf{U} :\; \|u - u_\alpha\|_{\mathbf{U}} < \rho\}.$$

Remark 4 If Assumption (E) and Assumption (S), 1. hold then $J''(\bar{u})$ is $\mathbf{U}$-coercive with $C_A = \sigma^{-1}$ and by Theorem 1 for all $\alpha > 0$ small enough Assumption (S), 2. holds for $C_A = 2\sigma^{-1}$.

Under Assumption (S), the elements of the generalized differential $\partial H(u)$ satisfy the following regularity property.

Lemma 2 *Let Assumption (S) hold. Then there exist $\delta > 0$ and $C_H > 0$ such that, for all $u \in \bar{\mathcal{U}}_\delta = \{u \in \mathbf{U} : \|u - u_\alpha\|_{\mathbf{U}} \leq \delta\}$ every element $Z \in \partial H(u)$ has a bounded inverse with*

$$\left\|Z^{-1}\right\|_{\mathbf{U}',\mathbf{U}} \leq C_H.$$

Proof Assumption (S) yields $\delta > 0$ such that $J''(u)$ is coercive with constant $1/C_A$ for all $u \in \bar{\mathcal{U}}_\delta$. By (33) we have for any $Z \in \partial H(u)$

$$Z = J''(u) + \alpha^{-1} B^* M B$$

and the definition of ∂S yields for all $v, w \in \mathbf{U}$, using $0 \leq m \leq 1$ on Γ_C)

$$\begin{aligned}\langle B^* M B v, v\rangle_{\mathbf{U}',\mathbf{U}} &= (MBv, Bv)_{\Gamma_C} = \|m^{\frac{1}{2}}\, Bv\|_{L^2(\Gamma_C)}^2 \geq 0,\\ \langle B^* M B v, w\rangle_{\mathbf{U}',\mathbf{U}} &\leq \|B\|_{\mathbf{U},L^2(\Gamma_C)}^2\, \|v\|_{\mathbf{U}}\, \|w\|_{\mathbf{U}}\,.\end{aligned}$$

Therefore, the operator $Z = J''(u) + \alpha^{-1} B^* M B \in \mathcal{L}(\mathbf{U}, \mathbf{U}')$ is continuous and coercive with constant $1/C_A$. Hence, Z^{-1} exists and $\|Z^{-1}\|_{\mathbf{U}',\mathbf{U}} \leq C_A$.

We consider now the following semismooth Newton method for (30).
Algorithm SN: *Semismooth Newton Method:*

1. *Choose an initial point $u_0 \in \mathbf{U}$ and set $k = 0$.*
2. *If $H(u_k) = 0$: STOP with solution u_k.*
3. *Choose $Z_k \in \partial H(u_k) = J''(u_k) + \partial G(u_k)$ with ∂G as in* (33) *and obtain $s_k \in \mathbf{U}$ from*

$$Z_k s_k = -H(u_k).$$

4. *Set $u_{k+1} = u_k + s_k$, increment k, and go to step 2.*

Superlinear convergence can be deduced from Theorem 4 and Lemma 2, see [33]:

Theorem 5 *Let Assumption (S) hold. Then there is $\rho > 0$ such that for all $u_0 \in \mathbf{U}$ with $\|u_0 - u_\alpha\|_{\mathbf{U}} \leq \rho$, Algorithm SN terminates finitely with solution u_α or converges q-superlinearly to u_α.*

5 Application of multigrid methods

We will now show that after discretization standard multigrid solvers can be applied.

Throughout this section we assume that Γ_D has positive measure such that by Poincaré's inequality there exists a constant $C_p > 0$ with

$$\|v\|_{H^1(\Omega)^d} \leq C_p \|v\|_{H^1(\Omega)^d} \quad \forall\, v \in \mathbf{U}.$$

Moreover, we assume that $J''(u)$ has the following structure.

Assumption (M)
Assumption (S) holds and $J''(u)$ has the structure

$$\begin{aligned}(J''(u)w, v)_{\mathbf{U}} = &:= \int_\Omega (\mathcal{C}\nabla w : \nabla v + D\nabla w : v + D\nabla v : w + Ew : v)\; dx \\ &=: c(w,v) + d_1(w,v) + d_2(w,v) + e(w,v) =: a(u; w, v)\end{aligned} \tag{34}$$

with $\mathcal{C} = (c_{ijkl}) \in W^{1,\infty}(\Omega)^{d^4}$, $D = (d_{ijk}) \in H^1(\Omega)^{d^3}$, $E = (e_{ij}) \in L^2(\Omega)^{d^2}$ uniformly bounded on bounded sets of $u \in \mathbf{U}$.

Assumption (M) holds for example if $J(u)$ has the form

$$J(u) = \frac{1}{2} \int_\Omega (\mathcal{C}\nabla u : \nabla u + d(u) : \nabla u + e(u))\; dx$$

and d, e satisfy appropriate conditions.

Under Assumption (M) there exists with Poincaré's inequality clearly a constant $M_a > 0$ with

$$|a(u; v, w)| \leq M_a \|v\|_{H^1(\Omega)^d} \|w\|_{H^1(\Omega)^d} \quad \forall\, v, w \in \mathbf{U}. \tag{35}$$

Moreover, Assumption (S) yields $C_A > 0$ and $\rho > 0$ with (ensured by (5) for the Signorini problem)

$$C_A a(u; v, v) \geq \|v\|^2_{H^1(\Omega)^d} \quad \forall\, v \in \mathbf{U}\;\; \forall\, u \in \mathcal{U}_\rho := \{u \in \mathbf{U} :\, \|u - u^*_\alpha\|_{\mathbf{U}} < \rho\}. \tag{36}$$

Let $A = A(u) := J''(u) \in \mathcal{L}(\mathbf{U}, \mathbf{U}')$ be the invertible operator corresponding to the coercive bilinear form $a(\cdot, \cdot; u)$.

In each semismooth Newton step of Algorithm SN we have in step 3 to solve an operator equation of the form

$$(A + \alpha^{-1} B^* M B)s = r \tag{37}$$

where we select $M = M(u) \in \partial S(\hat{y} + \alpha^{-1}(Bu - \psi))$ in (33) for simplicity as

$$M = m \cdot I, \quad m = \begin{cases} 1 & \text{if } \hat{y} + \alpha^{-1}(Bu - \psi) \geq 0, \\ 0 & \text{if } \hat{y} + \alpha^{-1}(Bu - \psi) < 0. \end{cases} \tag{38}$$

For notational convenience, we set

$$A_\alpha = A_\alpha(u) := A(u) + \alpha^{-1} B^* M(u) B.$$

We have seen in the proof of Lemma 2 that A_α is a lower order perturbation of A by a uniformly bounded, symmetric, positive semidefinite operator. Hence, A_α induces a symmetric, continuous, coercive bilinear form

$$a_\alpha(u; v, w) := \langle A_\alpha(u) v, w \rangle_{\mathbf{U}', \mathbf{U}} \quad \forall\, v, w \in \mathbf{U}.$$

For brevity, we suppress here the dependence of a_α on u. As shown in the proof of Lemma 2, there exist constants $c_1(\alpha), c_2 > 0$, independent of $u \in \mathcal{U}_\rho$, with

$$|a_\alpha(v,w)| \le c_1(\alpha)\,\|v\|_{\mathbf{U}}\,\|w\|_{\mathbf{U}} \quad \forall\, v,w \in \mathbf{U},$$
$$|a_\alpha(v,v)| \ge a(v,v) \ge c_2\,\|v\|_{\mathbf{U}}^2 \quad \forall\, v \in \mathbf{U}.$$

Thus, (37) has a unique solution and can briefly be written as

$$A_\alpha s = r. \tag{39}$$

We will prove that (39) can be solved efficiently by an appropriate application of standard multigrid methods after discretization. Since the term $\alpha^{-1}B^*MB$ does in general not admit H^2 regularity results as needed by classical multigrid theory, we will apply the multigrid theory reviewed by Yserentant in [38]. As we will see, this will require care in the cnstrauction of the multigrid spaces. We will select coarse grid spaces that are in the null space of the discrete approximation of the operator $\alpha^{-1}B^*MB$.

5.1 Discretization

We consider the following general framework for the discretization of the contact problem (1) which we write as before in the form (2).

For simplicity, let Ω_h be a polyhedral approximation of Ω. Let $\mathcal{T}_h$ be a simplicial triangulation of Ω_h. Moreover, let $\Gamma_{C,h} \subset \partial\Omega_h$ and $\Gamma_{D,h} \subset \partial\Omega_h$ be approximations of Γ_C and Γ_D, respectively, consisting of faces of elements in $\mathcal{T}_h$. Let $U_h \subset \{u \in H^1(\Omega_h) : u|_{\Gamma_{D,h}} = 0\}$ be a finite element space corresponding to $\mathcal{T}_h$, such that $\mathbf{U}_h := U_h^d$ is a finite element approximation of $\mathbf{U}$. Finally, let $V_h' \subset L^2(\Gamma_{C,h})$ be a finite element space for the Lagrange multipliers. Moreover, let $\chi_{h,i}^+$, $1 \le i \le N_{K_h}$ be a "positive" basis'(see 5.2) of V_h' such that with constants $0 < \kappa_1 \le \kappa_2$

$$\kappa_1\,\|v_h\|_{L^2(\Gamma_{C,h})} \le \left(\sum_{i=1}^{N_{K_h}} (\chi_{h,i}^+, v_h)_{L^2(\Gamma_{C,h})}^2\right)^{\frac{1}{2}} \le \kappa_2\,\|v_h\|_{L^2(\Gamma_{C,h})} \;\forall\, v_h \in V_h'. \tag{40}$$

We approximate K^+, K, and τ_C^n by K_h^+, K_h, and $\tau_{C,h}^n$ defined as follows:

$$K_h^+ := \left\{y_h \in V_h' : y_h = \sum_{i=1}^{N_{K_h}} y_i \chi_{h,i}^+,\;\; y_i \ge 0\right\},$$
$$K_h := \left\{v_h \in L^2(\Gamma_{C,h}) : (y_h, v_h)_{L^2(\Gamma_{C,h})} \ge 0 \;\;\forall\, y_h \in K_h^+\right\},$$
$$\tau_{C,h}^n : \mathbf{U}_h \to U_h \subset L^2(\Gamma_{C,h}),\;\text{ see (49) below.}$$

The nonpenetration condition $\tau_C^n(u) \le \psi$ is approximated by

$$(y_h, \psi - \tau_{C,h}^n(u_h))_{L^2(\Gamma_{C,h})} \ge 0 \quad \forall\, y_h \in K_h^+,$$

which is equivalent to

$$(\chi_{h,i}^+, \psi - \tau_{C,h}^n(u_h))_{L^2(\Gamma_{C,h})} \ge 0 \quad 1 \le i \le N_{K_h}.$$

We arrive at the following discrete approximation of (1).

$$\min_{u_h \in \mathbf{U}_h} J(u_h) \;\text{ subject to }\; (\chi_{h,i}^+, \psi - \tau_{C,h}^n(u_h))_{L^2(\Gamma_{C,h})} \ge 0,\;\; 1 \le i \le N_{K_h}. \tag{41}$$

Further, we introduce

$$B_h \in \mathcal{L}(\mathbf{U}_h, \mathbb{R}^{N_{K_h}}), \quad B_h u_h = \Big((\chi^+_{h,i}, \tau^n_{C,h}(u_h))_{L^2(\Gamma_C)} \Big)_{1\le i\le N_{K_h}},$$
$$\boldsymbol{\psi}_h \in \mathbb{R}^{N_{K_h}}, \quad \boldsymbol{\psi}_h = \Big((\chi^+_{h,i}, \psi)_{L^2(\Gamma_C)} \Big)_{1\le i\le N_{K_h}}.$$

Then (41) can be written in the form

$$\min_{u_h\in\mathbf{U}_h} J(u_h) \quad \text{subject to} \quad \boldsymbol{\psi}_h - B_h u_h \ge 0. \tag{42}$$

5.1.1 Discrete optimality conditions

The KKT-conditions read as follows. $u_h \in \mathbf{U}_h$ solves (41) iff there exists a vector $\mathbf{y}_h \in \mathbb{R}^{N_{K_h}}$, which is the coordinate vector of $y_h \in V_h'$ with respect to the nonnegative dual basis $\chi^+_{h,i}$, such that

$$\langle J'(u_h), w_k\rangle_{\mathbf{U}',\mathbf{U}} + \mathbf{y}_h^T B_h w_h = 0, \quad \forall\, w_h \in \mathbf{U}_h, \tag{43}$$
$$\mathbf{y}_h \ge 0, \quad \boldsymbol{\psi}_h - B_h u_h \ge 0, \quad \mathbf{y}_h^T(\boldsymbol{\psi}_h - B_h u_h) = 0. \tag{44}$$

Now let $(\cdot,\cdot)_0$ be an L^2-like inner product on $\mathbf{U}_h$, see (52) for our choice within the multilevel preconditioner. Introducing $J_h'(u_h) \in \mathbf{U}_h$, $A_h = A_h(u_h) \in \mathcal{L}(\mathbf{U}_h, \mathbf{U}_h)$ with

$$(J_h'(u_h), w_h)_0 = \langle J'(u_h), w_k\rangle_{\mathbf{U}',\mathbf{U}}, \quad (A_h v_h, w_h)_0 = a(u_h; v_h, w_h) \tag{45}$$

for all $u_h, v_h, w_h \in \mathbf{U}_h$ we can write (43) as $J_h'(u_h) + B_h^* \mathbf{y}_h = 0$. Here, $B_h^* \in \mathcal{L}(\mathbb{R}^{N_{K_h}}, {}^{\mathbf{U}_h})$ is defined by $\mathbf{z}_h^T B_h w_h = (B_h^* \mathbf{z}_h, w_h)_0$ for all $\mathbf{z}_h \in \mathbb{R}^{N_{K_h}}, w_h \in \mathbf{U}_h$. With the C-function $\phi(a,b) = \min(a, \alpha^{-1}b) = a - \max(0, \alpha^{-1}b - a)$, the discrete regularized optimality system can be written as.

$$J_h'(u_h) + B_h^* \mathbf{y}_h = 0,$$
$$\mathbf{y}_h - [\hat{\mathbf{y}}_h + \alpha^{-1}(B_h u_h - \boldsymbol{\psi}_h)]_+ = 0,$$

Inserting the second equation in the first one we arrive at the following discrete counterpart of (30)

$$J_h'(u_h) + B_h^*[\hat{\mathbf{y}}_h + \alpha^{-1}(B_h u_h - \boldsymbol{\psi}_h)]_+ = 0. \tag{46}$$

5.1.2 Discrete semismooth Newton system

Applying a semismooth Newton method to this discrete approximation (46) of (30), we obtain the following discrete counterpart of (37).

$$(A_h + \alpha^{-1} B_h^* M_h B_h) s_h = r_h, \tag{47}$$

where we select $M_h = M_h(u_h)$ by (38) as

$$M_h = \operatorname{diag}(m_1, \ldots, m_{N_{K_h}}), \quad m_i = \begin{cases} 1 & \text{if } (\hat{\mathbf{y}}_h + \alpha^{-1}(B_h u_h - \boldsymbol{\psi}_h))_i \ge 0, \\ 0 & \text{if } (\hat{\mathbf{y}}_h + \alpha^{-1}(B_h u_h - \boldsymbol{\psi}_h))_i < 0. \end{cases}$$

By the definition of A_h and B_h (47) is equivalent to

$$a(s_h, w_h) + \alpha^{-1}(B_h w_h)^T M_h B_h s_h = (r_h, w_h)_{L^2(\Omega)} \quad \forall\, w_h \in \mathbf{U}_h, \tag{48}$$

where a is the bilinear form corresponding to $J''(u_h)$.

We assume throughout that an analogue of Assumption (S) holds at the discrete solution u_h of (46), i.e., $J''(u_h)$ is $\mathbf{U}$-coercive. This follows from Assumption (S) if $\alpha > 0$ and the grid size of $\mathbf{U}_h$ is small enough by Theorem 1 if the finite element discretization is convergent; see e.g. in the case of elastic contact problems [16] for a convergence result of discretizations as we consider now in 5.2.

5.2 Choice of the finite element spaces

We consider the following finite element discretization, see for example [37]. Let U_h be the space of conforming linear finite elements, i.e.,

$$U_h = \left\{ v \in C(\bar{\Omega}_h) : v \text{ piecewise linear on } \mathcal{T}_h,\ v|_{\Gamma_{D,h}} = 0 \right\}.$$

Let $\mathcal{N}_h$ be the free nodes of $\mathcal{T}_h$, let $\{\phi_p : p \in \mathcal{N}_h\}$ be the nodal basis of U_h and set $\mathcal{N}_{C,h} := \mathcal{N}_h \cap \bar{\Gamma}_{C,h}$. Define at $p \in \mathcal{N}_{C,h}$ the approximate outer normals n_p for example by

$$n_p = \alpha_p \sum_{\text{Face } F \subset \text{supp}\, \phi_p} |F| n_F,$$

where n_F is the outer normal of F and $\alpha_p > 0$ is a weight such that $||n_p|| = 1$. Now define for $u_h \in \mathbf{U}_h$ the discrete normal trace by

$$\tau^n_{C,h}(u_h) = \sum_{p \in \mathcal{N}_{C,h}} n_p^T u_h(p)\, \phi_p. \tag{49}$$

Finally define as in [14,37] for $\mathcal{N}_{C,h} = \{p_1, \ldots, p_{N_{K_h}}\}$ the dual basis ψ_{h,p_i} with the same support as ϕ_{p_i} and satisfying the biorthogonality relation

$$\int_{\Gamma_{C,h}} \psi_{h,p} \phi_q \, dS(x) = \delta_{pq} \int_{\Gamma_{C,h}} \phi_q \, dS(x)\ \forall\, p, q \in \mathcal{N}_{C,h}.$$

Now we define our "nonnegative" basis by

$$\chi^+_{h,i} := \psi_{h,p_i} \frac{\|\phi_{p_i}\|_{L^2(\Gamma_{C,h})}}{\int_{\Gamma_{C,h}} \phi_{p_i} \, dS(x)}.$$

Then we obtain $(\chi^+_{h,i}, v_h)_{L^2(\Gamma_{C,h})} = v_h(p_i)\, \|\phi_{p_i}\|_{L^2(\Gamma_{C,h})}$ for all $v_h \in U_h$. Therefore, by estimating the corresponding quadratic forms we deduce

$$\sum_{i=1}^{N_{K_h}} (\chi^+_{h,i}, v_h)^2_{L^2(\Gamma_{C,h})} = \sum_{i=1}^{N_{K_h}} \int_{\Gamma_{C,h}} v_h(p_i)^2 \phi^2_{p_i} \, dS(x) \begin{cases} \le 2\, \|v_h\|^2_{L^2(\Gamma_{C,h})}, \\ \ge \frac{1}{2}\, \|v_h\|^2_{L^2(\Gamma_{C,h})}. \end{cases}$$

and thus (40) holds with $\kappa_1 = 1/\sqrt{2}$, $\kappa_2 = \sqrt{2}$. Similarly, we have for all $u_h \in \mathbf{U}_h$

$$(B_h u_h)_i = \sum_{q \in \mathcal{N}_{C,h}} n_q^T u_h(q) \int_{\Gamma_{C,h}} \chi^+_{h,i} \phi_q \, dS(x) = n_{p_i}^T u_h(p_i)\, \|\phi_{p_i}\|_{L^2(\Gamma_{C,h})}. \tag{50}$$

If we choose an appropriate nodal basis for $\mathbf{U}_h$, we can verify the following assumption. It uses a matrix $\mathbf{S}_\ell$, where ℓ appears for the first time. It denotes the number of grid levels and corresponds to the finest grid, given by $\mathcal{T}_h$.

Assumption (C)

(C1) There exists a constant $C_B > 0$ such that for all $M_h = \text{diag}(m_1, \ldots, m_{N_{K_h}})$, $m_i \in \{0, 1\}$, the matrix representation $\mathbf{S}_\ell$ of the operator $S_h := B_h^* M_h B_h$ corresponding to the nodal basis of $\mathbf{U}_h$ and its diagonal part $\text{diag}(\mathbf{S}_\ell)$ satisfy

$$C_B \mathbf{S}_\ell - \text{diag}(\mathbf{S}_\ell) \quad \text{positive semidefinite.}$$

(C2) There exists a constant $\kappa_3 > 0$ such that for all M_h as in (C1), there holds:

$$\inf_{\substack{w_h \in \mathbf{U}_h \\ (B_h w_h)^T M_h (B_h w_h) = 0}} \|v_h - w_h\|^2_{L^2(\Gamma_{C,h})} \le \kappa_3 (B_h v_h)^T M_h (B_h v_h) \quad \forall v_h \in \mathbf{U}_h.$$

We give an example for which (C1) and (C2) hold.

Lemma 3 *If we choose the nodal basis $\phi_{p,j} = e_{p,j}\phi_p$, $p \in \mathcal{N}_h$, $e_{p,j} \in \mathbb{R}^3$, $1 \le j \le 3$, of $\mathbf{U}_h$, with $(e_{p,1}, e_{p,2}, e_{p,3})$ orthonormal, $e_{p,3} = n_p$ for all $p \in \mathcal{N}_{C,h}$, then assumption (C1) holds for the above discretization with $C_B = 1$ and (C2) holds with some $\kappa_3 = 2$.*

Proof See appendix A.3.

5.3 General framework for multilevel preconditioners

As before let $\Omega_h \subset \mathbb{R}^3$ be a polygonal approximation of Ω and assume for simplicity that $\mathcal{T}_0$ is a coarse conforming simplicial triangulation of $\bar{\Omega}_h$ such that $\Gamma_{C,h}$ and $\Gamma_{D,h}$ are the union of certain boundary faces of simplices in $\mathcal{T}_0$. The triangulation $\mathcal{T}_0$ is refined several times by obeying the rules in [2] leading to a family of nested, possibly nonconforming triangulations $\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_\ell$, where $\mathcal{T}_h = \mathcal{T}_\ell$ is the finest grid. We require that in 3D the refinement strategy of [2] is implemented in such a way that $\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_\ell$ are locally quasiuniform. In addition let $\tilde{\mathcal{T}}_0, \tilde{\mathcal{T}}_1, \ldots, \tilde{\mathcal{T}}_\ell$ with $\tilde{\mathcal{T}}_0 = \mathcal{T}_0$ be a corresponding sequence of uniformly refined conforming meshes. Let

$$\mathcal{S}_k = \left\{v \in C(\bar{\Omega}_h) : v \text{ piecewise linear on } \mathcal{T}_k,\ v|_{\Gamma_{D,h}} = 0\right\}$$
$$\tilde{\mathcal{S}}_k = \left\{v \in C(\bar{\Omega}_h) : v \text{ piecewise linear on } \tilde{\mathcal{T}}_k,\ v|_{\Gamma_{D,h}} = 0\right\}.$$

We set $\mathcal{S}_k^3 = \mathcal{S}_k \times \mathcal{S}_k \times \mathcal{S}_k$ and $\tilde{\mathcal{S}}_k^3 = \tilde{\mathcal{S}}_k \times \tilde{\mathcal{S}}_k \times \tilde{\mathcal{S}}_k$. The free (non-Dirichlet) nodes of $\mathcal{T}_l$ and $\tilde{\mathcal{T}}_l$ are denoted by $\mathcal{N}_l$ and $\tilde{\mathcal{N}}_l$, respectively.

We want to solve the semismooth Newton equation (48) on the finest grid $\mathcal{T}_h = \mathcal{T}_\ell$ by using a multilevel method. We apply the general framework in [38].

5.3.1 General multigrid framework

Let $\mathbf{U}_h = \mathcal{S}_\ell^3$ be the finite element space on the finest grid and define on $\mathbf{U}_h$ the inner product induced by the bilinear form in (48)

$$a_\alpha(u_k; v, w) = a(u_h; v, w) + \alpha^{-1}(B_h v)^T M_h(u_h) B_h w, \quad v, w \in \mathbf{U}_h \tag{51}$$

with corresponding norm

$$\|v\| = a_\alpha(v, v)^{1/2}$$

and the L^2-like inner product

$$(v, w)_0 = \sum_{T \in \mathcal{T}_0} \frac{1}{\text{diam}(T)^2} \int_T v^T w \, dx \tag{52}$$

with induced norm

$$\|v\|_0 = (v,v)_0^{1/2}.$$

Define similar to (45) the operator $A_{\alpha,h} = A_{\alpha,h}(u_h) : \mathbf{U}_h \to \mathbf{U}_h$ by

$$(A_{\alpha,h}v, w)_0 = a_\alpha(u_h; v, w) \quad \forall\, v, w \in \mathbf{U}_h$$

and the right hand side $f_h \in \mathcal{S}$ by

$$(f_h, w)_0 = \langle r_h, w \rangle_{\mathbf{U}_h', \mathbf{U}_h} \quad \forall\, w \in \mathbf{U}_h.$$

Then $A_{\alpha,h}$ is symmetric positive definite with respect to the inner product $(\cdot,\cdot)_0$ and (48) is with $v = s_h$ equivalent to the linear equation

$$A_{\alpha,h} v = f_h. \tag{53}$$

To describe a general multiplicative multilevel method for (53) let $\mathcal{W}_0, \ldots, \mathcal{W}_\ell$ be a subspace decomposition such that any $v \in \mathbf{U}_h$ admits a possibly non-unique decomposition

$$v = w_0 + w_1 + \cdots + w_\ell, \quad w_l \in \mathcal{W}_l.$$

Define the projections $Q_l : \mathbf{U}_h \to \mathcal{W}_l$ by

$$(Q_l v, w_l)_0 = (v, w_l)_0 \quad \forall\, w_l \in \mathcal{W}_l$$

and the Ritz approximations $A_l : \mathcal{W}_l \to \mathcal{W}_l$ of $A_{\alpha,h}$ with respect to the spaces $\mathcal{W}_l$ by

$$(A_l v, w)_0 = (A_{\alpha,h} v, w)_0 = a_\alpha(v, w) \quad \forall\, v, w \in \mathcal{W}_l.$$

An exact subspace correction that makes the error a_α-orthogonal to $\mathcal{W}_l$ is given by

$$v \leftarrow v + A_l^{-1} Q_l (f_h - A_{\alpha,h} v).$$

In order to obtain an efficient algorithm, multilevel methods replace the exact subspace correction by an approximate subspace correction

$$v \leftarrow v + B_l^{-1} Q_l (f_h - A_{\alpha,h} v).$$

with a symmetric positive definite approximation $B_l : \mathcal{W}_l \to \mathcal{W}_l$ of A_l. Obviously, the approximate subspace correction can be implemented in the form

$$\begin{aligned} v \leftarrow v + d_l, \text{ where } d_l \in \mathcal{W}_l :\ & (B_l d_l, w_l)_0 = (f_h - A_{\alpha,h} v, w_l)_0 \quad \forall\, w_l \in \mathcal{W}_l \\ \iff\ & (B_l d_l, w_l)_0 = (f_h, w_l)_0 - a_\alpha(v, w_l) \quad \forall\, w_l \in \mathcal{W}_l. \end{aligned}$$

We consider now the following multilevel preconditioner.

Algorithm MPR: *Multilevel preconditioner*

Input: $f_h \in \mathbf{U}_h$, *starting point* $v \in \mathbf{U}_h$.
For $l = 0, 1, \ldots, \ell$:

$$v \leftarrow v + d_l,\ d_l \in \mathcal{W}_l :\ (B_l d_l, w_l)_0 = (f_h, w_l)_0 - a_\alpha(v, w_l) \quad \forall\, w_l \in \mathcal{W}_l. \tag{54}$$

If $\bar{v} \in \mathbf{U}_h$ denotes the solution of (53), the approximate subspace correction (54) leads to the update for the error

$$v - \bar{v} \leftarrow (I - T_l)(v - \bar{v})$$

with $T_l = B_l^{-1} Q_l A_{\alpha,h}$ and thus the preconditioner MPR results in the error update

$$v - \bar{v} \leftarrow E(v - \bar{v}), \quad E = (I - T_\ell) \cdots (I - T_1)(I - T_0).$$

We recall a general result of [38] to estimate $\|E\|$ that we will apply in the sequel. We will work with the following assumptions.

Assumption (A) There exists a subspace decomposition $\mathbf{U}_h = \mathcal{V}_0 \oplus \mathcal{V}_1 \oplus \ldots \oplus \mathcal{V}_\ell$ with $\mathcal{V}_l \subset \mathcal{W}_l$ such that the following holds.

(A1) The decomposition is stable in the sense that there exists a constant $K_1 > 0$ with

$$\sum_{l=0}^{\ell} (B_l v_l, v_l)_0 \leq K_1 \left\| \sum_{l=0}^{\ell} v_l \right\|^2 .$$

(A2) There exist constants $\gamma_{kl} = \gamma_{lk}$ with

$$\text{Spectral radius}((\gamma_{kl})_{0 \leq k,l \leq \ell}) \leq K_2$$
$$a_\alpha(w_k, v_l) \leq \gamma_{kl} (B_k w_k, w_k)_0^{1/2} (B_l v_l, v_l)_0^{1/2} \; \forall\, w_k \in \mathcal{W}_k,\; v_l \in \mathcal{V}_l,\; 0 \leq k \leq l \leq \ell.$$

(A3) There exists $0 < \omega < 2$ such that

$$a_\alpha(w_l, w_l) = (A_l w_l, w_l)_0 \leq \omega (B_l w_l, w_l)_0 \quad \forall\, w_l \in \mathcal{W}_l.$$

Then we have the following theorem, see [38, Thm. 5.1].

Theorem 6 *Let (A1), (A2), (A3) hold. Then the preconditioner MPR reduces the norm $\|v - \bar{v}\| = a_\alpha(v - \bar{v}, v - \bar{v})^{1/2}$ of the error at least by the factor $\|E\|$ where*

$$\|E\|^2 \leq 1 - \frac{2 - \omega}{K_1 (1 + K_2)^2}.$$

Note that the spaces $\mathcal{V}_l \subset \mathcal{W}_l$ do not enter the actual computations, they are a purely analytical tool.

Corollary 1 *Let the assumptions of Theorem 6 hold. Denote by M the multilevel preconditioner given by Algorithm MPR for starting point 0, i.e., $v = M f_h$. Then the spectral radius of $I - M A_{\alpha,h}$ is $\leq \sqrt{1 - \frac{2-\omega}{K_1(1+K_2)^2}}$.*

Proof Let $f_h = A_{\alpha,h} \bar{v}$. Then for $v = M f_h$ we have for starting point 0

$$(I - M A_{\alpha,h}) \bar{v} = \bar{v} - v = E(\bar{v} - 0) = E \bar{v}.$$

Hence, $E = I - M A_{\alpha,h}$ and the assertion follows from Theorem 6. □

5.4 A multilevel preconditioner for the semismooth Newton system

5.4.1 Finite element spaces

We choose $\mathcal{W}_\ell = \mathcal{S}_\ell^3 = \mathbf{U}_h$ and

$$\bar{\mathcal{W}}_\ell = \left\{ v \in \mathbf{U}_h : (B_h v)^T M_h B_h v = 0 \right\}.$$

We recall that $\mathcal{N}_h = \mathcal{N}_\ell$ are the nodes of $\mathcal{T}_\ell$ and ϕ_p, $p \in \mathcal{N}_h$ are the nodal basis of $\mathcal{S}_\ell$. Let as in 5.2 $\mathcal{N}_{C,h} := \mathcal{N}_h \cap \bar{\Gamma}_{C,h} =: \{p_1, \ldots, p_{N_{K_h}}\}$. With our choice of B_h in 5.2 we can easily obtain a basis of $\bar{\mathcal{W}}_\ell$ as follows.

We assume that the normals n_p, $p \in \mathcal{N}_{C,h}$, can be extended to a $W^{2,\infty}$-continuous function $n : \Omega_h \mapsto \{v \in \mathbb{R}^3 : v^T v = 1\}$ and that we can define a $W^{2,\infty}$-map

$$N : x \in \Omega_h \mapsto N(x) = (t_1, t_2, n)(x) \in \left\{Q \in \mathbb{R}^{3,3} : Q^T Q = I,\ \det(Q) = 1\right\}.$$

with some Lipschitz constant L_N. Since Γ_C is only a part of Γ, such a map can usually be constructed.

We know that $M_h(u_h) = \text{diag}(m_1, \ldots, m_{N_{K_h}})$, $m_i \in \{0,1\}$. We define the set of current contact nodes by $\Gamma_\ell = \Gamma_\ell(u_h) = \{p_k : 1 \le k \le N_{K_h},\ m_k = 1\}$. Let

$$\bar{\mathcal{W}}'_\ell := \{v \in \mathbf{U}_h : v_3(p) = 0 \text{ for all } p \in \Gamma_\ell\}$$

and define the linear operator

$$P : \bar{\mathcal{W}}'_\ell \mapsto \bar{\mathcal{W}}_\ell, \qquad Pv = \sum_{p \in \mathcal{N}_\ell} (N(p)v(p))\phi_p.$$

Then we have

$$\bar{\mathcal{W}}_\ell = P\,\bar{\mathcal{W}}'_\ell$$

and P transforms the nodal basis of $\bar{\mathcal{W}}'_\ell$ to a nodal basis of $\bar{\mathcal{W}}_\ell$.

Remark 5 Pv is nothing else but the linear interpolate of Nv on the finest mesh $\mathcal{T}_\ell$.

The definition of the coarse subspaces of $\bar{\mathcal{W}}_\ell$ uses the following extended contact set on the coarser meshes:

$$\Gamma_k = \Gamma_\ell \cup \bigcup_{F \text{ face of some } T \in \mathcal{T}_k \text{ with } \text{int}(F) \cap \Gamma_\ell \neq \emptyset} (\bar{F} \cap \mathcal{N}_k).$$

Setting

$$\bar{\mathcal{W}}'_k := \left\{v \in {\mathcal{S}_k}^3 : v_3(p) = 0,\ \ p \in \Gamma_k\right\} \tag{55}$$

we define the coarse subspaces

$$\bar{\mathcal{W}}_k = P\,\bar{\mathcal{W}}'_k \tag{56}$$

and P transforms the nodal basis of $\bar{\mathcal{W}}'_k$ to a nodal basis of $\bar{\mathcal{W}}_k$.

For the multilevel method we use now the spaces

$$\mathcal{W}_k = \bar{\mathcal{W}}_k,\ \ k = 0, \ldots, \ell - 1, \quad \mathcal{W}_\ell = \mathcal{S}_\ell^3 = \mathbf{U}_h. \tag{57}$$

The auxiliary spaces $\mathcal{V}_k \subset \mathcal{W}_k$ are obtained by an L^2-like subspace decomposition. To this end, we define the projections $\bar{Q}'_l : \mathbf{U}_h \to \bar{\mathcal{W}}'_l$, $0 \le l \le \ell$,

$$(\bar{Q}'_l v, w_l)_0 = (v, w_l)_0 \quad \forall\, w_l \in \bar{\mathcal{W}}'_l.$$

and, similarly to [38, §7], the subspaces $\bar{\mathcal{V}}'_l \subset \bar{\mathcal{W}}'_l$

$$\bar{\mathcal{V}}'_0 = \bar{\mathcal{W}}'_0, \quad \bar{\mathcal{V}}'_l = \left\{\bar{Q}'_l v - \bar{Q}'_{l-1} v : v \in \bar{\mathcal{W}}'_\ell\right\} \subset \bar{\mathcal{W}}'_l,\ \ 1 \le l \le \ell.$$

Now we set

$$\bar{\mathcal{V}}_l := P\bar{\mathcal{V}}'_l \subset \bar{\mathcal{W}}_l,\ \ 0 \le l \le \ell.$$

Finally, we define

$$\mathcal{V}_l := \bar{\mathcal{V}}_l,\ \ 0 \le l \le \ell - 1, \quad \mathcal{V}_\ell := \left\{P(v - \bar{Q}'_{\ell-1} v) : v \in \mathcal{W}_\ell\right\}.$$

5.4.2 The multilevel preconditioner

We apply now Algorithm MPR with the just defined subspaces $\mathcal{W}_l$ and use m symmetric Gauss-Seidel iterations as approximate subspace correction (54). On the coarsest level, we apply an exact subspace correction, i.e., we set $B_0 = A_0$:

To this end, we use now the nodal basis of $\mathcal{W}_l$ and denote for $w_l \in \mathcal{W}_l$ by $\mathbf{w}_l$ the coordinates (i.e., nodal values) of $w_l \in \mathcal{W}_l$ with respect to this basis. Denote by $\mathbf{A}_l$ the matrix representation of A_l corresponding to the nodal basis of $\mathcal{W}_l$, i.e.

$$\mathbf{w}_l^T \mathbf{A}_l \mathbf{d}_l = (A_l d_l, w_l)_0 \quad \forall\, d_l, w_l \in \mathcal{W}_l.$$

We assume that the choice of $\chi_{h,i}^+$ and $\tau_{C,h}^n$ ensures the property (C1) of the nonpenetration operator B_h.

As approximate subspace correction (54) we apply m symmetric Gauss-Seidel steps starting with 0 to the matrix representation of the exact subspace correction

$$(A_l d_l, w_l)_0 = (f_h, w_l)_0 - a_\alpha(v, w_l) \quad \forall\, w_l \in \mathcal{W}_l. \tag{58}$$

Let

$$\mathbf{A}_l \mathbf{d}_l = \mathbf{r}_l$$

be the corresponding matrix representation with respect to the nodal basis, where $\mathbf{r}_l$ represents the right hand side of (58) given by

$$\mathbf{r}_l^T \mathbf{w}_l = (f_h, w_l)_0 - a_\alpha(v, w_l) \quad \forall\, w_l \in \mathcal{W}_l. \tag{59}$$

We introduce the splitting

$$\mathbf{A}_l = \mathbf{D}_l - \mathbf{L}_l - \mathbf{L}_l^T$$

into diagonal, strictly lower and strictly upper triangular part, respectively. Then m symmetric Gauss-Seidel steps can be computed by

Algorithm SGS: *Symmetric Gauss-Seidel subspace correction*

Input: $\mathbf{r}_l \in \mathbb{R}^{\dim(\mathcal{W}_l)}$.
Output: $\mathbf{d}_l = \mathbf{B}_l^{-1}\mathbf{r}_l$ *after* m *Symmetric Gauss-Seidel iterations on* $\mathbf{A}_l\mathbf{d}_l = \mathbf{r}_l$.
Set $\mathbf{d}_l := 0$. *For* $i = 1, \ldots, m$: $\quad \mathbf{d}_l \leftarrow \mathbf{d}_l + (\mathbf{D}_l - \mathbf{L}_l)^{-1}(\mathbf{r}_l - \mathbf{A}_l\mathbf{d}_l)$
$\quad \mathbf{d}_l \leftarrow \mathbf{d}_l + (\mathbf{D}_l - \mathbf{L}_l^T)^{-1}(\mathbf{r}_l - \mathbf{A}_l\mathbf{d}_l)$.

For $l \geq 1$ we use Algorithm SGS with $\mathbf{r}_l$ given by (59) to implement the approximate subspace correction (54). On the coarse level $l = 0$ we compute an exact subspace correction, i.e., we use $B_0 = A_0$.

This leads to the following implementation of Algorithm MPR:

Algorithm MPRSN. *Multilevel preconditioner for semismooth Newton system* (48)

Input: $f_h \in \mathbf{U}_h$, *starting point* $v \in \mathbf{U}_h$.
Subspaces $\mathcal{W}_l$ *according to* (55)–(57).

For $l = 0, 1, \ldots, \ell$:

$$v \leftarrow v + d_l, \;\; d_l \in \mathcal{W}_l : \; (B_l d_l, w_l)_0 = (f_h, w_l)_0 - a_\alpha(v, w_l) \quad \forall\, w_l \in \mathcal{W}_l$$

by using $B_0 = A_0$ *for* $l = 0$ *and Algorithm SGS for* $l > 0$.

5.5 Convergence analysis of the multilevel preconditioner

We show next that Algorithm MPRSN satisfies assumptions (A1)–(A3), where the constants depend only on the mesh regularity of the initial mesh. Hence, Theorem 6 and Corollary 1 ensure a mesh independent $\kappa < 1$ such that the spectral radius of $I - MA_{\alpha,h}$ is $\leq \kappa$. We have the following result.

Theorem 7 *Let* $\mathcal{T}_j^* = \{T \in \mathcal{T}_j : \bar{T} \cap \Gamma_{C,h} \text{ contains a whole face}\}$.

1. *There exists a constant $C > 0$ only depending on the regularity of the initial mesh such that (A1) holds with*

$$K_1 = C\Big(1 + \max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{4^\ell \alpha \,\mathrm{diam}(T)}{\mathrm{diam}(T_0)^2}\Big).$$

2. *There exists a constant $C > 0$ only depending on the regularity of the initial mesh such that (A2) holds with*

$$\kappa_{kl} = C\left(\frac{1}{\sqrt{2}}\right)^{l-k}\Big(1 + \delta_{l\ell}(1-\delta_{k\ell}) \max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{2^\ell \sqrt{\alpha \,\mathrm{diam}(T)}}{\mathrm{diam}(T_0)}\Big)$$

   where $\delta_{l\ell}, \delta_{k\ell}$ are the usual Kronecker-symbols.
3. *The operators B_l, $l \geq 1$, corresponding to Algorithm SGS and $B_0 = A_0$ satisfy (A3) with $\omega = 1$.*

Thus (A3) holds with $\omega = 1$ and (A1), (A2) holds uniformly in ℓ as long as

$$\max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{2^\ell \sqrt{\alpha \,\mathrm{diam}(T)}}{\mathrm{diam}(T_0)}$$

is uniformly bounded.

The proof will be given in the rest of this section.

5.5.1 Auxiliary norm estimates

We prove now several estimates that will be used to verify assumptions (A1)–(A3).

By comparing the quadratic forms the following lemma is easy to verify.

Lemma 4 *Let T be a tetrahedron with nodes $p_1, \ldots, p_4$. Then for any linear function $v : T \to \mathbb{R}^d$*

$$\frac{1}{5}\|v\|^2_{L^2(T)} \leq \frac{|T|}{20}\sum_{i=1}^4 \|v(p_i)\|_2^2 \leq \|v\|^2_{L^2(T)}.$$

We will use the well known inverse inequality for linear finite elements, e.g. [2].

Lemma 5 *There exists a constant C_0 depending only on the shape regularity of the mesh $\tilde{T}_\ell$ such that*

$$|v|^2_{H^1(T)} \leq C_0 4^k \|v\|^2_{0,T} \quad \forall\, v \in \tilde{\mathcal{S}}_k, \quad \forall\, T \in \tilde{\mathcal{T}}_k, \;\; 0 \leq k \leq \ell,$$
$$|v|^2_{H^1(T)} \leq C_0 \frac{1}{\mathrm{diam}(T)^2} \|v\|^2_{L^2(T)} \quad \forall\, v \in \mathcal{S}_k, \quad \forall\, T \in \mathcal{T}_k, \;\; 0 \leq k \leq \ell.$$

Lemma 6 *Let T be a tetrahedron which is contained in a tetrahedron of $\mathcal{T}_0$. Let v be a linear function on T. Finally let $\bar{v}$ be a linear function that coincides with v in at least one vertex of T and vanishes in the remaining vertices of T. Then*

$$\|\bar{v}\|_{0,T} \leq \frac{9}{5} \|v\|_{0,T} \,.$$

Proof See appendix A.4.

Lemma 7 *There exist constants $C_P, C > 0$ only depending on the initial triangulation and the shape regularity of $\mathcal{T}_\ell$ such that for all $T \in \mathcal{T}_\ell$ and all $v \in \mathcal{S}_\ell^3$*

$$\begin{aligned}
&\frac{1}{5} \|v\|^2_{L^2(T)} \leq \|Pv\|^2_{L^2(T)} \leq 5 \|v\|^2_{L^2(T)} \,, \quad \frac{1}{5} \|v\|^2_{0,T} \leq \|Pv\|^2_{0,T} \leq 5 \|v\|^2_{0,T} \,, \\
&\|Pv\|^2_{H^1(T)} \leq (2 + C_P L_N^2) \|v\|^2_{H^1(T)} \,, \quad \|v\|^2_{H^1(T)} \leq (2 + C_P L_N^2) \|Pv\|^2_{H^1(T)} \,, \\
&\|Pv - Nv\|_{H^1(T)} \leq C \text{diam}(T)((1 + 2L_N) \|v\|_{H^1(T)} + |N|_{W^{2,\infty}(T)} \|v\|_{L^2(T)}).
\end{aligned}$$

Proof See appendix A.5.

The following fact plays an important role for the verification of (A1)–(A3).

Lemma 8 *There exists a constant C_1' not depending on ℓ and the particular $0-1$-structure of $M_h(u_h) = \text{diag}(m_1, \ldots, m_{N_{K_h}})$ such that the $\bar{\mathcal{V}}_k'$-decomposition of $\bar{\mathcal{W}}_\ell'$ admits the estimate*

$$|\bar{Q}_0' v|^2_{H^1(\Omega)} + \sum_{k=1}^{\ell} 4^k \left\|\bar{Q}_k' v - \bar{Q}_{k-1}' v\right\|_0^2 \leq C_1' |v|^2_{H^1(\Omega)} \quad \forall \, v \in \bar{\mathcal{W}}_\ell'. \tag{60}$$

Proof The proof is carried out for the 2D case in [2, Thm. 7.6] for the case that homogeneous boundary conditions are prescribed on a boundary part $\Gamma \subset \partial\Omega$ that is the union of boundary faces of the initial triangulation $\mathcal{T}_0$. As mentioned in this paper, the result holds with similar proof also for the 3D case. This result can be applied to the first and second component of the functions $v \in \bar{\mathcal{W}}_\ell'$, $\bar{Q}_0' v$, $\bar{Q}_k' v - \bar{Q}_{k-1}' v$. The third component satisfies in addition Dirichlet conditions on Γ_ℓ, Γ_k, and Γ_0, respectively. By extending the arguments in [2, Thm. 7.6] the estimate (60) can also be proven for this situation. □

By using the boundedness properties of the linear operator $P : \bar{\mathcal{W}}_k' \to \bar{\mathcal{W}}_k$ in Lemma 7 we obtain the following analogue for the $\bar{\mathcal{V}}_k$-decomposition of $\bar{\mathcal{W}}_\ell$.

Lemma 9 *There exists a constant C_1 not depending on ℓ and the particular $0-1$-structure of $M_h(u_h) = \text{diag}(m_1, \ldots, m_{N_{K_h}})$ such that the $\bar{\mathcal{V}}_k$-decomposition of $\bar{\mathcal{W}}_\ell$ admits the estimate*

$$|P\bar{Q}_0' v|^2_{H^1(\Omega)} + \sum_{k=1}^{\ell} 4^k \left\|P(\bar{Q}_k' v - \bar{Q}_{k-1}' v)\right\|_0^2 \leq C_1 |Pv|^2_{H^1(\Omega)} \quad \forall \, v \in \bar{\mathcal{W}}_\ell'. \tag{61}$$

Proof See appendix A.6.

The following estimate will be essential to verify assumption (A2).

Lemma 10 *There is a constant C depending only on the shape regularity of the tetrahedra such that for all $l > k$ the bilinear form $a(\cdot, \cdot)$ in* (4) *satisfies for all $l > k$*

$$a(w_k, v_l) \leq C \left(\frac{1}{\sqrt{2}}\right)^{l-k} \|w_k\| \, 2^l \, \|v_l\|_0 \quad \forall \, w_k \in \mathcal{W}_k, \; v_l \in \mathcal{W}_l.$$

Proof We have $w_k = P\bar{w}_k$, $v_l = P\bar{v}_l$ with $\bar{w}_k \in \bar{\mathcal{W}}_k'$ and $\bar{v}_l \in \bar{\mathcal{W}}_l'$ for $l < \ell$, $\bar{v}_l \in \mathcal{W}_l$ for $l = \ell$.

We consider all bilinear forms c, d, e in (34). We have

$$|c(w_k, v_l)| \le \sum_{i,j,r,s=1}^{3} \left| \int_\Omega c_{irjs}((P\bar{w}_k)_{x_i})_r((P\bar{v}_l)_{x_j})_s \, dx \right|.$$

Using $P\bar{w}_k = N\bar{w}_k + (P\bar{w}_k - N\bar{w}_k)$, $P\bar{v}_l = N\bar{v}_l + (P\bar{v}_l - N\bar{v}_l)$ we obtain

$$\begin{aligned} |c(w_k, v_l)| \le & \sum_{i,j,r,s=1}^{3} \left| \int_\Omega (c_{irjs}(N\bar{w}_k)_{x_i})_r((N\bar{v}_l)_{x_j})_s \, dx \right| \\ & + \sum_{T \in \mathcal{T}_\ell} (\|P\bar{w}_k - N\bar{w}_k\|_{H^1(T)} \|P\bar{v}_l\|_{H^1(T)} + \|N\bar{w}_k\|_{H^1(T)} \|P\bar{v}_l - N\bar{v}_l\|_{H^1(T)} \\ & \qquad + \|P\bar{w}_k - N\bar{w}_k\|_{H^1(T)} \|P\bar{v}_l - N\bar{v}_l\|_{H^1(T)}) =: R_1 + R_2. \end{aligned}$$

To estimate the first term we observe that

$$c_{irjs}((N\bar{w}_k)_{x_i})_r((N\bar{v}_l)_{x_j})_s = \sum_{1 \le \tilde{r}, \tilde{s} \le 3} c_{irjs}(N_{r,\tilde{r}}(\bar{w}_k)_{\tilde{r}})_{x_i}(N_{s,\tilde{s}}(\bar{v}_l)_{\tilde{s}})_{x_j}.$$

Hence, we have to estimate

$$\tilde{c}(w, v) := \int_\Omega c(aw)_{x_i}(bv)_{x_j} \, dx \tag{62}$$

with $a = N_{r,\tilde{r}}$, $b = N_{s,\tilde{s}}$, $c = c_{irjs}$, $v = (\bar{v}_l)_{\tilde{s}}$, $w = (\bar{w}_k)_{\tilde{r}}$.

For the 2D case it is shown in [38, Lem. 6.1] that for all $w \in \tilde{\mathcal{S}}_k, v \in \tilde{\mathcal{S}}_l, T \in \tilde{\mathcal{T}}_k$

$$\int_T c \, w_{x_i} v_{x_j} \, dx \le C(1 + \|c\|_{W^{1,\infty}(T)}) \sqrt{2}^{k-l} |w|_{H^1(T)} 2^l \|v\|_{0,T}.$$

It is easy to check that the proof can be extended to the 3D case and that it can be adapted to estimate (62). We only sketch the differences in the proof. Let as above $v = (\bar{v}_l)_{\tilde{s}}$, $w = (\bar{w}_k)_{\tilde{r}}$, $a = N_{r,\tilde{r}}$, $b = N_{s,\tilde{s}}$, $c = c_{irjs}$. Let $T \in \tilde{\mathcal{T}}_k$ be arbitrary. We want to estimate

$$\tilde{c}(w, v)|_T = \int_T c(aw)_{x_i}(bv)_{x_j} \, dx.$$

Denote the nodes of $\tilde{\mathcal{T}}_l$ by $\tilde{\mathcal{N}}_l$. We set $v = v_0 + v_1$, where $v_0 \in \tilde{\mathcal{S}}_l$ is defined by

$$v_0(x) = \begin{cases} v(x), & x \in \tilde{\mathcal{N}}_l \cap \partial T, \\ 0, & x \in \tilde{\mathcal{N}}_l \setminus \partial T. \end{cases}$$

We have $\tilde{c}(w, v)|_T = \tilde{c}(w, v_0)|_T + \tilde{c}(w, v_1)|_T$.

Since $v_1 = v - v_0$ vanishes on ∂T, integration by parts yields

$$\tilde{c}(w, v_1)|_T = -\int_T (ca_{x_i})_{x_j} w + ca_{x_i} w_{x_j} + (ca)_{x_j} w_{x_i}) b v_1 \, dx$$

by using that $w_{x_i x_j} = 0$ on T. Lemma 6 yields $\|v_1\|_{L^2(T)} \le \frac{9}{5} \|v\|_{L^2(T)}$ and thus

$$|\tilde{c}(w, v_1)|_T| \le \frac{9}{5} \|c\|_{W^{1,\infty}(T)} (|N|_{W^{2,\infty}(\Omega)} + 3L_N + 1) \|w\|_{H^1(T)} \|v\|_{L^2(T)}.$$

The function v_0 vanishes outside a boundary strip S of T with

$$|S| \le C_2 \frac{1}{2^{l-k}} |T|$$

with a constant $C_2 > 0$ depending only on the shape regularity of $\tilde{\mathcal{T}}_k$ and $\tilde{\mathcal{T}}_l$. Now

$$|\tilde{c}(w, v_0)|_T| \le \int_T |c(aw)_{x_i}(bv_0)_{x_j}|\,dx \le \|c\|_{L^\infty(T)} (1 + L_N)^2 \|w\|_{H^1(S)} \|v_0\|_{H^1(S)} .$$

As the restriction of w to $T \supset S$ is linear, we have

$$|w|^2_{H^1(S)} = |S||w|^2_{W^{1,\infty}(T)} = \frac{|S|}{|T|} |w|^2_{H^1(T)}$$

and similarly by using Lemma 4

$$\|w\|^2_{L^2(S)} \le |S| \|w\|^2_{L^\infty(T)} \le \frac{20|S|}{|T|} \|w\|^2_{L^2(T)} .$$

Hence,

$$\|w\|_{H^1(S)} \le \sqrt{\frac{20C_2}{2^{l-k}}} \|w\|_{H^1(T)} .$$

Now set $C_0' = \max_{T_0 \in \mathcal{T}_0} \operatorname{diam}(T_0)^2$. Then the inverse estimate of Lemma 5 yields with Lemma 6

$$\|v_0\|_{H^1(S)} \le \sqrt{C_0' + C_0 4^l} \|v_0\|_{0,T} \le \sqrt{C_0' + C_0 4^l} \frac{3}{\sqrt{5}} \|v\|_{0,T} .$$

In summary, there exists a constant $C_3 > 0$ with

$$|\tilde{c}(w, v)|_T| \le C_3 (1 + L_N + L_N^2 + |N|_{W^{2,\infty}(\Omega)}) \sqrt{2}^{k-l} \|w\|_{H^1(T)} 2^l \|v\|_{0,T} .$$

Using that $v = (\bar{v}_l)_{\tilde{s}}$, $w = (\bar{w}_k)_{\tilde{r}}$, summing these bounds for the terms in (62) yields a constant $C_4 > 0$ such that with $C_5 = \sqrt{2 + C_P L_N^2}$ holds

$$R_1 \le C_4 \sqrt{\frac{1}{2^{l-k}}} \|\bar{w}_k\|_{H^1(\Omega)} 2^l \|\bar{v}_l\|_0 \le C_4 C_5 \sqrt{\frac{5}{2^{l-k}}} \|w_k\|_{H^1(\Omega)} 2^l \|v_l\|_0 ,$$

where we have used Lemma 7. For the second term R_2 Lemma 7 and (80) yield

$$\begin{aligned}
R_2 \le\, & C(1 + 2L_N + \|N\|_{W^{2,\infty}(\Omega)}) \\
& \sum_{T \in \mathcal{T}_\ell} \operatorname{diam}(T) \left(\|\bar{w}_k\|_{H^1(T)} \|P\bar{v}_l\|_{H^1(T)} + \|N\bar{w}_k\|_{H^1(T)} \|\bar{v}_l\|_{H^1(T)} \right) \\
& + C^2 (1 + 2L_N + \|N\|_{W^{2,\infty}(\Omega)})^2 \sum_{T \in \mathcal{T}_\ell} \operatorname{diam}(T)^2 \|\bar{w}_k\|_{H^1(T)} \|\bar{v}_l\|_{H^1(T)} \\
& \le C_6 \sum_{T \in \mathcal{T}_\ell} \operatorname{diam}(T) \|\bar{w}_k\|_{H^1(T)} \|\bar{v}_l\|_{H^1(T)} .
\end{aligned}$$

Then we obtain by Lemma 5 and 7

$$\operatorname{diam}(T) \|\bar{v}_l\|_{H^1(T)} \le \sqrt{5(C_0' + C_0)} \|v_l\|_{L^2(T)} , \quad \|\bar{w}_k\|_{H^1(\Omega)} \le C_5 \|w_k\|_{H^1(\Omega)}$$

and summing the local estimate yields with $\|v_l\|_{L^2(\Omega)} \leq (\max_{T_0\in\mathcal{T}_0} \mathrm{diam}(T_0))\,\|v_l\|_0$

$$R_2 \leq \left(\max_{T_0\in\mathcal{T}_0} \mathrm{diam}(T_0)\right) C_6\sqrt{5(C_0'+C_0)}C_5\,\|w_k\|_{H^1(\Omega)}\,\|v_l\|_0\,.$$

Adding the bounds for R_1, R_2 yields the existence of a constant $C_7 > 0$ with

$$|c(w_k,v_l)| \leq R_1 + R_2 \leq C_7\sqrt{2}^{k-l}\,\|w_k\|_{H^1(\Omega)}\,2^l\,\|v_l\|_0\,.$$

Next consider $e(w_k,v_l) = \int_\Omega E w_k : v_l\,dx$. Since $H^1(\Omega) \hookrightarrow L^6(\Omega)$ for $d \leq 3$,

$$|e(w_k,v_l)| \leq C_8\,\|E\|_{L^2(\Omega)}\,\|w_k\|_{H^1(\Omega)}\,\|v_l\|^{1/2}_{L^2(\Omega)}\,\|v_l\|^{1/2}_{H^1(\Omega)}\,.$$

Now by Lemmas 5 and 7 we obtain for all $T \in \mathcal{T}_l$, since $\bar{v}_l$ is linear on T,

$$\begin{aligned}\|v_l\|^2_{H^1(T)} &\leq (2+C_PL_N^2)\,\|\bar{v}_l\|^2_{H^1(T)} \leq (2+C_PL_N^2)(C_0'+C_04^l)\,\|\bar{v}_l\|^2_{0,T}\\ &\leq 5(2+C_PL_N^2)(C_0'+C_04^l)\,\|v_l\|^2_{0,T}\,.\end{aligned}$$

Summing up yields a constant $C_9 > 0$ with

$$\|v_l\|^{1/2}_{H^1(\Omega)} \leq C_9\sqrt{2}^l\,\|v_l\|^{1/2}_0\,.$$

We conclude that

$$|e(w_k,v_l)| \leq C_{10}\sqrt{2}^l\,\|E\|_{L^2(\Omega)}\,\|w_k\|_{H^1(\Omega)}\,\|v_l\|_0\,.$$

The term $d_1(w_k,v_l) = \int_\Omega D\nabla w_k : v_l\,dx$ can be estimated similarly, since

$$|d_1(w_k,v_l)| \leq C_{11}\,\|D\|_{H^1(\Omega)}\,\|w_k\|_{H^1(\Omega)}\,\|v_l\|^{1/2}_{L^2(\Omega)}\,\|v_l\|^{1/2}_{H^1(\Omega)}$$

Now we can proceed axactly as for $e(w_k,v_l)$.

Finally, $d_2(w_k,v_l)$ can be treated similary as $b(w_k,v_l)$. Each term is of the form $\int_\Omega d_{irs}(P\bar{w}_k)_r((P\bar{v}_l)_{x_i})_s\,dx$. Using as above $P\bar{w}_k = N\bar{w}_k + (P\bar{w}_k - N\bar{w}_k)$, $P\bar{v}_l = N\bar{v}_l + (P\bar{v}_l - N\bar{v}_l)$, we obtain

$$\int_\Omega d_{irs}(P\bar{w}_k)_r((P\bar{v}_l)_{x_i})_s\,dx = \int_\Omega d_{irs}(N\bar{w}_k)_r((N\bar{v}_l)_{x_i})_s\,dx + R_3,$$

where R_3 can be estimated as R_2 above. It remains to consider

$$\tilde{d}_2(w,v) = \int_\Omega d(aw)(bv)_{x_i}\,dx$$

with $v = (\bar{v}_l)_{\tilde{s}}$, $w = (\bar{w}_k)_{\tilde{r}}$, $a = N_{r,\tilde{r}}$, $b = N_{s,\tilde{s}}$, $d = d_{irs}$.

Let $T \in \tilde{\mathcal{T}}_k$ be arbitrary. As above we we use the splitting $v = v_0 + v_1$. Since v_1 vanishes on ∂T, integration by parts yields

$$\tilde{d}_2(w,v_1)|_T = -\int_T (d_{x_i}(aw) + d(aw)_{x_i})bv_1\,dx.$$

Summing over all $T \in \tilde{\mathcal{T}}_k$ yields similarly as for d_1 and e

$$\begin{aligned}|\tilde{d}_2(w,v_1)| &= \left|\int_\Omega (d_{x_i}(aw) + d(aw)_{x_i})bv_1\,dx\right|\\ &\leq C_{12}\,\|d\|_{H^1(\Omega)}\,(1+L_N)^{3/2}\,\|w\|_{H^1(\Omega)}\,\|v_1\|^{1/2}_{L^2(\Omega)}\,\|v_1\|^{1/2}_{H^1(\Omega)}\,.\end{aligned}$$

Since v_1 is linear on all $\tilde{T} \in \tilde{\mathcal{T}}_l$, we have

$$\|v_1\|^2_{H^1(\tilde{T})} \le (C_0' + C_0 4^l)\,\|v_1\|^2_{0,\tilde{T}} \le \frac{9^2}{5^2}(C_0' + C_0 4^l)\,\|v\|^2_{0,\tilde{T}}$$

and thus $\|v_1\|^{1/2}_{H^1(\Omega)} \le C_{12}\sqrt{2^l}\,\|v\|_0$. This shows that

$$|\tilde{d}_2(w, v_1)| \le C_{13}\sqrt{2}^l\,\|w\|_{H^1(\Omega)}\,\|v\|_0\,.$$

Finally, since as above v_0 vanishes outside of the strip S we have

$$|\tilde{d}_2(w, v_0)|_T| \le \|d\|_{L^\infty(\Omega)}\,(1 + L_N)\,\|w\|_{L^2(S)}\,\|v_0\|_{H^1(S)}\,.$$

Now we can proceed for this term as for $\tilde{c}(w, v_0)|_T$. □

5.5.2 Application of the general convergence result

We check next that assumption (A3) is satisfied for Algorithm MPRSN. It is easy to check that each symmetric Gauss-Seidel step has the form

$$\mathbf{d}_l \leftarrow \mathbf{d}_l + \hat{\mathbf{B}}_l^{-1}(\mathbf{r}_l - \mathbf{A}_l\mathbf{d}_l), \quad \hat{\mathbf{B}}_l = (\mathbf{D}_l - \mathbf{L}_l)\mathbf{D}_l^{-1}(\mathbf{D}_l - \mathbf{L}_l)^T$$

and the m steps of Algorithm SGS lead to the approximate subspace correction

$$\mathbf{d}_l \leftarrow \mathbf{d}_l + \mathbf{B}_l^{-1}(\mathbf{r}_l - \mathbf{A}_l\mathbf{d}_l), \text{ where}$$
$$\mathbf{B}_l^{-1} = \sum_{j=0}^{m-1}(I - \hat{\mathbf{B}}_l^{-1}\mathbf{A}_l)^j\hat{\mathbf{B}}_l^{-1} = (I - (I - \hat{\mathbf{B}}_l^{-1}\mathbf{A}_l)^m)\mathbf{A}_l^{-1}. \tag{63}$$

$\mathbf{B}_l^{-1}$ is symmetric, since $\hat{\mathbf{B}}_l^{-1}$ and thus all matrices $(\hat{\mathbf{B}}_l^{-1}\mathbf{A}_l)^j\hat{\mathbf{B}}_l^{-1}$ are symmetric.

We are now ready to show that (A3) holds for $\omega = 1$ which is equivalent to

$$\mathbf{v}_l^T\mathbf{A}_l\mathbf{v}_l \le \mathbf{v}_l^T\mathbf{B}_l\mathbf{v}_l \quad \forall\,\mathbf{v}_l. \tag{64}$$

Lemma 11 *The operators B_l, $l \ge 1$, corresponding to Algorithm SGS and $B_0 = A_0$ satisfy (A3) with $\omega = 1$.*

Proof See appendix A.7.

We turn now to assumption (A2).

Lemma 12 *Let $\mathcal{T}_j^* = \{T \in \mathcal{T}_j : \bar{T} \cap \Gamma_{C,h}$ contains a whole face$\}$. Then there exists a constant $C > 0$ only depending on the regularity of the initial mesh such that*

$$a_\alpha(w_k, v_l) \le C\left(\frac{1}{\sqrt{2}}\right)^{l-k}\left(1 + \delta_{l\ell}(1 - \delta_{k\ell})\max_{T_0\in\mathcal{T}_0^*}\max_{\mathcal{T}_\ell^*\ni T\subset T_0}\frac{2^\ell\sqrt{\alpha\,\mathrm{diam}(T)}}{\mathrm{diam}(T_0)}\right)$$
$$(B_k w_k, w_k)_0^{1/2}(B_l v_l, v_l)_0^{1/2} \quad \forall\, w_k \in \mathcal{W}_k,\ v_l \in \mathcal{V}_l,\ 0 \le k \le l \le \ell,$$

where $\delta_{l\ell}, \delta_{k\ell}$ are the usual Kronecker-symbols.

Thus, (A2) holds uniformly in ℓ as long as

$$\max_{T_0\in\mathcal{T}_0^*}\max_{\mathcal{T}_\ell^*\ni T\subset T_0}\frac{2^\ell\sqrt{\alpha\,\mathrm{diam}(T)}}{\mathrm{diam}(T_0)}$$

is uniformly bounded.

Proof For $k = l$ we have by the Cauchy-Schwarz inequality and by Lemma 11

$$a_\alpha(w_k, v_l) \le a_\alpha(w_k, w_k)^{1/2} a_\alpha(v_l, v_l)^{1/2} \le (B_k w_k, w_k)_0^{1/2} (B_l v_l, v_l)_0^{1/2}$$

for all $w_k \in \mathcal{W}_k, v_l \in \mathcal{V}_l$. For $0 \le k < l \le \ell$ Lemma 10 yields

$$a_\alpha(w_k, v_l) = a(w_k, v_l) \le C\sqrt{2}^{k-l} \|w_k\| \, 2^l \|v_l\|_0 \quad \forall\, w_k \in \mathcal{W}_k, v_l \in \mathcal{V}_l. \tag{65}$$

By 5.4.1 we have $\mathcal{V}_l = \bar{\mathcal{V}}_l$ for $l < \ell$ and for $v_l \in \mathcal{V}_l$ there exists $v_l' \in \bar{\mathcal{V}}_l'$ with

$$v_l = P v_l' = P(\bar{Q}_l' v_l' - \bar{Q}_{l-1}' v_l').$$

Hence, Lemma 9 applied to $u = v_l' \in \bar{\mathcal{V}}_l'$ yields with (36)

$$4^l \|v_l\|_0^2 \le C_1 |v_l|_{H^1(\Omega)}^2 \le C_1 C_A a(v_l, v_l) = C_1 C_A a_\alpha(v_l, v_l) = C_1 C_A \|v_l\|^2 \;\; \forall\, l < \ell.$$

Therefore, we have with Lemma 11 and (65) for all $0 \le k < l < \ell$ and $w_k \in \mathcal{W}_k,\; v_l \in \mathcal{V}_l$ the estimate

$$a_\alpha(w_k, v_l) \le C\sqrt{C_1 C_A}\sqrt{2}^{k-l} (B_k w_k, w_k)_0^{1/2} (B_l v_l, v_l)_0^{1/2}.$$

It remains the case $0 \le k < l = \ell$. For $v_\ell \in \mathcal{V}_\ell$ we have by definition with $v_\ell' \in \mathcal{W}_\ell$

$$v_\ell = P(v_\ell' - \bar{Q}_{\ell-1}' v_\ell')$$

and without restriction we can choose v_ℓ' such that $v_\ell' = v_\ell' - \bar{Q}_{\ell-1}' v_\ell'$ (else use $v_\ell' - \bar{Q}_{\ell-1}' v_\ell'$ for v_ℓ'). Then we have

$$\begin{aligned} v_\ell = P v_\ell' &= P(v_\ell' - \bar{Q}_{\ell-1}' v_\ell') = P(v_\ell' - \bar{Q}_\ell' v_\ell') + P(\bar{Q}_\ell' - \bar{Q}_{\ell-1}') v_\ell' \\ &= P(v_\ell' - \bar{Q}_\ell' v_\ell') + P(\bar{Q}_\ell' - \bar{Q}_{\ell-1}') \bar{Q}_\ell' v_\ell'. \end{aligned}$$

Thus, (65) yields for all $0 \le k < l = \ell$ and all $w_k \in \mathcal{W}_k, v_\ell \in \mathcal{V}_\ell$

$$\begin{aligned} a_\alpha(w_k, v_\ell) &\le \sqrt{2}^{k-\ell} \|w_k\| \, 2^\ell \left\| P(v_\ell' - \bar{Q}_\ell' v_\ell') + P(\bar{Q}_\ell' - \bar{Q}_{\ell-1}') \bar{Q}_\ell' v_\ell' \right\|_0 \\ &\le C\sqrt{2}^{k-\ell} (B_k w_k, w_k)_0^{1/2} 2^\ell \left(\left\| P(v_\ell' - \bar{Q}_\ell' v_\ell') \right\|_0 + \left\| P(\bar{Q}_\ell' - \bar{Q}_{\ell-1}') \bar{Q}_\ell' v_\ell' \right\|_0 \right), \end{aligned} \tag{66}$$

since $a_\alpha(w_k, v_\ell) = a(w_k, v_\ell)$. Now Lemma 9 applied to $u = \bar{Q}_\ell' v_\ell'$ yields

$$4^\ell \left\| P(\bar{Q}_\ell' - \bar{Q}_{\ell-1}') \bar{Q}_\ell' v_\ell' \right\|_0^2 \le C_1 |P \bar{Q}_\ell' v_\ell'|_{H^1(\Omega)}^2. \tag{67}$$

All the following estimates until (70) hold generally for all $v_\ell' \in \mathcal{W}_\ell$. We have

$$\left\| v_\ell' - \bar{Q}_\ell' v_\ell' \right\|_0 = \inf_{v \in \bar{\mathcal{W}}_\ell'} \| v_\ell' - v \|_0 .$$

By assumption (C2) there exists a function $\bar{v}_\ell' \in \bar{\mathcal{W}}_\ell'$ with

$$\begin{aligned} &\| P v_\ell' - P \bar{v}_\ell' \|_{L^2(\Gamma_{C,h})}^2 \le \kappa_3 (B_h P v_\ell')^T M_h (B_h P v_\ell') \\ &= \kappa_3 (B_h P(v_\ell' - \bar{Q}_{\ell-1}' v_\ell'))^T M_h (B_h P(v_\ell' - \bar{Q}_{\ell-1}' v_\ell')) = \kappa_3 (B_h v_\ell)^T M_h (B_h v_\ell), \end{aligned} \tag{68}$$

where κ_3 does not depend on v_ℓ'. We have

$$\left\| v_\ell' - \bar{Q}_\ell' v_\ell' \right\|_0 \le \| v_\ell' - \bar{v}_\ell' \|_0 .$$

Let $\mathcal{T}_l' = \left\{ T \in \mathcal{T}_l : \bar{T} \cap \Gamma_{C,h} \ne \emptyset \right\}$ and $\mathcal{T}_l^* = \{ T \in \mathcal{T}_l :\, \bar{T} \cap \Gamma_{C,h} \text{ contains a whole face} \}$. Since $v_\ell - P\bar{v}_\ell' = P(v_\ell' - \bar{v}_\ell')$ vanishes on all nodes not contained in $\Gamma_{C,h}$ and all boundary nodes are only shared by a finite number of elements, there exists (cf. also [2]) a constant C'' independent of v'_ℓ with

$$\begin{aligned}\|P(v'_\ell - \bar{v}'_\ell)\|_0^2 &= \sum_{T_0\in\mathcal{T}'_0} \sum_{\mathcal{T}'_\ell \ni T\subset T_0} \frac{1}{\text{diam}(T_0)^2} \|P(v'_\ell - \bar{v}'_\ell)\|^2_{L^2(T)} \\ &\le C'' \sum_{T_0\in\mathcal{T}^*_0} \sum_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{1}{\text{diam}(T_0)^2} \|P(v'_\ell - \bar{v}'_\ell)\|^2_{L^2(T)}\,.\end{aligned}$$

Now it is obvious that for all $T \in \mathcal{T}^*_\ell$

$$\|P(v'_\ell - \bar{v}'_\ell)\|^2_{L^2(T)} \le \text{diam}(T)\, \|P(v'_\ell - \bar{v}'_\ell)\|^2_{L^2(T\cap\Gamma_{C,h})}\,.$$

This yields

$$\|P(v'_\ell - \bar{v}'_\ell)\|_0^2 \le C_2 \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{\text{diam}(T)}{\text{diam}(T_0)^2} \|P(v'_\ell - \bar{v}'_\ell)\|^2_{L^2(\Gamma_{C,h})}\,.$$

Using (68) and Lemma 7 we conclude that

$$\begin{aligned}\left\|P(v'_\ell - \bar{Q}'_\ell v'_\ell)\right\|_0^2 &\le 5\left\|v'_\ell - \bar{Q}'_\ell v'_\ell\right\|_0^2 \le 5\,\|v'_\ell - \bar{v}'_\ell\|_0^2 \le 25\,\|P(v'_\ell - \bar{v}'_\ell)\|_0^2 \\ &\le 25 C_2\kappa_3 \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{\alpha\text{diam}(T)}{\text{diam}(T_0)^2} \frac{1}{\alpha}(B_h v_\ell)^T M_h (B_h v_\ell) \\ &\le 25 C_2\kappa_3 \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{\alpha\text{diam}(T)}{\text{diam}(T_0)^2} a_\alpha(v_\ell, v_\ell) \\ &\le 25 C_2\kappa_3 \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{\alpha\text{diam}(T)}{\text{diam}(T_0)^2} (B_\ell v_\ell, v_\ell)_0,\end{aligned} \tag{69}$$

where we have used Lemma 11 in the last inequality. Finally, by using the inverse estimate of Lemma 5 we obtain again with Lemma 11

$$\begin{aligned}&|P\bar{Q}'_\ell v'_\ell|_{H^1(\Omega)} \le |Pv'_\ell|_{H^1(\Omega)} + |P(v'_\ell - \bar{Q}'_\ell v'_\ell)|_{H^1(\Omega)} \\ &\le |v_\ell|_{H^1(\Omega)} + C_0^{1/2} 2^\ell \left\|P(v'_\ell - \bar{Q}'_\ell v'_\ell)\right\|_0 \\ &\le C_A^{1/2}\|v_\ell\| + 5(C_0C_2\kappa_3)^{1/2} \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{2^\ell\sqrt{\alpha\,\text{diam}(T)}}{\text{diam}(T_0)} a_\alpha(v_\ell, v_\ell)^{1/2} \\ &\le \left(C_A^{1/2} + 5(C_0C_2\kappa_3)^{1/2} \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{2^\ell\sqrt{\alpha\,\text{diam}(T)}}{\text{diam}(T_0)}\right)(B_\ell v_\ell, v_\ell)_0^{1/2}.\end{aligned} \tag{70}$$

Combining (66), (67), (69), and (70) we obtain for all $w_k \in \mathcal{W}_k, v_\ell \in \mathcal{V}_\ell, 0 \le k < \ell$

$$\begin{aligned}&a_\alpha(w_k, v_\ell) \le C\sqrt{2}^{k-\ell}\Big(5(C_2\kappa_3)^{1/2}((C_0C_1)^{1/2} + 1) \\ &\cdot \max_{T_0\in\mathcal{T}^*_0} \max_{\mathcal{T}^*_\ell \ni T\subset T_0} \frac{2^\ell\sqrt{\alpha\,\text{diam}(T)}}{\text{diam}(T_0)} + (C_1C_A)^{1/2}\Big)(B_k w_k, w_k)_0^{1/2}(B_\ell v_\ell, v_\ell)_0^{1/2}.\end{aligned}$$

□

We turn finally to the verification of assumption (A1). We start with the following auxiliary lemmas.

Lemma 13 *Let $B \in \mathbb{R}^{n,n}$ be an arbitrary matrix such that all entries have absolute value ≤ 1. Moreover, let N_{nb} be an upper bound for the nonzero entries in each row and column of B. Then the spectral norm of B is bounded by*

$$\|B\|_2 \leq N_{nb}.$$

Proof Let $x \in \mathbb{R}^d$. Then

$$((Bx)_i)^2 \leq \big(\sum_{b_{ij}\neq 0} |x_j|\big)^2 \leq N_{nb} \sum_{b_{ij}\neq 0} x_j^2.$$

Since any x_j appears in at most N_{nb} rows, summation yields $\|Bx\|_2^2 \leq N_{nb}^2 \|x\|_2^2$. □

Lemma 14 *Let $\mathcal{T}_\ell^* = \big\{T \in \mathcal{T}_\ell : \bar{T} \cap \Gamma_{C,h}$ contains a whole face$\big\}$. Then there exists a constant $C > 0$ only depending on the regularity of the initial mesh such that*

$$\sum_{l=0}^{\ell} (B_l v_l, v_l)_0 \leq C\Big(1 + \max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{4^\ell \alpha \operatorname{diam}(T)}{\operatorname{diam}(T_0)^2}\Big) \left\|\sum_{l=0}^{\ell} v_l\right\|^2 \qquad \forall\, v_l \in \mathcal{V}_l.$$

Thus, (A1) holds uniformly in ℓ as long as

$$\max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{2^\ell \sqrt{\alpha \operatorname{diam}(T)}}{\operatorname{diam}(T_0)}$$

is uniformly bounded.

Proof We know by (35) that there exists a constant M_a with

$$|a(u,v)| \leq M_a |u|_{H^1(\Omega)}\, |v|_{H^1(\Omega)}.$$

Denote as before $\mathbf{A_l}$ the matrix representation of the bilinear form (51) on $\mathcal{W}_l$. Let N_{nb} be a uniform upper bound for the number of nonzero entries in the rows of $\mathbf{A_l}$. N_{nb} can be chosen only depending on the initial triangulation.

We recall that $B_0 = A_0$. Therefore, we have

$$(B_0 v_0, v_0)_0 = a_\alpha(v_0, v_0) = a(v_0, v_0) \leq M_a |v_0|^2_{H^1(\Omega)} \quad \forall\, v_0 \in \mathcal{V}_0. \tag{71}$$

For $l \geq 1$ the operator B_l is the result of symmetric Gauss-Seidel iterations. As we have seen in the proof of Lemma 11, the symmetric matrices $\hat{\mathbf{B}}_l^{-1}$ corresponding to a single symmetric Gauss-Seidel step and $\mathbf{B}_l^{-1}$ for m iterations have the same eigenvectors and the eigenvalue $\lambda \in (0,1)$ of $\hat{\mathbf{B}}_l^{-1}$ corresponds to the eigenvalue $1-(1-\lambda)^m \in (\lambda, 1)$ of $\mathbf{B}_l^{-1}$. Therefore, $\hat{\mathbf{B}}_\mathbf{l}$ and $\mathbf{B_l}$ have the same orthonormal basis of eigenvectors and the eigenvalues satisfy $\lambda_i(\mathbf{B}_l) \in (0, \lambda_i(\hat{\mathbf{B}}_l)]$. Thus,

$$\mathbf{v}_l^T \mathbf{B}_l \mathbf{v}_l \leq \mathbf{v}_l^T \hat{\mathbf{B}}_l \mathbf{v}_l. \tag{72}$$

Now $\hat{\mathbf{B}}_l = \mathbf{A}_l + \mathbf{L}_l \mathbf{D}_l^{-1} \mathbf{L}_l^T = \mathbf{A}_l + \mathbf{L}_l \mathbf{D}_l^{-1} \mathbf{D}_l \mathbf{D}_l^{-1} \mathbf{L}_l^T$. Since $\mathbf{A}_l$ is positive definite, $\mathbf{L}_l \mathbf{D}_l^{-1}$ has entries with absolute value ≤ 1. Applying Lemma 13 yields $\left\|\mathbf{L}_l \mathbf{D}_l^{-1}\right\|_2, \left\|\mathbf{D}_l^{-1} \mathbf{L}_l^T\right\|_2 \leq N_{nb}$. Hence, together with (72) we obtain

$$\mathbf{v}_l^T \mathbf{B}_l \mathbf{v}_l \leq \mathbf{v}_l^T \hat{\mathbf{B}}_l \mathbf{v}_l \leq \mathbf{v}_l^T \mathbf{A}_l \mathbf{v}_l + N_{nb}^2 \mathbf{v}_l^T \mathbf{D}_l \mathbf{v}_l, \quad l \geq 1. \tag{73}$$

Let $\{\phi^l_{p_i,j}\}_{1\leq i\leq N_l, 1\leq j\leq 3}$ be the nodal basis of $\mathcal{S}_l^3$ corresponding to the free nodes p_i of $\mathcal{T}_l$. To estimate the second term we observe that

$$\mathbf{v}_l^T \mathbf{D}_l \mathbf{v}_l = \sum_{i=1}^{N_l} \sum_{j=1}^{3} (\mathbf{v}_l)^2_{i,j} a_\alpha(\phi^l_{p_i,j}, \phi^l_{p_i,j}).$$

We consider first the coarser levels $1 \le l < \ell$. We obtain for all $v_l \in \mathcal{W}_l$, $1 \le l < \ell$, with a constant M_a by using Lemma 5 and 6

$$\mathbf{v}_l^T \mathbf{A}_l \mathbf{v}_l = a_\alpha(v_l, v_l) = a(v_l, v_l) \le M_a |v_l|^2_{H^1(\Omega)} \le M_a C_0 4^l \|v_l\|_0^2$$

as well as

$$\begin{aligned}
\mathbf{v}_l^T \mathbf{D}_l \mathbf{v}_l = & \sum_{\substack{i=1 \\ 1\le j\le 3}}^{N_l} (\mathbf{v}_l)^2_{i,j} a(\phi^l_{p_i,j}, \phi^l_{p_i,j}) \le \sum_{\substack{i=1 \\ 1\le j\le 3}}^{N_l} M_a (\mathbf{v}_l)^2_{i,j} \|\phi^l_{p_i,j}\|^2_{H^1(\Omega)} \\
& \le \sum_{T\in\mathcal{T}_l} \sum_{p_i \in T^{cl}} \sum_{1\le j\le 3} (\mathbf{v}_l)^2_{i,j} M_a C_0 4^l \left\|\phi^l_{p_i,j}\right\|^2_{0,T} \\
& \le \sum_{T\in\mathcal{T}_l} \sum_{p_i \in T^{cl}} \sum_{1\le j\le 3} M_a C_0 4^l \frac{9}{5} \|v_{l,j}\|^2_{0,T} \le 4 M_a C_0 4^l \frac{9}{5} \|v_l\|_0^2 .
\end{aligned}$$

We conclude with (72) and (73) that for all $v_l \in \mathcal{W}_l, 1 \le l < \ell$

$$(B_l v_l, v_l)_0 \le \mathbf{v}_l^T \hat{\mathbf{B}}_l \mathbf{v}_l \le M_a C_0 (1 + N_{nb}^2 36/5) 4^l \|v_l\|_0^2 =: C_4 4^l \|v_l\|_0^2 . \tag{74}$$

Now consider $v_\ell \in \mathcal{W}_\ell$. We obtain similarly as above

$$\begin{aligned}
&\mathbf{v}_\ell^T \mathbf{A}_\ell \mathbf{v}_\ell = a_\alpha(v_\ell, v_\ell) = a(v_\ell, v_\ell) + \alpha^{-1} (v_\ell, B_h^* M_h(u_h) B_h v_\ell)_0 \\
&\le M_a |v_\ell|^2_{H^1(\Omega)} + \alpha^{-1}(v_\ell, S_h v_\ell)_0 \le M_a C_0 4^\ell \|v_\ell\|_0^2 + \alpha^{-1}(v_\ell, S_h v_\ell)_0,
\end{aligned}$$

where $S_h = B_h^* M_h(u_h) B_h$ as in Assumption (C1). We have $\mathbf{A}_\ell = \tilde{\mathbf{A}}_\ell + \alpha^{-1}\mathbf{S}_\ell$ with the matrix representations $\tilde{\mathbf{A}}_\ell$ of $a(\cdot,\cdot)$ and $\mathbf{S}_\ell$ of S_h on $\mathcal{W}_\ell$. Let

$$\mathbf{D}_\ell = \tilde{\mathbf{D}}_\ell + \alpha^{-1} \operatorname{diag}(\mathbf{S}_\ell)$$

be the corresponding splitting of the diagonal part. As above we can show that

$$\mathbf{v}_\ell^T \tilde{\mathbf{D}}_\ell \mathbf{v}_\ell \le 4 M_a C_0 4^\ell \frac{9}{5} \|v_\ell\|_0^2 \quad \forall\, v_\ell \in \mathcal{W}_\ell.$$

Thus, we have by (73) for all $v_\ell \in \mathcal{V}_\ell$

$$\begin{aligned}
(B_\ell v_\ell, v_\ell)_0 &\le \mathbf{v}_\ell^T A_\ell \mathbf{v}_\ell + N_{nb}^2 (\mathbf{v}_\ell^T \tilde{\mathbf{D}}_\ell \mathbf{v}_\ell + \alpha^{-1} \mathbf{v}_\ell^T \operatorname{diag}(\mathbf{S}_\ell) \mathbf{v}_\ell) \\
&\le M_a C_0 4^\ell (1 + N_{nb}^2 \frac{36}{5}) \|v_\ell\|_0^2 + \alpha^{-1}(1 + N_{nb}^2 C_B)(v_\ell, S_h v_\ell),
\end{aligned}$$

where we have used Assumption (C1) in the last term. We conclude that with $C_5 := (1 + N_{nb}^2 C_B)$

$$(B_\ell v_\ell, v_\ell)_0 \le C_4 4^\ell \|v_\ell\|_0^2 + \alpha^{-1} C_5 (v_\ell, S_h v_\ell). \tag{75}$$

Now let $v = \sum_{l=0}^{\ell} v_l$ with $v_l \in \mathcal{V}_l$ be arbitrary. Then there exists a unique $v' \in \mathcal{W}_\ell = \mathcal{S}_\ell^3 = \mathbf{U}_h$ with

$$v = P v' = P(v' - \bar{Q}'_{\ell-1} v') + \sum_{l=1}^{\ell-1} P(\bar{Q}'_l - \bar{Q}'_{l-1}) v' + P \bar{Q}'_0 v'$$

By the definition of $\mathcal{V}_l$ we have therefore

$$\begin{aligned}
v_0 &= P\bar{Q}'_0 v' = P\bar{Q}'_0 \bar{Q}'_\ell v', \\
v_l &= P(\bar{Q}'_l - \bar{Q}'_{l-1}) v' = P(\bar{Q}'_l - \bar{Q}'_{l-1}) \bar{Q}'_\ell v', \;\; 1 \le l < \ell, \\
v_\ell &= P(v' - \bar{Q}'_{\ell-1} v') = P(v' - \bar{Q}'_\ell v') + P(\bar{Q}'_\ell - \bar{Q}'_{\ell-1}) \bar{Q}'_\ell v'.
\end{aligned}$$

Using (71), (74) and (75) we obtain

$$\sum_{l=0}^{\ell}(B_l v_l, v_l)_0 \leq M_a |v_0|^2_{H^1(\Omega)} + \sum_{l=1}^{\ell} C_4 4^l \|v_l\|_0^2 + \alpha^{-1} C_5(v_\ell, S_h v_\ell). \tag{76}$$

By the definition of the spaces $\mathcal{V}_l$ for $0 \leq l < \ell$ we have

$$\alpha^{-1} C_5(v_\ell, S_h v_\ell) = \alpha^{-1} C_5(v, S_h v) \leq C_5 a_\alpha(v, v) = C_5 \|v\|^2 .$$

Moreover,

$$\begin{aligned} &\|v_0\|_0^2 = \left\|P\bar{Q}_0'\bar{Q}_\ell' v'\right\|_0^2, \quad \|v_l\|_0^2 = \left\|P(\bar{Q}_l' - \bar{Q}_{l-1}')\bar{Q}_\ell' v'\right\|_0^2, \;\; 1 \leq l < \ell, \\ &\|v_\ell\|_0^2 \leq 2\left\|P(v' - \bar{Q}_\ell' v')\right\|_0^2 + 2\left\|P(\bar{Q}_\ell' - \bar{Q}_{\ell-1}')\bar{Q}_\ell' v'\right\|_0^2 . \end{aligned}$$

By using (76) and Lemma 9 (with $u = \bar{Q}_\ell' v'$) we obtain

$$\begin{aligned} &\sum_{l=0}^{\ell}(B_l v_l, v_l)_0 \leq M_a |v_0|^2_{H^1(\Omega)} + \sum_{l=0}^{\ell} C_4 4^l \|v_l\|_0^2 + C_5 \|v\|^2 \\ &\leq M_a |P\bar{Q}_0'\bar{Q}_\ell' v'|^2_{H^1(\Omega)} + 2C_4 \sum_{l=1}^{\ell} 4^l \left\|P(\bar{Q}_l' - \bar{Q}_{l-1}')\bar{Q}_\ell' v'\right\|_0^2 \\ &+ 2C_4 4^\ell \left\|P(v' - \bar{Q}_\ell' v')\right\|_0^2 + C_5 \|v\|^2 \\ &\leq \max(M_a, 2C_4) C_1 |P\bar{Q}_\ell' v'|^2_{H^1(\Omega)} + 2C_4 4^\ell \left\|P(v' - \bar{Q}_\ell' v')\right\|_0^2 + C_5 \|v\|^2 . \end{aligned}$$

We recall that (69) and (70) hold for all $v_\ell' \in \mathcal{W}_\ell$ and thus also for v'. This yields

$$\begin{aligned} &\sum_{l=0}^{\ell}(B_l v_l, v_l)_0 \leq 50 C_4 C_2 \kappa_3 \max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{\alpha 4^\ell \mathrm{diam}(T)}{\mathrm{diam}(T_0)^2} \|v\|^2 + C_5 \|v\|^2 \\ &+ 2\max(M_a, 2C_4) C_1 \left(C_A + 25 C_0 C_2 \kappa_3 \max_{T_0 \in \mathcal{T}_0^*} \max_{\mathcal{T}_\ell^* \ni T \subset T_0} \frac{4^\ell \alpha\, \mathrm{diam}(T)}{\mathrm{diam}(T_0)^2} \right) \|v\|^2 . \end{aligned}$$

□

6 Numerical results: Hertzian contact problem

In this section, we use a 3D Hertzian contact problem to demonstrate the efficiency of the proposed multilevel semismooth Newton method. The example to be considered is an unilateral contact problem consisting of an elastic ball and a rigid planar surface.

For this example under the assumption of small deformation the normal contact stress distribution can be computed analytically. The problem description of the 3D Hertzian contact problem and an analytical solution can be found [10] and [31]. The ball in this example has a radius of $R = 8$ mm and the material behavior is assumed to be linear elastic with Young's modulus $E = 210$ GPa and Poisson's ratio $\nu = 0.3$ from steel (correspond to Lamé coefficients $\lambda = 1.2115e + 05$ and $\mu = 8.0769e + 04$). The symmetry of the problem requires only one-eight of the ball to be modeled. A constant pressure $p = 2\,\frac{N}{mm^2}$ is applied to the top surface of the one-eight of the ball. On the side surfaces of the one-eight of the ball, symmetry conditions are imposed. The maximum normal contact stress y_{max}, the radius of the contact

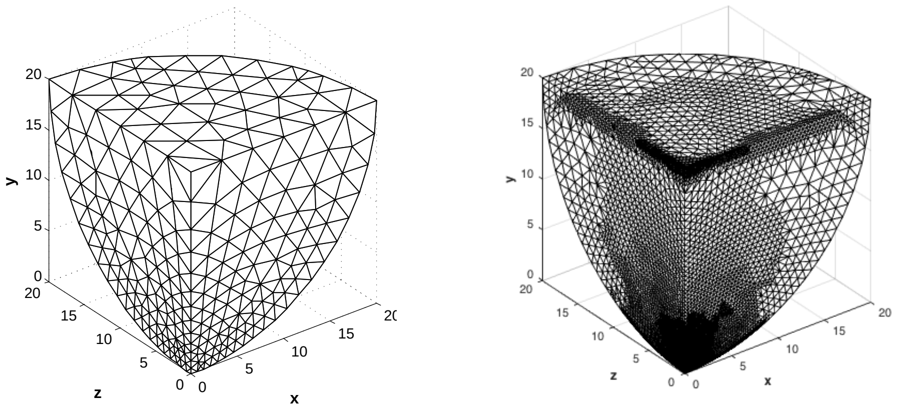

Fig. 1 Finite element discretization, (l) coarse mesh - 3993 elements, (r) finest mesh - 4120119 elements

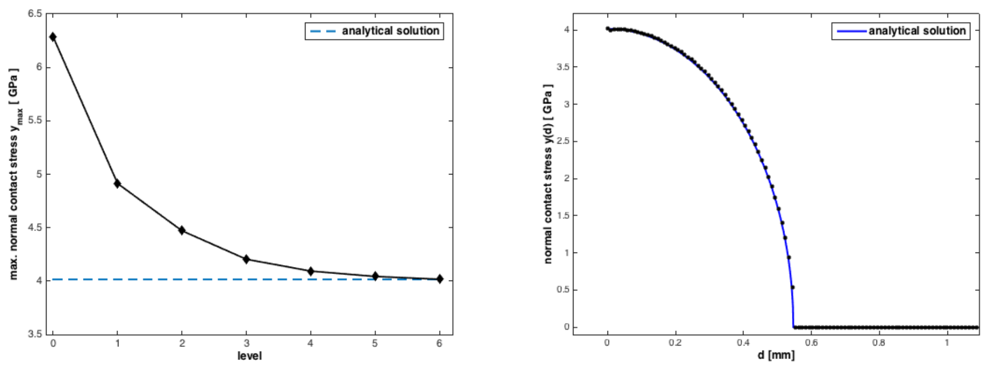

Fig. 2 (l): Maximal contact normal stresses on level 0,. . . , 6, (r): Normal contact stress distribution in the x-y plane

zone a and the normal contact stress distribution on the boundary $y(d)$ are given by

$$y_{max} = \frac{3pR^2}{2a^2}, \qquad a = R\sqrt[3]{\frac{3(1-\nu)^2 p\pi}{4E}}, \qquad y(d) = y_{max}\sqrt{1 - \frac{d^2}{a^2}}.$$

with the constant pressure p and the distance d from the midpoint of the contact zone. For the given set of parameters, we obtain $y_{max} \approx 4016$ MPa and $a \approx 0.5467$ mm.

The coarse grid can be seen in Figure 1(a) and has been generated using the commercial software package ABAQUS. An error estimator for linear elasticity and a uniform refinement of the potential contact zone are used for the adaptive refinement process. During the adaptive refinement process, new boundary nodes are moved to their positions on the surface of the ball.

We apply the semismooth Newton method and on each level the nonlinear discrete system is solved up to a given relative error of $\varepsilon_N = 10^{-6}$. The prolongated solution from level $l-1$ is used as initial solution on level l. For the regularization parameter we chose $\alpha = 10^{-8}$. The linear system in a Newton step is solved by multigrid preconditioned conjugate gradient method up to a given relative error of $\varepsilon_{pcg} = 10^{-2}, 10^{-4}$ and 10^{-8}, respectively, using the proposed multilevel preconditioner with a V-Cycle and two presmoothing and two postsmoothing steps.

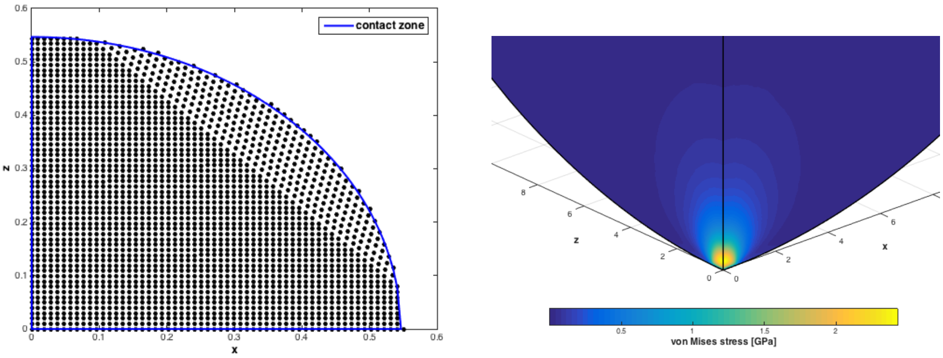

Fig. 3 (l): contact zone, (r): von Mises stress distribution

The convergence of y_{max} to the analytical solution is shown Figure 2(a). The approximated normal contact stresses along the distance d agree with the analytical solution (blue line), see Figure 2(b).

The active contact zone and the von Mises stress distribution are illustrated in Figure 3(a) and Figure 3(b). The numerical solution shows a very good agreement with the analytical circular shape of the contact zone.

Table 1 shows the required iterations of the semismooth Newton method and the average pcg iterations on each level. The pcg-multigrid method is able to solve the Newton system to a relative accuracy of $\varepsilon_{pcg} = 10^{-2}, 10^{-4}$ and 10^{-8} in about $4, 8$ and 14 pcg iterations on the finest grid. The slight dependence of the pcg iterations on the mesh size is in agreement with the results of Theorem 6, Corollary 1 and Theorem 7 that smaller values of $\alpha/\text{diam}(T)$ lead to better contraction factors of the multigrid scheme. It turns out that also for $\varepsilon_{pcg} = 10^{-2}$ the number of semismooth Newton iterations does not exceed 5 and therefore smaller values of $\varepsilon_{pcg} = 10^{-2}$ are not beneficial for the runtime.

			$\alpha = 10^{-8}, \varepsilon_{pcg} = 10^{-2}$		$\alpha = 10^{-8}, \varepsilon_{pcg} = 10^{-4}$		$\alpha = 10^{-8}, \varepsilon_{pcg} = 10^{-8}$	
l	n_l	$n_{C,l}$	it_{Newt}	avg-it_{pcg}	it_{Newt}	avg-it_{pcg}	it_{Newt}	avg-it_{pcg}
0	922	69	3	1.00	3	1.00	3	1.00
1	1793	245	6	2.33	4	4.00	4	7.50
2	4827	929	5	2.40	4	5.00	3	8.667
3	16456	3621	5	3.00	4	6.25	3	10.67
4	61711	14257	5	3.76	4	7.00	4	11.75
5	237300	56612	5	3.80	4	7.50	4	12.75
6	928152	225563	5	4.00	4	7.75	4	13.75

Table 1 Convergence history semismooth Newton method with pcg-multigrid solver: l: Level, n_l: number of grid points, $n_{C,l}$: number of contact nodes, it_{Newt}: number of semismooth Newton iterations, avg-it_{pcg} average number of pcg iterations per Newton iteration

We have tested the method also for other values of α and it turns out to be very robust. The iteration numbers are very similar for α ranging from 10^{-6} to 10^{-10}.

To illustrate the convergence of the pure multigrid method, which follows from Theorem 6, Corollary 1 and Theorem 7, we show in Table 2 the corresponding results if the pure multigrid solver is used instead of using it as preconditioner in pcg.

	$\alpha=10^{-8}, \varepsilon_{mg}=10^{-2}$				$\alpha=10^{-8}, \varepsilon_{mg}=10^{-4}$	
l	n_l	$n_{C,l}$	it_{Newt}	avg-it_{mg}	it_{Newt}	avg-it_{mg}
0	922	69	3	1.00	3	1.00
1	1793	245	6	6.67	5	13.65
2	4827	929	5	6.40	4	16.00
3	16456	3621	6	12.17	4	32.00
4	61711	14257	6	13.50	4	33.00
5	237300	56612	6	11.50	4	32.75
6	928152	225563	5	9.6	4	32.25

Table 2 Convergence history semismooth Newton method with multigrid solver: l: Level, n_l: number of grid points, $n_{C,l}$: number of contact nodes, it_{Newt}: number of semismooth Newton iterations, avg-it_{mg} average number of multigrid iterations per Newton iteration

A Appendix

A.1 Proof of Theorem 3

. We need the following Lemma:

Lemma 15 *Under Assumption (E), with the Lagrange function L and the regularized Lagrange function L_α defined in* (6) *and* (15)*, respectively, there holds:*
a) *$\bar{u}$ minimizes $L_\alpha(\cdot,\bar{y})$ on $\bar{\mathcal{U}}_\delta$ for all $\alpha \geq 0$.*
b) *u_α minimizes $L_\beta(\cdot,y_\alpha)$ on $\bar{\mathcal{U}}_\delta$ for all $\alpha > 0$, $\beta \geq 0$.*
c) *$\bar{y}$ maximizes $L(\bar{u},\cdot)$ on $L^2(\Gamma_C)_+$.*
d) *y_α maximizes $L_\alpha(u_\alpha,\cdot)$ on $L^2(\Gamma_C)_+$ for all $\alpha > 0$.*

Proof a): For all $u \in \bar{\mathcal{U}}_\delta$ we have

$$\begin{aligned} L_\beta(u,\bar{y}) - L_\beta(\bar{u},\bar{y}) &= J(u) - J(\bar{u}) + \langle \bar{y}, B(u-\bar{u})\rangle_{\Gamma_C} \\ &\geq \langle J'(\bar{u}) + B^*\bar{y}, u-\bar{u}\rangle + \tfrac{\sigma^2}{2}\|u-\bar{u}\|_{\mathbf{U}}^2 = \tfrac{\sigma^2}{2}\|u-\bar{u}\|_{\mathbf{U}}^2 \geq 0. \end{aligned}$$

b): For all $u \in \bar{\mathcal{U}}_\delta$, using (22), there holds:

$$\begin{aligned} L_\beta(u,y_\alpha) - L_\beta(u_\alpha,y_\alpha) &= J(u) - J(u_\alpha) + \langle y_\alpha, B(u-u_\alpha)\rangle_{\Gamma_C} \\ &\geq \langle J'(u_\alpha) + B^*y_\alpha, u-u_\alpha\rangle + \tfrac{\sigma^2}{2}\|u-u_\alpha\|_{\mathbf{U}}^2 \geq \tfrac{\sigma^2}{2}\|u-u_\alpha\|_{\mathbf{U}}^2 \geq 0. \end{aligned}$$

c): This follows directly from the linearity of $L_y(\bar{u},\cdot)$ and the complementarity condition: $\bar{y} \geq 0$, $L_y(\bar{u},\bar{y}) = B\bar{u}-\psi \leq 0$, $\langle L_y(\bar{u},\bar{y}),\bar{y}\rangle_{\Gamma_C} = \langle B\bar{u}-\psi,\bar{y}\rangle_{\Gamma_C} = 0$.
d): By (21), there holds $y_\alpha = \alpha^{-1}[r_\alpha]_+ \geq 0$ and

$$(L_\alpha)_y(u_\alpha,y_\alpha) = Bu_\alpha - \psi - \alpha(y_\alpha - \hat{y}) = r_\alpha - [r_\alpha]_+ \leq 0.$$

Hence, by the concavity of $L_\alpha(u_\alpha,\cdot)$, we get for all $y \in L^2(\Gamma_C)_+$:

$$\begin{aligned} L_\alpha(u_\alpha,y) - L_\alpha(u_\alpha,y_\alpha) &\leq \langle (L_\alpha)_y(u_\alpha,y_\alpha), y-y_\alpha\rangle_{\Gamma_C} \\ &\leq -\langle r_\alpha - [r_\alpha]_+, \alpha^{-1}[r_\alpha]_+\rangle_{\Gamma_C} = 0. \end{aligned}$$

Proof of Theorem 3:
We know from Theorem 1 that $u_\alpha \to \bar{u}$ in $\mathbf{U}$ as $\alpha \to 0^+$. From Theorem 2, we obtain the uniform boundedness of y_α in $L^2(\Gamma_C)$:

$$\|y_\alpha\|_{L^2} = \|\alpha^{-1}v_\alpha\|_{L^2} \leq \sqrt{2}\,(\|\bar{y}\|_{L^2} + \|\bar{y}-\hat{y}\|_{L^2}).$$

Hence any sequence (y_{α_k}), $\alpha_k \to 0^+$, has a subsequence $(y_{\alpha_{k_l}})$ that converges weakly in $L^2(\Gamma_C)$. Since $y_\alpha \to \bar{y}$ in V' as $\alpha \to 0^+$, see Theorem 1 c), there holds $y_{\alpha_{k_l}} \to \bar{y}$ weakly in $L^2(\Gamma_C)$. This shows $y_\alpha \to \bar{y}$ weakly in $L^2(\Gamma_C)$ as $\alpha \to 0^+$.
We continue by using the results of Lemma 15:

$$\begin{aligned} L(\bar{u},\bar{y}) &\geq L(\bar{u},y_\alpha) \geq L(u_\alpha,y_\alpha) = L_\alpha(u_\alpha,y_\alpha) + \frac{\alpha}{2}\|y_\alpha-\hat{y}\|_{L^2}^2 \\ &\geq L_\alpha(u_\alpha,\bar{y}) + \frac{\alpha}{2}\|y_\alpha-\hat{y}\|_{L^2}^2 = L(u_\alpha,\bar{y}) - \frac{\alpha}{2}\|\bar{y}-\hat{y}\|_{L^2}^2 + \frac{\alpha}{2}\|y_\alpha-\hat{y}\|_{L^2}^2 \\ &\geq L(\bar{u},\bar{y}) - \frac{\alpha}{2}\|\bar{y}-\hat{y}\|_{L^2}^2 + \frac{\alpha}{2}\|y_\alpha-\hat{y}\|_{L^2}^2. \end{aligned}$$

This shows $\|y_\alpha - \hat{y}\|_{L^2} \le \|\bar{y} - \hat{y}\|_{L^2}$ as well as

$$0 \le L(\bar{u}, \bar{y}) - L(u_\alpha, y_\alpha) \le \frac{\alpha}{2}\|\bar{y} - \hat{y}\|_{L^2}^2 - \frac{\alpha}{2}\|y_\alpha - \hat{y}\|_{L^2}^2. \tag{77}$$

From the weak convergence of $y_\alpha - \hat{y}$ to $\bar{y} - \hat{y}$ in $L^2(\Gamma_C)$ and the weak lower semicontinuity of $\|\cdot\|_{L^2}$ we obtain

$$\liminf_{\alpha \to 0^+} \|y_\alpha - \hat{y}\|_{L^2} \ge \|\bar{y} - \hat{y}\|_{L^2}.$$

Together with $\|y_\alpha - \hat{y}\|_{L^2} \le \|\bar{y} - \hat{y}\|_{L^2}$ this shows $\|y_\alpha - \hat{y}\|_{L^2} \to \|\bar{y} - \hat{y}\|_{L^2}$. Since $L^2(\Gamma_C)$ is a Hilbert space, this and weak convergence imply strong convergence $y_\alpha \to \bar{y}$ in $L^2(\Gamma_C)$ as $\alpha \to 0^+$.

With (19) and the optimality conditions (7), (8) we conclude:

$$\begin{aligned}
&L(\bar{u}, \bar{y}) - L(u_\alpha, y_\alpha) = J(\bar{u}) + \langle \bar{y}, B\bar{u} - \psi\rangle_{\Gamma_C} - J(u_\alpha) - \langle y_\alpha, Bu_\alpha - \psi\rangle_{\Gamma_C} \\
&= J(\bar{u}) - J(u_\alpha) + \langle B^* y_\alpha, \bar{u} - u_\alpha\rangle + \langle \bar{y} - y_\alpha, B\bar{u} - \psi\rangle_{\Gamma_C} \\
&\ge \langle J'(u_\alpha) + B^* y_\alpha, \bar{u} - u_\alpha\rangle + \frac{\sigma}{2}\|\bar{u} - u_\alpha\|_{\mathbf{U}_2}^2 - \langle y_\alpha, B\bar{u} - \psi\rangle_{\Gamma_C} \\
&\ge \frac{\sigma}{2}\|\bar{u} - u_\alpha\|_{\mathbf{U}_2}^2 - \langle y_\alpha, B\bar{u} - \psi\rangle_{\Gamma_C} \ge \frac{\sigma}{2}\|\bar{u} - u_\alpha\|_{\mathbf{U}_2}^2.
\end{aligned}$$

Thus, using (77), we obtain (28):

$$\begin{aligned}
\|\bar{u} - u_\alpha\|_{\mathbf{U}}^2 &\le \frac{2}{\sigma}(L(\bar{u}, \bar{y}) - L(u_\alpha, y_\alpha)) \\
&\le \frac{\alpha}{\sigma}(\|\bar{y} - \hat{y}\|_{L^2}^2 - \|y_\alpha - \hat{y}\|_{L^2}^2) = o(\alpha).
\end{aligned}$$

Since B is surjective, the assertion (29) follows from the open mapping theorem and

$$J'(\bar{u}) + B^*\bar{y} = 0, \quad J'(u_\alpha) + B^* y_\alpha = 0.$$

□

A.2 Proof of Lemma 2

Proof For arbitrary $(u, y) \in \mathbf{U} \times L^2(\Gamma_C)$ and $Z \in \partial H(u, y)$, Z has the form

$$Z = \begin{pmatrix} A & B^* \\ -\alpha^{-1} MB & I \end{pmatrix}, \quad M \in \partial S(\alpha^{-1} Bu + y_r - \alpha^{-1}\psi).$$

Now let $r = \binom{r_1}{r_2} \in \mathbf{U}' \times L^2(\Gamma_C)$ be arbitrary and consider the equation $Zs = r$, where $s = \binom{s_u}{s_y}$. Elimination with the second row yields the equivalent system

$$\begin{pmatrix} A + \alpha^{-1} B^* MB & 0 \\ -\alpha^{-1} MB & I \end{pmatrix} \binom{s_u}{s_y} = \binom{r_1 - B^* r_2}{r_2}. \tag{78}$$

By the definition of ∂S we have for all $u, v \in \mathbf{U}$, using $0 \le m \le 1$ on Γ_C):

$$\begin{aligned}
&\langle B^* MBu, u\rangle_{\mathbf{U}', \mathbf{U}} = (MBu, Bu)_{\Gamma_C} = \|m^{\frac{1}{2}}\, Bu\|_{L^2(\Gamma_C)}^2 \ge 0, \\
&\langle B^* MBu, v\rangle_{\mathbf{U}', \mathbf{U}} \le \|B\|_{\mathbf{U}, L^2(\Gamma_C)}^2 \|u\|_{\mathbf{U}} \|v\|_{\mathbf{U}}.
\end{aligned}$$

Therefore, the operator $A_\alpha := A + \alpha^{-1} B^* MB \in \mathcal{L}(\mathbf{U}, \mathbf{U}')$ is continuous and coercive with constants independent of M. Hence, A_α^{-1} exists and $\|A_\alpha^{-1}\|_{\mathbf{U}', \mathbf{U}} \le C_A$, since the coercivity constant of A is also valid for A_α. The bound $\|s\|_{\mathbf{U} \times L^2(\Gamma_C)} \le C_H \|r\|_{\mathbf{U}' \times L^2(\Gamma_C)}$ now follows easily. □

A.3 Proof of Lemma 3

Proof Let as above p_k, $1 \le k \le N_{K_h}$, be the nodes in $\mathcal{N}_{C,h}$. By (50) we have

$$(B_h\phi_{p,i})_k = n_{p_k}^T \phi_{p,i}(p_k)\, \|\phi_{p_k}\|_{L^2(\Gamma_{C,h})} = \delta_{pp_k}\delta_{i1}\, \|\phi_{p_k}\|_{L^2(\Gamma_{C,h})},$$

where $\delta_{pp_k} = 1$ if $p = p_k$ and $\delta_{pp_k} = 0$, otherwise. We thus obtain

$$\begin{aligned}(\phi_{p,i}, B_h^* M_h B_h \phi_{q,j})_{\mathbf{U}_h',\mathbf{U}_h} &= (B_h\phi_{p,i})^T M_h (B_h\phi_{q,j}) \\ &= \begin{cases} \|\phi_{p_k}\|^2_{L^2(\Gamma_{C,h})} & \text{if } \exists\, k \le N_{K_h} \colon p = q = p_k,\, m_k = 1,\, i = j = 3, \\ 0 & \text{otherwise.}\end{cases}\end{aligned} \tag{79}$$

Thus, $\mathbf{S}_\ell$ is diagonal with nonnegative entries. Hence (C1) holds with $C_B = 1$.

Now let $v_h \in \mathbf{U}_h$ be arbitrary and let $v_h|_{\Gamma_{C,h}} = \sum_{k=1}^{N_{K_h}} \sum_{i=1}^{3} v_{p_k,i}\phi_{p_k,i}$ be the corresponding basis representation. We now choose $w_h \in \mathbf{U}_h$ with

$$w_h|_{\Gamma_{C,h}} = \sum_{k:m_k=0} \sum_{i=1}^{3} v_{p_k,i}\phi_{p_k,i}|_{\Gamma_{C,h}} + \sum_{k:m_k=1} \sum_{i=1}^{2} v_{p_k,i}\phi_{p_k,i}|_{\Gamma_{C,h}}.$$

Then (79) yields $(B_h w_h)^T M_h (B_h w_h) = 0$ and

$$\begin{aligned}(B_h v_h)^T M_h (B_h v_h) &= \sum_{k:m_k=1} v_{p_k,3}^2\, \|\phi_{p_k}\|^2_{L^2(\Gamma_{C,h})} \\ &= \sum_{k=1}^{N_{K_h}} \sum_{i=1}^{3} \int_{\Gamma_{C,h}} (v_{p_k,i} - w_{p_k,i})^2 \phi_{p_k}^2 \, dS(x) \ge \frac{1}{\kappa_3}\, \|v_h - w_h\|^2_{L^2(\Gamma_{C,h})},\end{aligned}$$

where $\kappa_3 = 2$ for linear finite elements. □

A.4 Proof of Lemma 6

Proof One can show that

$$\|v\|^2_{L^2(T)} = \frac{\text{area}(T)}{20}(v_0^2 + v_1^2 + v_2^2 + v_3^2 + (v_0 + v_1 + v_2 + v_3)^2),$$

where v_0, v_1, v_2, v_3 are the values of v at the vertices of T. Now the proof can be obtained by considering the quadratic forms when one, two or three of the v_i are set to zero, see [2] for the 2D-case. □

A.5 Proof of Lemma 7

Proof Let $\hat{N}$ be the linear interpolate of the function N. Then $\hat{N}$ admits the same Lipschitz constant as N. We recall that Pv is the linear interpolate of Nv and also of the function $\hat{N}v$. Therefore, by standard approximation results in 3D, see e.g. [3], [6], there exist constants C_m only depending on the shape regularity of $\mathcal{T}_\ell$ such that

$$\|Pv - \hat{N}v\|_{H^m(T)} \le C_m \text{diam}(T)^{2-m} |\hat{N}v|_{H^2(T)}, \quad m = 0, 1, \ \forall\, T \in \mathcal{T}_\ell.$$

In particular,

$$\|Pv\|_{H^m(T)} \le \|\hat{N}v\|_{H^m(T)} + C_m \text{diam}(T)^{2-m} |\hat{N}v|_{H^2(T)}, \quad m = 0, 1.$$

Now $\|\hat{N}\|_{L^\infty(T)} \le 1$ and $|\hat{N}v|_{W^{2,\infty}(T)} = 0$, $|v|_{H^2(T)} = 0$ by linearity on T yield

$$\|\hat{N}v\|_{L^2(T)} \le \|v\|_{L^2(T)}, \ |\hat{N}v|_{H^1(T)} \le L_N \|v\|_{L^2(T)} + |v|_{H^1(T)}, \ |\hat{N}v|_{H^2(T)} \le 2L_N |v|_{H^1(T)},$$

Thus, we obtain

$$\|Pv\|_{H^1(T)} \le \|v\|_{H^1(T)} + L_N(\|v\|_{L^2(T)} + 2C_1 \text{diam}(T) |v|_{H^1(T)}).$$

Vice versa, v is the linear interpolate of $\hat{N}^T Pv$. Therefore, we can use the same arguments with v, $\hat{N}^T Pv$ instead of Pv, $\hat{N}v$ and obtain

$$\|v\|_{H^1(T)} \le \|Pv\|_{H^1(T)} + L_N(\|Pv\|_{L^2(T)} + 2C_1 \mathrm{diam}(T)|Pv|_{H^1(T)}).$$

Moreover, since $(Pv)_T$ is a linear function, Lemma 4 yields

$$\|Pv\|^2_{L^2(T)} \le \frac{|T|}{4}\sum_{i=1}^{4}\|N(p_i)v(p_i)\|_2^2 = \frac{|T|}{4}\sum_{i=1}^{4}\|v(p_i)\|_2^2 \le 5\,\|v\|^2_{L^2(T)}\,.$$

On the other hand

$$\|v\|^2_{L^2(T)} \le \frac{|T|}{4}\sum_{i=1}^{4}\|v(p_i)\|_2^2 = \frac{|T|}{4}\sum_{i=1}^{4}\|N(p_i)v(p_i)\|_2^2 \le 5\,\|Pv\|^2_{L^2(T)}\,.$$

Finally, since Pv is also the linear interpolate of Nv we have

$$\|Pv - Nv\|_{H^1(T)} \le C_1 \mathrm{diam}(T)\,\|Nv\|_{H^2(T)}\,.$$

Now we have similarly as above

$$\begin{aligned} &\|Nv\|_{L^2(T)} \le \|v\|_{L^2(T)}\,, \quad |Nv|_{H^1(T)} \le L_N\,\|v\|_{L^2(T)} + |v|_{H^1(T)}, \\ &|Nv|_{H^2(T)} \le 2L_N|v|_{H^1(T)} + |N|_{W^{2,\infty}(T)}\,\|v\|_{L^2(T)}\,. \end{aligned} \tag{80}$$

We obtain

$$\|Pv - Nv\|_{H^1(T)} \le C_1 \mathrm{diam}(T)((1+2L_N)\,\|v\|_{H^1(T)} + |N|_{W^{2,\infty}(T)}\,\|v\|_{L^2(T)}).$$

□

A.6 Proof of Lemma 9

Proof By the Poincaré inequality and Lemma 7 there exists a constant $C > 0$ with

$$|P\bar{Q}_0'v|^2_{H^1(\Omega)} \le (2 + C_P L_N^2)\left\|\bar{Q}_0'v\right\|^2_{H^1(\Omega)} \le C(2 + C_P L_N^2)|\bar{Q}_0'v|^2_{H^1(\Omega)}.$$

Now Lemma 7 and 8 yield

$$\begin{aligned} &|P\bar{Q}_0'v|^2_{H^1(\Omega)} + \sum_{k=1}^{\ell} 4^k \left\|P(\bar{Q}_k'v - \bar{Q}_{k-1}'v)\right\|_0^2 \le \\ &\le C(2 + C_P L_N^2)|\bar{Q}_0'v|^2_{H^1(\Omega)} + \sum_{k=1}^{\ell} 4^k 5\left\|\bar{Q}_k'v - \bar{Q}_{k-1}'v\right\|_0^2 \\ &\le C_1'C(5 + C_P L_N^2)|v|^2_{H^1(\Omega)} \le C_1'C^2(5 + C_P L_N^2)(2 + C_P L_N^2)|Pv|^2_{H^1(\Omega)}. \end{aligned}$$

□

A.7 Proof of Lemma 11

Proof For $l = 0$ this is trivial, since $B_0 = A_0$. Now let $l \ge 1$. We have to show (64) for $\mathbf{B}_l$ in (63). Since

$$\hat{\mathbf{B}}_l = (\mathbf{D}_l - \mathbf{L}_l)\mathbf{D}_l^{-1}(\mathbf{D}_l - \mathbf{L}_l)^T = \mathbf{A}_l + \mathbf{L}_l\mathbf{D}_l^{-1}\mathbf{L}_l^T,$$

(64) holds for $\hat{\mathbf{B}}_l$ instead of $\mathbf{B}_l$ (i.e., for $m = 1$). (64) is equivalent to $\mathbf{v}_l^T\mathbf{v}_l \le \mathbf{v}_l^T\mathbf{A}_l^{-1/2}\mathbf{B}_l\mathbf{A}_l^{-1/2}\mathbf{v}_l$ with the symmetric positive definite root of $\mathbf{A}_l^{-1}$. The latter is equivalent to $\sigma(\mathbf{A}_l^{1/2}\mathbf{B}_l^{-1}\mathbf{A}_l^{1/2}) \subset (0,1]$ with the spectrum $\sigma(\cdot)$. Since (64) holds for $\hat{\mathbf{B}}_l$ instead of $\mathbf{B}_l$, we know that $\sigma(\mathbf{A}_l^{1/2}\hat{\mathbf{B}}_l^{-1}\mathbf{A}_l^{1/2}) \subset (0,1]$. We have by (63)

$$\mathbf{A}_l^{\frac{1}{2}}\mathbf{B}_l^{-1}\mathbf{A}_l^{\frac{1}{2}} = I - \mathbf{A}_l^{\frac{1}{2}}(I - \hat{\mathbf{B}}_l^{-1}\mathbf{A}_l)^m\mathbf{A}_l^{-\frac{1}{2}} = \sum_{k=1}^{m}(-1)^{k+1}\binom{m}{k}(\mathbf{A}_l^{\frac{1}{2}}\hat{\mathbf{B}}_l^{-1}\mathbf{A}_l^{\frac{1}{2}})^k$$

and thus any eigenvector of $\mathbf{A}_l^{1/2}\hat{\mathbf{B}}_l^{-1}\mathbf{A}_l^{1/2}$ for an eigenvalue λ is also eigenvector of $\mathbf{A}_l^{1/2}\mathbf{B}_l^{-1}\mathbf{A}_l^{1/2}$ for the eigenvalue

$$\sum_{k=1}^{m}(-1)^{k+1}\binom{m}{k}\lambda^k = 1 - (1-\lambda)^m \in [\lambda, 1], \quad \text{since } \lambda \in (0,1].$$

This shows that $\sigma(\mathbf{A}_l^{1/2}\mathbf{B}_l^{-1}\mathbf{A}_l^{1/2}) \subset (0,1]$ and the proof is complete. □

References

1. L. BADEA, X. C. TAI, AND J. WANG, *Convergence rate analysis of a multiplicative Schwarz method for variational inequalities*, SIAM J. Numer. Anal., 41 (2003), pp. 1052–1073.
2. F. BORNEMANN AND H. YSERENTANT, *A basic norm equivalence for the theory of multilevel methods*, Numer. Math., 64 (1993), pp. 455–476.
3. S. C. BRENNER AND L. R. SCOTT, *The mathematical theory of finite element methods*, Springer, New York, 2008.
4. S. BRUNSSEN, F. SCHMID, M. SCHÄFER, AND B. WOHLMUTH, *A fast and robust iterative solver for nonlinear contact problems using a primal-dual active set strategy and algebraic multigrid*, Internat. J. Numer. Methods Engrg., 69 (2007), pp. 524–543.
5. P. W. CHRISTENSEN AND J.-S. PANG, *Frictional contact algorithms based on semismooth Newton methods*, in Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997), vol. 22 of Appl. Optim., Kluwer Acad. Publ., Dordrecht, 1999, pp. 81–116.
6. P. G. CIARLET, *The finite element method for elliptic problems*, vol. 40 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].
7. Z. DOSTÁL, *Optimal quadratic programming algorithms*, vol. 23 of Springer Optimization and Its Applications, Springer, New York, 2009.
8. F. FACCHINEI AND J.-S. PANG, *Finite-dimensional variational inequalities and complementarity problems I & II*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.
9. C. GRÄSER AND R. KORNHUBER, *Multigrid methods for obstacle problems*, J. Comput. Math., 27 (2009), pp. 1–44.
10. H. HERTZ, *Über die Berührung fester elastischer Körper*, Journal für Mathematik, 92 (1982), pp. 156–171.
11. M. HINTERMÜLLER, K. ITO, AND K. KUNISCH, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888 (electronic) (2003).
12. M. HINTERMÜLLER, V. A. KOVTUNENKO, AND K. KUNISCH, *Generalized Newton methods for crack problems with nonpenetration condition*, Numer. Methods Partial Differential Equations, 21 (2005), pp. 586–610.
13. I. HLAVÁČEK AND J. NEČAS, *On inequalities of Korn's type, parts I & II*, Arch. Rational Mech. Anal., 36 (1970), pp. 305–311, 312–334.
14. S. HÜEBER, M. MAIR, AND B. I. WOHLMUTH, *A priori error estimates and an inexact primal-dual active set strategy for linear and quadratic finite elements applied to multibody contact problems*, Appl. Numer. Math., 54 (2005), pp. 555–576.
15. S. HÜEBER, G. STADLER, AND B. I. WOHLMUTH, *A primal-dual active set algorithm for three-dimensional contact problems with Coulomb friction*, SIAM J. Sci. Comput., 30 (2008), pp. 572–596.
16. S. HÜEBER AND B. I. WOHLMUTH, *An optimal a priori error estimate for nonlinear multibody contact problems*, SIAM J. Numer. Anal., 43 (2005), pp. 156–173.
17. ———, *A primal-dual active set strategy for non-linear multibody contact problems*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 3147–3166.
18. K. ITO AND K. KUNISCH, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, 2008.
19. N. KIKUCHI AND J. T. ODEN, *Contact problems in elasticity: a study of variational inequalities and finite element methods*, vol. 8 of SIAM Studies in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1988.
20. D. KINDERLEHRER, *Remarks about Signorini's problem in linear elasticity*, Ann. Sc. Norm. Super. Cl. Sci. IV, 8 (1981), pp. 605–645.
21. R. KORNHUBER, *Monotone multigrid methods for elliptic variational inequalities. I*, Numer. Math., 69 (1994), pp. 167–184.
22. ———, *Monotone multigrid methods for elliptic variational inequalities. II*, Numer. Math., 72 (1996), pp. 481–499.
23. R. KORNHUBER AND R. KRAUSE, *Adaptive multigrid methods for Signorini's problem in linear elasticity*, Comput. Vis. Sci., 4 (2001), pp. 9–20.
24. R. KRAUSE, *From inexact active set strategies to nonlinear multigrid methods*, in Analysis and Simulation of Contact Problems, Springer, 2006, pp. 13–21.
25. J. MANDEL, *A multilevel iterative method for symmetric, positive definite linear complementarity problems*, Appl. Math. Opt., 11 (1984), pp. 77–95.
26. J. NEČAS, *On regularity of solutions to nonlinear variational inequalities for second-order elliptic systems*, Rend. Mat., 8 (1975), pp. 481–498.
27. A. POPP, M. W. GEE, AND W. A. WALL, *A finite deformation mortar contact formulation using a primal-dual active set strategy*, Internat. J. Numer. Methods Engrg., 79 (2009), pp. 1354–1391.
28. L. QI AND J. SUN, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.
29. G. STADLER, *Infinite-dimensional semi-smooth Newton and augmented Lagrangian methods for friction and contact problems in elasticity*, phd thesis, Universität Graz, 2004.

30. X. C. TAI, *Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities*, Numer. Math., 93 (2003), pp. 755–786.
31. S. TIMOSHENKO AND J. N. GOODIER, *Theory of Elasticity*, McGraw-Hill, New York, 1951.
32. M. ULBRICH, *Nonsmooth Newton-like methods for variational inequalities and constrained optimization problems in function spaces*, habilitation, Technische Universität München, 2001.
33. M. ULBRICH, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2002), pp. 805–842.
34. ———, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, SIAM, Philadelphia, 2011.
35. A. WEISS AND B. WOHLMUTH, *Multigrid methods for unresolved Dirichlet boundaries*, Preprint 2005/011, IANS, University of Stuttgart, 2005.
36. B. WOHLMUTH, *Variationally consistent discretization schemes and numerical algorithms for contact problems*, Acta Numer., 20 (2011), pp. 569–734.
37. B. I. WOHLMUTH AND R. H. KRAUSE, *Monotone multigrid methods on nonmatching grids for nonlinear multibody contact problems*, SIAM J. Sci. Comput., 25 (2003), pp. 324–347 (electronic).
38. H. YSERENTANT, *Old and new convergence proofs for multigrid methods*, Acta Numer., 2 (1993), pp. 285–326.