Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 203 (2007) 533–547
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Path-following and augmented Lagrangian methods for contact problems in linear elasticity☆

Georg Stadler

Center of Mathematics, University of Coimbra, Apartado 3008, 3001-454 Coimbra, Portugal

Received 20 December 2004; received in revised form 12 June 2005

Abstract

A certain regularization technique for contact problems leads to a family of problems that can be solved efficiently using infinite-dimensional semismooth Newton methods, or in this case equivalently, primal–dual active set strategies. We present two procedures that use a sequence of regularized problems to obtain the solution of the original contact problem: first-order augmented Lagrangian, and path-following methods. The first strategy is based on a multiplier-update, while path-following with respect to the regularization parameter uses theoretical results about the path-value function to increase the regularization parameter appropriately. Comprehensive numerical tests investigate the performance of the proposed strategies for both a 2D as well as a 3D contact problem.

MSC: 49M15; 74M15; 49M37; 65K05

Keywords: Contact problems; Path-following; Semismooth Newton methods; Active sets; Augmented Lagrangians; Primal–dual methods

1. Introduction

In contact problems (also known as Signorini problems), one is concerned with the deformation of an elastic body whose surface or boundary possibly hits a rigid foundation; in advance it is not known which part of the body's surface will be in contact with this obstacle. Theoretically, sound and efficient algorithms for the solution of these problems are still a very active field of research; we refer to the selected contributions [1,8,10,13,18–21,25] and the references given therein.

In this paper, we focus on the combination of semismooth Newton (SSN) methods [9,22] (which are in certain cases equivalent to primal–dual active set strategies [9]) with augmented Lagrangian and path-following methods in function space. The resulting algorithms are applied for the solution of contact problems in linear elasticity. Since we are interested in these methods in a function space framework, a regularization technique for the Signorini problem is introduced that allows an infinite-dimensional analysis. Aside from the theoretical interest, such a function space approach is also of significant practical importance, e.g., it is well known that the performance of a numerical algorithm is related to the structure of the infinite-dimensional problem. To be precise, regularity of Lagrange multipliers and smoothing properties of the involved operators can influence the performance of a numerical algorithm dealing with

☆ Support from the Austrian Science Fund FWF under SRC 03 "Optimization and Control" is gratefully acknowledged.
E-mail address: georgst@mat.uc.pt.

0377-0427/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2006.04.017

discretizations of continuous problems. Moreover, results such as mesh-independence heavily rely on the availability of the method in infinite dimensions. Only recently, these facts have led to the consequent investigation of numerical algorithms in function space (see, e.g., [9,11,15,22–24]).

The regularization chosen for the Signorini problem results in a family of problems having more favorable properties than the original one. It is motivated by augmented Lagrangians [3,14] and SSN methods [9,15,22]. Clearly, there exist numerous smoothing techniques for the Signorini and related problems (see, e.g., [12]). However, our choice results in first-order optimality conditions that can be written as semismooth equations in function space. This allows the application of an infinite-dimensional Newton-type method leading to an algorithm that converges locally with superlinear rate. Having this method for the solution of the regularized problem at hand, the question arises how to obtain a solution to the original problem. Here, we first consider the first-order augmented Lagrangian method which is an update strategy for the multiplier. The second approach relies on a path-following idea taken from the recent article [11]. Once the path-value function (or a sufficient good approximation of the path-value function) is available it can serve as basis for updating strategies of the path (or regularization) parameter. The proposed updating strategy has been inspired by path-following and interior point methods in finite-dimensional spaces [6,26].

Concerning the application of SSN (and primal–dual active set) methods for contact problems, an approach related to the one proposed in this work is followed in [10,13]. The techniques used in these papers are inherently finite-dimensional and, thus, can only be applied to discretizations of continuous problems. Restricting oneself to finite-dimensional problems allows to apply SSN methods without the need of any regularization. This gain opposes the advantages of an infinite-dimensional approach mentioned above.

Let us now give a brief outline of this paper. In Section 2, the contact problem in linear elasticity is stated and basic results are summarized. In Section 3, the regularized problem is introduced and a SSN algorithm for its solution is presented and analyzed. In the first part of Section 4, an augmented Lagrangian method is presented, while the second part of Section 4 is concerned with exact and inexact path-following strategies. In Section 5, the performance of our algorithms is investigated for two examples.

2. Problem statement

Let $\Omega \subset \mathbb{R}^n$, $n \geqslant 2$ be an open bounded domain with $\mathscr{C}^{1,1}$-boundary $\Gamma := \partial\Omega$ (see Fig. 1). We define the product space $\mathbf{H}^1(\Omega) := \prod_{j=1}^n H^1(\Omega)$ and analogously use bold notation for the product spaces $\mathbf{L}^2(\Omega)$, $\mathbf{H}^{1/2}(\Gamma)$, ... and their duals. The set of admissible deformations is

$$\mathbf{Y} := \{\boldsymbol{v} \in \mathbf{H}^1(\Omega) : \boldsymbol{\tau v} = 0 \text{ a.e. on } \Gamma_d\},$$

where $\Gamma_d \subset \Gamma$ is open, nonempty and $\boldsymbol{\tau} : \mathbf{H}^1(\Omega) \to \mathbf{H}^{1/2}(\Gamma)$ denotes the (componentwise) trace operator. Furthermore, we denote by $\Sigma := \text{int}(\Gamma \backslash \Gamma_d)$ the interior of $\Gamma \backslash \Gamma_d$, by $\Gamma_c \subset \Sigma$ the nonempty open region of possible contact and by $\Gamma_n := \text{int}(\Sigma \backslash \Gamma_c)$ the (possibly empty) set with given Neumann conditions. We assume that $\partial\Gamma_c, \partial\Sigma \subset \Gamma$ are smooth

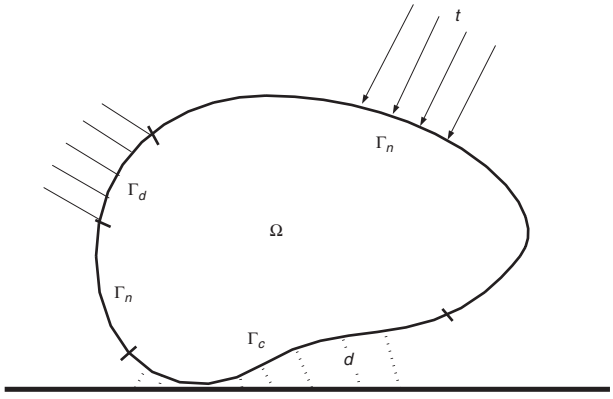

Fig. 1. Elastic body with rigid obstacle.

and, for simplicity, that $\overline{\Gamma}_c \subset \Sigma$. The latter assumption implies that

$$\mathbf{H}^{1/2}(\Gamma_c) = \{(\tau \boldsymbol{v})_{|\Gamma_c} : \boldsymbol{v} \in \mathbf{H}^1(\Omega)\}.$$

The elements in $\mathbf{H}^{1/2}(\Gamma_c)$ can be decomposed into their normal and tangential components [16], i.e., there exists the continuous linear normal trace mapping $\tau_N : \mathbf{Y} \to H^{1/2}(\Gamma_c)$ defined by $\tau_N \boldsymbol{v} := (\tau \boldsymbol{v})^\top \boldsymbol{v}$ for $v \in \mathbf{Y}$. Here, $\boldsymbol{v}$ denotes the unit outward normal vector along the boundary. The corresponding tangential trace mapping $\boldsymbol{\tau}_T : \mathbf{Y} \to \{\boldsymbol{v} \in \mathbf{H}^{1/2}(\Sigma) : \tau_N \boldsymbol{v} = 0\}$ is defined by $\boldsymbol{\tau}_T \boldsymbol{v} := \boldsymbol{v} - (\tau_N \boldsymbol{v})\boldsymbol{v}$.

As usual in linear elasticity we define for $\boldsymbol{y} \in \mathbf{Y}$ the components of the strain tensor $\underline{\boldsymbol{\epsilon}}(\boldsymbol{y}) \in (L^2(\Omega))^{n\times n}$ as

$$\underline{\boldsymbol{\epsilon}}(\boldsymbol{y}) := \tfrac{1}{2}\left(\nabla \boldsymbol{y} + (\nabla \boldsymbol{y})^\top\right).$$

The stress tensor $\underline{\boldsymbol{\sigma}}(\boldsymbol{y}) \in (L^2(\Omega))^{n\times n}$ is given by

$$\underline{\boldsymbol{\sigma}}(\boldsymbol{y}) := (\lambda \operatorname{tr}(\underline{\boldsymbol{\epsilon}}(\boldsymbol{y}))I + 2\mu \underline{\boldsymbol{\epsilon}}(\boldsymbol{y})),$$

where I denotes the $n \times n$-identity matrix and $\operatorname{tr}(\cdot)$ the trace of a matrix. The Lamé constants λ, μ are given by $\mu = E/(2(1+\nu))$ and $\lambda = E\nu/((1+\nu)(1-2\nu))$, with Young's modulus $E > 0$ and the Poisson ratio $\nu \in (0, \frac{1}{2})$.

For $\boldsymbol{y}, \boldsymbol{z} \in \mathbf{Y}$ we define the symmetric bilinear form

$$a(\boldsymbol{y}, \boldsymbol{z}) := \int_\Omega \underline{\boldsymbol{\epsilon}}(\boldsymbol{y}) : \underline{\boldsymbol{\sigma}}(\boldsymbol{z})\, \mathrm{d}x,$$

where ":" denotes the matrix L^2-product. For a given volume force $\boldsymbol{f} \in \mathbf{L}^2(\Omega)$ and surface tractions $\boldsymbol{t} \in \mathbf{L}^2(\Gamma_n)$ we define the continuous linear form $L : \mathbf{Y} \to \mathbb{R}$ by

$$L(\boldsymbol{y}) := \int_\Omega \boldsymbol{f}\boldsymbol{y}\, \mathrm{d}x + \int_{\Gamma_n} \boldsymbol{t}\,(\tau \boldsymbol{y})\, \mathrm{d}x \quad \text{for } \boldsymbol{y} \in \mathbf{Y}.$$

To model a possible gap between the elastic body and the rigid foundation we introduce $d \in H^{1/2}(\Gamma_c)$, with $d \geqslant 0$ pointwise almost everywhere. After these preparations the contact problem can be written as minimization of the potential energy over the set of admissible deformations, i.e., as

$$\min_{\tau_N \boldsymbol{y} \leqslant d} J(\boldsymbol{y}) := \frac{1}{2} a(\boldsymbol{y}, \boldsymbol{y}) - L(\boldsymbol{y}), \tag{$\mathscr{P}$}$$

or equivalently as elliptic variational inequality of the first kind [7]:

$$\begin{cases} \text{Find } \boldsymbol{y} \text{ with } \tau_N \boldsymbol{y} \leqslant d \text{ such that} \\ a(\boldsymbol{y}, \boldsymbol{z} - \boldsymbol{y}) \geqslant L(\boldsymbol{z} - \boldsymbol{y}) \text{ for all } \boldsymbol{z} \in \mathbf{Y} \text{ with } \tau_N \boldsymbol{z} \leqslant d. \end{cases} \tag{1}$$

It follows from the Korn inequality (see, e.g., [16]) that the problem ($\mathscr{P}$) or equivalently (1) admits a unique solution. From the surjectivity of the mapping τ_N we obtain that the solution to ($\mathscr{P}$) is characterized by the existence of $\bar{\lambda} \in H^{-1/2}(\Gamma_c)$ such that

$$a(\bar{\boldsymbol{y}}, \boldsymbol{z}) - L(\boldsymbol{z}) + \langle \bar{\lambda}, \tau_N \boldsymbol{z} \rangle_{\Gamma_c} = 0 \quad \text{for all } \boldsymbol{z} \in \mathbf{Y}, \tag{2a}$$

$$\langle \bar{\lambda}, \tau_N \boldsymbol{z} \rangle_{\Gamma_c} \leqslant 0 \quad \text{for all } \boldsymbol{z} \in \mathbf{Y} \text{ with } \tau_N \boldsymbol{z} \leqslant 0, \tag{2b}$$

$$\langle \bar{\lambda}, \tau_N \bar{\boldsymbol{y}} - d \rangle_{\Gamma_c} = 0. \tag{2c}$$

Above, $\langle \cdot, \cdot \rangle_{\Gamma_c}$ denotes the duality pairing between elements in $H^{-1/2}(\Gamma_c)$ and $H^{1/2}(\Gamma_c)$. Note that using (2a) it can be shown that the Lagrange multiplier $\bar{\lambda}$ coincides with the negative normal stress on the boundary [16].

Formally, the conditions $\tau_N \bar{\boldsymbol{y}} \leqslant d$, (2b) and (2c) can be rewritten as

$$\bar{\lambda} = \max(0, \bar{\lambda} + \sigma(\tau_N \bar{\boldsymbol{y}} - d)) \quad \text{for any fixed } \sigma > 0, \tag{3}$$

as one can verify by a short computation. However, in general $\bar{\lambda}$ is only an element in $H^{-1/2}(\Gamma_c)$. Thus, it is not pointwise almost everywhere defined and (3) has no meaning. Moreover, the pointwise max-functional appearing in (3) is not differentiable. This causes difficulties if one intends to develop Newton-like methods for the solution of the system consisting of (2a) and (3).

3. A family of regularized problems and a SSN method for their solution

Due to the problems mentioned at the end of the previous section we replace (3) by

$$\lambda = \max(0, \hat{\lambda} + \gamma(\tau_N \boldsymbol{y} - d)), \tag{4}$$

where $\hat{\lambda} \in L^2(\Gamma_c)$ and $\gamma > 0$ are given. For $\hat{\lambda} \equiv 0$, this results in a penalty-type regularization. The introduction of $\hat{\lambda}$ is motivated by augmented Lagrangians (see, e.g., [14,15] and the discussion in the next section). Moreover, for certain problems, $\hat{\lambda}$ can be used to control the feasibility of the solution of the regularized problems (see [15]). Another reason for the choice of (4) are SSN methods for the solution of complementarity systems in function space: considering in (4) the variable $\boldsymbol{y}$ as a function of $\lambda \in L^2(\Gamma_c)$ defined by means of (2a), we observe that $\tau_N \boldsymbol{y}$ is smoother than λ. In other words, the expression inside the max-functional is the image of λ under an affine mapping with a compact linear operator. Such a property is necessary for semismoothness of the max-function (see [9,22]). In the original problem (3) we cannot expect any smoothing of the expression inside the max-function due to the explicit appearance of λ.

We remark that the optimization problem corresponding to the smoothed system (2a) and (4) is

$$\min_{\boldsymbol{y} \in \mathbf{Y}} J(\gamma, \boldsymbol{y}) := \frac{1}{2} a(\boldsymbol{y}, \boldsymbol{y}) - L(\boldsymbol{y}) + \frac{1}{2\gamma} \| \max(0, \hat{\lambda} + \gamma(\tau_N \boldsymbol{y} - d)) \|^2_{\Gamma_c}. \tag{$\mathscr{P}_\gamma$}$$

Due to the uniform convexity of $J(\gamma, \cdot)$ for all $\gamma > 0$, the system (2a), (4) admits a unique solution. To highlight the dependence on γ, this solution is denoted by $\boldsymbol{y}_\gamma$ and the corresponding multiplier by λ_γ (the dependence on $\bar{\lambda}$ is neglected in the notation). In the next theorem, we show that $\boldsymbol{y}_\gamma$ converges to $\bar{\boldsymbol{y}}$ as $\gamma \to \infty$; for a proof we refer to Appendix A. Aside from the fact that this result justifies our regularization, it motivates a continuation procedure for the parameter γ (see Section 4.2).

Theorem 1. *For every $\hat{\lambda} \in L^2(\Gamma_c)$, the solutions $\boldsymbol{y}_\gamma$ of ($\mathscr{P}_\gamma$) converge to the solution $\bar{\boldsymbol{y}}$ of ($\mathscr{P}$) strongly in $\mathbf{Y}$, and the corresponding multipliers λ_γ converge to $\bar{\lambda}$ weakly in $H^{-1/2}(\Gamma_c)$ as $\gamma \to \infty$.*

We can now show how a SSN method can be utilized for the solution of ($\mathscr{P}_\gamma$), i.e., the solution of (2a) and (4). To be precise, we apply the method to the mapping $F : L^2(\Gamma_c) \to L^2(\Gamma_c)$ given by

$$F(\lambda) := \lambda - \max(0, \hat{\lambda} + \gamma(\tau_N \boldsymbol{y}(\lambda) - d)),$$

where $\boldsymbol{y}(\lambda) \in \mathbf{Y}$ is the unique solution $\boldsymbol{y}$ of (2a) for given $\lambda \in L^2(\Gamma_c)$. Note that $\tau_N \boldsymbol{y} \in H^{1/2}(\Gamma_c)$, which embeds continuously into $L^q(\Gamma_c)$ for every $q < \infty$ in the case $n = 2$ and for $q = 2(n-1)/(n-2)$ if $n \geqslant 3$. Thus, for each dimension n, $\tau_N \boldsymbol{y} \in L^q(\Gamma_c)$ for some $q > 2$ and we obtain the norm gap required for semismoothness of the max-function according to [9, Proposition 4.1]. Hence, we can apply a generalized Newton method for the solution of $F(\lambda) = 0$, where a generalized derivative of the max-function is given by

$$G_m(v)(x) = \begin{cases} 1 & \text{if } v(x) > 0, \\ 0 & \text{if } v(x) \leqslant 0. \end{cases}$$

This results in the following algorithm, in which $\chi_{\mathscr{S}}$ denotes the characteristic function for a set $\mathscr{S} \subset \Gamma_c$.

Algorithm (SSN).

(1) Choose $\boldsymbol{y}^0 \in \mathbf{Y}$ and set $k := 0$.
(2) Determine

$$\mathscr{A}^{k+1} = \{x \in \Gamma_c : \hat{\lambda} + \gamma(\tau_N \boldsymbol{y}^k - d) > 0\},$$
$$\mathscr{I}^{k+1} = \Gamma_c \backslash \mathscr{A}^{k+1}.$$

(3) If $k \geqslant 1$ and $\mathscr{A}^{k+1} = \mathscr{A}^k$ stop, else

(4) Solve

$$a(\mathbf{y}^{k+1}, z) - L(z) + (\hat{\lambda} + \gamma(\tau_N \mathbf{y}^{k+1} - d), \chi_{\mathscr{A}^{k+1}} \tau_N z)_{\Gamma_c} = 0$$
$$\text{for all } z \in \mathbf{Y}, \text{ set } \lambda^{k+1} = \begin{cases} \hat{\lambda} + \gamma(\tau_N \mathbf{y}^{k+1} - d) & \text{on } \mathscr{A}^{k+1}, \\ 0 & \text{on } \mathscr{I}^{k+1}, \end{cases} \tag{5}$$

and $k := k + 1$; Go to Step 2.

Note that the solution to (5) is unique, since (5) represents the necessary and sufficient optimality condition for the auxiliary problem

$$\min_{\mathbf{y} \in \mathbf{Y}} \frac{1}{2} a(\mathbf{y}, \mathbf{y}) - L(\mathbf{y}) + \frac{1}{2\gamma} \|\hat{\lambda} + \gamma(\tau_N \mathbf{y} - d)\|^2_{\mathscr{A}^{k+1}},$$

which is uniquely solvable. Observe that (SSN) obeys the form of an active set strategy. This is caused by the (generalized) derivative of the pointwise max-function appearing in (3). Thus, codes for the solution of elasticity problems not taking into account contact conditions can—relatively easily—be adopted for the solution of contact problems. Note that a similar strategy for finite-dimensional problems is investigated in [1]. Next, we summarize properties of (SSN). For the proofs and related results we refer to [15,21].

Lemma 2. *If Algorithm* (SSN) *stops, the last iterate* $\mathbf{y}^k$ *is the solutions to* $(\mathscr{P}_\gamma)$, *i.e.*, $(\mathbf{y}^k, \lambda^k)$ *solves* (2a) *and* (4).

Theorem 3. *Let* $\hat{\lambda} \in L^2(\Gamma_c)$ *and* $\gamma > 0$ *be arbitrary given. Provided that* $\|\lambda_0 - \lambda_\gamma\|_{\Gamma_c}$ *is sufficiently small, the iterates* $(\mathbf{y}^k, \lambda^k)$ *of* (SSN) *converge to* $(\mathbf{y}_\gamma, \lambda_\gamma)$ *superlinearly in* $\mathbf{Y} \times L^2(\Gamma_c)$.

4. Towards the solution of the original contact problem

In this section we present two approaches in function space that utilize a sequence of regularized problems $(\mathscr{P}_\gamma)$ for the solution of the original contact problem $(\mathscr{P})$.

4.1. Augmented Lagrangian methods

Augmented Lagrangian methods apply for the solution of contact problems provided the solution multiplier $\bar{\lambda}$ is in $L^2(\Gamma_c)$. Following regularity results for the Signorini problem (see [17, Theorem 2.2]), a sufficient condition for $\bar{\lambda} \in L^2(\Gamma_c)$ is that the active set at the solution, i.e.,

$$\mathscr{A}(\bar{\mathbf{y}}) = \{x \in \Gamma_c : \tau_N \bar{\mathbf{y}} - d = 0 \text{ a.e.}\}$$

is strictly contained in Γ_c, that is

$$\overline{\mathscr{A}(\bar{\mathbf{y}})} \subset \Gamma_c. \tag{A}$$

This condition holds for many contact problems. It can often be verified a priori from considering the geometry of the problem. For the rest of this section we assume that (A) holds.

The first-order augmented Lagrangian method is, such as the Uzawa algorithm, an update strategy for the multiplier in $(\mathscr{P})$. It can be considered as an implicit version of the Uzawa method (see [14]). Its main advantage, compared to the latter strategy, is its unconditional convergence for all penalty (or regularization) parameters $\gamma > 0$ while the Uzawa algorithm only converges for sufficiently small $\gamma > 0$ possibly leading to extremely slow convergence. However, the drawback of the augmented Lagrangian method is that in every iteration step it requires to solve a nonlinear problem compared to the linear problem in every iteration of the Uzawa algorithm. Since this nonlinear problem is exactly of the form $(\mathscr{P}_\gamma)$, we can apply (SSN) for its solution. The whole method is specified next.

Algorithm (ALM).

(1) Choose $\lambda^0 \in L^2(\Gamma_c)$ and set $l := 0$.
(2) Choose $\gamma^{l+1} > 0$ and solve $(\mathscr{P}_\gamma)$ with $\hat{\lambda} := \lambda^l$, i.e., determine $(\mathbf{y}^{l+1}, \lambda^{l+1})$ such that

$$a(\mathbf{y}^{l+1}, z) - L(z) + (\lambda^{l+1}, \tau_N z)_{\Gamma_c} = 0 \quad \text{for all } z \in \mathbf{Y},$$
$$\lambda^{l+1} - \max(0, \lambda^l + \gamma^{l+1}(\tau_N \mathbf{y}^{l+1} - d)) = 0.$$

(3) Update $l := l + 1$ and go to Step 2.

In the next theorem global convergence of (ALM) is stated. For the proof we refer to Appendix B.

Theorem 4. *For every choice of parameters* $0 < \gamma^0 \leqslant \gamma^1 \leqslant \gamma^2 \leqslant \cdots$ *the iterates* λ^l *of* (ALM) *converge weakly to* $\bar{\lambda}$ *in* $L^2(\Gamma_c)$. *Furthermore, the corresponding iterates* $\mathbf{y}^l$ *converge strongly to* $\bar{\mathbf{y}}$ *in* $\mathbf{Y}$.

Let us comment on the role of the parameters γ^l in (ALM). Due to possible difficulties in the numerical treatment of $(\mathscr{P}_\gamma)$ for large penalty parameters γ one may start with a moderate value for γ in Step 2 of (ALM) and increase this value during the iteration. However, note that λ^l converges to $\bar{\lambda}$ without the requirement that γ tends to infinity such as in Theorem 1 where $\hat{\lambda}$ is kept fixed.

4.2. Path-following methods

Following Theorem 1, the solutions $\mathbf{y}_\gamma$ of $(\mathscr{P}_\gamma)$ converge to $\bar{\mathbf{y}}$ as $\gamma \to \infty$. Starting immediately with a huge value for γ may lead to a badly conditioned problem that is difficult to solve. Therefore, it appears advantageous to use a continuation procedure with respect to γ. Such a procedure has already been applied for obstacle and related problems [15,21], where the adaptation of γ was heuristic. In the recent article [11], a path-following method has been developed that allows appropriate steering of this parameter. Aside from the infinite-dimensional theory devised in [11], promising numerical experiments for obstacle problems are presented. Furthermore, both a feasible as well as an infeasible regularization approach are presented. However, to be able to divide the discussion into these two cases it is required that the involved operator satisfies a maximum principle. This is, in general, not the case for the operator in linear elasticity.

We briefly summarize the approach for $\hat{\lambda} \equiv 0$ (called "infeasible method" in [11]). Although our operators do not satisfy a maximum principle, we numerically experience a very similar behavior as for obstacle problems. To be precise, we observe that, for $\hat{\lambda} \equiv 0$, the solutions of the regularized problem $(\mathscr{P}_\gamma)$ are typically *infeasible*.

We set $\hat{\lambda} \equiv 0$ and consider the primal–dual path $\gamma \to (\mathbf{y}_\gamma, \lambda_\gamma)$ with $\gamma \in [0, \infty)$. Similarly, as in [11] for obstacle problems one can show that this path is bounded in $\mathbf{H}^1(\Omega) \times H^{-1/2}(\Gamma_c)$. We introduce the primal–dual path-value functional by

$$\gamma \to V(\gamma) := J(\gamma, \mathbf{y}_\gamma), \tag{6}$$

with $J(\cdot, \cdot)$ as defined in $(\mathscr{P}_\gamma)$. Now, we focus on an appropriate model function $m(\gamma)$ approximating $V(\cdot)$. In deriving such a model, knowledge about the derivatives of V turns out to be useful. As in [11], it can be shown that $V(\cdot)$ is differentiable with first derivative

$$\dot{V}(\gamma) = \frac{1}{2} \int_{\Gamma_c} |\max(0, \tau_N \mathbf{y}_\gamma - d)|^2 \, \mathrm{d}x = \frac{1}{2\gamma^2} \int_{\Gamma_c} |\lambda_\gamma|^2 \, \mathrm{d}x. \tag{7}$$

Remarkably, the derivative $\dot{V}(\gamma)$ is characterized without recourse to $\dot{\mathbf{y}}_\gamma$. Thus, having calculated $(\mathbf{y}_\gamma, \lambda_\gamma)$ not only $V(\gamma)$ but also $\dot{V}(\gamma)$ is available. Taking into account second derivative information it turns out (see [11]) that an appropriate model function for $V(\cdot)$ is

$$m(\gamma) = C_1 - \frac{C_2}{E + \gamma}, \tag{8}$$

with $C_1 \in \mathbb{R}$, $C_2 \geqslant 0$ and $E > 0$. In [11] further evidence for the choice of (8) as model function is given.

To adjust the parameters C_1, C_2 and E we choose $\gamma^\star \in (0, \infty)$ and compute the values for $V(0)$ and $V(\gamma^\star)$. Due to (7), the value for $\dot{V}(\gamma^\star)$ is available as well. Using $V(0)$, $V(\gamma^\star)$ and $\dot{V}(\gamma^\star)$ we can fix the parameters in (8) by setting $m(0) = V(0)$, $m(\gamma^\star) = V(\gamma^\star)$ and $\dot{m}(\gamma^\star) = \dot{V}(\gamma^\star)$, which results in the following equations:

$$\begin{aligned} E &= (\gamma^\star)^2 \dot{V}(\gamma^\star)(V(\gamma^\star) - V(0) - \gamma^\star \dot{V}(\gamma^\star))^{-1}, \\ C_2 &= (\gamma^\star)^{-1} E(E + \gamma^\star)(V(\gamma^\star) - V(0)), \\ C_1 &= V(0) + C_2 E^{-1}. \end{aligned} \tag{9}$$

Clearly, if more data points than $V(0)$ and $V(\gamma^\star)$, $\dot{V}(\gamma^\star)$ for a certain $\gamma^\star$ are available one can use, e.g., a least-squares approach to fit the model parameters C_1, C_2 and E. However, our numerical tests show that the model function calculated from the data $V(0)$ and $V(\gamma^\star)$, $\dot{V}(\gamma^\star)$ already gives a remarkable correspondence with the path-value functional. Thus, for every choice of γ^k we obtain a slightly different model function denoted by $m^k(\gamma) = C_1^k - C_2^k/(E^k + \gamma)$.

Next, we describe how knowledge about the primal–dual path-value functional can be used in numerical practice. We introduce the notation $\bar{V} := \lim_{\gamma \to \infty} V(\gamma)$ and assume given a sequence τ^k with $\tau^k \in (0, 1)$ and $\tau^k \downarrow 0$. Taking into account the superlinear convergence of (SSN) motivates the following strategy [11]: provided $V(\gamma^k)$ is available, the new path parameter γ^{k+1} should ideally satisfy

$$|\bar{V} - V(\gamma^{k+1})| \leqslant \tau^k |\bar{V} - V(\gamma^k)|. \tag{10}$$

Since $\bar{V}$ and $V(\gamma^{k+1})$ are unknown, we use our model m^k to obtain $\bar{V} \approx C_1^k$ and $V(\gamma^{k+1}) \approx m^k(\gamma^{k+1})$. Hence, (10) is replaced by

$$|C_1^k - m^k(\gamma^{k+1})| \leqslant \tau^k |C_1^k - V(\gamma^k)| =: \beta^k. \tag{11}$$

Solving the equation $C_1^k - m^k(\gamma^{k+1}) = \beta^k$ yields

$$\gamma^{k+1} = C_2^k/\beta^k - E^k. \tag{12}$$

Summarizing, we obtain the following algorithm for exact path-following:

Algorithm (PF).

(1) Derive $V(0)$, choose $\gamma^0 > 0$ and set $l := 0$.
(2) Solve $(\mathscr{P}_\gamma)$ with $\hat{\lambda} := 0$ and $\gamma := \gamma^l$, i.e., determine $(\mathbf{y}^{l+1}, \lambda^{l+1})$ such that

$$\begin{aligned} &a(\mathbf{y}^{l+1}, z) - L(z) + (\lambda^{l+1}, \tau_N z)_{\Gamma_c} = 0 \quad \text{for all } z \in \mathbf{Y}, \\ &\lambda^{l+1} - \max(0, \gamma^l(\tau_N \mathbf{y}^{l+1} - d)) = 0. \end{aligned}$$

(3) Use the values $V(0)$, $V(\gamma^l)$, $\dot{V}(\gamma^l)$ to derive the model function $m^k(\cdot)$. Calculate the new path parameter γ^{l+1} according to (12).
(4) Update $l := l + 1$ and go to Step 2.

Of course, in the above algorithm the previous iterate is used as initialization for (SSN) when solving the system in Step 2 of (PF). So far, we have a simple rule for an automatic update of the path parameter γ if every auxiliary problem $(\mathscr{P}_\gamma)$ is solved exactly. From finite-dimensional interior point and path-following methods one can expect that the overall number of iterations is reduced if the auxiliary problems are only solved approximately. That is, not for every γ^k the corresponding point on the primal–dual path is calculated exactly, which leads to *inexact path-following methods*. Typically, it is required that the primal–dual iterates stay within a certain neighborhood of the path with the goal to keep the number of iterations low while still obtaining reliable convergence. In order to define an appropriate neighborhood of the primal–dual path we introduce the residuals

$$r_\gamma^1(\mathbf{y}, \lambda) = \sup_{z \in \mathbf{H}^1(\Omega)} \frac{1}{\|z\|_{\mathbf{H}^1}} \{a(\mathbf{y}, z) - L(z) + (\lambda, \tau_N z)\}, \tag{13a}$$

$$r_\gamma^2(\mathbf{y}, \lambda) = \| \max(0, \gamma(\tau_N \mathbf{y} - d))\|_{L^2}. \tag{13b}$$

Note that $r_\gamma^1(\boldsymbol{y}, \lambda)$ represents the norm in the dual of $\mathbf{H}^1(\Omega)$. For a fixed parameter τ we define the following neighborhood of the primal–dual path:

$$\mathcal{N}_\gamma := \{(\boldsymbol{y}, \lambda) \in \mathbf{H}^1(\Omega) \times L^2(\Omega) : \|r_\gamma^1(\boldsymbol{y}, \lambda), r_\gamma^2(\boldsymbol{y}, \lambda)\|_2 \leqslant \tau/\sqrt{\gamma}\}. \tag{14}$$

In our inexact path-following method we require the iterates to stay within $\mathcal{N}_\gamma$. The question arises about an update strategy for the path-parameter γ. Increasing γ too slow causes that we follow the path very closely leading to slow convergence. Too aggressive γ-updates lead to points which are far away from the path and many SSN iterations are necessary to produce iterates lying within $\mathcal{N}_\gamma$. As in [11] we introduce the *primal infeasibility measure* ρ_{F}, and the *complementarity measure* ρ_{C} for the $(k+1)$th iterate as follows:

$$\rho_{\mathrm{F}}^{k+1} := \int_{\Gamma_c} \max(0, \tau_N \boldsymbol{y}^{k+1} - d)\, \mathrm{d}x,$$
$$\rho_{\mathrm{C}}^{k+1} := \int_{\mathcal{I}^{k+1}} \max(0, \tau_N \boldsymbol{y}^{k+1} - d)\, \mathrm{d}x + \int_{\mathcal{A}^{k+1}} -\min(0, \tau_N \boldsymbol{y}^{k+1} - d)\, \mathrm{d}x.$$

These measures are used in the following rule for choosing the new path parameter γ^{k+1}:

$$\gamma^{k+1} \geqslant \max\left(\gamma^k \max(\tau_1, \rho_{\mathrm{F}}^{k+1}/\rho_{\mathrm{C}}^{k+1}), 1/(\max(\rho_{\mathrm{F}}^{k+1}, \rho_{\mathrm{C}}^{k+1}))^q\right), \tag{15}$$

with $\tau_1 > 1$ and $q \geqslant 1$. To safeguard the γ-updates in order to avoid too large changes of γ we use our model functions m^k. Namely, we reduce the value of γ^{k+1} as long as γ^{k+1} is much larger than γ^k and until

$$|t^k(\gamma^{k+1}) - m^k(\gamma^{k+1})| \leqslant \tau_3 |J(\gamma^k, \boldsymbol{y}^{k+1}) - J(\gamma^{k-1}, \boldsymbol{y}^k)|, \tag{16}$$

with $\tau_3 > 0$ and $t^k(\gamma) = J(\gamma^k, \boldsymbol{y}^{k+1}) + (\partial J/\partial\gamma)(\gamma^k, \boldsymbol{y}^{k+1})(\gamma - \gamma^k)$. Note that $m^k(\gamma^k) = J(\gamma^k, \boldsymbol{y}^{k+1})$ due to the definition of our model functions. Observe that the right-hand side in (16) is independent of γ^{k+1} while the left-hand side tends to 0 as γ^{k+1} goes to γ^k. Thus, if (16) is not satisfied immediately, it is likely to be satisfied already after a few backtracking-like steps. We remark that for small γ there might be a big difference between the tangent t^k and the model m^k. However, the right-hand side in (16) is large as well because we expect a relatively large change in the function value. For large γ both sides in (16) tend to be small.

5. Numerical results

Here we present results of our numerical tests for a 2D as well as a 3D contact problem. In the first subsection, we introduce our test examples and discuss the implementations before we report on results obtained for the algorithms presented in the previous sections.

5.1. Implementation and presentation of the examples

Our implementation adopts the MATLAB-code for elasticity problems published in [2]. We use linear finite elements to discretize our problems and solve the arising linear systems exactly. The SSN method is always initialized with the solution of the unconstrained problem (i.e., the solution of (2a) with $\lambda = 0$). As initialization for the first-order augmented Lagrangian method we always use $\lambda^0 = 0$. Unless otherwise specified we stop the algorithms (ALM) and (PF) if the L^2-norm of the nonlinear residual drops below 10^{-10}. Now, we present our test examples.

Example 1. In Fig. 2, the geometry used for this example is sketched. For reasons of graphical presentation the gap function d is multiplied by a factor of 20. The exact data are as follows: $\Omega = [0, 3] \times [0, 1]$, $\Gamma_d = \{0\} \times (0, 1)$ and $\Gamma_c = (0, 3) \times \{0\}$. Furthermore, $\boldsymbol{f} = 0$ and $\boldsymbol{t} = (0, -2)^\top$ on $\{3\} \times (0, 1)$, where $\boldsymbol{t} \equiv 0$ on the rest of Γ_n. The distance to the obstacle is given by $d(x_1) = 0.003(x_1 - 1.5)^2 + 0.001$. We use materials described by $E = 5000$ and various values for the Poisson ratio $\nu \in (0, \frac{1}{2})$.

Example 2. In this example a 3D rectangular elastic body is pressed onto a solid hemisphere. The geometry of the body is given by $\Omega = [-0.5, 0.5] \times [-1, 1] \times [-0.5, 0.5]$ and the obstacle is a half-ball with a radius of 0.5 and

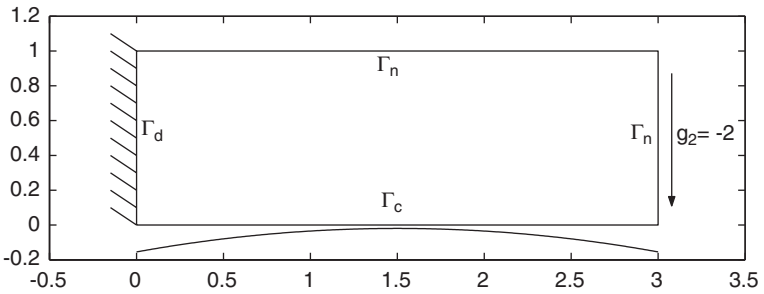

Fig. 2. Geometry for Example 1.

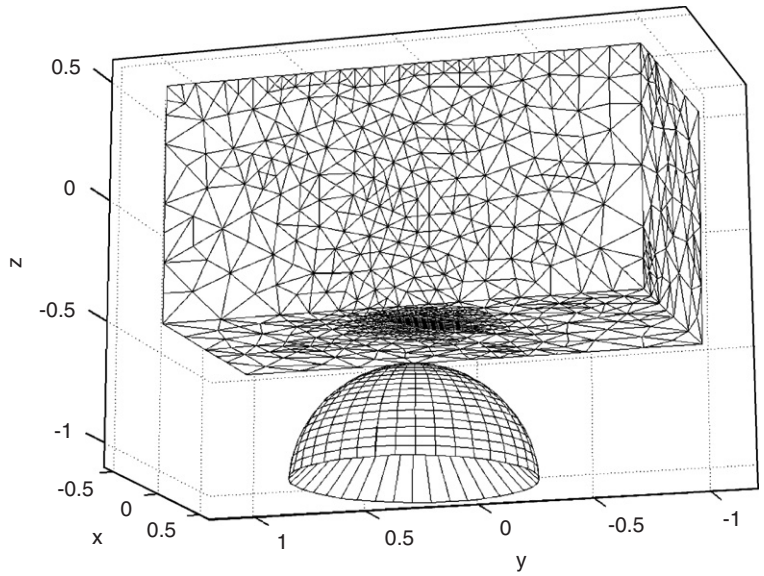

Fig. 3. Geometry and obstacle for Example 2.

midpoint $(-0.3, 0, -1)^\top$. For the upper surface $(-0.5, 0.5) \times (-1, 1) \times \{0.5\}$ we prescribe the Dirichlet condition $\boldsymbol{y} = (0, 0, -0.2)^\top$. The side surfaces are assumed to be traction-free, i.e., $\boldsymbol{t} \equiv 0$. The contact surface is the lower face given by $(-0.5, 0.5) \times (-1, 1) \times \{-0.5\}$. The material parameters are $E = 1e6$ and $\nu = 0.45$. The problem is discretized with tetrahedral elements. The initial mesh is generated using FEMLAB's [5] mesh generator. This initial mesh is adaptively refined using an averaging a posteriori error estimator [4]. The resulting locally refined mesh and the rigid obstacle are shown in Fig. 3, where for reasons of graphical representation the obstacle has been moved slightly downwards (i.e., in negative z-direction). The mesh consists of 31 453 elements and has automatically been refined in and around the contact zone.

5.2. Results for Example 1

First, we briefly discuss (SSN) for the solution of the regularized problem ($\mathscr{P}_\gamma$). For all choices of ν and regularization parameters the algorithm converges after a few iterations. Moreover, we observe monotonicity in the sense that the size of the active set decreases in every iteration. Note that monotonicity of iterates can be proved rigorously if the operator satisfies a maximum principle [15].

The deformed mesh for $\gamma = 10^{10}$ and $\nu = 0.49$ is shown in Fig. 4, where the displacement $\boldsymbol{y}_\gamma$ as well as the gap-function d are magnified by the factor 20. The gray values visualize the elastic shear energy density [2]. In Table 1, the numbers of iterations for $\nu = 0.4, 0.49, 0.499, 0.4999$ and various values of γ on a mesh of 120×40 elements are shown. We observe that for ν-values very close to the threshold $\frac{1}{2}$ the number of iterations increase significantly as the

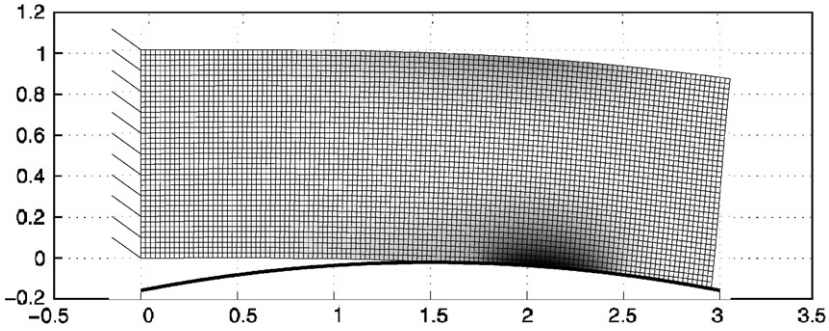

Fig. 4. Example 1: deformed mesh, gray tones visualize the elastic shear energy density, $\nu = 0.49$, $\gamma = 10^{10}$.

Table 1
Example 1: number of iterations for different values of ν and γ

ν	γ			
	10^2	10^4	10^6	10^{10}
0.4	4	6	7	7
0.49	3	7	8	8
0.499	4	8	12	12
0.4999	4	9	25	29

regularization parameter γ becomes large. Especially, for these cases either an augmented Lagrangian or a continuation procedure with respect to γ appears worth using.

Our tests for the first-order augmented Lagrangian method show a relatively slow convergence behavior if the parameter γ is kept fixed. For example, using $\gamma^l = 10^3$ for all l, overall 27 iterations for $\nu = 0.4$ are necessary. Increasing γ simultaneously with the update of the multiplier speeds up the convergence significantly. Clearly, such a strategy is related to a pure continuation strategy and thus to the path-following methods we turn to next.

First, we numerically investigate the quality of the model function (8) by comparing it to the primal–dual value function. To calculate the unknown parameters in $m(\cdot)$ we use $V(0)$, $V(\gamma^\star)$ and $\dot{V}(\gamma^\star)$, where we derive the model functions for $\gamma^\star = 10^2, 10^3, 10^5$. The results for $\nu = 0.4$ and 0.4999 are shown in Fig. 5. We observe that the model functions $m(\cdot)$ reflect the behavior of $V(\cdot)$ very well, especially in a neighborhood around the γ-value used to derive the parameters E, C_1, C_2 according to (9). These excellent approximation properties of our model functions can be used to obtain an automatic update of the path parameter γ as described in Section 4.2. We shall now present our numerical results for both path-following as well as inexact path-following.

To initialize the model function we first solve the problem for $\gamma := \gamma^0 = 0$, i.e., we solve the unrestricted problem. Since the corresponding optimality system is linear only one system solve is necessary. As second step to derive a model function we solve the problem for some $\gamma^1 > 0$. Here, $\gamma^1 = 10^2$ is chosen. Using (9) in (8) we obtain the model function $m^1(\cdot)$. As described at the end of Section 4.2, $m^1(\cdot)$ can be used to derive the new path parameter γ^2, where we choose $\tau^k := 10^{-k}$. The resulting data are summarized in Table 2 for the two values $\nu = 0.4$ and 0.4999.

We also tested less aggressive γ-updates, e.g., using $\tau^k = 0.2^k$. This leads to fewer inner iterations (i.e., fewer iterations of (SSN)). However, the number of outer iterations, i.e., steps where γ is updated increases. For $\tau_k = 0.2^k$ this leads to six and seven outer, and to overall 17 and 18 inner iterations for $\nu = 0.4$ and 0.4999, respectively.

We now turn to tests for inexact path-following. Here, iterates are not required to lie on the primal–dual path exactly, but one stops the inner iteration as soon as a point in the neighborhood $\mathcal{N}_\gamma$ (for its definition see (14)) is found. We choose the parameters $\tau_1 = 5$, $q = 1.5$ and $\tau = 10^6$. In Fig. 6, we show the primal–dual path-value function together with the value function evaluated at the iterates. Note that a logarithmic scale for the γ-values is used. Since $\mathcal{N}_\gamma$ is rather large for γ^1 and γ^2 only one (SSN)-iteration is needed for these values of the path parameter. For γ^3 and γ^4 three and two inner iterations are needed, respectively, to find a point that is sufficiently close to the path. The inexact strategy requires overall 11 and 15 iterations for $\nu = 0.4$ and 0.4999, respectively.

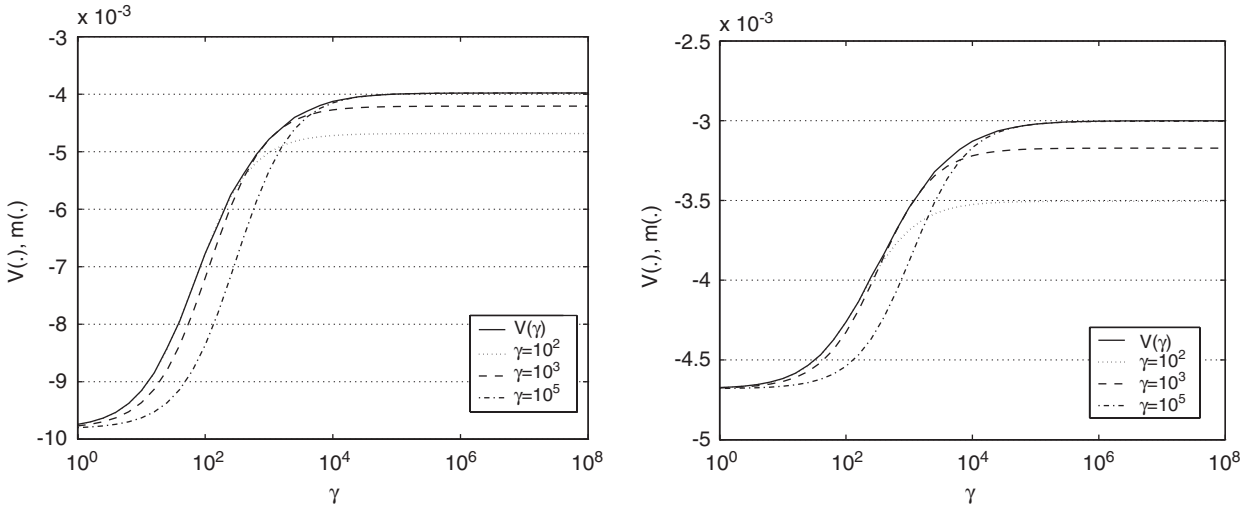

Fig. 5. Example 1: primal–dual value function $V(\cdot)$ versus model functions $m(\cdot)$ for $\gamma = 10^2, 10^3, 10^5$ for $v = 0.4$ (left plot) and $v = 0.4999$ (right plot).

Table 2
Example 1: number of inner iterations (#(SSN)-iter) and value for path-parameter γ per path-following step (path-iter.) for $v = 0.4$ and 0.4999

	Path-iter. l	1	2	3	4	5
$v = 0.4$	γ^l	1.00e2	1.62e3	1.75e5	1.75e8	1.76e12
	#(SSN)-iter	3	+3	+6	+1	+1
$v = 0.4999$	γ^l	1.00e2	2.80e3	3.23e5	3.24e8	3.24e12
	#(SSN)-iter	3	+4	+7	+3	+1

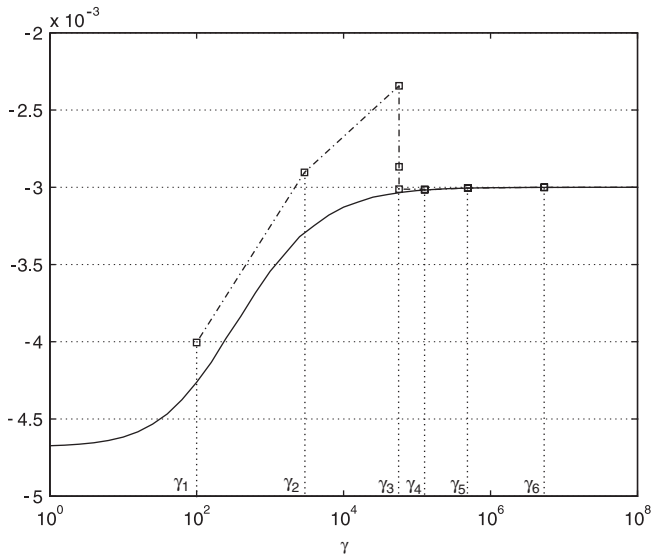

Fig. 6. Example 1: primal–dual path-value function (solid) and path-value functional evaluated at iterates produced using inexact path-following.

5.3. Results for Example 2

In the second example, we start with investigating how well our model functions $m(\cdot)$ approximate the primal–dual path-value function $V(\cdot)$. Similarly as in the previous example we use different γ's to fit the parameters C_1, C_2, E in $m(\cdot)$; the remarkable match between $V(\cdot)$ and $m(\cdot)$ can be seen from Fig. 7, where $V(\cdot)$ is compared with the models using the γ-values 10^6, 10^8, 10^{10} as support points. Note that in Example 2 it is not necessary to calculate the solution of $(\mathscr{P}_\gamma)$ for $\gamma = 0$: since $\boldsymbol{t} \equiv 0$ and $\boldsymbol{f} \equiv 0$, the linear form L in $(\mathscr{P}_\gamma)$ is the zero-functional. Thus, for $\gamma = 0$ the minimum in $(\mathscr{P}_\gamma)$ is obtained for $\boldsymbol{y} = (0, 0, -0.2)^{\text{top}}$, which implies that $V(0) = 0$.

We apply our inexact path-following strategy with the parameters $\tau = 10^8$, $\tau_1 = 5$, $q = 1.5$ for the solution of the problem. The algorithm stops after seven iterations with a residual of 3.85e−12 and the path parameter $\gamma = 1.02\text{e}10$. The deformed mesh is depicted in Fig. 8. Again, the gray values visualize the elastic shear energy density [2].

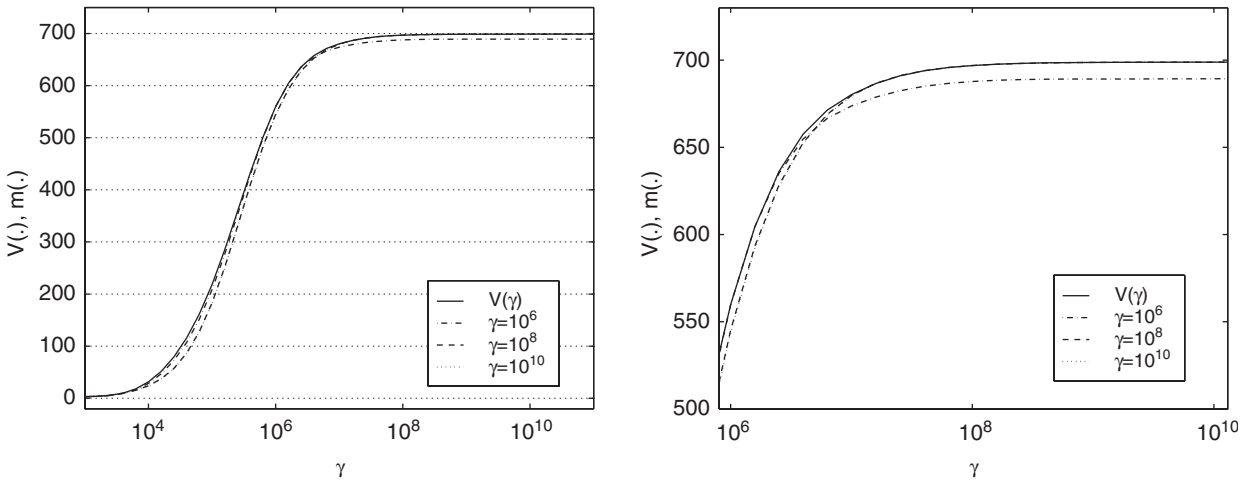

Fig. 7. Example 2, left: primal–dual path-value function $V(\cdot)$ versus model functions $m(\cdot)$ using the support points $\gamma^\star = 10^6, 10^8, 10^{10}$; right: detail of left plot.

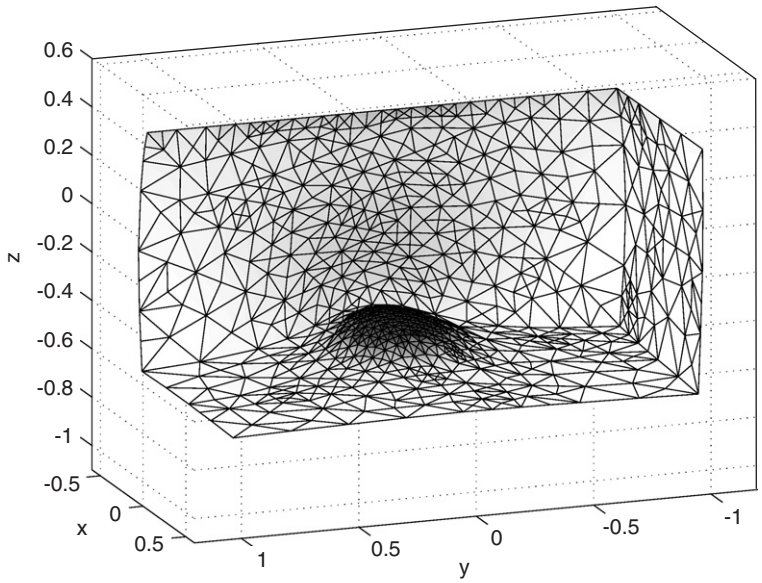

Fig. 8. Deformed mesh for Example 2, gray values visualize the stress density.

Appendix A. Proof of Theorem 1

Recall that both the variables $(\boldsymbol{y}_\gamma, \lambda_\gamma)$ and $(\bar{\boldsymbol{y}}, \bar{\lambda})$ satisfy Eq. (2a). Moreover, (2b) and (2c) hold for $(\bar{\boldsymbol{y}}, \bar{\lambda})$, while $(\boldsymbol{y}_\gamma, \lambda_\gamma)$ satisfies

$$\lambda_\gamma = \max(0, \hat{\lambda} + \gamma(\tau_N \boldsymbol{y}_\gamma - d)). \tag{A.1}$$

Setting $z := \boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}$ in (2a) results in

$$a(\boldsymbol{y}_\gamma, \boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}) - L(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}) + (\lambda_\gamma, \tau_N(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}))_{\Gamma_c} = 0. \tag{A.2}$$

Next, we estimate

$$\begin{aligned}(\lambda_\gamma, \tau_N(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}))_{\Gamma_c} &= (\lambda_\gamma, \tau_N \boldsymbol{y}_\gamma - d)_{\Gamma_c} - (\lambda_\gamma, \tau_N \bar{\boldsymbol{y}} - d)_{\Gamma_c} \\ &\geqslant \gamma^{-1}(\lambda_\gamma, \hat{\lambda} + \gamma(\tau_N \boldsymbol{y}_\gamma - d))_{\Gamma_c} - \gamma^{-1}(\lambda_\gamma, \hat{\lambda})_{\Gamma_c},\end{aligned}$$

where $(\lambda_\gamma, \tau_N \bar{\boldsymbol{y}} - d) \leqslant 0$ a.e. on Γ_c was used. Thus,

$$\begin{aligned}(\lambda_\gamma, \tau_N(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}))_{\Gamma_c} &\geqslant \gamma^{-1}(\lambda_\gamma, \max(0, \hat{\lambda} + \gamma(\tau_N \boldsymbol{y}_\gamma - d)))_{\Gamma_c} - \gamma^{-1}(\lambda_\gamma, \hat{\lambda})_{\Gamma_c} \\ &= \gamma^{-1}\|\lambda_\gamma\|^2_{\Gamma_c} - \gamma^{-1}(\lambda_\gamma, \hat{\lambda})_{\Gamma_c} \qquad (A.3) \\ &= \frac{1}{2\gamma}\|\lambda_\gamma - \hat{\lambda}\|^2_{\Gamma_c} + \frac{1}{2\gamma}\|\lambda_\gamma\|^2_{\Gamma_c} - \frac{1}{2\gamma}\|\hat{\lambda}\|^2_{\Gamma_c} \geqslant -\frac{1}{2\gamma}\|\hat{\lambda}\|^2_{\Gamma_c}. \qquad (A.4)\end{aligned}$$

Eqs. (A.2) and (A.3) imply that

$$a(\boldsymbol{y}_\gamma, \boldsymbol{y}_\gamma) + \frac{1}{\gamma}\|\lambda_\gamma\|^2_{\Gamma_c} \leqslant a(\boldsymbol{y}_\gamma, \bar{\boldsymbol{y}}) + \frac{1}{\gamma}(\lambda_\gamma, \hat{\lambda})_{\Gamma_c} + L(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}). \tag{A.5}$$

Using the coercivity (with constant $c > 0$) and the continuity of $a(\cdot\,, \cdot)$ in (A.5) shows that

$$c\|\boldsymbol{y}_\gamma\|_{\mathbf{Y}} + \frac{1}{\gamma}\|\lambda_\gamma\|_{\Gamma_c}$$

is uniformly bounded with respect to $\gamma \geqslant 1$. Hence $\boldsymbol{y}_\gamma$ is bounded in $\mathbf{Y}$ and λ_γ in $H^{-1/2}(\Gamma_c)$ from (2a). Consequently, there exist $(\tilde{\boldsymbol{y}}, \tilde{\lambda}) \in \mathbf{Y} \times H^{-1/2}(\Gamma_c)$ and a sequence γ^k with $\lim_{k\to\infty} \gamma^k = \infty$ such that

$$\boldsymbol{y}_{\gamma^k} \rightharpoonup \tilde{\boldsymbol{y}} \text{ weakly in } \mathbf{Y} \quad \text{and} \quad \lambda_{\gamma^k} \rightharpoonup \tilde{\lambda} \text{ weakly in } H^{-1/2}(\Gamma_c). \tag{A.6}$$

In the sequel, we dismiss the subscript "k" with γ^k. Note that, due to the definition of λ_γ,

$$\frac{1}{\gamma}\|\lambda_\gamma\|^2_{\Gamma_c} = \gamma\left\|\max\left(0, \frac{1}{\gamma}\hat{\lambda} + \tau_N \boldsymbol{y}_\gamma - d\right)\right\|^2_{\Gamma_c}. \tag{A.7}$$

Since $H^{1/2}(\Gamma_c)$ embeds compactly into $L^2(\Gamma_c)$, $\tau_N \boldsymbol{y}_\gamma$ converges to $\tau_N \tilde{\boldsymbol{y}}$ almost everywhere on Γ_c. Thus, (A.7) implies that $\tau_N \tilde{\boldsymbol{y}} - d \leqslant 0$ a.e. on Γ_c.

Subtracting Eq. (2a) for $(\boldsymbol{y}_\gamma, \lambda_\gamma)$ from the same equation for $(\bar{\boldsymbol{y}}, \bar{\lambda})$ and setting $z := \boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}$ yields

$$a(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}, \boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}) = -\langle \lambda_\gamma - \bar{\lambda}, \tau_N(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}})\rangle_{\Gamma_c}. \tag{A.8}$$

Using (A.4), the coercivity of $a(\cdot\,, \cdot)$, (A.8) and (2c) shows that

$$\begin{aligned}0 &\leqslant \limsup_{\gamma\to\infty} c\|\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}}\|^2_{\mathbf{Y}} \leqslant \lim_{\gamma\to\infty} \langle \bar{\lambda}, \tau_N(\boldsymbol{y}_\gamma - \bar{\boldsymbol{y}})\rangle_{\Gamma_c} \\ &= \langle \bar{\lambda}, \tau_N \tilde{\boldsymbol{y}} - d\rangle_{\Gamma_c} - \langle \bar{\lambda}, \tau_N \bar{\boldsymbol{y}} - d\rangle_{\Gamma_c} \\ &= \langle \bar{\lambda}, \tau_N \tilde{\boldsymbol{y}} - d\rangle_{\Gamma_c} \leqslant 0,\end{aligned}$$

where $\tau_N\tilde{\boldsymbol{y}} - d \leqslant 0$ a.e. on Γ_c is used. From the above estimate follows that $\boldsymbol{y}_\gamma \to \bar{\boldsymbol{y}}$ strongly in $\mathbf{Y}$ and thus $\tilde{\boldsymbol{y}} = \bar{\boldsymbol{y}}$. Passing to the limit in

$$a(\boldsymbol{y}_\gamma, \boldsymbol{z}) - L(\boldsymbol{z}) + (\lambda_\gamma, \tau_N \boldsymbol{z})_{\Gamma_c} = 0 \quad \text{for all } \boldsymbol{z} \in \mathbf{Y}$$

yields

$$a(\bar{\boldsymbol{y}}, \boldsymbol{z}) - L(\boldsymbol{z}) + \langle \tilde{\lambda}, \tau_N \boldsymbol{z} \rangle_{\Gamma_c} = 0 \quad \text{for all } \boldsymbol{z} \in \mathbf{Y}. \tag{A.9}$$

Comparing (A.9) with (2a) shows that $\hat{\lambda}$ and $\bar{\lambda}$ satisfy the same equation for all $\boldsymbol{z} \in \mathbf{Y}$, and thus $\tilde{\lambda} = \bar{\lambda}$ in $H^{-1/2}(\Gamma_c)$. Due to the uniqueness of the solution variables $\bar{\boldsymbol{y}}, \bar{\lambda}$ this implies that the whole family $\{(\boldsymbol{y}_\gamma, \lambda_\gamma)\}$ converges as stated in the theorem.

Appendix B. Proof of Theorem 4

Let us denote by $\delta_y^l := \boldsymbol{y}^l - \bar{\boldsymbol{y}} \in \mathbf{Y}$ and $\delta_\lambda^l := \lambda^l - \bar{\lambda} \in L^2(\Gamma_c)$, where $\bar{\boldsymbol{y}}$ and $\bar{\lambda}$ denote the solution of $(\mathscr{P})$ and the corresponding multiplier, respectively. From the fact that $(\bar{\boldsymbol{y}}, \bar{\lambda})$ satisfies (2a) we have for $l \geqslant 1$

$$a(\bar{\boldsymbol{y}}, \delta_y^{l+1}) - L(\delta_y^{l+1}) + (\bar{\lambda}, \tau_N \delta_y^{l+1})_{\Gamma_c} = 0 \tag{B.1}$$

and from Step 2 of (ALM) that

$$a(\boldsymbol{y}^{l+1}, \delta_y^{l+1}) - L(\delta_y^{l+1}) + (\lambda^{l+1}, \tau_N \delta_y^{l+1})_{\Gamma_c} = 0. \tag{B.2}$$

Subtracting (B.1) from (B.2) results in

$$0 = a(\delta_y^{l+1}, \delta_y^{l+1}) + (\delta_\lambda^{l+1}, \tau_N \delta_y^{l+1})_{\Gamma_c}. \tag{B.3}$$

Note that

$$\lambda^{l+1} = P(\lambda^l + \gamma^{l+1}(\tau_N \boldsymbol{y}^{l+1} - d)) \quad \text{and} \quad \bar{\lambda} = P(\bar{\lambda} + \gamma^{l+1}(\tau_N \bar{\boldsymbol{y}} - d)), \tag{B.4}$$

where $P : L^2(\Gamma_c) \to L^2(\Gamma_c)$ denotes the pointwise projection onto the convex cone of nonnegative functions. Thus, we obtain

$$(\lambda^{l+1} - \bar{\lambda}, (\lambda^l + \gamma^{l+1}(\tau_N \boldsymbol{y}^{l+1} - d)) - \lambda^{l+1}) - (\bar{\lambda} + \gamma^{l+1}(\tau_N \bar{\boldsymbol{y}} - d) - \bar{\lambda}))_{\Gamma_c} \geqslant 0.$$

This implies that

$$\begin{aligned}
(\delta_\lambda^{l+1}, \tau_N \delta_y^{l+1})_{\Gamma_c} &= (\gamma^{l+1})^{-1}(\lambda^{l+1} - \bar{\lambda}, (\lambda^l + \gamma^{l+1}(\tau_N \boldsymbol{y}^{l+1} - d)) - (\bar{\lambda} + \gamma^{l+1}(\tau_N \bar{\boldsymbol{y}} - d)))_{\Gamma_c} \\
&\quad - (\gamma^{l+1})^{-1}(\lambda^{l+1} - \bar{\lambda}, \lambda^l - \bar{\lambda})_{\Gamma_c} \\
&\geqslant (\gamma^{l+1})^{-1} \|\lambda^{l+1} - \bar{\lambda}\|_{\Gamma_c}^2 - (\gamma^{l+1})^{-1}(\lambda^{l+1} - \bar{\lambda}, \lambda^l - \bar{\lambda})_{\Gamma_c} \\
&\geqslant \frac{1}{2\gamma^{l+1}} \|\delta_\lambda^{l+1}\|_{\Gamma_c}^2 - \frac{1}{2\gamma^{l+1}} \|\delta_\lambda^l\|_{\Gamma_c}^2.
\end{aligned} \tag{B.5}$$

Utilizing the above estimate and (B.3) yield that

$$\frac{1}{2\gamma^{l+1}} \|\delta_\lambda^{l+1}\|_{\Gamma_c}^2 - \frac{1}{2\gamma^{l+1}} \|\delta_\lambda^l\|_{\Gamma_c}^2 + a(\delta_y^{l+1}, \delta_y^{l+1}) \leqslant 0. \tag{B.6}$$

Using the assumption that $\gamma^l \leqslant \gamma^{l+1}$ for all $l \geqslant 1$ and summing over all l shows that $\lim_{l\to\infty} a(\delta_y^l, \delta_y^l) = 0$. Consequently, $\boldsymbol{y}^l \to \bar{\boldsymbol{y}}$ strongly in $\mathbf{Y}$ and from Step 2 of (ALM) $\lambda^l \to \bar{\lambda}$ weakly in $L^2(\Gamma_c)$.

References

[1] P. Alart, A. Curnier, A mixed formulation for frictional contact problems prone to Newton like solution methods, Comput. Methods Appl. Mech. Eng. 92 (1991) 353–375.
[2] J. Alberty, C. Carstensen, S.A. Funken, R. Klose, Matlab implementation of the finite element method in elasticity, Computing 69 (3) (2002) 239–263.
[3] D.P. Bertsekas, Constrained optimization and Lagrange multiplier methods. Computer Science and Applied Mathematics, Academic Press Inc., New York, 1982.
[4] C. Carstensen, S.A. Funken, Averaging technique for FE—a posteriori error control in elasticity. I. Conforming FEM, Comput. Methods Appl. Mech. Eng. 190 (18–19) (2001) 2483–2498.
[5] COMSOL Group. FEMLAB: Multiphysics Modeling. ⟨http://www.femlab.com⟩.
[6] A. Forsgren, P.E. Gill, M.H. Wright, Interior methods for nonlinear optimization, SIAM Rev. 44 (4) (2003) 525–597.
[7] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, in: Springer Series in Computational Physics, Springer, New York, 1984.
[8] J. Haslinger, Z. Dostál, R. Kučera, On a splitting type algorithm for the numerical realization of contact problems with Coulomb friction, Comput. Methods Appl. Mech. Eng. 191 (21–22) (2002) 2261–2281.
[9] M. Hintermüller, K. Ito, K. Kunisch, The primal–dual active set strategy as a semismooth Newton method, SIAM J. Optim. 13 (3) (2003) 865–888.
[10] M. Hintermüller, V. Kovtunenko, K. Kunisch, Semismooth Newton methods for a class of unilaterally constrained variational inequalities, Adv. Math. Sci. Appl. 14 (2) (2004) 513–535.
[11] M. Hintermüller, K. Kunisch, Path-following methods for a class of constrained minimization problems in function space, Technical Report TR04-11, CAAM, Rice University, 2004, SIAM J. Optim. 17 (1) (2006) 159–187.
[12] H.C. Huang, W. Han, J.S. Zhou, The regularization method for an obstacle problem, Numer. Math. 69 (2) (1994) 155–166.
[13] S. Hüeber, B. Wohlmuth, A primal–dual active set strategy for non-linear multibody contact problems, Comput. Methods Appl. Mech. Eng. 194 (2005) 3147–3166.
[14] K. Ito, K. Kunisch, Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces, Nonlinear Anal., Ser. A: Theory Methods 41 (5–6) (2000) 591–616.
[15] K. Ito, K. Kunisch, Semi-smooth Newton methods for variational inequalities of the first kind, M2AN Math. Model. Numer. Anal. 37 (1) (2003) 41–62.
[16] N. Kikuchi, J.T. Oden, Contact problems in elasticity: a study of variational inequalities and finite element methods, SIAM Studies in Applied Mathematics, vol. 8, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1988.
[17] D. Kinderlehrer, Remarks about Signorini's problem in linear elasticity, Ann. Scuola Norm. Sup. Pisa 4 (8) (1981) 605–645.
[18] R. Kornhuber, R. Krause, Adaptive multigrid methods for Signorini's problem in linear elasticity, Comp. Vis. Sci. 4 (1) (2001) 9–20.
[19] R. Krause, Monotone multigrid methods for Signorini's problem with friction, Ph.D. Thesis, FU Berlin, 2001.
[20] J. Schöberl, Efficient contact solvers based on domain decomposition techniques, Comput. Math. Appl. 42 (8–9) (2001) 1217–1228 (Numerical methods and computational mechanics (Miskolc, 1998)).
[21] G. Stadler, Infinite-dimensional semi-smooth Newton and augmented Lagrangian methods for friction and contact problems in elasticity, Ph.D. Thesis, University of Graz, 2004.
[22] M. Ulbrich, Semismooth Newton methods for operator equations in function spaces, SIAM J. Optim. 13 (3) (2003) 805–842.
[23] M. Weiser, Function space complementarity methods for optimal control problems, Ph.D. Thesis, Freie Universität Berlin, 2001.
[24] M. Weiser, Interior point methods in function space, SIAM J. Control Optim. 44 (5) (2005) 1766–1786.
[25] P. Wriggers, Computational Contact Mechanics, Wiley, New York, 2002.
[26] Y. Ye, Interior Point Algorithms, in: Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1997.