SEMISMOOTH NEWTON AND AUGMENTED LAGRANGIAN METHODS FOR A SIMPLIFIED FRICTION PROBLEM[*]

GEORG STADLER[†]

Abstract. In this paper a simplified friction problem and iterative second-order algorithms for its solution are analyzed in infinite dimensional function spaces. Motivated from the dual formulation, a primal-dual active set strategy and a semismooth Newton method for a regularized problem as well as an augmented Lagrangian method for the original problem are presented and their close relation is analyzed. Local as well as global convergence results are given. By means of numerical tests, we discuss among others convergence properties, the dependence on the mesh, and the role of the regularization and illustrate the efficiency of the proposed methodologies.

Key words. friction problem, semismooth Newton method, augmented Lagrangians, primal-dual active set algorithm

AMS subject classifications. 74M10, 65K05, 35J85, 49M15

DOI. 10.1137/S1052623403420833

1. Introduction. This paper is devoted to the convergence analysis of iterative algorithms for the solution of mechanical problems involving friction. As a model problem we consider a simplified friction problem that can be stated as the minimization of the nondifferentiable functional

$$\text{(P)} \qquad \begin{cases} J(y) := \frac{1}{2}a(y,y) - (f,y)_{L^2(\Omega)} + g\int_{\Gamma_f} |\tau y(x)|\, dx \\ \text{over the set } Y := \{y \in H^1(\Omega) : \tau y = 0 \text{ a.e. on } \Gamma_0\}, \end{cases}$$

where $\Omega \subset \mathbb{R}^n$ is an open bounded domain with Lipschitz continuous boundary Γ, $\Gamma_0 \subset \Gamma$ is a possibly empty open set, $\Gamma_f := \Gamma \setminus \Gamma_0$, further $g > 0$, $f \in L^2(\Omega)$, τ denotes the trace operator, and $a(\cdot,\cdot)$ denotes a coercive bilinear form on $Y \times Y$. Introducing for $y \in H^1(\Omega)$ the abbreviation

$$j(y) := g\int_{\Gamma_f} |\tau y(x)|\, dx,$$

it is well known (cf. [12]) that (P) can be written as the elliptic variational inequality of the second kind

$$\text{(1.1)} \qquad \begin{cases} \text{Find } y \in Y \text{ such that} \\ a(y, z-y) + j(z) - j(y) \geq (f, z-y)_{L^2(\Omega)} \text{ for all } z \in Y. \end{cases}$$

While usually in engineering papers finite dimensional discretizations of (P) or (1.1) and related problems are studied, little attention has been paid to their infinite dimensional counterparts, specifically to Newton-type methods. This contribution focuses on the formulation and analysis of second-order solution algorithms for (P) in a function space framework. Such an infinite dimensional analysis gives more

---

[†]Institut für Mathematik, Karl-Franzens-Universität Graz, Heinrichstraße 36, A-8010 Graz, Austria (ge.stadler@uni-graz.at).

insight into the problem, which is also of significant practical importance since the performance of a numerical algorithm is closely related to the infinite dimensional problem structure. In particular, it is desirable that the numerical method can be considered as the discrete version of a well-defined and well-behaved algorithm for the continuous problem. A finite dimensional approach misses important features as, for example, the regularity of Lagrange multipliers and its consequences as well as smoothing and uniform definiteness properties of the involved operators. It is well accepted that these properties significantly influence the behavior of numerical algorithms.

In principal there are two approaches to overcome the difficulty associated with the nondifferentiability in (P). One is based on resolving the derivative of the absolute value function introducing a Lagrange multiplier; the other one is based on an appropriate smoothing of the nondifferentiable term $j(\cdot)$.

An overrelaxation method and the Uzawa algorithm are proposed in the monographs [12, 11] for the solution of (P), and convergence results for these first-order methods are given. The Uzawa method is also suggested for a variational inequality of the second kind in [14]; however, in that paper no numerical results are given. In [19] iterative techniques for the solution of friction contact problems are presented and further developed in [15]. Those methods require minimization of a nondifferentiable functional over a convex set in every iteration step, which also motivates our investigation of problem (P).

In [7, 5, 6] a generalized differentiability concept (Pang's B-differential) is used that allows application of a Newton-like method for discretizations of friction contact problems, whereas algorithm formulation and analysis are done in finite dimensional spaces and only few convergence rate results are given. The authors of those contributions report on good numerical results and, in [6], an almost mesh independent behavior of the algorithm is observed, which suggests that the finite dimensional method is induced by an infinite dimensional one. A different approach towards numerical realization of discrete elliptic variational inequalities of the second kind was followed in [26, 25], where monotone multigrid methods are employed to derive an efficient solution method.

For a smoothed variational inequality of the second kind, again in [11] the Uzawa method is proposed. More recent contributions apply classical Newton methods to the smoothed finite dimensional problems (see, e.g., [29]).

While there is a large literature on finite dimensional constrained and nondifferential optimization techniques (see, e.g., [10, 31, 30] for finite dimensional semismooth Newton methods), the systematic analysis of these methods in continuous function spaces started only rather recently [16, 33]. The methods proposed in this paper are related to the primal-dual active set strategy for the solution of constrained optimal control problems [1, 2]. This algorithm is closely related to semismooth Newton methods as shown in [16]. The papers [1, 2, 16, 22] apply these methodologies to unilateral pointwise constrained optimization problems; the convergence analysis for bilaterally constrained problems (as is the dual of (P)) involves additional problems as will come out in this contribution (see also [18, 17]). The first-order augmented Lagrangian method for nonsmooth optimization is investigated within a Hilbert space framework in [21].

This paper is organized as follows: In section 2 the dual problem for (P) and the extremality conditions are determined, and further possible generalizations of the model problem to vector-valued friction problems in linear elasticity are discussed. Section 3 is devoted to a regularization procedure for the dual formulation, the

corresponding primal problem, and the convergence of the regularized problems. In section 4 we state algorithms for the solution of the regularized and the original friction problem and investigate their relation. Section 5 analyzes these algorithms and gives local as well as global convergence results. Finally, section 6 summarizes our numerical testing.

2. The dual problem. In this section we summarize basic results for (P); further we calculate the dual problem and corresponding extremality conditions and discuss the relation between friction problems in elasticity and the model problem (P).

To simplify notation we use the trace operator $\tau_f : Y \longrightarrow L^2(\Gamma_f)$ defined by $\tau_f y = (\tau y)_{|\Gamma_f}$. Before we calculate the dual problem, we state the following existence and uniqueness result.

THEOREM 2.1. *Problem* (P) *or equivalently* (1.1) *admits a unique solution* $\bar{y} \in Y$.

Note that in [8] conditions are stated that guarantee existence and uniqueness of a solution to (P) in the case that $a(\cdot, \cdot)$ is not coercive. To get a deeper insight into problem (P) we next calculate the corresponding dual problem. We start with rewriting problem (P) as

$$(2.1) \qquad \inf_{y \in Y} \{F(y) + G(\tau_f y)\}$$

with the convex functionals

$$F(y) = \frac{1}{2}a(y, y) - (f, y)_{L^2(\Omega)} \ \text{ and } \ G(\tau_f y) = g \int_{\Gamma_f} |\tau_f y(x)| \, dx.$$

Following [9], the dual problem can be written as

$$(2.2) \qquad \sup_{\lambda \in L^2(\Gamma_f)} \left\{ -F^*(-\tau_f^* \lambda) - G^*(\lambda) \right\},$$

where $F^* : Y^* \longrightarrow \mathbb{R}$ and $G^* : L^2(\Gamma_f) \longrightarrow \mathbb{R}$ denote the convex conjugate functionals to F and G, and $\tau_f^* \in \mathcal{L}(L^2(\Gamma_f), Y^*)$ is the adjoint of τ_f. In (2.2) we already identified $L^2(\Gamma_f)$ with its dual. Calculating (2.2) explicitly results in the following dual problem.

$$(P^*) \qquad \sup_{|\lambda| \leq g \text{ a.e. on } \Gamma_f} J^*(\lambda) := -\frac{1}{2}a(w(\lambda), w(\lambda)),$$

where $w(\lambda)$ denotes the unique solution of

$$(2.3) \qquad a(w, v) - (f, v)_{L^2(\Omega)} + (\lambda, \tau_f v)_{L^2(\Gamma_f)} = 0 \ \text{ for all } v \in Y$$

for given $\lambda \in L^2(\Gamma_f)$. From the convexity of F and G, further from the fact that there exists a $y_0 \in Y$ such that $F(y_0) < \infty$, $G(\tau_f y_0) < \infty$, and from the continuity of G at $\tau_f y_0$, it follows that

$$\inf_{y \in Y} J(y) = - \sup_{|\lambda| \leq g \text{ a.e. on } \Gamma_f} J^*(\lambda);$$

that is, no duality gap occurs (see [9]). Existence of a solution $\bar{\lambda} \in L^2(\Gamma_f)$ for the dual problem (P*) follows from Fenchel's duality theorem. Thus, by means of duality theory we have transformed (P), the unconstrained minimization of a nondifferentiable functional, into (P*), the constrained maximization of a smooth functional.

Following duality theory the solutions $\bar{y}, \bar{\lambda}$ of the primal problem (P) and the dual problem (P*), respectively, are connected by the extremality conditions

$$-\tau_f^* \bar{\lambda} \in \partial F(\bar{y}), \ \ \bar{\lambda} \in \partial G(\tau_f \bar{y}),$$

with ∂ denoting the subdifferential. These conditions, which also characterize primal and dual solution (cf. [9]), result in

$$(2.4) \qquad a(\bar{y}, v) - (f, v)_{L^2(\Omega)} + (\bar{\lambda}, \tau_f v)_{L^2(\Gamma_f)} = 0 \text{ for all } v \in Y$$

and in the complementarity condition

$$(2.5) \qquad \begin{cases} \tau_f \bar{y} \leq 0 & \text{a.e. on} \quad \mathcal{A}_- \quad := \quad \{x \in \Gamma_f : \bar{\lambda} = -g \text{ a.e. on } \Gamma_f\}, \\ \tau_f \bar{y} = 0 & \text{a.e. on} \quad \mathcal{I} \quad := \quad \{x \in \Gamma_f : |\bar{\lambda}| < g \text{ a.e. on } \Gamma_f\}, \\ \tau_f \bar{y} \geq 0 & \text{a.e. on} \quad \mathcal{A}_+ \quad := \quad \{x \in \Gamma_f : \bar{\lambda} = g \text{ a.e. on } \Gamma_f\}. \end{cases}$$

The next lemma states an equivalent expression for condition (2.5), which will be used frequently. The equivalence follows from results in convex analysis but can also easily be verified by a direct computation.

LEMMA 2.2. *Condition* (2.5) *can equivalently be expressed as*

$$(2.6) \qquad \tau_f \bar{y} = \max(0, \tau_f \bar{y} + \sigma(\bar{\lambda} - g)) + \min(0, \tau_f \bar{y} + \sigma(\bar{\lambda} + g))$$

for every $\sigma > 0$.

Concerning a numerical solution of (P) in [11], two methods are proposed. First, an overrelaxation method for the discretized problem is described and tested; second, Uzawa's algorithm is applied to (P) and convergence results are given. Since we focus on higher-order methods, and further since for (P*) the iterates of the algorithms presented in section 4 are not contained in spaces of square integrable functions, we introduce a regularization procedure for (P*) that allows the statement and analysis of our algorithms in infinite dimensional Hilbert spaces and will be shown to be closely related to augmented Lagrangians.

We conclude this section with a brief discussion of possible generalizations of our model problem (P) to vector-valued friction problems in linear elasticity (see [8, 24]). Let $\Omega \subset \mathbb{R}^n$ be occupied by the elastic body in its nondeformed state and denote by $\boldsymbol{y} \in \boldsymbol{Y} := Y^n$ the deformation of the body that is fixed on Γ_0, subject to a given force on Ω and to friction conditions on Γ_f. For the material law we take, e.g., the Navier–Lamé equations and denote the corresponding bilinear form on $\boldsymbol{Y} \times \boldsymbol{Y}$ by $\boldsymbol{a}(\cdot, \cdot)$. Then, similarly as for (P), we get the following condition for $\boldsymbol{y}$:

$$(2.7) \qquad \boldsymbol{a}(\boldsymbol{y}, \boldsymbol{v}) - (\boldsymbol{f}, \boldsymbol{v})_{\boldsymbol{L}^2(\Omega)} + (\boldsymbol{\lambda}, \boldsymbol{v}_T)_{\boldsymbol{L}^2(\Gamma_f)} = 0 \text{ for all } \boldsymbol{v} \in \boldsymbol{Y},$$

where $\boldsymbol{L}^2(\Omega) := (L^2(\Omega))^n$, $\boldsymbol{L}^2(\Gamma_f) = (L^2(\Gamma_f))^n$, $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega)$ is a given force, $(\cdot)_T$ denotes the tangential part along Γ_f, and $\boldsymbol{\lambda} \in \boldsymbol{L}^2(\Gamma_f)$. Let us assume that the region of contact between the body and the rigid foundation is fixed (bilateral contact), i.e., the normal component $\boldsymbol{y}_N$ of $\boldsymbol{y}$ along Γ_f is zero, and let us take Tresca's law to model friction on Γ_f. This leads to the conditions $\boldsymbol{y}_N = 0$ and $||\boldsymbol{\lambda}|| \leq g$ on Γ_f (here $|| \cdot ||$ denotes the Euclidean norm in $\mathbb{R}^n$) and to

$$(2.8) \qquad \begin{cases} \boldsymbol{y}_T = 0 & \text{a.e. on} \quad \mathcal{I} \quad := \quad \{x \in \Gamma_f : ||\boldsymbol{\lambda}|| < g \text{ a.e. on } \Gamma_f\}, \\ \boldsymbol{y}_T = \frac{1}{g}||\boldsymbol{y}_T||\boldsymbol{\lambda} & \text{a.e. on} \quad \mathcal{A} \quad := \quad \{x \in \Gamma_f : ||\boldsymbol{\lambda}|| = g \text{ a.e. on } \Gamma_f\}. \end{cases}$$

The above complementarity system models the assumption that slip begins if a certain magnitude of friction traction is exceeded. Note the similarity between (2.4), (2.5) and (2.7), (2.8). In case of planar elasticity (i.e., $n = 2$), one has that $\boldsymbol{y}_T = y\boldsymbol{t}$, where $\boldsymbol{t}$ is the unit tangential vector along Γ_f and y is a scalar-valued function. This allows us to replace (2.8) by (2.5); i.e., this paper covers friction in planar elasticity. For $n \geq 3$, however, compared to (2.5) the expression in (2.8) includes an inherent nonlinearity on $\mathcal{A}$. We plan to extend the results presented in this paper for the simplified model problem (P) to elasticity problems with friction for $n \geq 3$ in our future research.

3. Regularization. In this section we introduce a regularization procedure to overcome the difficulty associated with the nondifferentiability of the functional J in (P). Therefore we consequently utilize results from duality theory and discuss relations between the regularization and the primal and dual problem.

For the term $\int_{\Gamma_f} |\tau y(x)| \, dx$ in (P), which involves the absolute value function, many ways to construct sequences of differentiable approximations are possible (cf., e.g., [13, 20]). While these regularizations are mainly motivated by the primal problem, our approximation is motivated by considering the dual problem and by results in the context of semismooth Newton methods [16, 22, 33] and augmented Lagrangians [21]. In the corresponding primal problem the regularization turns out to be a very natural one that is related to those used in [11, 14, 28].

This section is organized as follows. After presenting the regularization for the dual problem, we calculate the corresponding primal problem and the optimality system, argue the connection with [11, 14, 22, 28], and investigate the convergence as the regularization parameter tends to infinity.

3.1. Regularization for (P*). For fixed $\gamma > 0$ and $\hat{\lambda} \in L^2(\Gamma_f)$, we consider

$$(P_\gamma^*) \qquad \sup_{|\lambda| \leq g \text{ a.e. on } \Gamma_f} J_\gamma^*(\lambda) := -\frac{1}{2} a(w(\lambda), w(\lambda)) - \frac{1}{2\gamma} \|\lambda - \hat{\lambda}\|_{L^2(\Gamma_f)}^2 + \frac{1}{2\gamma} \|\hat{\lambda}\|_{L^2(\Gamma_f)}^2,$$

where again $w(\lambda)$ denotes the solution to (2.3) for given λ. This regularized problem, which has the form of the auxiliary problem in proximal point methods, is obtained from (P*) by adding

$$(3.1) \qquad -\frac{1}{2\gamma} \|\lambda - \hat{\lambda}\|_{L^2(\Gamma_f)}^2 + \frac{1}{2\gamma} \|\hat{\lambda}\|_{L^2(\Gamma_f)}^2$$

to the objective functional. Standard arguments show that (P_γ^*) admits a unique solution λ_γ for every $\gamma > 0$. The second term in (3.1), which is a constant, can be neglected from the optimizational point of view; it has been introduced to get a simple connection with the corresponding primal problem (see Theorem 3.2).

In what follows we shall use $e : Y \times L^2(\Gamma_f) \longrightarrow Y^*$ defined by

$$\langle e(y, \lambda), z \rangle_{Y^*, Y} = a(y, z) - (f, z)_{L^2(\Omega)} + (\lambda, \tau_f z)_{L^2(\Gamma_f)}.$$

This allows us to write (2.3) as $e(y, \lambda) = 0$ in Y^*. We now derive the first-order optimality conditions for the constrained optimization problem (P_γ^*) using Lagrange multipliers.

THEOREM 3.1. *Let $\lambda_\gamma \in L^2(\Gamma_f)$ be the unique solution of (P_γ^*). Then there exist $y_\gamma \in Y$ and $\xi_\gamma \in L^2(\Gamma_f)$ such that*

$$(3.2a) \qquad\qquad\qquad e(y_\gamma, \lambda_\gamma) = 0 \ \ in \ Y^*,$$

$$(3.2b) \qquad\qquad \tau_f y_\gamma + \gamma^{-1}(\hat{\lambda} - \lambda_\gamma) - \xi_\gamma = 0 \ \ in \ L^2(\Gamma_f),$$

$$(3.2c) \quad \xi_\gamma - \max(0, \xi_\gamma + \sigma(\lambda_\gamma - g)) - \min(0, \xi_\gamma + \sigma(\lambda_\gamma + g)) = 0 \ \ in \ L^2(\Gamma_f)$$

hold for every $\sigma > 0$.

Note that (3.2c) is equivalent to

$$
\text{(3.3)} \quad
\begin{cases}
\xi_\gamma \leq 0 & \text{a.e. on} \quad \mathcal{A}_{\gamma,-} \quad := \quad \{x \in \Gamma_f : \lambda_\gamma = -g \text{ a.e. on } \Gamma_f\}, \\
\xi_\gamma = 0 & \text{a.e. on} \quad \mathcal{I}_\gamma \quad := \quad \{x \in \Gamma_f : |\lambda_\gamma| < g \text{ a.e. on } \Gamma_f\}, \\
\xi_\gamma \geq 0 & \text{a.e. on} \quad \mathcal{A}_{\gamma,+} \quad := \quad \{x \in \Gamma_f : \lambda_\gamma = g \text{ a.e. on } \Gamma_f\}.
\end{cases}
$$

3.2. Corresponding primal problem. Next we turn our attention to the primal formulation of problem (P^*_γ). For $\alpha \in \mathbb{R}$ we define

$$
h(x, \alpha) =
\begin{cases}
\frac{g}{\gamma} |\gamma x + \alpha| - \frac{g^2}{2\gamma} & \text{if} \quad |\gamma x + \alpha| > g, \\
\frac{1}{2\gamma} (\gamma x + \alpha)^2 & \text{if} \quad |\gamma x + \alpha| \leq g.
\end{cases}
$$

The function h is a continuously differentiable smoothing of the absolute value function. We can now define problem (P_γ):

$$
\text{(P}_\gamma\text{)} \qquad \min_{y \in Y} J_\gamma(y) := \frac{1}{2} a(y, y) - (f, y)_{L^2(\Omega)} + \int_{\Gamma_f} h(\tau_f y(x), \hat{\lambda}(x)) \, dx.
$$

Note that the functional J_γ is uniformly convex and continuously differentiable. The next theorem clarifies the connection between (P_γ) and (P^*_γ). The proof uses standard arguments from duality theory.

THEOREM 3.2. *Problem* (P^*_γ) *is the dual problem of* (P_γ) *and we have* $J^*_\gamma(\lambda_\gamma) = -J_\gamma(y_\gamma)$, *where* λ_γ *and* y_γ *denote the solutions of* (P^*_γ) *and* (P_γ), *respectively. Furthermore, if one introduces the variable* $\xi_\gamma := \tau_f y_\gamma + \gamma^{-1}(\hat{\lambda} - \lambda_\gamma) \in L^2(\Gamma_f)$, *the extremality conditions yield* (3.2a)–(3.2c) *and these conditions are sufficient for* λ_γ *and* y_γ *to be the solution of* (P^*_γ) *and* (P_γ), *respectively.*

We next discuss by what means the above regularization is related to those used in other papers. First consider the case that $\hat{\lambda} \equiv 0$. Then the smoothing of the absolute value function in (P_γ) results in

$$
h(x) =
\begin{cases}
g|x| - \frac{g^2}{2\gamma} & \text{if } |x| \geq \frac{g}{\gamma}, \\
\frac{\gamma}{2} x^2 & \text{if } |x| < \frac{g}{\gamma}.
\end{cases}
$$

This approximation of the absolute value function has also been studied and used in [11, 14, 28] for the numerical solution of related problems. Let us now argue the relation of the above regularization to the one in [22], where $\hat{\lambda}$ is now arbitrary. For this purpose we choose $\sigma := \gamma^{-1}$ in the complementarity condition (3.2c) and eliminate the variable ξ_γ using (3.2b). This gives

(3.4)
$$
\tau_f y_\gamma + \gamma^{-1}(\hat{\lambda} - \lambda_\gamma) - \max(0, \tau_f y_\gamma + \gamma^{-1}(\hat{\lambda} - g)) - \min(0, \tau_f y_\gamma + \gamma^{-1}(\hat{\lambda} + g)) = 0.
$$

One can see that the specific choice of $\sigma := \gamma^{-1}$ results in eliminating the variable λ_γ in the max- and min-function, which is of interest regarding semismooth Newton methods, as will become clear in the next section. In [22] a formulation related to (3.4) was successfully used to construct an effective algorithm for unilaterally constrained variational problems of the first kind. However, in the case of (P^*_γ), which is a bilaterally constrained optimization problem, (3.4) may mislead us to an algorithm, which is less efficient. Obviously, from the theoretical point of view, the two formulations are equivalent, but the splitting of (3.4) into (3.2b) and (3.2c) contains the additional parameter σ and thus motivates a slightly different algorithm, as will be discussed in section 4. Next we investigate the convergence as $\gamma \to \infty$.

3.3. Convergence of the regularized problems. We conclude this section with a convergence result with respect to the regularization parameter γ (for related results see [11, 22]).

THEOREM 3.3. *For any $\hat{\lambda} \in L^2(\Gamma_f)$ the solutions y_γ of the regularized problems (P_γ) converge to the solution $\bar{y}$ of the original problem (P) strongly in Y as $\gamma \to \infty$. Furthermore, the solutions λ_γ of the dual problems (P_γ^*) converge to the solution $\bar{\lambda}$ of (P^*) weakly in $L^2(\Gamma_f)$.*

Proof. Recall the complementarity conditions and the definition of the active and inactive sets (2.5) for the original and (3.3) for the regularized problem. Note that for all $\gamma > 0$ we have $|\lambda_\gamma| \leq g$ a.e. on Γ_f. We now choose an arbitrary sequence γ_n such that $\gamma_n \to \infty$ for $n \to \infty$. From the weak compactness of the unit sphere in a Hilbert space, we can infer the existence of $\lambda^* \in L^2(\Gamma_f)$ and a subsequence $\lambda_{\gamma_{n_k}}$ in $L^2(\Gamma_f)$ such that

$$\lambda_{\gamma_{n_k}} \rightharpoonup \lambda^* \text{ weakly in } L^2(\Gamma_f).$$

Since closed convex sets in Hilbert spaces are weakly closed, we have $|\lambda^*| \leq g$ a.e. in $L^2(\Gamma_f)$. The weak convergence of $(\lambda_{\gamma_{n_k}})$ in $L^2(\Gamma_f)$ implies $y_{\gamma_{n_k}} \rightharpoonup y^*$ weakly in Y for some $y^* \in Y$ and that the pair (y^*, λ^*) also satisfies $e(y^*, \lambda^*) = 0$. We henceforth drop the subscript n_k with γ_{n_k}. It follows that

$$(3.5) \qquad a(y_\gamma - \bar{y}, y_\gamma - \bar{y}) = (\tau_f \bar{y}, \lambda_\gamma - \bar{\lambda})_{L^2(\Gamma_f)} + (\tau_f y_\gamma, \bar{\lambda} - \lambda_\gamma)_{L^2(\Gamma_f)}.$$

We are now going to estimate the above two terms separately. Let us first turn our attention to the term $(\tau_f \bar{y}, \lambda_\gamma - \bar{\lambda})_{L^2(\Gamma_f)}$. We have that

$$\tau_f \bar{y}(\lambda_\gamma - \bar{\lambda}) = \tau_f \bar{y}(\lambda_\gamma - g) \leq 0 \text{ a.e. on } \mathcal{A}_+,$$

since $\tau_f \bar{y} \geq 0$ and $\lambda_\gamma \leq g$. Similarly we find that $\tau_f \bar{y}(\lambda_\gamma - \bar{\lambda}) \leq 0$ on $\mathcal{A}_-$ utilizing $\tau_f \bar{y} \leq 0$ and $\lambda_\gamma \geq -g$. Finally, on $\mathcal{I}$ we have $\tau_f \bar{y} = 0$ which yields, since $\Gamma_f = \mathcal{A}_- \cup \mathcal{A}_+ \cup \mathcal{I}$, that

$$(3.6) \qquad (\tau_f \bar{y}, \lambda_\gamma - \bar{\lambda})_{L^2(\Gamma_f)} \leq 0.$$

Next we consider $\tau_f y_\gamma(\bar{\lambda} - \lambda_\gamma)$ on the sets $\mathcal{A}_{\gamma,-}, \mathcal{A}_{\gamma,+}$, and $\mathcal{I}_\gamma$, which also form a disjoint splitting of Γ_f. On $\mathcal{A}_{\gamma,-}$ the variable λ_γ is equal to $-g$ and $\gamma \tau_f y_\gamma + \hat{\lambda} \leq -g$ holds. This implies

$$(3.7) \qquad \tau_f y_\gamma(\bar{\lambda} - \lambda_\gamma) = \tau_f y_\gamma(\bar{\lambda} + g) \leq \gamma^{-1}(-g - \hat{\lambda})(\bar{\lambda} + g) \text{ a.e. on } \mathcal{A}_{\gamma,-}.$$

By a similar calculation one finds

$$(3.8) \qquad \tau_f y_\gamma(\bar{\lambda} - \lambda_\gamma) \leq \gamma^{-1}(g - \hat{\lambda})(\bar{\lambda} - g) \text{ a.e. on } \mathcal{A}_{\gamma,+}.$$

On $\mathcal{I}_\gamma$ we have $\lambda_\gamma = \gamma \tau_f y_\gamma + \hat{\lambda}$ and thus $|\gamma \tau_f y_\gamma + \hat{\lambda}| < g$, which shows that a.e.

$$(3.9) \quad \begin{aligned} \tau_f y_\gamma(\bar{\lambda} - \lambda_\gamma) &= \tau_f y_\gamma(\bar{\lambda} - \gamma \tau_f y_\gamma - \hat{\lambda}) = -\gamma|\tau_f y_\gamma|^2 + \tau_f y_\gamma(\bar{\lambda} - \hat{\lambda}) \\ &\leq -\gamma|\tau_f y_\gamma|^2 + |\tau_f y_\gamma||\bar{\lambda} - \hat{\lambda}| \leq -\gamma|\tau_f y_\gamma|^2 + \gamma^{-1}(g + |\hat{\lambda}|)|\bar{\lambda} - \hat{\lambda}|. \end{aligned}$$

Hence, using (3.7), (3.8), and (3.9), one gets

$$(3.10) \qquad (\tau_f y_\gamma, \bar{\lambda} - \lambda_\gamma)_{L^2(\Gamma_f)} \leq \gamma^{-1}(g + |\hat{\lambda}|, |\bar{\lambda}| + |\hat{\lambda}| + g)_{L^2(\Gamma_f)}.$$

Using (3.5), (3.6), (3.10), and the coercivity (with constant $\nu > 0$) of $a(\cdot, \cdot)$ on Y, we can estimate

$$
\begin{aligned}
0 \le \limsup_{\gamma \to \infty} \nu \|y_\gamma - \bar{y}\|^2_{H^1(\Omega)} &\le \limsup_{\gamma \to \infty} a(y_\gamma - \bar{y}, y_\gamma - \bar{y}) \\
&\le \lim_{\gamma \to \infty} (\tau_f y_\gamma, \bar{\lambda} - \lambda_\gamma)_{L^2(\Gamma_f)} \le \lim_{\gamma \to \infty} \gamma^{-1}(g + |\hat{\lambda}|, |\bar{\lambda}| + |\hat{\lambda}| + g)_{L^2(\Gamma_f)} = 0.
\end{aligned}
$$

(3.11)

It follows that $y_\gamma \to \bar{y}$ strongly in Y and hence $y^* = \bar{y}$. Passing to the limit in $e(y_\gamma, \lambda_\gamma) = 0$ and using that weak limits are unique imply $\lambda^* = \bar{\lambda}$. Thus we have proved that every sequence γ_n with $\gamma_n \to \infty$ for $n \to \infty$ contains a subsequence γ_{n_k} such that $\lambda_{\gamma_{n_k}} \rightharpoonup \bar{\lambda}$ in $L^2(\Gamma_f)$ and $y_{\gamma_{n_k}} \to \bar{y}$ in Y. Since $(\bar{y}, \bar{\lambda})$ is the unique solution to (3.2a)–(3.2c), the whole family $\{(y_\gamma, \lambda_\gamma)\}$ converges in the sense given in the statement of the theorem. □

As a corollary to the proof of Theorem 3.3 one obtains a convergence rate of y_γ to $\bar{y}$.

COROLLARY 3.4. *Let y_γ and $\bar{y}$ be solutions of (P_γ) and (P), respectively. Then there exists a $C > 0$ independent of γ such that*

$$
(3.12) \qquad\qquad \|y_\gamma - \bar{y}\|_{H^1(\Omega)} \le \frac{C}{\sqrt{\gamma}}.
$$

Proof. The inequality follows from (3.11) and the coercivity of $a(\cdot, \cdot)$ on Y. □

4. Algorithms for the solution of (P^*_γ) and (P^*). In this section we present iterative algorithms to solve (P^*_γ) and (P^*). To simplify notation we drop the subscript γ for the iterates $(y^k_\gamma, \lambda^k_\gamma, \xi^k_\gamma)$ of the algorithms. The solution variables of the regularized problem are still denoted by $(y_\gamma, \lambda_\gamma, \xi_\gamma)$.

4.1. Primal-dual active set algorithm for (P^*_γ). The primal-dual active set algorithm (PDAS) is related to the algorithms in [1, 2, 18, 17, 22] in the context of constrained optimal control and obstacle problems. It is an iterative algorithm which uses the current variables λ^k, ξ^k for (P^*_γ) to predict new active sets $\mathcal{A}^{k+1}_-, \mathcal{A}^{k+1}_+$ for the constrained optimization problem (P^*_γ), whereas this prediction is motivated from expressing the complementarity condition in the form (3.2c). On these active sets the variable λ^{k+1} is fixed. Thus in each iteration step the method requires solving the equality constrained problem

$$
(4.1) \qquad \sup_{\lambda \in L^2(\Gamma_f)} J^*_\gamma(\lambda) \quad \text{s.t.} \quad \lambda = -g \text{ on } \mathcal{A}^{k+1}_-, \ \lambda = g \text{ on } \mathcal{A}^{k+1}_+,
$$

which admits a unique solution. Note that, compared to inequality constrained optimization, equality constrained problems are significantly easier to handle, both theoretically and numerically. The algorithm is specified next.

Algorithm 1: (PDAS)

1. Choose $y^0 \in \{y \in Y : \frac{\partial y}{\partial n}|_{\Gamma_f} \in L^2(\Gamma_f)\}$, $\sigma > 0$ and set $\lambda^0 := -\frac{\partial y^0}{\partial n}|_{\Gamma_f}$, $\xi^0 := \tau_f y^0 + \gamma^{-1}(\hat{\lambda} - \lambda^0)$, $k := 0$.

2. Determine

$$
\begin{aligned}
\mathcal{A}^{k+1}_- &= \{x \in \Gamma_f : \xi^k + \sigma(\lambda^k + g) < 0\}, \\
\mathcal{A}^{k+1}_+ &= \{x \in \Gamma_f : \xi^k + \sigma(\lambda^k - g) > 0\}, \\
\mathcal{I}^{k+1} &= \Gamma_f \setminus (\mathcal{A}^{k+1}_- \cup \mathcal{A}^{k+1}_+).
\end{aligned}
$$

3. If $k \ge 1$, $\mathcal{A}^{k+1}_- = \mathcal{A}^k_-$, and $\mathcal{A}^{k+1}_+ = \mathcal{A}^k_+$ stop, else

4. Solve problem (4.1) for λ^{k+1} on $\mathcal{I}^{k+1}$ and the corresponding $y^{k+1} \in Y$ and update

$$\xi^{k+1} = \begin{cases} \tau_f y^{k+1} + \gamma^{-1}(\hat{\lambda} + g) \text{ on } \mathcal{A}_-^{k+1}, \\ \tau_f y^{k+1} + \gamma^{-1}(\hat{\lambda} - g) \text{ on } \mathcal{A}_+^{k+1}, \\ 0 \text{ on } \mathcal{I}^{k+1}, \end{cases}$$

$k := k + 1$ and goto step 2.

Note that ξ^{k+1} is the Lagrange multiplier for the equality constraints in (4.1). The justification of the stopping criterion in step 3 of (PDAS) is given in the following lemma (see also [2]). The proof relies on the fact that, if the active sets coincide for two consecutive iterations, the quantities ξ and λ satisfy the sign-structure as required by the complementarity conditions (3.3).

LEMMA 4.1. *If Algorithm (PDAS) stops, the last iterate is the solution to system* (3.2a)–(3.2c).

We now discuss the influence of the parameter σ on the iteration sequence for $k \geq 1$. On $\mathcal{I}^k$ we have that $\xi^k = 0$ and thus σ has no influence when determining the new active and inactive sets. On $\mathcal{A}_-^k$ we have $\lambda^k = -g$ and distinguish two cases: The set where $\xi^k < 0$ belongs to $\mathcal{A}_-^{k+1}$ for the next iteration independently from σ. In case $\xi^k > 0$ we have $\xi^k + \sigma(\lambda^k - g) = \xi^k - 2\sigma g$. The set where $\xi^k - 2\sigma g \leq 0$ moves to $\mathcal{I}^{k+1}$ if $\xi^k - 2\sigma g > 0$ to $\mathcal{A}_+^{k+1}$ for the next iteration. Hence, only in this case σ influences the sequence of iterates. Smaller values for σ make it more likely that points belong to $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1}$. A similar observation as for $\mathcal{A}_-^k$ holds true for $\mathcal{A}_+^k$, which shows that with $\sigma > 0$ one can control the probability that points are shifted from one active set to the other within one iteration. We also remark that, if for some $\sigma := \sigma_1 > 0$ one has $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1} = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} = \emptyset$, then for every $\sigma \geq \sigma_1$ also $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1} = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} = \emptyset$ and the sets $\mathcal{A}_-^{k+1}, \mathcal{A}_+^{k+1}$, and $\mathcal{I}^{k+1}$ are the same for all $\sigma \geq \sigma_1$. This observation will be of interest regarding our local convergence analysis of (PDAS).

4.2. Semismooth Newton method for ($\mathbf{P}_\gamma^*$). This section applies an infinite-dimensional semismooth Newton method (SS) to (P_γ^*). We use the differentiability concept as introduced in [16], which we recall for the reader's convenience later in this section.

We start with writing the optimality system (3.2a)–(3.2c) as one nonlinear operator equation utilizing (3.4). For this purpose we denote by $\tilde{y}$ the solution to the problem

$$a(y, v) - (f, v)_{L^2(\Omega)} = 0 \text{ for all } v \in Y.$$

Further we introduce $B^{-1} \in \mathcal{L}(H^{-\frac{1}{2}}, Y)$, the solution mapping for the variational equality

$$a(y, v) - \langle \lambda, \tau_f v \rangle_{H^{-\frac{1}{2}}, H^{\frac{1}{2}}} = 0 \text{ for all } v \in Y$$

for given $\lambda \in H^{-\frac{1}{2}}(\Gamma_f)$. We can now define the Neumann-to-Dirichlet operator

(4.2) $$C := \tau_f B^{-1}_{|L^2(\Gamma_f)} \in \mathcal{L}(L^2(\Gamma_f), L^2(\Gamma_f))$$

and summarize some of its properties in the next lemma.

LEMMA 4.2. *The Neumann-to-Dirichlet operator defined in* (4.2) *is self-adjoint, positive definite, injective, and compact.*

Proof. Self-adjointness, positive definiteness, and injectivity follow easily from the properties of $a(\cdot,\cdot)$. Sobolev's embedding theorem implies the compactness of C. $\quad\square$

With the help of the operators B^{-1} and C one can write the solution y to $e(y,\lambda) = 0$ in Y^* for given $\lambda \in L^2(\Gamma_f)$ as $y = -B^{-1}\lambda + \tilde{y}$ and $\tau_f y$ as $-C\lambda + \tau_f \tilde{y}$. This allows elimination of the variable $\tau_f y$ in (3.4). We introduce the mapping $\bar{F} : L^2(\Gamma_f) \longrightarrow L^2(\Gamma_f)$ by

$$(4.3) \quad \bar{F}(\lambda) = C\lambda - \tau_f\tilde{y} - \gamma^{-1}(\hat{\lambda}-\lambda) \quad + \max(0, -C\lambda + \tau_f\tilde{y} + \gamma^{-1}(\hat{\lambda}-g)) \\ + \min(0, -C\lambda + \tau_f\tilde{y} + \gamma^{-1}(\hat{\lambda}+g)).$$

Note that $\bar{F}(\lambda) = 0$ characterizes λ as the solution of (P^*_γ). In what follows we utilize for $S \subset \Gamma_f$ the extension-by-zero operator $E_S : L^2(S) \longrightarrow L^2(\Gamma_f)$ defined by

$$(4.4) \qquad\qquad E_S(g)(x) := \begin{cases} g(x) \text{ if } x \in S, \\ 0 \text{ else.} \end{cases}$$

Its adjoint operator $E_S^* : L^2(\Gamma_f) \longrightarrow L^2(S)$ is the restriction operator onto S. Writing the optimality system as done in (4.3) suggests applying a semismooth Newton method to solve $\bar{F}(\lambda) = 0$. We briefly summarize those facts on semismooth Newton methods which are relevant for the following results. Let X, Y, and Z be Banach spaces and $F : D \subset X \longrightarrow Z$ be a nonlinear mapping with open domain D.

DEFINITION 4.3. *The mapping $F : D \subset X \longrightarrow Z$ is called* Newton differentiable *on the open subset $U \subset D$ if there exists a mapping $G : U \longrightarrow \mathcal{L}(X,Z)$ such that*

$$(4.5) \qquad\qquad \lim_{h\to 0} \frac{1}{\|h\|} \|F(x+h) - F(x) - G(x+h)h\| = 0$$

for every $x \in U$.

The mapping G in the above definition is referred to as *generalized derivative*. The following convergence result for a generalized Newton method holds (see [4]).

THEOREM 4.4. *Suppose that $x^* \in D$ is a solution to $F(x) = 0$ and that F is Newton differentiable in an open neighborhood U containing x^* and that $\{\|G(x)^{-1}\| : x \in U\}$ is bounded. Then the Newton-iteration*

$$(4.6) \qquad\qquad x^{k+1} = x^k - G(x^k)^{-1}F(x^k)$$

converges superlinearly to x^* provided that $\|x^0 - x^*\|$ is sufficiently small.

To apply a Newton iteration to the mapping $\bar{F}$, we need to consider Newton differentiability of the max- and min-operator. For this purpose let X denote a function space of real-valued functions on some $\Omega \subset \mathbb{R}^n$, and further $\max(0,y)$ and $\min(0,y)$ the pointwise max- and min-operations, respectively. We now introduce candidates for the generalized derivatives

$$(4.7) \quad G_{max}(y)(x) = \begin{cases} 1 & \text{if } y(x) \geq 0, \\ 0 & \text{if } y(x) < 0, \end{cases} \qquad G_{min}(y)(x) = \begin{cases} 1 & \text{if } y(x) \leq 0, \\ 0 & \text{if } y(x) > 0. \end{cases}$$

Then we have the following result (see [16]).

THEOREM 4.5. *The mappings $\max(0,\cdot) : L^q(\Omega) \longrightarrow L^p(\Omega)$ and $\min(0,\cdot) : L^q(\Omega) \longrightarrow L^p(\Omega)$ with $1 \leq p < q < \infty$ are Newton differentiable on $L^q(\Omega)$ with generalized derivatives G_{max} and G_{min}, respectively.*

Note that Theorem 4.5 requires a norm gap (i.e., $p < q$) to hold true. In [16] it is shown that the functions defined in (4.7) cannot serve as generalized gradients if $p \geq q$. We now quote a chain rule for Newton differentiability (for a proof see [22]).

THEOREM 4.6. *Let $F_2 : Y \longrightarrow X$ be an affine mapping with $F_2 y = By + b$, $B \in \mathcal{L}(Y, X)$, $b \in X$, and assume that $F_1 : D \subset X \longrightarrow Z$ is Newton differentiable on the open subset $U \subset D$ with generalized derivative G. If $F_2^{-1}(U)$ is nonempty, then $F = F_1 \circ F_2$ is Newton differentiable on $F_2^{-1}(U)$ with generalized derivative given by $G(By + b)B \in \mathcal{L}(Y, Z)$ for $y \in F_2^{-1}(U)$.*

We can now apply the above results to the mapping $\bar{F}$. Observe that $Rg(C) \subset H^{\frac{1}{2}}(\Gamma_f)$ and that

$$(4.8) \qquad H^{\frac{1}{2}}(\Gamma_f) \hookrightarrow L^q(\Gamma_f) \text{ for } \begin{cases} q = \frac{2(n-1)}{n-2} & \text{if } n \geq 3, \\ q < \infty & \text{if } n = 2, \end{cases}$$

where $n \geq 2$ denotes the dimension of Ω. Note that $q > 2$ for all $n \geq 2$. From Theorems 4.5 and 4.6 it follows that $\bar{F}$ is Newton differentiable on $L^2(\Gamma_f)$. A generalized derivative of $\bar{F}$ is given by

$$(4.9) \qquad G_{\bar{F}}(\lambda)(\delta) = \left(E_{\mathcal{I}} E_{\mathcal{I}}^* C + \frac{1}{\gamma} \right) \delta$$

with the following definition for $\mathcal{A}_-$, $\mathcal{A}_+$, $\mathcal{I}$:

$$\mathcal{A}_- = \{ x \in \Gamma_f : -A\lambda + \tau_f \tilde{y} + \gamma^{-1}(\hat{\lambda} - g) \geq 0 \},$$
$$\mathcal{A}_+ = \{ x \in \Gamma_f : -A\lambda + \tau_f \tilde{y} + \gamma^{-1}(\hat{\lambda} + g) \leq 0 \},$$
$$\mathcal{I} = \Gamma_f \setminus (\mathcal{A}_- \cup \mathcal{A}_+).$$

Calculating (4.6) explicitly results in a semismooth Newton-iteration step for the solution of $\bar{F}(\lambda) = 0$ that is equal to one iteration step of (PDAS) with $\sigma = \gamma^{-1}$.

An analogous result for unilaterally constrained optimal control problems was established in [16]. Note that the norm gap required for Newton differentiability of the max- and min-function results from directly exploiting the smoothing property of the operator C. This has become possible since we chose $\sigma := \gamma^{-1}$ in (3.2c) which allowed elimination of the explicit appearance of λ in the max- and min-function. Taking advantage of this fact, the above semismooth Newton method does not require a smoothing step, as do the semismooth Newton methods in [33].

We now investigate whether (PDAS) with arbitrary $\sigma > 0$ can also be interpreted as a Newton method. We introduce $F : Y \times L^2(\Gamma_f) \times L^2(\Gamma_f) \longrightarrow Y^* \times L^2(\Gamma_f) \times L^2(\Gamma_f)$ by

$$(4.10) \qquad F(y, \lambda, \xi) := \begin{pmatrix} e(y, \lambda) \\ \tau_f y + \gamma^{-1}(\hat{\lambda} - \lambda) - \xi \\ \xi - \max(0, \xi + \sigma(\lambda - g)) - \min(0, \xi + \sigma(\lambda + g)) \end{pmatrix}$$

and observe that $F(y, \lambda, \xi) = 0$ characterizes y and λ as solutions to (P_γ) and (P_γ^*), respectively. Applying now the Newton iteration (4.6) with the generalized derivative of the max- and min-function as given in (4.7) to the mapping F results in Algorithm (PDAS) which can be seen similarly as for (SS). In section 5.2 it is shown that for certain problems, (PDAS) converges locally superlinearly without the necessity of a smoothing step as used in [33] to get local superlinear convergence of semismooth Newton methods.

We conclude this section with some remarks on the development of semismooth Newton methods. The application of generalized Newton methods for semismooth problems in finite dimensions has a rather long history (see, e.g., [10, 31, 30] and the references given there). Recently, in [4, 16, 27, 33] concepts for generalized derivatives in infinite dimensions were introduced. Our work uses the notion of *slant differentiability in a neighborhood* as proposed in [16], which is a slight adaptation of the terminology introduced in [4], where the term *slant differentiability at a point* is also introduced. A similar concept is proposed in [27], where the name "Newton map" is coined. Applications of such pointwise approaches to Newton's method, however, presuppose knowledge of the solution. The differentiability concept in [16] coincides with a specific application of the theory developed in [33]; we refer to the discussion on this relationship in [16]. As in [22] and also motivated by [27], we use instead of the notion slant differentiability in a neighborhood the name *Newton differentiability*.

4.3. Augmented Lagrangian methods for (P*). Augmented Lagrangian methods (ALMs) combine ordinary Lagrangian methods and penalty methods without suffering of the disadvantages of these methods. For instance, the augmented Lagrangian method converges without requiring that the penalty parameter tends to infinity. For a detailed discussion of these methods we refer to [3, 21].

To argue the close relation of the regularization for (P*) to augmented Lagrangians recall that (3.2b), (3.2c) can equivalently be expressed as (3.4) and, after multiplication with γ, as

$$(4.11) \qquad \lambda_\gamma = \gamma \tau_f y_\gamma + \hat\lambda - \max(0, \gamma \tau_f y_\gamma + \hat\lambda - g) - \min(0, \gamma \tau_f y_\gamma + \hat\lambda + g).$$

The augmented Lagrangian method is an iterative algorithm for the calculation of λ in (P*).

It sets $\hat\lambda := \lambda^l$ in (4.11) and determines λ^{l+1} from the solution of (4.11), (3.2a). Note that the augmented Lagrangian method can be seen as an implicit version of Uzawa's algorithm (see [21]). The whole method is specified next.

Algorithm 3: (ALM)
1. Choose $\gamma > 0$, $\lambda^0 \in L^2(\Gamma_f)$ and set $l := 0$.
2. Solve for $(y^{l+1}, \lambda^{l+1}, \xi^{l+1}) \in Y \times L^2(\Gamma_f) \times L^2(\Gamma_f)$ system (3.2a)–(3.2c) with $\hat\lambda := \lambda^l$.
3. Update $l := l + 1$ and goto step 2.

The auxiliary problem in step 2 of (ALM) has exactly the form of our regularized problem and can thus efficiently be solved using (PDAS) or (SS). The question arises concerning the precision to which the system in step 2 is solved. Several strategies are possible, such as solving the system exactly for all l or performing only one iteration step of the semismooth Newton method in each iteration. We tested several strategies and report on them in section 6. Note that in (ALM) the regularization parameter γ plays the role of a penalty parameter, which is not necessarily taken to infinity; nevertheless (ALM) detects the solution of (P*), as will be shown in section 5.4.

5. Convergence analysis. In this section we present local convergence results for (SS) and (PDAS) as well as a conditional global convergence result for (SS) and unconditional global convergence of (ALM).

5.1. Local superlinear convergence of (SS). In this section we give a local convergence result for algorithm (SS) for the solution of the regularized friction problem.

THEOREM 5.1. *If $\|\lambda^0 - \lambda_\gamma\|_{L^2(\Gamma_f)}$ is sufficiently small, then the iterates λ^k of (SS) converge to (λ_γ) superlinearly in $L^2(\Gamma_f)$. Furthermore, the corresponding primal iterates y^k converge superlinearly to y_γ in Y.*

Proof. We have only to show superlinear convergence of λ^k to λ_γ in $L^2(\Gamma_f)$; then superlinear convergence of y^k to y_γ in $Y \subset H^1(\Omega)$ follows since B^{-1} is continuous.

We already argued Newton differentiability of $\bar{F} \in \mathcal{L}(L^2(\Gamma_f), L^2(\Gamma_f))$. To apply Theorem 4.4 it remains to verify that the generalized gradients $G_{\bar{F}} \in \mathcal{L}(L^2(\Gamma_f), L^2(\Gamma_f))$ of $\bar{F}$ have uniformly bounded inverses. Recall the definition of the extension-by-zero operator E and its adjoint E^* as given in (4.4). Let $(h_{\mathcal{A}_-}, h_{\mathcal{A}_+}, h_{\mathcal{I}}) \in L^2(\mathcal{A}_-) \times L^2(\mathcal{A}_+) \times L^2(\mathcal{I})$ and consider the equation

(5.1) $$G_{\bar{F}}(\lambda)(\delta) = G_{\bar{F}}(\lambda)(\delta_{\mathcal{A}_-}, \delta_{\mathcal{A}_+}, \delta_{\mathcal{I}}) = (h_{\mathcal{A}_-}, h_{\mathcal{A}_+}, h_{\mathcal{I}}).$$

Recalling the explicit form (4.9) of $G_{\bar{F}}$, we get from (5.1) that $\delta_{\mathcal{A}_-} = \gamma h_{\mathcal{A}_-}$ and $\delta_{\mathcal{A}_+} = \gamma h_{\mathcal{A}_+}$ must hold; further

$$\left(\frac{1}{\gamma} + E_{\mathcal{I}}^* C E_{\mathcal{I}}\right)\delta_{\mathcal{I}} = h_{\mathcal{I}} - \gamma E_{\mathcal{I}}^* C E_{\mathcal{A}_-} h_{\mathcal{A}_-} - \gamma E_{\mathcal{I}}^* C E_{\mathcal{A}_+} h_{\mathcal{A}_+}.$$

Due to the positivity of C we can define a new scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ on $L^2(\mathcal{I})$ by

$$\langle\langle x, y \rangle\rangle := \left\langle \left(\frac{1}{\gamma} + E_{\mathcal{I}}^* C E_{\mathcal{I}}\right) x, y \right\rangle_{L^2(\mathcal{I})} \quad \text{for } x, y \in L^2(\mathcal{I}).$$

Utilizing the positivity of C we have that the product $\langle\langle \cdot, \cdot \rangle\rangle$ is coercive with constant γ^{-1} independently from $\mathcal{I}$. Applying the Lax–Milgram lemma, one finds not only that (5.1) admits a unique solution $\delta_{\mathcal{I}}$, but also that

$$\|\delta_{\mathcal{I}}\|_{L^2(\mathcal{I})} \leq \gamma \|h_{\mathcal{I}}\|_{L^2(\mathcal{I})} + \gamma^2 \|C\|_{\mathcal{L}(L^2(\Gamma_f))} \left\{ \|h_{\mathcal{A}_-}\|_{L^2(\mathcal{A}_-)} + \|h_{\mathcal{A}_+}\|_{L^2(\mathcal{A}_+)} \right\}.$$

This proves the uniform boundedness of $G_{\bar{F}}(\lambda)^{-1}$ with respect to $\lambda \in L^2(\Gamma_f)$ and ends the proof. $\square$

5.2. Local superlinear convergence of (PDAS). As observed at the end of section 4, algorithm (PDAS) cannot directly be interpreted as a locally superlinear convergent semismooth Newton method if no smoothing steps are used. However, exploiting the role of σ in (PDAS) (see the discussion after Lemma 4.1), it turns out that local superlinear convergence holds for (PDAS) as well, provided the dimension n of Ω is 2.

COROLLARY 5.2. *Assume that $n = 2$ and $\Gamma_0 \subset \Gamma$ is a sufficiently regular subset. If $\|\lambda^0 - \lambda_\gamma\|_{L^2(\Gamma_f)}$ is sufficiently small, the iterates (y^k, λ^k) of (PDAS) with $\sigma \geq \gamma^{-1}$ converge superlinearly in $Y \times L^2(\Gamma_f)$.*

Proof. The idea of this proof is to show that in a neighborhood of the solution λ_γ the iterates λ^k of (PDAS) coincide with λ^k from (SS), which allows application of Theorem 5.1 also for (PDAS).

Step 1. We first consider only (SS) and denote by $\delta > 0$ the convergence radius of this semismooth Newton method. We introduce a δ_0 with $0 < \delta_0 \leq \delta$, which will be further specified below, and choose $\lambda^0 \in L^2(\Gamma_f)$ such that $\|\lambda^0 - \lambda_\gamma\|_{L^2(\Gamma_f)} \leq \delta_0$.

Since $\delta_0 \leq \delta$ the method converges and $\|\lambda^k - \lambda^{k+1}\|_{L^2(\Gamma_f)} \leq 2\delta_0$ for $k \geq 1$. Note that the difference of the corresponding variables $y^k - y^{k+1}$ solves

$$a(y^k - y^{k+1}, v) + (\lambda^k - \lambda^{k+1}, \tau_f v)_{L^2(\Gamma_f)} = 0 \text{ for all } v \in Y.$$

It thus follows from regularity results for mixed elliptic problems [32] that

$$\|y^k - y^{k+1}\|_{C^0(\bar{\Omega})} \leq C\|\lambda^k - \lambda^{k+1}\|_{L^2(\Gamma_f)} \text{ for some } C > 0.$$

For the corresponding traces we have

(5.2) $$\|\tau_f(y^k - y^{k+1})\|_{C^0(\Gamma_f)} \leq C\|\lambda^k - \lambda^{k+1}\|_{L^2(\Gamma_f)} \leq 2C\delta_0.$$

We now show that for δ_0 sufficiently small $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1} = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} = \emptyset$. We prove this claim by contradiction; i.e., we assume that $J = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} \neq \emptyset$. Then, almost everywhere on J we have

$$\tau y^{k-1} + \gamma^{-1}(\hat{\lambda} - g) > 0 \text{ and } \tau y^k + \gamma^{-1}(\hat{\lambda} + g) < 0,$$

which implies $\tau(y^{k+1} - y^k) > 2g\gamma^{-1}$. Thus, utilizing (5.2)

(5.3) $$2g\gamma^{-1} < \|\tau(y^{k+1} - y^k)\|_{C^0(\Gamma_f)} \leq 2C\delta_0.$$

If we choose $\delta_0 \leq \frac{g}{C\gamma}$, relation (5.3) cannot hold true and therefore $J = \emptyset$. An analogous observation holds true for $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1}$, which shows that

(5.4) $$\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1} = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} = \emptyset \text{ if } \delta_0 \leq \frac{g}{C\gamma}.$$

Step 2. Recall that the iterates of (PDAS) with $\sigma = \gamma^{-1}$ coincide with those of (SS). Thus, if $\|_{L^2(\Gamma_f)}\|\lambda^0 - \lambda_\gamma \leq \delta_0$, then $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1} = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} = \emptyset$ for (PDAS) with $\sigma = \gamma^{-1}$. It follows from the discussion after Lemma 4.1 that for the active sets calculated from (PDAS) using $\sigma \geq \gamma^{-1}$ also $\mathcal{A}_-^k \cap \mathcal{A}_+^{k+1} = \mathcal{A}_+^k \cap \mathcal{A}_-^{k+1} = \emptyset$ holds. This shows that (SS) and (PDAS) determine the same iterates for the variable λ_γ provided that $\|\lambda^0 - \lambda_\gamma\|_{L^2(\Gamma_f)} < \delta_0$. Hence, superlinear L^2-convergence for λ^k determined from (PDAS) holds. For the variables y^k superlinear convergence in Y follows from the continuity of the solution mapping B^{-1}. □

5.3. Conditional global convergence of (SS). Our global convergence result is based on an appropriately defined functional which decays when evaluated along the iterates of the algorithm. A related strategy to prove global convergence (i.e., convergence from arbitrary initialization) is used in [23] in the context of optimal control problems. In what follows we use the notation from (PDAS) with $\sigma := \gamma^{-1}$ for (SS). For $(\lambda, \xi) \in L^2(\Gamma_f) \times L^2(\Gamma_f)$ we define the functional

(5.5) $$M(\lambda, \xi) := \frac{1}{\gamma^2} \int_{\Gamma_f} |(\lambda - g)^+|^2 + |(\lambda + g)^-|^2 \, dx + \int_{\mathcal{A}_+^*} |\xi^-|^2 \, dx + \int_{\mathcal{A}_-^*} |\xi^+|^2 \, dx,$$

where $\mathcal{A}_+^* = \{x \in \Gamma_f : \lambda(x) \geq g\}$ and $\mathcal{A}_-^* = \{x \in \Gamma_f : \lambda(x) \leq -g\}$. By $(\cdot)^+$ and $(\cdot)^-$ we denote the positive and negative part, i.e., $(\cdot)^+ := \max(0, \cdot)$ and $(\cdot)^- := -\min(0, \cdot)$. As a preparatory step for the following estimates we prove a lemma that can easily be verified using the spectral theorem for compact and positive definite operators.

LEMMA 5.3. *Let X be a real Hilbert space with inner product $(\cdot\,,\cdot)$ and $C \in \mathcal{L}(X)$ be injective, self-adjoint, positive, and compact. Then*

$$(y, y) \leq \|C\|_{\mathcal{L}(X)}(C^{-1}y, y)$$

for all $y \in Rg(C)$.

Following Lemma 4.2 the Neumann-to-Dirichlet mapping C, as given in (4.2), fulfills the conditions of Lemma 5.3. Utilizing the operator C step 4 of (PDAS) implies

$$C^{-1}\tau_f y^{k+1} = -\lambda^{k+1} = -\begin{cases} g & \text{on } \mathcal{A}_+^{k+1}, \\ \gamma\tau_f y^{k+1} + \hat{\lambda} & \text{on } \mathcal{I}^{k+1}, \\ -g & \text{on } \mathcal{A}_-^{k+1}, \end{cases}$$

$$(5.6) \qquad \tau_f y^{k+1} + \gamma^{-1}(\hat{\lambda} - \lambda^{k+1}) - \xi^{k+1} = 0.$$

With the above notation we get

$$(5.7) \quad C^{-1}(\tau_f(y^k - y^{k+1})) = \lambda^{k+1} - \lambda^k = \begin{cases} R_{\mathcal{A}+}^k & \text{on } \mathcal{A}_+^{k+1}, \\ \gamma(\tau_f(y^{k+1} - y^k)) + R_{\mathcal{I}}^k & \text{on } \mathcal{I}^{k+1}, \\ R_{\mathcal{A}-}^k & \text{on } \mathcal{A}_-^{k+1}, \end{cases}$$

where

$$R_{\mathcal{A}_+}^k = \begin{cases} 0 & \text{on } \mathcal{A}_+^{k+1} \cap \mathcal{A}_+^k, \\ g - \lambda^k < 0 & \text{on } \mathcal{A}_+^{k+1} \cap \mathcal{I}^k, \\ 2g < \gamma\xi^k & \text{on } \mathcal{A}_+^{k+1} \cap \mathcal{A}_-^k, \end{cases}$$

$$R_{\mathcal{I}}^k = \begin{cases} \gamma\tau_f y^k + \hat{\lambda} - g = \gamma\xi^k \leq 0 & \text{on } \mathcal{I}^{k+1} \cap \mathcal{A}_+^k, \\ 0 & \text{on } \mathcal{I}^{k+1} \cap \mathcal{I}^k, \\ \gamma\xi^k \geq 0 & \text{on } \mathcal{I}^{k+1} \cap \mathcal{A}_-^k, \end{cases}$$

$$R_{\mathcal{A}_+}^k = \begin{cases} -2g > \gamma\xi^k & \text{on } \mathcal{A}_-^{k+1} \cap \mathcal{A}_+^k, \\ -g - \lambda^k > 0 & \text{on } \mathcal{A}_-^{k+1} \cap \mathcal{I}^k, \\ 0 & \text{on } \mathcal{A}_-^{k+1} \cap \mathcal{A}_-^k. \end{cases}$$

Let us denote by R^k the function defined on Γ_f, whose restrictions to $\mathcal{A}_-^{k+1}$, $\mathcal{I}^{k+1}$, and $\mathcal{A}_+^{k+1}$ coincide with $R_{\mathcal{A}_-}^k$, $R_{\mathcal{I}}^k$, and $R_{\mathcal{A}_-}^k$, respectively. Note that, from the definition of R^k, we have

$$(5.8) \qquad \|R^k\|_{L^2(\Gamma_f)}^2 \leq \gamma^2 M(\lambda^k, \xi^k).$$

To shorten the notation we introduce $\delta_y^k := \tau_f(y^{k+1} - y^k)$. Multiplying (5.7) by $-\delta_y^k$ results in

$$C^{-1}(\delta_y^k)(\delta_y^k) = -R^k\delta_y^k - \chi_{\mathcal{I}^{k+1}}\gamma(\delta_y^k)^2$$

a.e. on Γ_f, where $\chi_{\mathcal{I}^{k+1}}$ denotes the characteristic function for $\mathcal{I}^{k+1}$. Thus,

$$(5.9) \quad (C^{-1}\delta_y^k, \delta_y^k)_{L^2(\Gamma_f)} = \int_{\Gamma_f} -R^k\delta_y^k \, dx - \gamma\int_{\mathcal{I}^{k+1}}(\delta_y^k)^2 \, dx \leq \|R^k\|_{L^2(\Gamma_f)}\|\delta_y^k\|_{L^2(\Gamma_f)},$$

where we used the Cauchy–Schwarz inequality and $\int_{\mathcal{I}^{k+1}}(\delta_y^k)^2\,dx \geq 0$. Utilizing Lemma 5.3 for the Neumann-to-Dirichlet mapping C and (5.9) yields

$$(5.10) \qquad \|\delta_y^k\|_{L^2(\Gamma_f)} \leq \|C\|_{\mathcal{L}(L^2(\Gamma_f))}\|R^k\|_{L^2(\Gamma_f)}.$$

We can now prove the following convergence theorem for (SS), or equivalently for (PDAS), with $\sigma = \gamma^{-1}$.

THEOREM 5.4. *If $\gamma < \|C\|_{\mathcal{L}(L^2(\Gamma_f))}^{-1}$, then*

$$M(\lambda^{k+1},\xi^{k+1}) < M(\lambda^k,\xi^k)$$

for $k = 0, 1, 2, \ldots$ with $(\lambda^k,\xi^k) \neq (\lambda_\gamma,\xi_\gamma)$, where (λ^k,ξ^k) denote the iterates of (PDAS) with $\sigma = \gamma^{-1}$. Moreover, (y^k,λ^k,ξ^k) converges to $(y_\gamma,\lambda_\gamma,\xi_\gamma)$ strongly in $Y \times L^2(\Gamma_f) \times L^2(\Gamma_f)$.

Proof. Recall that from the definition of (PDAS) with $\sigma = \gamma^{-1}$ one gets

$$\begin{aligned}
\lambda^{k+1} &= & \gamma\tau_f y^{k+1} + \hat{\lambda} & \text{ on } \mathcal{I}^{k+1}, \\
\xi^{k+1} &= & \tau_f y^{k+1} + \gamma^{-1}(\hat{\lambda} + g) & \text{ on } \mathcal{A}_-^{k+1}, \\
\xi^{k+1} &= & \tau_f y^{k+1} + \gamma^{-1}(\hat{\lambda} - g) & \text{ on } \mathcal{A}_+^{k+1}.
\end{aligned}$$

We therefore have

$$\xi^{k+1} = \delta_y^k + \tau_f y^k + \gamma^{-1}(\hat{\lambda} - g) = \delta_y^k + \begin{cases} \xi^k - \gamma^{-1}(g - \lambda^k) > 0 & \text{on } \mathcal{A}_+^{k+1} \cap \mathcal{A}_-^k, \\ \gamma^{-1}(\lambda^k - g) > 0 & \text{on } \mathcal{A}_+^{k+1} \cap \mathcal{I}^k, \\ \xi^k > 0 & \text{on } \mathcal{A}_+^{k+1} \cap \mathcal{A}_+^k. \end{cases}$$

Thus, $|(\xi^{k+1})^-| := |\max(0, -\xi^{k+1})| \leq |\delta_y^k|$ a.e. on $\mathcal{A}_+^{k+1}$. Note that

$$\mathcal{A}_+^{*,k+1} := \{x \in \Gamma_f : \lambda^{k+1}(x) \geq g\} = \mathcal{A}_+^{k+1} \cup \{x \in \mathcal{I}^{k+1} : \lambda^{k+1}(x) \geq g\},$$

which implies, using $\xi^{k+1} = 0$ on $\mathcal{I}^{k+1}$, that

$$(5.11) \qquad |(\xi^{k+1})^-| \leq |\delta_y^k| \text{ a.e. on } \mathcal{A}_+^{*,k+1}.$$

Analogously, it follows that

$$(5.12) \qquad |(\xi^{k+1})^+| \leq |\delta_y^k| \text{ a.e. on } \mathcal{A}_-^{*,k+1},$$

where $\mathcal{A}_-^{*,k+1} := \{x \in \Gamma_f : \lambda^{k+1}(x) \leq -g\}$. Moreover, on $\mathcal{I}^{k+1}$

$$\lambda^{k+1} - g = \gamma\delta_y^k + \gamma\tau_f y^k + \hat{\lambda} - g = \gamma\delta_y^k + \begin{cases} \gamma(\xi^k + \gamma^{-1}(\lambda^k - g)) \leq 0 & \text{on } \mathcal{I}^{k+1} \cap \mathcal{A}_-^k, \\ \lambda^k - g \leq 0 & \text{on } \mathcal{I}^{k+1} \cap \mathcal{I}^k, \\ \gamma\xi^k \leq 0 & \text{on } \mathcal{I}^{k+1} \cap \mathcal{A}_+^k. \end{cases}$$

The above estimate shows that

$$(5.13) \qquad |(\lambda^{k+1} - g)^+| \leq \gamma|\delta_y^k| \text{ a.e. on } \mathcal{I}^{k+1},$$

and analogously one can show

$$(5.14) \qquad |(\lambda^{k+1} + g)^-| \leq \gamma|\delta_y^k| \text{ a.e. on } \mathcal{I}^{k+1}.$$

Since on active sets λ^{k+1} is set to either g or $-g$, we have on $\mathcal{A}_-^{k+1} \cup \mathcal{A}_+^{k+1}$ that $(\lambda^{k+1} - g)^+ = (\lambda^{k+1} + g)^- = 0$. Further, at most one of the expressions at a.a. $x \in \Gamma_f$

$$|(\lambda^{k+1} - g)^+|, \; |(\lambda^{k+1} + g)^-|, \; |(\xi^{k+1})^-|, \; |(\xi^{k+1})^+|$$

can be strictly positive, which shows, combining (5.11)–(5.14), that

$$(5.15) \qquad\qquad M(\lambda^{k+1}, \xi^{k+1}) \le \|\delta_y^k\|_{L^2(\Gamma_f)}^2.$$

Combining (5.8) and (5.10) with (5.15) shows that

$$(5.16) \qquad\qquad M(\lambda^{k+1}, \xi^{k+1}) \le \gamma^2 \|C\|_{\mathcal{L}(L^2(\Gamma_f))}^2 M(\lambda^k, \xi^k).$$

Our assumption on γ implies that

$$(5.17) \qquad\qquad \|C\|_{\mathcal{L}(L^2(\Gamma_f))}^2 \gamma^2 < 1,$$

which shows that

$$M(\lambda^{k+1}, \xi^{k+1}) < M(\lambda^k, \xi^k)$$

unless $(\lambda^k, \xi^k) = (\lambda_\gamma, \xi_\gamma)$. Combining (5.8), (5.10), (5.15), and (5.16) it follows that

$$\|\delta_y^k\|_{L^2(\Gamma_f)}^2 \le \|C\|_{\mathcal{L}(L^2(\Gamma_f))}^2 \|R^k\|_{L^2(\Gamma_f)}^2 \le \gamma^2 \|C\|_{\mathcal{L}(L^2(\Gamma_f))}^2 M(\lambda^k, \xi^k)$$

$$(5.18) \qquad \le \gamma^2 \|C\|_{\mathcal{L}(L^2(\Gamma_f))}^2 \|\delta_y^{k-1}\|_{L^2(\Gamma_f)}^2 \le \left(\gamma \|C\|_{\mathcal{L}(L^2(\Gamma_f))}\right)^{2(k+1)} M(\lambda^0, \xi^0),$$

which shows, utilizing (5.17), that $\lim_{k\to\infty} M(\lambda^k, \xi^k) = \lim_{k\to\infty} \|R^k\|_{L^2(\Gamma_f)} = 0$. Further, summing up (5.18) over k and utilizing (5.17) results in $\sum_{k=1}^\infty \|\delta_y^k\|_{L^2(\Gamma_f)}^2 < \infty$, which shows that there exists a $z \in L^2(\Gamma_f)$ such that $\lim_{k\to\infty} \tau_f y^k = z$ in $L^2(\Gamma_f)$. Using (5.7) results in

$$\|\lambda^{k+1} - \lambda^k\|_{L^2(\Gamma_f)} \le \gamma \|\delta_y^k\|_{L^2(\Gamma_f)} + \|R^k\|_{L^2(\Gamma_f)}$$

$$\le \left(\gamma \|C\|_{\mathcal{L}(L^2(\Gamma_f))}\right)^k \left(\gamma^2 \|C\|_{\mathcal{L}(L^2(\Gamma_f))} + \gamma\right) M(\lambda^0, \xi^0)^{\frac{1}{2}},$$

and thus there exists $\check\lambda \in L^2(\Gamma_f)$ with $\lim_{k\to\infty} \lambda^k = \check\lambda$ in $L^2(\Gamma_f)$. From (5.6) one gets $\lim_{k\to\infty} \xi^k = \check\xi \in L^2(\Gamma_f)$ for $\check\xi \in L^2(\Gamma_f)$. Since $0 = \lim_{k\to\infty} M(\lambda^k, \xi^k) = M(\check\lambda, \check\xi)$, the pair $(\check\lambda, \check\xi)$ satisfies condition (3.2c). From (3.2a) follows the existence of $\tilde y \in Y$ such that $\lim_{k\to\infty} y^k = \check y$ in Y. Note that, since (y^k, λ^k, ξ^k) satisfies (3.2a), (3.2b), this is also the case for $(\check y, \check\lambda, \check\xi)$. Hence $(\check y, \check\lambda, \check\xi) = (y_\gamma, \lambda_\gamma, \xi_\gamma)$ due to the uniqueness of a solution to (3.2a)–(3.2c), which ends the proof. $\square$

Note that for unilaterally constrained problems global convergence results can possibly be gained using monotonicity properties of the involved operators (see, e.g., [22]), where the maximum principle for the Laplace operator is used to prove global convergence. For bilaterally constrained problems (such as (P_γ^*)), however, monotonicity of the iterates does not hold, even if the operator satisfies a maximum principle.

5.4. Global convergence of (ALM). The next theorem states global convergence of (ALM) for all $\gamma > 0$ and shows that large γ increases the speed of convergence. In the statement of the next theorem we denote the coercivity constant of $a(\cdot, \cdot)$ on Y by $\nu > 0$.

THEOREM 5.5. *The iterates λ^l of (ALM) and the corresponding variables y^l satisfy*

$$(5.19) \qquad \nu\|y^{l+1} - \bar{y}\|^2_{H^1(\Omega)} + \frac{1}{2\gamma}\|\lambda^{l+1} - \bar{\lambda}\|^2_{L^2(\Gamma_f)} \leq \frac{1}{2\gamma}\|\lambda^l - \bar{\lambda}\|^2_{L^2(\Gamma_f)}$$

and

$$(5.20) \qquad \nu\sum_{k=1}^{\infty}\|y^{l+1} - \bar{y}\|^2_{H^1(\Omega)} \leq \frac{1}{2\gamma}\|\lambda^0 - \bar{\lambda}\|^2_{L^2(\Gamma_f)},$$

which implies that $y^l \longrightarrow \bar{y}$ strongly in Y and $\lambda^l \rightharpoonup \bar{\lambda}$ weakly in $L^2(\Gamma_f)$.

Proof. From the fact that $\bar{y}$ and $\bar{\lambda}$ are the solutions to (P) and (P*), respectively, it follows that

$$(5.21) \qquad a(\bar{y}, y^{l+1} - \bar{y}) - (f, y^{l+1} - \bar{y})_{L^2(\Omega)} + (\bar{\lambda}, \tau_f(y^{l+1} - \bar{y}))_{L^2(\Gamma_f)} = 0,$$

and since y^{l+1} and λ^{l+1} solve (P$_\gamma$) and (P$^*_\gamma$) with $\hat{\lambda} := \lambda^l$, we infer

$$(5.22) \qquad a(y^{l+1}, y^{l+1} - \bar{y}) - (f, y^{l+1} - \bar{y})_{L^2(\Omega)} + (\lambda^{l+1}, \tau_f(y^{l+1} - \bar{y}))_{L^2(\Gamma_f)} = 0.$$

Subtracting (5.21) from (5.22) results in

$$(5.23) \qquad a(y^{l+1} - \bar{y}, y^{l+1} - \bar{y}) + (\lambda^{l+1} - \bar{\lambda}, \tau_f(y^{l+1} - \bar{y}))_{L^2(\Gamma_f)} = 0.$$

Note that one can write (4.11) and (2.6) as

$$(5.24) \qquad \lambda^{l+1} = P(\gamma\tau_f y^{l+1} + \lambda^l) \text{ and } \bar{\lambda} = P(\gamma\tau_f\bar{y} + \bar{\lambda}),$$

where $P : L^2(\Gamma_f) \rightarrow L^2(\Gamma_f)$ denotes the pointwise projection onto the convex set $\{v \in L^2(\Gamma_f) : |v| \leq g \text{ a.e. on } L^2(\Gamma_f)\}$. Thus we get

$$\begin{aligned}(\lambda^{l+1} - \bar{\lambda}, \tau_f(y^{l+1} - \bar{y}))_{L^2(\Gamma_f)} &= \gamma^{-1}(\lambda^{l+1} - \bar{\lambda}, (\gamma\tau_f y^{l+1} + \lambda^l) - (\gamma\tau_f\bar{y} + \bar{\lambda}))_{L^2(\Gamma_f)} \\ &\quad - \gamma^{-1}(\lambda^{l+1} - \bar{\lambda}, \lambda^l - \bar{\lambda})_{L^2(\Gamma_f)} \\ &\geq \gamma^{-1}\|\lambda^{l+1} - \bar{\lambda}\|^2_{L^2(\Gamma_f)} - \gamma^{-1}(\lambda^{l+1} - \bar{\lambda}, \lambda^l - \bar{\lambda})_{L^2(\Gamma_f)},\end{aligned}$$

where we used that

$$(\lambda^{l+1} - \bar{\lambda}, (\gamma\tau_f y^{l+1} + \lambda^l - \lambda^{l+1}) - (\gamma\tau_f\bar{y} + \bar{\lambda} - \bar{\lambda}))_{L^2(\Gamma_f)} \geq 0,$$

which holds using (5.24) and since P is a projection onto a convex set. Using (5.23) and the coercivity of $a(\cdot, \cdot)$ on Y, we get

$$\begin{aligned}\nu\|y^{l+1} - \bar{y}\|^2_{H^1(\Omega)} &\leq a(y^{l+1} - \bar{y}, y^{l+1} - \bar{y}) = -(\lambda^{l+1} - \bar{\lambda}, \tau_f(y^{l+1} - \bar{y}))_{L^2(\Gamma_f)} \\ &\leq -\frac{1}{\gamma}\|\lambda^{l+1} - \bar{\lambda}\|^2_{L^2(\Gamma_f)} + \frac{1}{\gamma}(\lambda^{l+1} - \bar{\lambda}, \lambda^l - \bar{\lambda})_{L^2(\Gamma_f)} \\ &\leq -\frac{1}{2\gamma}\|\lambda^{l+1} - \bar{\lambda}\|^2_{L^2(\Gamma_f)} + \frac{1}{2\gamma}\|\lambda^l - \bar{\lambda}\|^2_{L^2(\Gamma_f)},\end{aligned}$$

which proves (5.19). Summing up (5.19) with respect to l, we obtain (5.20). □

TABLE 6.1
Number of iterations for different values of γ for Example 1.

γ	10^2	10^4	10^6	10^8
No. of iterations	2	3	3	3

6. Numerical tests. In this section we present three test examples for the algorithms proposed in section 5 for the solution of (P_γ^*) and (P^*). For simplicity we use for all examples the unit square as domain, i.e., $\Omega = (0,1) \times (0,1)$ and the bilinear form

$$a(y,z) := (\nabla y, \nabla z)_{L^2(\Omega)} + \mu(y,z)_{L^2(\Omega)} \qquad \text{for } y, z \in Y,$$

which is coercive if $\mu > 0$ or $\Gamma_0 \subset \Gamma$ has positive measure. For our calculations we utilize a finite difference discretization with the usual five-point stencil approximation to the Laplace operator. The discretization of the normal derivative is based on one-sided differences and all linear systems are solved exactly. We denote by N the number of gridpoints in one space dimension, which means we work on $N \times N$-grids. To investigate convergence properties we frequently report on

$$(6.1) \qquad d_\lambda^l := \|\bar\lambda - \lambda^l\|_{L^2(\Gamma_f)},$$

where $\bar\lambda := \lambda_{10^{16}}$ is the solution of (P_γ^*) with $\gamma = 10^{16}$ and λ^l denotes the actual iterate. We compare our results with those obtained using the Uzawa algorithm, which can be interpreted as an explicit form of the augmented Lagrangian method [21]. While (ALM) converges for every $\gamma > 0$ the Uzawa method converges for only $\gamma \in [\alpha_1, \alpha_2]$ with $0 < \alpha_1 < \alpha_2$, where α_1 and α_2 are in general not known [11]. We initialize the Uzawa method with $\lambda^0 := 0$ and report on the number of iterations (one iteration requires one linear solve), where the iteration is stopped if $d_\lambda^l < 10^{-4}$. Unless otherwise specified, we use $\hat\lambda = 0$, and $\sigma = 1$ for (PDAS). As initialization for (PDAS), (SS), and (ALM), the solution to (3.2a)–(3.2c) with $\xi^0 = 0$ is used, which corresponds to the solution of (P_γ^*) neglecting the constraints on λ.

6.1. Example 1. This example is taken from [11]. The data are as follows: $\Gamma_f = ([0,1] \times \{0\}) \cup ([0,1] \times \{1\})$, $g = 1.5$, $\mu = 0$, and

$$f(x) = \begin{cases} 10 & \text{if } x \in (0, \frac{1}{2}) \times (0,1), \\ -10 & \text{if } x \in [\frac{1}{2}, 1) \times (0,1). \end{cases}$$

Choosing $N = 80$ the Uzawa algorithm requires 32 iterations for $\gamma = 10$ and 17 iterations for $\gamma = 20$ and does not converge for $\gamma = 30$. In our tests for (PDAS) and (SS) we vary the value for the regularization parameter γ and investigate the convergence as $\gamma \to \infty$. Table 6.1 reports on the number of iterations needed by (SS) for various values of γ and $N = 80$. It can be seen that the algorithm requires only very few iterations to find the solution and that increasing the regularization parameter for this example does not increase the number of iterations required to detect the solution (cf. [22], where a different behavior for obstacle problems is observed). We remark that for (SS) no points are shifted from the lower active set to the upper or conversely within one iteration and thus (PDAS) determines the same iterates as (SS) for all $\sigma \geq \gamma^{-1}$.
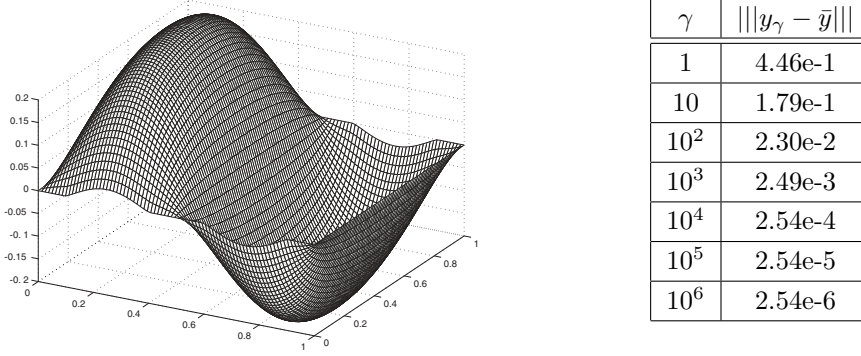
γ	$|||y_\gamma - \bar{y}|||$
1	4.46e-1
10	1.79e-1
10^2	2.30e-2
10^3	2.49e-3
10^4	2.54e-4
10^5	2.54e-5
10^6	2.54e-6

FIG. 6.1. *Solution y_γ for $\gamma = 10^8$ (left) and convergence with respect to γ (right) for Example* 1.
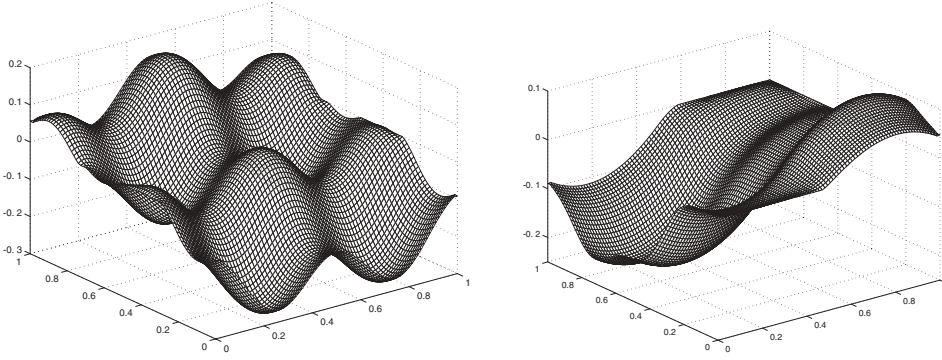

FIG. 6.2. *Solution y_γ for Example* 2 *(left) and Example* 3 *(right) for $\gamma = 10^{10}$.*

The primal solution y_γ for $\gamma = 10^8$ is shown in Figure 6.1 (left). We also investigate the convergence as $\gamma \to \infty$. The table in Figure 6.1 reports on the value of

$$|||y_\gamma - \bar{y}||| := (a(y_\gamma - \bar{y}, y_\gamma - \bar{y}))^{\frac{1}{2}}$$

for various γ, where we take $\bar{y} := y_{10^{16}}$ as approximation to the solution of the simplified friction problem and use $N = 160$. Note that $||| \cdot |||$ is a norm equivalent to the usual one in $H^1(\Omega)$. The result suggests the convergence rate γ^{-1}, while from Corollary 3.4 we get only the rate $\gamma^{-\frac{1}{2}}$.

6.2. Example 2. This example investigates the behavior of (PDAS), (SS), and (ALM) for a more complicated structure of the active and inactive sets (see the solution y_γ for $\gamma = 10^{10}$ in Figure 6.2 (left)). Uzawa's method faces serious troubles with this example: In all test examples with $\gamma > 0.2$ the method did not converge. For $\gamma = 0.2$ we stopped our test after 400 iterations at $d_\lambda^{400} = 1.03\mathrm{e}\text{-}2$. We choose $\Gamma_f = \partial\Omega$, which implies $Y = H^1(\Omega)$; and further $\mu = 0.5$, $g = 0.4$, and the external force $f(x) = 10(\sin 4\pi x + \cos 4\pi x)$.

We investigate superlinear convergence of the iterates, properties of (PDAS) and (SS), and the behavior of algorithm (ALM). Table 6.2 reports on

$$q_\lambda^k := \frac{\|\lambda^k - \lambda_\gamma\|_{L^2(\Gamma_f)}}{\|\lambda^{k-1} - \lambda_\gamma\|_{L^2(\Gamma_f)}}, \ k = 1, 2, \dots,$$

TABLE 6.2
Values for q_λ^k in Example 2, $\gamma = 50$.

k	1	2	3	4	5
q_λ^k	0.58	0.45	0.25	0.18	0.00

TABLE 6.3
Number of iterations of (SS) and (PDAS) for different values of γ for Example 2, $N = 160$.

γ	3	5	10	50	100	150	160	10^3	10^{10}
#iter$_{\text{SS}}$	3	4	4	5	6	7	div	div	div
#iter$_{\text{PD}}$	3	4	4	5	6	5	5	6	6

TABLE 6.4
Tests for (ALM) with exact solve of the auxiliary problem for Example 2, $N = 160$.

	l	1	2	3	4	5	6	7
$\gamma = 10^2$	#iter$_{PD}$	6	3	2	1	1	1	1
	d_λ^l	2.55e-2	8.54e-3	3.62e-3	1.62e-3	7.45e-4	3.58e-4	1.77e-4
$\gamma = 10^4$	#iter$_{PD}$	6	1	1	1	1		
	d_λ^l	4.00e-4	3.06e-6	2.67e-8	2.62e-10	2.84e-12		

for $\gamma = 50$. We observe superlinear convergence of the iterates determined with (SS).

Table 6.3 shows the number of iterations $\#iter_{\text{SS}}$ and $\#iter_{\text{PD}}$ required by (SS) and (PDAS), respectively, to find the solution for different values of γ. We observe a slight increase in the number of iterations as γ increases. For $\gamma \geq 160$ (SS) does not detect the solution, whereas (PDAS) does. Using (SS) for these examples we can observe the following behavior: Points in Γ_f move from $\mathcal{A}_-^{k-1}$ to $\mathcal{A}_+^k$, and then from $\mathcal{A}_+^k$ back to $\mathcal{A}_-^{k+1}$ for some $k \geq 2$, and due to this scattering the algorithm does not find the solution (cf. the remark on the role of σ in (PDAS) after Lemma 4.1). Algorithm (PDAS) with $\sigma = 1$ does not experience such problems and finds the solution after a few iterations for all tested $\gamma > 0$.

To avoid possible difficulties due to the local convergence of (SS), we test two globalization strategies: First we use a continuation procedure with respect to γ, motivated from the local convergence result for (SS). We solve for $\gamma = 150$ and use the solution as initialization for the algorithm with larger γ. This procedure turns out to be successful for only a moderate increase in γ. Increasing γ moderately, typically only one or two more iterations are needed to find the solution for larger γ. However, this method appears inconvenient and costly. Next we test backtracking with J_γ as merit function to globalize (SS). This strategy works successfully, but in particular for larger γ, several backtracking steps are necessary in each iteration. The resulting stepsize is very small and thus overall up to 50 iterations are needed to find the solution. This behavior becomes more distinct for large γ.

We also apply algorithm (ALM) for the solution of this example. In a first attempt we solve the auxiliary problem in (ALM) exactly using (PDAS), whereas this method is initialized with the solution of the auxiliary problem in the previous iteration step of (ALM). Due to the local superlinear convergence of (PDAS), the auxiliary problem is solved in very few iterations, as can be seen in Table 6.4, where for $\gamma = 10^2$ and $\gamma = 10^4$ we report on the number of iterations $\#iter_{PD}$ required by (PDAS)

TABLE 6.5
Tests for (ALM) with inexact solve of the auxiliary problem for Example 2, $N = 160$.

	l	1	2	3	4	5	6	7
$\gamma = 10^2$	#iter$_{PD}$	1	2	2	2	1	1	1
	d_λ^l	2.96e-1	5.03e-2	7.60e-3	1.89e-3	9.99e-4	5.36e-4	2.91e-4
$\gamma = 10^4$	#iter$_{PD}$	1	2	1	2	1	1	1
	d_λ^l	3.81e-1	1.46e-1	3.79e-2	1.87e-4	2.20e-6	2.67e-8	3.28e-10
$\gamma = 10^l$	#iter$_{PD}$	1	3	2	1	1	1	
	d_λ^l	1.41e-1	2.30e-2	2.16e-3	1.11e-5	1.04e-8	1.14e-12	

TABLE 6.6
Number of iterations for different values of γ (first line) and N (first column) for Example 3.

	5	10	30	50	10^2	10^3	10^5	10^{10}
20	3	3	3	3	3	3	3	3*
40	3	3	4	5	4	4	4	4*
80	3	4	4	5	5	6	6	6*
160	3	3	5	5	6	8	7*	7*
320	3	4	5	6	7	7	8*	8*

in every step l of (ALM). Further we report on d_λ^l as defined in (6.1). We observe a monotone decrease of d_λ^l and a faster convergence in the case that $\gamma = 10^4$. In a second approach we test an inexact version of the augmented Lagrangian method: We stop the (PDAS)-iterations for the auxiliary problem as soon as the initial residual in (3.2c) has been reduced at least by a factor of 2. The results for $\gamma = 10^2, 10^4$ are shown in Table 6.5, where we report on the number of (PDAS)-iterations and the value of d_λ^l. We observe that for $\gamma = 10^2$ the first iterates present a better approximation to $\bar\lambda$ than for $\gamma = 10^4$, whereas then the case $\gamma = 10^4$ shows a faster convergence behavior. This leads to the idea of increasing the parameter γ in every step of (ALM). Choosing $\gamma = 10^l$ in the lth (ALM)-iteration, the impressing results shown in the last line of Table 6.5 were obtained.

6.3. Example 3. The solution $y_{10^{10}}$ for this last example is shown in Figure 6.2 (right). Again the Uzawa algorithm only converges for small γ which results in an extremely slow convergence. The data are as follows: $\Gamma_f = \partial\Omega$, $\mu = 0.5$, $g = 0.3$, and

$$f(x) = |3x - 1| + 2\text{sgn}(2y - 1) + 2\text{sgn}(x - 0.75) + 5\sin(6\pi x).$$

We investigate the number of iterations required by (PDAS) and (SS) for various values of γ and N; further we report on results obtained with (ALM). In a series of test runs (Table 6.6) we investigate the number of iterations for various grids and different values for γ. We observe the low number of iterations for all mesh-sizes and choices for γ. For the calculations in Table 6.6 we use (SS), except for those indicated by *. For these examples with a rather large value of γ, (SS) starts to chatter due to effects described in the previous example. Thus we utilize (PDAS) to solve these problems, which is always successful. Utilizing (PDAS) also for the examples with smaller γ yields the same number of iterations as (SS). Further, note from Table 6.6 that the number of iterations increases very weakly as the mesh becomes finer.

TABLE 6.7
Tests for (ALM) with inexact solve of the auxiliary problem for Example 3, $N = 160$.

	l	1	2	3	4	5	6	7
$\gamma = 10^2$	#iter$_{PD}$	1	2	1	1	1	1	2
	d_λ^l	1.62-1	1.99e-2	6.94e-3	1.40e-3	4.67e-4	1.91e-4	9.20e-5
$\gamma = 10^4$	#iter$_{PD}$	1	3	2	1	1	1	
	d_λ^l	2.14e-1	1.46e-2	4.11e-4	9.06e-7	8.20e-9	8.43e-11	
$\gamma = 10^l$	#iter$_{PD}$	1	2	2	1	1	1	
	d_λ^l	8.54e-2	2.62e-2	2.40e-3	4.21e-4	9.08e-8	8.23e-12	

We again test (ALM) with (PDAS) for the auxiliary problem. In the case that this inner problem is solved exactly, the overall number of system solves is 16 for $\gamma = 50$ and between 10 and 20 for other tested values of γ, where we used $d_\lambda^l < 10^{-4}$ as stopping criterion and $N = 160$. The results for the case where the auxiliary problem is solved approximately in every (ALM)-iteration are summarized in Table 6.7, where we again report on the number of (PDAS)-iterations and on d_λ^l as defined in (6.1). Again increasing γ as in Example 6.2 turns out to be very efficient.

6.4. Summary of the numerical results. In our numerical testing we observe a remarkable efficiency of algorithms (SS) and (PDAS) for the solution of the regularized simplified friction problem (P_γ^*) and of (ALM) for the solution of (P^*). For moderate values of γ, the iterates for (SS) and (PDAS) coincide and these algorithms converge superlinearly. For large values of γ, (SS) may start to chatter, while (PDAS) always detects the solution. The two tested globalization strategies for (SS) turn out to be successful but inconvenient. The number of iterations of the semismooth Newton methods increases only slightly for finer grid and larger regularization parameter. The efficiency of (SS) and (PDAS) is interesting also with respect to augmented Lagrangian methods since these algorithms present a powerful tool to solve or approximately solve the auxiliary problem in (ALM). Our tests show that, solving the auxiliary problem in (ALM) only approximately using (SS) or (PDAS), the overall number of system solves for (ALM) is rather the same as for the semismooth Newton methods with large γ. However, (ALM) has the advantage that it detects the solution of the dual (P^*) of the original simplified friction problem without requiring that $\gamma \to \infty$. Finally, we remark the advantage of (PDAS) with large γ that one has a simple stopping criterion at hand that guarantees that the exact solution of (P_γ^*) is detected.

REFERENCES

[1] M. BERGOUNIOUX, M. HADDOU, M. HINTERMÜLLER, AND K. KUNISCH, *A comparison of a Moreau–Yosida-based active set strategy and interior point methods for constrained optimal control problems*, SIAM J. Optim., 11 (2000), pp. 495–521.
[2] M. BERGOUNIOUX, K. ITO, AND K. KUNISCH, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37 (1999), pp. 1176–1194.
[3] D. P. BERTSEKAS, *Constraint Optimization and Lagrange Multipliers*, Academic Press, Paris, 1982.
[4] X. CHEN, Z. NASHED, AND L. QI, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.
[5] P. W. CHRISTENSEN, A. KLARBRING, J. S. PANG, AND N. STRÖMBERG, *Formulation and comparison of algorithms for frictional contact problems*, Internat. J. Numer. Methods Engrg., 42 (1998), pp. 145–173.

[6] P. W. Christensen and J. S. Pang, *Frictional contact algorithms based on semismooth Newton methods*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Appl. Optim. 22, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 81–116.

[7] P. W. Christensen, *A nonsmooth Newton method for elastoplastic problems*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 1189–1219.

[8] G. Duvaut and J.-L. Lions, *Inequalities in Mechanics and Physics*, Springer-Verlag, Berlin, 1976.

[9] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Classics Appl. Math. 28, SIAM, Philadelphia, 1999.

[10] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol.* 1 *and* 2, Springer Ser. Oper. Res., Springer-Verlag, New York, 2003.

[11] R. Glowinski, J.-L. Lions, and R. Trémolières, *Numerical Analysis of Variational Inequalities*, North–Holland, Amsterdam, 1981.

[12] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, 1984.

[13] W. Han, *A regularization procedure for a simplified friction problem*, Math. Comput. Modelling, 15 (1991), pp. 65–70.

[14] W. Han, *On the numerical approximation of a frictional contact problem with normal compliance*, Numer. Funct. Anal. Optim., 17 (1996), pp. 307–321.

[15] J. Haslinger, Z. Dostál, and R. Kučera, *On a splitting type algorithm for the numerical realization of contact problems with Coulomb friction*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 2261–2281.

[16] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.

[17] M. Hintermüller and G. Stadler, *A semi-smooth Newton method for constrained linear-quadratic control problems*, ZAMM Z. Angew. Math. Mech., 83 (2003), pp. 219–237.

[18] M. Hintermüller, *A primal-dual active set algorithm for bilaterally control constrained optimal control problems*, Quart. Appl. Math., 61 (2003), pp. 131–160.

[19] I. Hlaváček, J. Haslinger, J. Nečas, and J. Lovíšek, *Solution of Variational Inequalities in Mechanics*, Appl. Math. Sci. 66, Springer-Verlag, New York, 1988.

[20] H. C. Huang, W. Han, and J. S. Zhou, *The regularization method for an obstacle problem*, Numer. Math., 69 (1994), pp. 155–166.

[21] K. Ito and K. Kunisch, *Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces*, Nonlinear Anal., 41 (2000), pp. 591–616.

[22] K. Ito and K. Kunisch, *Semi-smooth Newton methods for variational inequalities of the first kind*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 41–62.

[23] K. Ito and K. Kunisch, *The primal-dual active set method for nonlinear optimal control problems with bilateral constraints*, SIAM J. Control Optim., 43 (2004), pp. 357–376.

[24] N. Kikuchi and J. T. Oden, *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods*, SIAM Stud. Appl. Math. 8, SIAM, Philadelphia, 1988.

[25] R. Kornhuber, *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems*, Adv. Numer. Math., Teubner, Stuttgart, 1997.

[26] R. Kornhuber, *Monotone multigrid methods for elliptic variational inequalities. II*, Numer. Math., 72 (1996), pp. 481–499.

[27] B. Kummer, *Generalized Newton and NCP-methods: Convergence, regularity, actions*, Discuss. Math. Differ. Incl. Control Optim., 20 (2000), pp. 209–244.

[28] C.-Y. Lee and J. T. Oden, *A priori error estimation of hp-finite element approximations of frictional contact problems with normal compliance*, Internat. J. Engrg. Sci., 31 (1993), pp. 927–952.

[29] A. Y. T. Leung, G. Chen, and W. Chen, *Smoothing Newton method for solving two- and three-dimensional frictional contact problems*, Internat. J. Numer. Methods Engrg., 41 (1998), pp. 1001–1027.

[30] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.

[31] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

[32] G. Savaré, *Regularity and perturbation results for mixed second order elliptic problems*, Comm. Partial Differential Equations, 22 (1997), pp. 869–899.

[33] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2003), pp. 805–841.