A GLOBALIZED INEXACT SEMISMOOTH NEWTON METHOD FOR STRONGLY CONVEX OPTIMAL CONTROL PROBLEMS

Daniel Wachsmuth*

Abstract   We investigate a globalized inexact semismooth Newton method applied to strongly convex optimization problems in Hilbert spaces. Here, the semismooth Newton method is appplied to the dual problem, which has a continuously differentiable objective. We prove global strong convergence of iterates as well as transition to local superlinear convergence. The latter needs a second-order Taylor expansion involving semismooth derivative concepts. The convergence of the globalized method is demonstrated in numerical examples, for which the local unglobalized method diverges.

1 INTRODUCTION

The goal of this article is to develop a globalized semismooth Newton method to solve the following strongly convex optimization problem:

$$(1.1) \qquad \min_{u \in U} \frac{1}{2}\|Su - z\|_Y^2 + \frac{\alpha}{2}\|u\|_U^2 + g(u)$$

where U, Y are Hilbert spaces, $S \in \mathcal{L}(U, Y)$, $z \in Y$, $\alpha > 0$, $g : U \to \bar{\mathbb{R}}$ is proper, convex, lower semicontinuous with $0 \in \mathrm{dom}\, g$.

Problem (1.1) is a nice convex problem and its properties are well understood. However, the local (unglobalized) semismooth Newton method converges globally only for relatively large values of α.

Following [18, 42, 25], we propose to apply the semismooth Newton method to the dual problem of (1.1), where the globalization uses the dual objective as merit function. Due to the structure of (1.1), the dual problem has a continuously differentiable objective functional, which enables a particularly nice convergence analysis. As this analysis crucially relies on the strong convexity of the primal problem, it cannot be extended to non-convex or not strongly convex problems. Hence, the proposed method is intended to be used as a subproblem solver in, e.g., proximal point methods [25] or iterated Bregman methods [28].

*Institut für Mathematik, Universität Würzburg, 97074 Würzburg, Germany, daniel.wachsmuth@uni-wuerzburg.de,
ORCID: 0000-0001-7828-5614

Interestingly, in the implementation only knowledge about the Moreau-Yosida envelope and the proximal operator of g is needed. Using the proximal operator, one can prove that $u \in U$ solves (1.1) if and only if $u \in U$ solves the non-smooth equation

$$(1.2) \qquad u = \text{prox}_{g/\alpha}\left(-\frac{1}{\alpha}S^*(Su - z)\right),$$

where S^* denotes the Hilbert space adjoint of S. In order to solve (1.1) numerically, the most common approach is to apply the semismooth Newton method to the non-smooth equation (1.2). In this paper, we propose to solve the equation

$$(1.3) \qquad \xi - z + S\,\text{prox}_{g/\alpha}(S^*(\xi/\alpha)) = 0,$$

which is related to (1.2) via $\xi = z - Su$. The advantage of this approach is that the derivative of the operator appearing in (1.3) is a self-adjoint map, which is not the case for (1.2). In addition, (1.3) is equivalent to $\nabla\Phi(\xi) = 0$, where Φ is the dual function to (1.1), which will be used as merit function.

A third approach to solve (1.1) is based on the so-called normal map [33], where one solves

$$(1.4) \qquad p - S^*(S\,\text{prox}_{g/\alpha}(-p/\alpha) - z) = 0$$

for $p \in U$. Clearly, p is related to ξ by $p = -S^*\xi$. Again the derivative of the operator appearing in (1.4) is not self-adjoint.

This article is written 25 years after the publication of the seminal contributions [39] (submitted 2000) and [14] (submitted 2001) on semismooth Newton methods applied to optimal control problems. Still the literature on globalized semismooth Newton methods for infinite-dimensional problems is rather scarce. In fact, to the best of our knowledge, only the following contributions develop a global convergent method applicable to (1.1). In [18] the globalization based on (1.3) was investigated for the special case, where g is the indicator function associated with box constraints. In the monograph [40] a trust-region globalization of the semi-smooth Newton method applied to the non-smooth equation (1.2) is analyzed. Since continuous differentiability of g is required, it is not applicable to choices like $g(u) = \|u\|_{L^1(\Omega)}$. A similar method was analyzed for the special case of a parabolic control problem in [20]. In [8] a globalized method is introduced, where the convergence analysis relies on strong assumptions on the sequence of chosen derivatives. For ℓ^1-regularized problems, a globalized semismooth Newton method can be found in [12]. One major drawback of the convergence analysis in [40, 8, 12] is that they assume the existence of strong accumulation points of the sequence of iterates. While it is natural to expect that the sequence of iterates is bounded, this only guarantees the existence of *weakly* converging subsequences. However, *strong* convergence of iterates has to be verified in order to prove transition to fast local convergence.

Let us comment on related work on globalized semismooth Newton methods. The recent work [1] introduces a semismooth Newton method to solve contractive fixed point problems. Due to the contraction assumption, no damping or globalization is needed. In [29, 30] globalized proximal Newton methods that solve minimization problems in Hilbert spaces are investigated. The use of the normal map equation was proposed in [17] together with a damped Newton method without convergence analysis. Global convergence of the unregularized semismooth Newton method applied to problems with box constraints was proven in [3, 14] for discretized problems and [22] in the infinite-dimensional setting provided α is large relative to $\|S\|$. For a review of early results on (globalized) semismooth Newton methods we refer to [31]. Recent contributions on globalized semismooth Newton methods to minimize the sum of a differentiable and a non-smooth function in finite dimensions can be found in [26, 9].

The goal of this article is to give an almost self-contained analysis of the global convergence of the globalized semismooth Newton method. A few facts about subdifferentials and proximal operators

are collected in Section 2.1. The dual problem to (1.1) is introduced in Section 2.2. Due to the problem structure, no duals have to be computed for the implementation, only the proximal operator of g is needed. For the transition to fast convergence, a second order Taylor expansion is essential, which is introduced in Section 3. As in [42, 25] we use a conjugate gradient method as solver for the Newton equations. In order to prove that all iterates of the CG method are descent directions, one elementary result is needed, which is provided in Section 4. The main section is Section 5, which contains the globalized semismooth Newton method Algorithm 2. We prove *strong* convergence of the iterates in Lemma 5.10, and the transition to fast local convergence in Theorem 5.13. Section 6 shows that commonly studied examples admit a second order Taylor expansion. As we show in Section 7, the application of the semismooth Newton method to (1.3) enables the use of the variational discretization [15] similar to the arguments in [17]. The paper concludes with numerical experiments in Section 8. The results indicate that the variational discretization leads to better performance of the semismooth Newton method for small values of α in some examples.

2 PRELIMINARIES ON THE PRIMAL AND DUAL PROBLEMS

Lemma 2.1. *The problem* (1.1) *is uniquely solvable. An element* $\bar{u} \in U$ *is a solution of* (1.1) *if and only if*

$$(2.1) \qquad\qquad 0 \in S^*(S\bar{u} - z) + \alpha\bar{u} + \partial g(\bar{u}).$$

Proof. Existence can be proven by classical arguments, see, e.g., [2, Theorem 11.10]. Let $\bar{u}$ be a solution of (1.1). Since the first two addends of (1.1) are continuous, we can apply the subgradient sum-rule twice to obtain (2.1). The sufficiency of (2.1) follows from convexity. □

2.1 CONJUGATE FUNCTIONS AND PROXIMAL OPERATORS

Definition 2.2. Let X be a Hilbert space, $h : X \to \bar{\mathbb{R}}$ proper, convex, lower semicontinuous. The convex conjugate h^* of h is defined by

$$h^*(\zeta) := \sup_{x \in X} \langle \zeta, x \rangle - h(x).$$

The Moreau-Yosida envelope of h is defined by

$$\mathrm{env}_h(\zeta) := \min_{x \in X} \frac{1}{2}\|x - \zeta\|_X^2 + h(x),$$

and the proximal operator associated to h is defined by

$$\mathrm{prox}_h(\zeta) := \arg\min_{x \in X} \frac{1}{2}\|x - \zeta\|_X^2 + h(x).$$

For properties of Moreau-Yosida envelopes and proximal operators, we refer to [2, Section 12.4]. Convex conjugates are developed in [2, Chapter 13].

Proposition 2.3 ([2, Proposition 12.26]). *Let X be a Hilbert space, $h : X \to \bar{\mathbb{R}}$ proper, convex, lower semicontinuous. Let $x, \zeta \in X$. Then $x = \mathrm{prox}_h(\zeta)$ if and only if $0 \in x - \zeta + \partial h(x)$.*

Proposition 2.4 ([2, Proposition 12.28]). *Let X be a Hilbert space, $h : X \to \bar{\mathbb{R}}$ proper, convex, lower semicontinuous. Let $\zeta_1, \zeta_2 \in X$. Then $\|\mathrm{prox}_h(\zeta_1) - \mathrm{prox}_h(\zeta_2)\|_X \le \|\zeta_1 - \zeta_2\|_X$.*

Proposition 2.5 ([2, Proposition 12.30]). *Let X be a Hilbert space, $h : X \to \bar{\mathbb{R}}$ proper, convex, lower semicontinuous. Then $\mathrm{env}_h : X \to \mathbb{R}$ is Fréchet differentiable with*

$$\nabla \mathrm{env}_h(\zeta) = \zeta - \mathrm{prox}_h(\zeta).$$

Lemma 2.6. *Let X be a Hilbert space, $h : X \to \bar{\mathbb{R}}$ proper, convex, lower semicontinuous. Define*

$$H^* := \left(\frac{1}{2} \| \cdot \|_X^2 + h \right)^*.$$

Then $H^* : X \to \mathbb{R}$ satisfies

$$H^*(\zeta) = \frac{1}{2} \|\zeta\|_X^2 - \mathrm{env}_h(\zeta)$$

and is Fréchet differentiable with

$$\nabla H^*(\zeta) = \mathrm{prox}_h(\zeta).$$

If $0 \in \mathrm{dom}\, h$ then H^* is bounded from below.

Proof. By definition, we have

$$
\begin{aligned}
H^*(\zeta) &= \sup_{x \in X} \langle \zeta, x \rangle - \frac{1}{2} \|x\|_X^2 - h(x) \\
&= - \inf_{x \in X} \left(\frac{1}{2} \|x - \zeta\|_X^2 + h(x) - \frac{1}{2} \|\zeta\|_X^2 \right) \\
&= - \mathrm{env}_h(\zeta) + \frac{1}{2} \|\zeta\|_X^2.
\end{aligned}
$$

Due to Proposition 2.5, it follows that H^* is Fréchet differentiable, and the expression of ∇H^* follows. Boundedness from below is a direct consequence of the definition of H^*. $\qquad\square$

2.2 OPTIMALITY CONDITIONS FOR THE DUAL PROBLEM

We define the functions $F : Y \to \mathbb{R}$ and $G : U \to \bar{\mathbb{R}}$ by

$$F(y) = \frac{1}{2} \|y - z\|_Y^2, \quad G(u) := \frac{\alpha}{2} \|u\|_U^2 + g(u).$$

Let us introduce the Fenchel dual problem to (1.1), see, e.g., [2, Section 15.3]. The dual problem of (1.1) is given by

$$(2.2) \qquad \min_{\xi \in Y} F^*(-\xi) + G^*(S^*\xi) =: \Phi(\xi)$$

where

$$F^*(\zeta) = \frac{1}{2} \|\zeta + z\|_Y^2 - \frac{1}{2} \|z\|_Y^2$$

and

$$G^*(u) = \frac{1}{2\alpha} \|u\|_U^2 - \alpha \, \mathrm{env}_{g/\alpha}(u/\alpha).$$

Due to the structure of the problem both functions F^* and G^* can be evaluated without the need to compute conjugates, only the Moreau-Yosida envelope of g/α is needed.

Lemma 2.7. *The dual problem (2.2) is uniquely solvable. Let $\bar{\xi} \in Y$. Then $\bar{\xi}$ solves (2.2) if and only if*

$$\nabla \Phi(\bar{\xi}) = \bar{\xi} - z + S \, \mathrm{prox}_{g/\alpha}(S^*\bar{\xi}/\alpha) = 0.$$

Proof. The function Φ is continuous and convex. Since F^* is coercive, existence of solutions can be proven easily. Uniqueness of minimizers is a consequence of the strict convexity of F^*. The dual problem (2.2) is the minimization problem of the convex and Fréchet differentiable function Φ. Then $\nabla \Phi(\bar{\xi}) = 0$ is necessary and sufficient for optimality of $\bar{\xi}$. The expression of the derivative is a consequence of Proposition 2.5. $\qquad\square$

Lemma 2.8. *If $\bar{u} \in U$ solves the primal problem* (1.1) *then $\bar{\xi} := -(S\bar{u} - z)$ solves* (2.2). *Conversely, if $\bar{\xi} \in Y$ solves* (2.2) *then $\bar{u} := \mathrm{prox}_{g/\alpha}(S^*\bar{\xi}/\alpha)$ solves* (1.1).

Proof. Due to Proposition 2.3, the optimality condition (2.1) is equivalent to

$$\bar{u} = \mathrm{prox}_{g/\alpha}\left(-\frac{1}{\alpha}S^*(S\bar{u} - z)\right).$$

Then the claim follows by elementary computations.   □

3 SECOND-ORDER SEMISMOOTHNESS

For the proof of transition from global convergence to fast local convergence, a second-order Taylor expansion of the function G^* is indispensible [7]. Following [29], we introduce the notion of second-order semismoothness.

Let X be a Banach space. We will denote the space of bilinear forms from $X \times X$ to $\mathbb{R}$ by $\mathcal{L}^{(2)}(X, \mathbb{R})$.

Definition 3.1. Let $T : X \to \mathbb{R}$ be continuously differentiable with derivative T'. Let $\partial^2 T$ be a set-valued map $\partial^2 T : X \rightrightarrows \mathcal{L}^{(2)}(X, \mathbb{R})$ with non-empty images. Then T is called second-order semismooth with respect to $\partial^2 T$ if for all $x \in X$

$$(3.1) \qquad \sup_{M \in \partial^2 T(x+h)} \left| T(x + h) - T(x) - T'(x)h - \frac{1}{2}M(h, h) \right| = o(\|h\|_X^2)$$

and

$$(3.2) \qquad \sup_{M \in \partial^2 T(x+h)} \|T'(x + h) - T'(x) - M(h, \cdot)\|_{X^*} = o(\|h\|_X)$$

for $\|h\|_X \to 0$, where $M(h, \cdot)$ denotes the element of X^* given by $v \mapsto M(h, v)$.

Here, second-order semismoothness refers to the availability of a second-order Taylor expansion (3.1), which is different to the use in [40], where higher-order semismoothness refers to stronger estimates of the first-order expansion (3.2) replacing $o(\|h\|_X)$ for $O(\|h\|_X^{1+\alpha})$ with $\alpha \in (0, 1]$.

If X is a Hilbert space, we can use operators $M \in \mathcal{L}(X, X)$ instead of bilinear forms, and replace $M(h, h)$ and $M(h, \cdot)$ by $\langle Mh, h \rangle$ and Mh. In [29] an example is presented to show that (3.1) does not imply (3.2). Let us prove a sufficient condition for (3.1) which is only based on assumptions on T'.

Lemma 3.2. *Let $T : X \to \mathbb{R}$ be continuously differentiable with Bouligand differentiable T', i.e., T' is directionally differentiable and for all $x \in X$ it holds*

$$\|T'(x + h) - T'(x) - T''(x; h)\|_{X^*} = o(\|h\|_X).$$

Let $\partial^2 T : X \rightrightarrows \mathcal{L}^{(2)}(X, \mathbb{R})$ be given such that (3.2) *is satisfied for all $x \in X$. Then* (3.1) *is satisfied for all $x \in X$, and hence, T is second-order semismooth with respect to $\partial^2 T$.*

Proof. For $M \in \partial^2 T(x + h)$, we have the expansion

$$T(x + h) - T(x) - T'(x)h - \frac{1}{2}M(h, h) =$$

$$\left(T(x + h) - T(x) - \frac{1}{2}[T'(x) + T'(x + h)]h\right)$$

$$+ \frac{1}{2}\left(T'(x + h)h - T'(x)h - M(h, h)\right).$$

Due to (3.2), the second addend is of order $o(\|h\|_X^2)$ for $h \to 0$. Hence, it suffices to prove

$$\left| T(x + h) - T(x) - \frac{1}{2}[T'(x) + T'(x + h)]h \right| = o(\|h\|_X^2).$$

Take $x \in X$. Let $\epsilon > 0$. Then there is $\rho > 0$ such that $\|T'(x + h) - T'(x) - T''(x; h)\|_{X^*} \le \epsilon \|h\|_X$ for all h with $\|h\|_X < \rho$. Take h with $\|h\|_X < \rho$. Denote

$$R_t := T'(x + th) - T'(x) - tT''(x; h),$$

which implies $\|R_t\|_{X^*} \le \epsilon t \|h\|_X$ for all $t \in [0, 1]$. Then we can write

$$
\begin{aligned}
T(x + h) - T(x) - \frac{1}{2}[T'(x) + T'(x + h)]h &= \int_0^1 [T'(x + sh) - T'(x)]h \, ds - \frac{1}{2}[T'(x + h) - T'(x)]h \\
&= \int_0^1 [sT''(x; h) + R_s]h \, ds - \frac{1}{2}T''(x; h) - \frac{1}{2}R_1 h \\
&= \int_0^1 R_s h \, ds - \frac{1}{2}R_1 h
\end{aligned}
$$

so that

$$\left| T(x + h) - T(x) - \frac{1}{2}[T'(x) + T'(x + h)]h \right| \le \epsilon \|h\|_X^2$$

which finishes the proof. $\qquad\square$

In the smooth case, we have the following result.

Corollary 3.3 ([29, Proposition 5]). *Let $T : X \to \mathbb{R}$ be twice continuously differentiable. Then T is second-order semismooth with respect to $\partial^2 T(x) := \{T''(x)\}$.*

4 CONJUGATE GRADIENTS

Let us recall the celebrated conjugate gradient method to solve linear systems with positive definite operators. Let X be a Hilbert space, $A \in \mathcal{L}(X, X)$ self-adjoint and positive definite, and $b \in X$. The CG algorithm to solve $Ax = b$ in the notation from [38, Section 38] can be found in Algorithm 1.

Algorithm 1 CG algorithm to solve $Ax = b$ with tolerance δ_{CG}

Input: tolerance $\delta_{\mathrm{CG}} > 0$.
Set $x_0 = 0$, $r_0 = b$, $p_0 = r_0$, $k = 0$.
while $\|r_k\| > \delta_{\mathrm{CG}}$ **do**
$\quad \alpha_k = \frac{\|r_k\|_X^2}{\langle Ap_k, p_k \rangle}$
$\quad x_{k+1} = x_k + \alpha_k p_k$
$\quad r_{k+1} = r_k - \alpha_k Ap_k$
$\quad \beta_k = \frac{\|r_{k+1}\|_X^2}{\|r_k\|_X^2}$
$\quad p_{k+1} = r_{k+1} + \beta_k p_k$
$\quad$ Set $k := k + 1$
end while

The classical orthogonality relations and convergence guarantees transfer to the Hilbert space case, see, e.g., [6]. If $A = I + K$ with K compact we have r-superlinear convergence of $x_k \to A^{-1}b$ [41]. These results imply that for positive tolerance δ_{CG}, algorithm Algorithm 1 terminates after finitely many steps.

In the following, we need an estimate of $\langle x_k, b\rangle$. A similar result can be found in [42, Lemma 3.1]. We give a simplified proof.

Lemma 4.1. *Let $A \in \mathcal{L}(X,X)$ be self-adjoint and positive definite, and $b \in X$. Let (x_k) be iterates of Algorithm 1. Then*

$$\langle x_k, b\rangle \geq \frac{1}{\|A\|_{\mathcal{L}(X,X)}} \|b\|_X^2.$$

Proof. For $k = 1$, we have $x_0 = 0$ and $p_0 = b$ so that

$$\langle x_1, b\rangle = \alpha_0 \|b\|_X^2 = \frac{\|b\|_X^4}{\langle Ab, b\rangle} \geq \frac{1}{\|A\|_{\mathcal{L}(X,X)}} \|b\|_X^2.$$

Due to the orthogonality relations of the residuals, [38, (38.4)], we have $\langle r_k, b\rangle = \langle r_k, r_0\rangle = 0$ for all $k \geq 1$. Then we get for the conjugate directions

$$\langle p_{k+1}, b\rangle = \langle r_{k+1}, b\rangle + \beta_k \langle p_k, b\rangle = \beta_k \langle p_k, b\rangle,$$

so that $\langle p_{k+1}, b\rangle = \prod_{i=0}^k \beta_i \langle p_0, b\rangle$. Since $p_0 = r_0 = b$ and $\beta_i > 0$, this implies $\langle p_k, b\rangle \geq 0$ for all $k \geq 0$. Using this in the update formula for x_{k+1}, we find

$$\langle x_{k+1}, b\rangle = \langle x_k, b\rangle + \alpha_k \langle p_k, b\rangle \geq \langle x_k, b\rangle,$$

so that the sequence $(\langle x_k, b\rangle)$ is monotonically increasing.   □

5 SEMISMOOTH NEWTON

Here, we want to investigate a semismooth Newton method to minimize the function Φ in (2.2), which can be written as

$$(5.1) \qquad \Phi(\xi) = \frac{1}{2}\|\xi - z\|_Y^2 - \frac{1}{2}\|z\|_Y^2 + \frac{1}{2\alpha}\|S^*\xi\|_U^2 - \alpha\,\mathrm{env}_{g/\alpha}(S^*\xi/\alpha).$$

Its gradient is given by Lemma 2.7 as

$$(5.2) \qquad \nabla\Phi(\xi) = \xi - z + S\,\mathrm{prox}_{g/\alpha}(S^*\xi/\alpha).$$

Corollary 5.1. *The map Φ is continuously differentiable. In addition, there is $L > 0$ such that $\|\nabla\Phi(y_1) - \nabla\Phi(y_2)\|_Y \leq L\|y_1 - y_2\|_Y$ for all $y_1, y_2 \in Y$. Consequently,*

$$|\Phi(y + h) - \Phi(y) - \Phi'(y)h| \leq \frac{L}{2}\|h\|_Y^2.$$

Proof. Differentiability of Φ is a consequence of the differentiability of $\mathrm{env}_{g/\alpha}$ by Proposition 2.5. By Proposition 2.4, the proximal operator $\mathrm{prox}_{g/\alpha}$ is Lipschitz continuous. Hence, $\nabla\Phi$ is Lipschitz continuous. The claimed estimate follows is then a consequence of this Lipschitz property.   □

Let us define for convenience

$$(5.3) \qquad \Psi(\xi) := \frac{1}{2\alpha}\|S^*\xi\|_U^2 - \alpha\,\mathrm{env}_{g/\alpha}(S^*\xi/\alpha),$$

so that $\nabla\Psi(\xi) = S\,\mathrm{prox}_{g/\alpha}(S^*\xi/\alpha)$. Throughout this section, we assume that the following assumption is satisfied.

Assumption 1. There is a set-valued map $\partial^2\Psi : Y \rightrightarrows \mathcal{B}$, where

$$\mathcal{B} := \{M \in \mathcal{L}(Y,Y) : \ M = M^*, \ \langle My, y\rangle \geq 0 \ \forall y \in Y\},$$

such that Ψ given by (5.3) is second-order semismooth with respect to $\partial^2\Psi$, see Definition 3.1. In addition, $\partial^2\Psi$ is bounded, i.e., there is $c > 0$ such that for all $y \in Y$ and $M \in \partial^2\Psi(y)$ it holds

$$\|M\|_{\mathcal{L}(Y,Y)} \leq c.$$

Let us emphasize that in the algorithm we will not need to know the precise values of the Lipschitz constant L or the bound c. For the particular case of Nemyzki operators on L^2 spaces, validity of Assumption 1 is discussed in Section 6.

Corollary 5.2. Φ *is second-order semismooth with respect to* $\partial^2\Phi := I + \partial^2\Psi$.

Corollary 5.3. *Let* $M \in \partial^2\Phi(y)$ *for some* $y \in Y$. *Then* $M \in \mathcal{L}(Y,Y)$ *is continuously invertible with* $\|M\|_{\mathcal{L}(Y,Y)} \leq 1 + c$ *and* $\|M^{-1}\|_{\mathcal{L}(Y,Y)} \leq 1$. *Moreover,* $\langle Md, d\rangle \geq \|d\|_Y^2$ *for all* $d \in Y$.

The core of the semismooth Newton methods is the Newton equation to solve a linearization of $\nabla\Phi(\xi) = 0$. Given an iterate $\xi_k \in Y$, we choose $M_k \in \partial^2\Psi(\xi_k)$, and determine $d_k \in Y$ as a solution of the Newton equation

$$(5.4) \qquad (I + M_k)d_k = -\nabla\Phi(\xi_k) = -\left(\xi_k - z + S\operatorname{prox}_{g/\alpha}(S^*\xi_k/\alpha)\right).$$

In the upcoming convergence analysis, we will use both remainder estimates (3.1) and (3.2). The first-order estimate is an ingredient of the classical local convergence proof of semismooth Newton [40], while the second-order estimate is needed in the proof of transition to fast local convergence with full steps ($t_k = 1$).

Algorithm 2 Globalized inexact semismooth Newton method

    Choose $\delta_{\text{tol}} > 0$, $\sigma \in (0, 1/2)$, $\beta \in (0,1)$, $\eta \geq 0$, $\tau \in (0,1]$.   ▷ Parameters

    Choose $\xi_0 \in Y$, set $k := 0$.   ▷ Initialization

    while $\|\nabla\Phi(\xi_k)\|_Y > \delta_{\text{tol}}$ **do**

        Choose $M_k \in \partial^2\Phi(\xi_k)$, see Corollary 5.2.

        Compute $d_k \in Y$ such that   ▷ Inexact solve of Newton equation

$$(5.5) \qquad\qquad \|M_k d_k + \nabla\Phi(\xi_k)\|_Y \leq \eta\|\nabla\Phi(\xi_k)\|_Y^{1+\tau}$$

        using Algorithm 1 with tolerance   ▷ CG

$$\delta_{\text{CG}} := \eta\|\nabla\Phi(\xi_k)\|_Y^{1+\tau}.$$

        Compute the smallest $l_k \in \mathbb{N}_0$ such that   ▷ Armijo linesearch

$$(5.6) \qquad\qquad \Phi(\xi_k + \beta^{l_k}d_k) - \Phi(\xi_k) \leq \sigma\beta^{l_k}\langle d_k, \nabla\Phi(\xi_k)\rangle$$

        Set $t_k := \beta^{l_k}$, $\xi_{k+1} := \xi_k + t_k d_k$, $k := k + 1$.

    end while

Algorithm 2 is basically [25, Algorithm 4.1]. The convergence proof in the finite-dimensional case is given in [25, Theorem 4.1]. For the special case of box constraints and exact Newton solves ($\eta = 0$), Algorithm 2 is equivalent to [18, Algorithm 4.2].

Lemma 5.4. *Algorithm 2 is well-defined.*

Proof. Suppose $\nabla\Phi(\xi_k) \neq 0$. Due to the special form of the Newton derivatives $M_k \in \partial^2\Phi(\xi_k)$ in Corollary 5.2, it follows directly that the operators M_k are self-adjoint, positive definite, and invertible. The direction $-M_k^{-1}\nabla\Phi(\xi_k)$, satisfies (5.5). Then Algorithm 1 terminates after finitely many steps. Due to Lemma 4.1 the returned element d_k is a descent direction to the Fréchet differentiable function Φ. Under this condition, the Armijo linesearch terminates after finitely many steps. $\square$

The termination criterion is motivated by the following bound, see also [18, Lemma 4.3 (iii)].

Lemma 5.5. *Let $\bar{\xi} \in Y$ be such that $\nabla\Phi(\bar{\xi}) = 0$. Let $y \in Y$. Then*

$$\|y - \bar{\xi}\|_Y \leq \|\nabla\Phi(y)\|_Y.$$

Proof. Since $y \mapsto S\,\mathrm{prox}_{g/\alpha}(S^*\bar{\xi}/\alpha)$ is monotone as the derivative of the convex function Ψ, it follows

$$\langle \nabla\Phi(y) - \nabla\Phi(\bar{\xi}), y - \bar{\xi}\rangle \geq \|\bar{\xi} - y\|_Y^2,$$

which proves the claim using the Cauchy-Schwarz inequality. $\square$

The inexact Newton directions are descent directions of Φ. In addition, we have the following estimate on the directional derivative.

Corollary 5.6. *Let (ξ_k) be a sequence of iterates of Algorithm 2 together with the search directions (d_k) satisfying (5.5). Then it holds*

$$(5.7) \qquad \langle d_k, \nabla\Phi(\xi_k)\rangle \leq -\frac{1}{1+c}\|\nabla\Phi(\xi_k)\|_Y^2,$$

where c is from Assumption 1.

Proof. Lemma 4.1 implies $\langle d_k, -\nabla\Phi(\xi_k)\rangle \geq \frac{1}{\|M_k\|_{\mathcal{L}(Y,Y)}}\|\nabla\Phi(\xi_k)\|_Y^2$, and the claim follows with Corollary 5.3. $\square$

In the sequel, we assume that the algorithm performs infinitely many steps and generates a sequence of iterates (ξ_k). We start the convergence analysis with the following basic lemma, which uses the properties of the Armijo linesearch.

Lemma 5.7. *Let (ξ_k) be a sequence of iterates of Algorithm 2 together with the step sizes (t_k) and search directions (d_k). Then*

$$t_k\langle d_k, \nabla\Phi(\xi_k)\rangle \to 0.$$

Proof. Due to Lemma 2.6, G^*, Ψ, and Φ are bounded from below, hence $(\Phi(\xi_k))$ is monotonically decreasing and converging. The claim is now a consequence of (5.6). $\square$

Since we have not proven until now that (ξ_k) has strongly converging subsequences, we cannot conclude like in finite-dimensions, where $\nabla\Phi(\xi_k) \to 0$ along a strongly converging subsequence follows from uniform continuity of $\nabla\Phi$.

Using Corollary 5.1, we will prove a lower bound of t_k next. There, we will use the global Lipschitz continuity of $\nabla\Phi$. If we would know that (ξ_k) has strongly converging subsequences then local Lipschitz continuity of $\nabla\Phi$ would be enough.

Lemma 5.8. *Let (ξ_k) be a sequence of iterates of Algorithm 2 together with the step sizes (t_k) and search directions (d_k). Then there is a constant $\sigma' > 0$ such that $t_k < 1$ implies*

$$t_k \geq \sigma'\frac{|\langle d_k, \nabla\Phi(\xi_k)\rangle|}{\|d_k\|_Y^2}.$$

Proof. Let $t > 0$. Then

$$\Phi(\xi_k + t d_k) - \Phi(\xi_k) - \sigma t \langle d_k, \nabla\Phi(\xi_k) \rangle \le (1 - \sigma) t \langle d_k, \nabla\Phi(\xi_k) \rangle + \frac{L t^2}{2} \|d_k\|_Y^2.$$

The expression on the right-hand side is non-positive for all t satisfying

$$0 \le t \le (1 - \sigma) \frac{|\langle d_k, \nabla\Phi(\xi_k) \rangle|}{\frac{L}{2} \|d_k\|_Y^2}.$$

If $t_k < 1$ satisfies the Armijo condition (5.6) then $l_k \ge 1$, and $\beta^{l_k - 1} = t_k/\beta$ violates the Armijo condition (5.6). Consequently, t_k/β is larger than the quantity on the right-hand side in the above inequality, which implies the claim. $\qquad\square$

Now we are in the position to prove $\nabla\Phi(\xi_k) \to 0$.

Lemma 5.9. *Let (ξ_k) be a sequence of iterates of Algorithm 2. Then it holds $\liminf_{k\to\infty} t_k > 0$ and $\nabla\Phi(\xi_k) \to 0$ in Y.*

Proof. By Lemma 5.7, we have $t_k \langle d_k, \nabla\Phi(\xi_k) \rangle \to 0$. If $\liminf_{k\to\infty} t_k > 0$ then (5.7) implies $\|\nabla\Phi(\xi_k)\|_Y \to 0$. Consider the case $\liminf_{k\to\infty} t_k = 0$. By extracting a subsequence, we can assume $t_k \to 0$ and $t_k < 1$ for all k. Then from $t_k \to 0$ and $t_k \langle d_k, \nabla\Phi(\xi_k) \rangle \to 0$ we get with the help of Lemma 5.8

$$\frac{|\langle d_k, \nabla\Phi(\xi_k) \rangle|}{\|d_k\|_Y^2} + \frac{|\langle d_k, \nabla\Phi(\xi_k) \rangle|^2}{\|d_k\|_Y^2} \to 0,$$

so that (5.7) implies

$$\frac{\|\nabla\Phi(\xi_k)\|_Y^2}{\|d_k\|_Y^2} + \frac{\|\nabla\Phi(\xi_k)\|_Y^4}{\|d_k\|_Y^2} \to 0.$$

From the inexactness condition (5.5) and $\|M_k^{-1}\|_{\mathcal{L}(Y,Y)} \le 1$, we get

$$\|d_k\|_Y \le \|\nabla\Phi(\xi_k)\|_Y + \eta \|\nabla\Phi(\xi_k)\|_Y^{1+\tau}.$$

Dividing by $\|d_k\|_Y$ and passing to the limit $k \to \infty$ leads to the contradiction $1 \le 0$. Here, we used $\tau \le 1$, and $\frac{a^{1+\tau}}{b} = \left(\frac{a^2}{b}\right)^\tau \left(\frac{a}{b}\right)^{1-\tau}$. $\qquad\square$

Due to the structure of the problem, we have strong convergence of iterates.

Lemma 5.10. *Let (ξ_k) be a sequence of iterates of Algorithm 2. Then $\xi_k \to \bar{\xi}$ in Y, where ξ is the unique solution of $\Phi(\xi) = 0$. In addition, $d_k \to 0$ in Y.*

Proof. Let $\bar{\xi}$ be such that $\Phi(\bar{\xi}) = 0$. The existence of such $\bar{\xi}$ follows from Lemma 2.7. Lemma 5.5 implies $\|\bar{\xi} - \xi_k\|_Y \le \|\nabla\Phi(\xi_k)\|_Y$, the latter quantity tends to zero by Lemma 5.9. $\qquad\square$

The next step is to prove that the step size $t_k = 1$ is accepted for all $k \ge k_0$. Here, we follow the exposition in [7]. First, we establish superlinear convergence of the local inexact semismooth Newton method.

Lemma 5.11. *Let $\bar{\xi}$ be the unique solution of $\Phi(\bar{\xi}) = 0$. Let (ξ_k) be a sequence of iterates of Algorithm 2 with step sizes (t_k). Let $r_k := M_k d_k + \nabla\Phi(\xi_k)$. Then*

$$\|r_k\|_Y = o(\|\xi_k - \bar{\xi}\|_Y), \qquad \|\xi_k + d_k - \bar{\xi}\|_Y = o(\|\xi_k - \bar{\xi}\|_Y)$$

for $k \to \infty$.

Proof. Using the inexact Newton equation (5.5) and Corollary 5.1, we get

$$\|r_k\|_Y \leq \eta \|\nabla\Phi(\xi_k)\|_Y^{1+\tau} \leq \eta L \|\xi_k - \bar{\xi}\|_Y^{1+\tau} = o(\|\xi_k - \bar{\xi}\|_Y).$$

Using again (5.5), we find

$$\begin{aligned}
\|\xi_k + d_k - \bar{\xi}\|_Y &= \|M_k^{-1}(M_k(\xi_k - \bar{\xi}) - \nabla\Phi(\xi_k) + r_k)\|_Y \\
&\leq \|M_k^{-1}\|_{\mathcal{L}(Y,Y)} \|M_k(\xi_k - \bar{\xi}) - \nabla\Phi(\xi_k)\|_Y + \|r_k\|_Y \\
&= o(\|\xi_k - \bar{\xi}\|_Y),
\end{aligned}$$

where we have used Corollary 5.3 and the semismoothness of $\nabla\Phi$. □

Now, we will prove that the step size $t_k = 1$ will be accepted for large k. This proof crucially relies on the second-order expansion in the definition of second-order semismoothness (Definition 3.1).

Lemma 5.12. *Let $\bar{\xi}$ be the unique solution of $\Phi(\bar{\xi}) = 0$. Let (ξ_k) be a sequence of iterates of Algorithm 2 with step sizes (t_k). Then there is $k_0 > 0$ such that $t_k = 1$ for all $k \geq k_0$.*

Proof. See also [7, Theorem 3.3]. Using the simple inequalities

$$\|\xi_k - \bar{\xi}\|_Y \leq \|d_k\|_Y + \|\xi_k + d_k - \bar{\xi}\|_Y \leq \|\xi_k - \bar{\xi}\|_Y + 2\|\xi_k + d_k - \bar{\xi}\|_Y,$$

we get by Lemma 5.11

$$\|d_k\|_Y = \|\xi_k - \bar{\xi}\|_Y + o(\|\xi_k - \bar{\xi}\|_Y).$$

Similarly, we get from (5.5) an estimate of the directional derivative

$$\begin{aligned}
\langle d_k, \nabla\Phi(\xi_k)\rangle &\leq -\langle M_k d_k, d_k\rangle + \|r_k\|_Y \|d_k\|_Y \\
&\leq -\|d_k\|_Y^2 + o(\|\xi_k - \bar{\xi}\|_Y^2) \\
&= -\|\xi_k - \bar{\xi}\|_Y^2 + o(\|\xi_k - \bar{\xi}\|_Y^2).
\end{aligned}$$

Since $\xi_k + d_k \to \bar{\xi}$ in Y, we have the second-order expansion

$$\begin{aligned}
\Phi(\xi_k + d_k) - \Phi(\bar{\xi}) - \langle\nabla\Phi(\bar{\xi}), \xi_k + d_k - \bar{\xi}\rangle &= -\frac{1}{2}\langle H_k(\xi_k + d_k - \bar{\xi}), \xi_k + d_k - \bar{\xi}\rangle + o(\|\xi_k + d_k - \bar{\xi}\|_Y^2) \\
&= o(\|\xi_k - \bar{\xi}\|_Y^2),
\end{aligned}$$

where $H_k \in \partial^2\Phi(\bar{\xi}_k + d_k)$, which is bounded due to Corollary 5.2. Similarly, we prove using $\nabla\Phi(\bar{\xi}) = 0$

$$\Phi(\xi_k) - \Phi(\bar{\xi}) = -\frac{1}{2}\langle M_k(\xi_k - \bar{\xi}), \xi_k - \bar{\xi}\rangle + o(\|\xi_k - \bar{\xi}\|_Y^2).$$

Collecting all these estimates yields

$$\begin{aligned}
\Phi(\xi_k + d_k) - \Phi(\xi_k) - \frac{1}{2}\langle\nabla\Phi(\xi_k), d_k\rangle \\
= \frac{1}{2}\langle M_k(\xi_k - \bar{\xi}), \xi_k - \bar{\xi}\rangle - \frac{1}{2}\langle\nabla\Phi(\xi_k), d_k\rangle + o(\|\xi_k - \bar{\xi}\|_Y^2) \\
= \frac{1}{2}\langle M_k(\xi_k - \bar{\xi}), \xi_k + d_k - \bar{\xi} - d_k\rangle - \frac{1}{2}\langle\nabla\Phi(\xi_k), d_k\rangle + o(\|\xi_k - \bar{\xi}\|_Y^2) \\
= -\frac{1}{2}\langle M_k(\xi_k - \bar{\xi}) + \nabla\Phi(\xi_k) - \nabla\Phi(\bar{\xi}), d_k\rangle + o(\|\xi_k - \bar{\xi}\|_Y^2) \\
= o(\|\xi_k - \bar{\xi}\|_Y^2).
\end{aligned}$$

Using this in the descent condition (5.6), we get

$$\Phi(\xi_k + d_k) - \Phi(\xi_k) - \sigma\langle d_k, \nabla\Phi(\xi_k)\rangle = \left(\frac{1}{2} - \sigma\right)\langle d_k, \nabla\Phi(\xi_k)\rangle + o(\|\xi_k - \bar{\xi}\|_Y^2)$$

$$= -\left(\frac{1}{2} - \sigma\right)\|\xi_k - \bar{\xi}\|_Y^2 + o(\|\xi_k - \bar{\xi}\|_Y^2).$$

The right-hand side is less than zero for k large enough, which proves the claim. $\qquad\square$

Theorem 5.13. *Let Assumption 1 be satisfied. Let $\bar{\xi}$ be the unique solution of $\nabla\Phi(\bar{\xi}) = 0$. Let (ξ_k) be a sequence of iterates of Algorithm 2. Then $\xi_k \to \bar{\xi}$ q-superlinearly.*

Proof. Due to Lemma 5.12, there is $k_0 > 0$ such that $t_k = 1$ for all $k \geq k_0$. Then Lemma 5.11 implies $\|\xi_k + d_k - \bar{\xi}\|_Y = \|\xi_{k+1} - \bar{\xi}\|_Y = o(\|\xi_k - \bar{\xi}\|_Y)\|$ for $k \geq k_0$, which is the claim. $\qquad\square$

Let us conclude this section with several remarks.

Remark 5.14. The proof of transition to fast convergence in [18, Lemma 4.4] relied on stronger assumptions. There a condition on the measure of almost active sets was used, which implies strict complementarity. In addition, the operator S^* has to be continuous with values in $C^1(\bar{\Omega})$. Under these conditions $\partial^2\Psi$ is single-valued and Hölder continuous in a neighborhood of the solution [18, Lemma 3.4], which allows to prove the higher-order convergence rate $\|\xi_{k+1} - \bar{\xi}\|_Y \leq c\|\xi_k - \bar{\xi}\|_Y^{3/2}$.

Remark 5.15 (Newton equation in primal form). The element $\bar{\xi} \in Y$ solves

$$0 = \nabla\Phi(\bar{\xi}) = \bar{\xi} - z + S\operatorname{prox}_{g/\alpha}(S^*\bar{\xi}/\alpha)$$

if and only if $\bar{u} = \operatorname{prox}_{g/\alpha}(S^*\bar{\xi}/\alpha)$ solves

$$(5.8) \qquad\qquad 0 = \bar{u} - \operatorname{prox}_{g/\alpha}(-S^*(S\bar{u} - z)/\alpha).$$

Conversely, if $\bar{u}$ is a solution of (5.8) then $\bar{\xi} = z - S\bar{u}$ satisfies $\nabla\Phi(\bar{\xi}) = 0$. Let $u_k \in U$ be given. Applying the semismooth Newton method to (5.8) amounts to choosing $H_k \in \partial\operatorname{prox}_{g/\alpha}(-S^*(Su_k - z)/\alpha)$ and solving

$$(5.9) \qquad\qquad \delta u + \frac{1}{\alpha}H_k S^* S\delta u = -\left(u_k - \operatorname{prox}_{g/\alpha}(-S^*(Su_k - z)/\alpha)\right).$$

Let us set $\xi_k := z - Su_k$. If δu solves (5.9) then $\delta\xi := -S\delta u$ solves

$$\delta u - \frac{1}{\alpha}H_k S^*\delta\xi = -\left(u_k - \operatorname{prox}_{g/\alpha}(S^*\xi_k/\alpha)\right).$$

Multiplying by $-S$, we get

$$\delta\xi + \frac{1}{\alpha}S H_k S^*\delta\xi = -\left(\xi_k - z + S\operatorname{prox}_{g/\alpha}(S^*\xi_k/\alpha)\right),$$

which is the Newton equation (5.4) with $M_k := S^* H_k S$. Equation (5.9) is an equation with an unsymmetric operator. If H_k is suitably structured, one can transform the equation into an equation with a symmetric operator.

Remark 5.16 (Semismooth Newton applied to normal map). It is easy to show that $\bar{u} = \operatorname{prox}_{g/\alpha}(-\bar{p}/\alpha)$ is a solution of (1.1) if and only if $\bar{p} \in U$ solves

$$0 = \bar{p} - S^*(S\operatorname{prox}_{g/\alpha}(-\bar{p}/\alpha) - z).$$

The map $p \mapsto p - S^*(S \operatorname{prox}_{g/\alpha}(-p/\alpha) - z)$ is called normal map after [33]. Semismooth Newton methods based on this equation are considered in [27, 17]. In the recent article [26], the convergence analysis of a trust-region globalization of a semismooth Newton method applied to the normal map equation is developed in finite dimensions, where globalization is based on the residual in the normal equation.

Remark 5.17. In [10] a globalization of the active-set strategy [3] to solve discretized box constrained optimal control problems was proposed. However, the analysis does not carry over to the infinite-dimensional setting, as certain operators are not invertible in infinite dimensions [10, Remark 3].

Remark 5.18 (Non-quadratic functional). Let us briefly consider the following generalization of (1.1):

$$\min_{u \in X} F(Su) + \frac{\alpha}{2} \|u\|_U^2 + g(u),$$

where $F : Y \to \mathbb{R}$ is convex. Let F^* denote the convex conjugate of F. Under the assumption that F is strongly convex and twice continuously differentiable, we get that F^* is also strongly convex and twice continuously differentiable. Let us denote with $\nabla F^* := \nabla(F^*)$ its gradient. Then, we have $\nabla F(\nabla F^*(y)) = y$ for all $y \in Y$ by [2, Theorem 16.29]. Differentiating this identity yields $\nabla^2 F(\nabla F^*(y)) \nabla^2 F^*(y) = \mathrm{id}$ or $\nabla^2 F^*(y) = \nabla^2 F(\nabla F^*(y))^{-1}$. The resulting second-order derivative of Φ is given by $\partial^2 \Phi = \nabla^2 F^* + \partial^2 \Psi$. Then all the convergence analysis for the corresponding semismooth Newton algorithm remains true, only the constants in Lemma 5.5 and Corollary 5.3 have to be adapted.

6 SEMISMOOTHNESS OF NEMYZKI OPERATORS

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. In many optimization problems, the function g in (1.1) is defined as

$$(6.1) \qquad\qquad g(u) = \int_\Omega \tilde{g}(u(x)) \, d\mu$$

with a convex function $\tilde{g} : \mathbb{R} \to \bar{\mathbb{R}}$. Here, we follow the convention to set $g(u) := +\infty$ if $x \mapsto \tilde{g}(u(x))$ is not integrable. The resulting proximal operator is then a superposition operator on $U = L^2(\Omega)$.

Corollary 6.1. Let $\tilde{g} : \mathbb{R} \to \bar{\mathbb{R}}$ be proper, convex, and lower semicontinuous. Let g be defined in (6.1). Suppose there is $u \in L^2(\Omega)$ with $g(u) < +\infty$. Then $\operatorname{prox}_g : L^2(\Omega) \to L^2(\Omega)$ is given by

$$\operatorname{prox}_g(v)(x) = \operatorname{prox}_{\tilde{g}}(v(x)) \quad \text{for } \mu\text{-a.a. } x \in \Omega.$$

Proof. This follows from the results of [34, Section 3], in particular the characterization of $\partial g(u)$ as

$$\partial g(u) = \{d \in L^2(\Omega) : \ d(x) \in \partial \tilde{g}(u(x)) \text{ for } \mu\text{-almost all } x \in \Omega\},$$

see [34, Corollary 3E]. $\qquad\qquad\qquad \square$

6.1 PIECEWISE SMOOTH PROXIMALS

We will now work with piecewise smooth proximal operators.

Assumption 2. The function $\operatorname{prox}_{\tilde{g}/\alpha} : \mathbb{R} \to \mathbb{R}$ is piecewise smooth in the following sense: there are real numbers $v_1 < v_2 < \cdots < v_N$ and continuously differentiable functions $p_k : \mathbb{R} \to \mathbb{R}$, $k = 0 \dots N$, such that

$$\operatorname{prox}_{\tilde{g}/\alpha} = p_k \text{ on } [v_k, v_{k+1}],$$

where we set $v_0 = -\infty$, $v_{N+1} := +\infty$.

Lemma 6.2. *Let* $\text{prox}_{\tilde{g}/\alpha}$ *satisfy Assumption 2. Define the single-valued map*

$$\partial \text{prox}_{\tilde{g}/\alpha}(v) := p_k'(v) \ if \ v \in [v_k, v_{k+1}) \ for \ some \ k = 0 \dots N.$$

Then for all $v \in \mathbb{R}$, *we have*

$$\left| \text{prox}_{\tilde{g}/\alpha}(v + h) - \text{prox}_{\tilde{g}/\alpha}(v) - \partial \text{prox}_{\tilde{g}/\alpha}(v + h)h \right| = o(|h|)$$

and

$$| \text{prox}_{\tilde{g}/\alpha}(v + h) - \text{prox}_{\tilde{g}/\alpha}(v) - \text{prox}_{\tilde{g}/\alpha}'(v; h)| = o(|h|).$$

Proof. Take $v \in \mathbb{R}$, and h such that $\text{prox}_{\tilde{g}/\alpha}$ is continuously differentiable between v and $v + h$. Then both claims follow easily. $\qquad\square$

Clearly, this result can be generalized for $\text{prox}_{\tilde{g}/\alpha} : \mathbb{R}^n \to \mathbb{R}^n$ with infinitely many pieces, see also [40, Section 2.5.3].

In the next step, we prove that the derivative of the proximal operator of the integral functional g is semismooth from $L^q(\Omega)$ to $L^p(\Omega)$ for $p < q$. The proof of the following result is adapted from the proof of [29, Proposition 6], more general results can be found in [40, Section 3.3].

Lemma 6.3. *Let* $1 \le p < q \le \infty$ *be given. Let* $\mu(\Omega) < \infty$. *Let* $\text{prox}_{\tilde{g}/\alpha}$ *satisfy Assumption 2. Let* $v \in L^q(\Omega)$. *Define the measurable function* $\partial \text{prox}_{g/\alpha}(v)$ *for* μ-*almost all* $x \in \Omega$ *by*

$$\partial \text{prox}_{g/\alpha}(v)(x) := \partial \text{prox}_{\tilde{g}/\alpha}(v(x)).$$

Then we have $\partial \text{prox}_{g/\alpha}(v) \in L^\infty(\Omega)$,

$$\left\| \text{prox}_{g/\alpha}(v + h) - \text{prox}_{g/\alpha}(v) - \partial \text{prox}_{g/\alpha}(v + h) \cdot h \right\|_{L^p(\Omega)} = o(\|h\|_{L^q(\Omega)})$$

and

$$\| \text{prox}_{g/\alpha}(v + h) - \text{prox}_{g/\alpha}(v) - \text{prox}_{g/\alpha}'(v; h)\|_{L^p(\Omega)} = o(\|h\|_{L^q(\Omega)}).$$

Proof. Take $v \in L^q(\Omega)$. Let us define for $x \in \Omega$ and $h \in \mathbb{R} \setminus \{0\}$

$$r(x, h) := \frac{1}{h} \left(\text{prox}_{\tilde{g}/\alpha}(v(x) + h) - \text{prox}_{\tilde{g}/\alpha}(v(x)) - \partial \text{prox}_{\tilde{g}/\alpha}(v(x) + h)h \right).$$

In addition, we set $r(x, 0) = 0$ for $x \in \Omega$. Then $x \mapsto r(x, h(x))$ is μ-measurable for all measurable $h : \Omega \to \mathbb{R}$. In addition, $\lim_{h \to 0} r(x, h) = 0$ for almost all x by Lemma 6.2. Since $\text{prox}_{\tilde{g}/\alpha}$ is Lipschitz (Proposition 2.4) and consequently $|\partial \text{prox}_{\tilde{g}/\alpha}| \le 1$, we have for almost all x and all h the estimate $|r(x, h)| \le 2$.

Let (h_n) be a sequence in $L^q(\Omega)$ such that $h_n \to 0$ in $L^q(\Omega)$. Then there is a pointwise μ-a.e. converging subsequence (h_{n_k}). It follows $r(x, h_{n_k}(x)) \to 0$ for almost all x. Since r is uniformly bounded and $\mu(\Omega) < \infty$, we get $\|r(\cdot, h_{n_k})\|_{L^s(\Omega)} \to 0$ for all $s \in [1, \infty)$ by dominated convergence. It follows

$$\left\| \text{prox}_{g/\alpha}(v + h_{n_k}) - \text{prox}_{g/\alpha}(v) - \partial \text{prox}_{g/\alpha}(v + h_{n_k}) \cdot h_{n_k} \right\|_{L^p(\Omega)}$$
$$= \|r(\cdot, h_{n_k})h_{n_k}\|_{L^p(\Omega)} \le \|r(\cdot, h_{n_k})\|_{L^s(\Omega)} \|h_{n_k}\|_{L^q(\Omega)}$$

from the generalized Hölder inequality with $s < \infty$ such that $\frac{1}{p} = \frac{1}{s} + \frac{1}{q}$. Consequently,

$$\frac{\left\| \text{prox}_{g/\alpha}(v + h_{n_k}) - \text{prox}_{g/\alpha}(v) - \partial \text{prox}_{g/\alpha}(v + h_{n_k}) \cdot h_{n_k} \right\|_{L^p(\Omega)}}{\|h_{n_k}\|_{L^q(\Omega)}} \to 0.$$

Since the limit is independent of the chosen subsequence, the convergence along the whole sequence (h_k) follows.

The second claim can be proven with similar arguments, observing that $|\operatorname{prox}'_{\tilde{g}/\alpha}(v; h)| \le |h|$. $\qquad \square$

Note that we cannot get rid of the norm gap in the previous result. The necessity of this gap has been discussed in [40, Example 3.57], [19, Example 8.14] for semismoothness and in [11, Remark 4.4], [24] for Bouligand differentiability.

Example 6.4. Without the assumption $\mu(\Omega) < \infty$, the result of Lemma 6.3 is not true. Let

$$\tilde{g}(v) = \begin{cases} +\infty & \text{if } v < 0, \\ 0 & \text{if } v \ge 0. \end{cases}$$

Then $\operatorname{prox}_{\tilde{g}}(v) = \max(v, 0)$. Let $\Omega = (0, +\infty)$, and let μ be the Lebesgue measure. Take $v(x) := -\exp(-x)$, so that $v \in L^p(\mathbb{R})$ for all $p \in [1, +\infty]$. Define $h_n := \chi_{[n, +\infty)}(-2v)$. Then $\|h_n\|_{L^p(\Omega)} = \exp(-n)\|v\|_{L^p(\Omega)}$, so that it follows $\lim_{n \to \infty} \|h_n\|_{L^p(\Omega)} \to 0$ for all $p \in [1, +\infty]$. In addition,

$$\operatorname{prox}_g(v + h_n) - \operatorname{prox}_g(v + h_n) - \operatorname{prox}'_g(v; h + h_n) = \chi_{[n, +\infty)}(v + h_n) - 0 - 0 = \frac{1}{2}h_n.$$

It is easy to check that $\|h_n\|_{L^p(\Omega)} / \|h_n\|_{L^q(\Omega)}$ does not converge to zero for any choice of p, q.

Due to the norm gap appearing in Lemma 6.3, we cannot expect that $\operatorname{prox}_{g/\alpha}$ is semismooth as a mapping on $L^2(\Omega)$. In order that Assumption 1 is satisfied, we will need to rely on smoothing properties of S or S^*.

Lemma 6.5. Assume $\mu(\Omega) < \infty$. Set $U := L^2(\Omega)$. Let $\operatorname{prox}_{\tilde{g}/\alpha}$ satisfy Assumption 2. Let Y be a Hilbert space, $S \in \mathcal{L}(L^2(\Omega), Y)$ be given such that the Hilbert space adjoint S^* belongs to $\mathcal{L}(Y, L^q(\Omega))$ for some $q > 2$.

Then the map $\Psi : Y \to Y$ defined by (5.3) is second-order semismooth with respect to $\partial^2 \Psi : Y \rightrightarrows \mathcal{L}(Y, Y)$ defined by

$$(6.2) \qquad \qquad \partial^2 \Psi(\xi) := \left\{ \frac{1}{\alpha} S \partial \operatorname{prox}_{g/\alpha}(S^* \xi / \alpha) S^* \right\},$$

where $\partial \operatorname{prox}_{g/\alpha}$ is defined in Lemma 6.3. Moreover, Assumption 1 is satisfied.

In (6.2), the function $\partial \operatorname{prox}_{g/\alpha}$ is meant to act as multiplication operator on S^*, i.e.,

$$\left(\frac{1}{\alpha} S \partial \operatorname{prox}_{g/\alpha}(S^* \xi / \alpha) S^* \right)(v) = \frac{1}{\alpha} S \left(\partial \operatorname{prox}_{g/\alpha} \cdot S^* v \right),$$

where $\cdot$ denotes the pointwise multiplication of $\partial \operatorname{prox}_{g/\alpha} \in L^\infty(\Omega)$ and $S^* v \in L^q(\Omega)$.

Proof of Lemma 6.5. We show that $\partial^2 \Psi(\xi)$ from (6.2) satisfies the assumptions of Lemma 3.2. Let us recall $\nabla \Psi(\xi) = S \operatorname{prox}_{g/\alpha}(S^* \xi / \alpha)$. Let $v, h \in Y$ be given. Then Lemma 6.3 with $p = 2$ and the assumption on S^* imply

$$\left\| \operatorname{prox}_{g/\alpha}(S^*(v + h)) - \operatorname{prox}_{g/\alpha}(S^* v) - \partial \operatorname{prox}_{g/\alpha}(S^*(v + h)) \cdot S^* h \right\|_{L^2(\Omega)}$$
$$= o(\|S^* h\|_{L^q(\Omega)}) = o(\|h\|_Y).$$

Clearly this implies that $\partial^2 \Psi(\xi)$ satisfies (3.2). The Bouligand property of the directional derivative can be proven with analogous arguments. $\qquad \square$

6.2 EXAMPLES

Box constraints. Let $R > 0$ be given. Let us define $\tilde{g}$ by

$$\tilde{g}_{\text{box}}(u) = \begin{cases} 0 & \text{if } |u| \leq R, \\ +\infty & \text{if } |u| > R. \end{cases}$$

Then $\int_\Omega \tilde{g}_{\text{box}}(u)\,\mathrm{d}x < +\infty$ if and only if $|u(x)| \leq R$ μ-almost everywhere, that is if u satisfies the box constraints. In this case, the proximal operator is equal to the projection onto the associated feasible set,

$$\operatorname{prox}_{\tilde{g}_{\text{box}}}(v) = \max(-R, \min(+R, v)).$$

Clearly, this function is piecewise affine linear, and Assumption 2 is satisfied. For this particular choice of $\tilde{g}$, we have $\tilde{g}_{\text{box}}/\alpha = \tilde{g}_{\text{box}}$.

L^1-norm. Let us now consider the case

$$\tilde{g}_{L^1}(u) = \beta|u|.$$

Then $\int_\Omega \tilde{g}_{L^1}(u)\,\mathrm{d}\mu = \beta\|u\|_{L^1(\Omega)}$. The corresponding proximal operator is the so-called soft-thresholding operator given by

$$\operatorname{prox}_{\tilde{g}_{L^1}}(v) = \max(0, v - \beta) + \min(0, v + \beta) = \begin{cases} v + \beta & \text{if } v < -\beta, \\ 0 & \text{if } |v| \leq \beta, \\ v - \beta & \text{if } v > \beta. \end{cases}$$

This is again a piecewise affine linear map, so that Assumption 2 is satisfied. The proximal operator of $\tilde{g}_{L^1}/\alpha$ is obtained by replacing β with β/α in the formula for $\operatorname{prox}_{\tilde{g}_{L^1}}$.

Box constraints plus L^1-norm. If $\tilde{g} = \tilde{g}_{L^1}(u) + \tilde{g}_{\text{box}}(u)$, then it is not hard to check that

$$\operatorname{prox}_{\tilde{g}}(v) = \max(0, \min(v - \beta, R)) + \min(0, \max(v + \beta, -R)) = \begin{cases} -R & \text{if } v < -\beta - R \\ v + \beta & \text{if } -\beta - R \leq v < -\beta, \\ 0 & \text{if } |v| \leq \beta, \\ v - \beta & \text{if } \beta \leq v < \beta + R, \\ R & \text{if } v > \beta + R, \end{cases}$$

and hence, Assumption 2 is satisfied. Again, the proximal operator of $(\tilde{g}_{L^1} + \tilde{g}_{\text{box}})/\alpha$ is obtained by replacing β with β/α in the formula above.

L^2-ball constraint. Let us now consider g to be the indicator function of a ball in $L^2(\Omega)$ with radius $\gamma > 0$:

$$g(u) = \begin{cases} 0 & \text{if } \|u\|_{L^2(\Omega)} \leq \gamma \\ +\infty & \text{if } \|u\|_{L^2(\Omega)} > \gamma. \end{cases}$$

Again, the corresponding proximal map is the projection on the ball with radius γ, $\operatorname{prox}_g(v) = v \max(1, \frac{\gamma}{\|v\|_{L^2(\Omega)}})$. Let us choose $\partial \operatorname{prox}_g(v) \in \mathcal{L}(L^2(\Omega), L^2(\Omega))$ by

$$\partial \operatorname{prox}_g(v)h := \begin{cases} h & \text{if } \|v\|_{L^2(\Omega)} \leq \gamma \\ \dfrac{\gamma h}{\|v\|_{L^2(\Omega)}} - \dfrac{\gamma v\langle v, h\rangle_{L^2(\Omega)}}{\|v\|_{L^2(\Omega)}^3} & \text{if } \|q\|_{L^2(\omega)} > \gamma. \end{cases}$$

Then $T' = \operatorname{prox}_g$ and $T'' = \partial \operatorname{prox}_g$ satisfy condition (3.2) by [23, Proposition 4.1]. In addition, prox_g is Bouligand differentiable due to [23, Proposition 3.4].

Directional sparsity.   Here, $U = L^2(\Omega; L^2(I))$ where $I = (0, T)$.

$$g(u) = \beta \int_\Omega \left(\int_I |u(x, t)|^2 \, \mathrm{d}t \right)^{1/2} \mathrm{d}x$$

The proximal operator associated to g is given by

$$\mathrm{prox}_{g/\alpha}(u)(x, t) = \max\left(0, \|u(x, \cdot)\|_{L^2(I)} - \frac{\beta}{\alpha} \right) \frac{u(x, t)}{\|u(x, \cdot)\|_{L^2(I)}}.$$

Here, we used the notation

$$\|u(x, \cdot)\|_{L^2(I)} := \left(\int_I |u(x, t)|^2 \, \mathrm{d}t \right)^{1/2}.$$

This operator is semismooth from $L^p(\Omega; L^2(I))$ to $L^2(\Omega; L^2(I))$ for all $p > 2$, [27, Lemma 3.22], [13, Lemma 3.2]. In addition, it is easy to check that $\mathrm{prox}_{g/\alpha}$ is Bouligand differentiable using the chain rule for the Bouligand derivative [32, Corollary A.4].

Further examples.   Additional examples that satisfy these assumptions are multi-bang problems [4], where g is given as the convex hull of

$$u \mapsto \begin{cases} \beta u^2 & \text{if } u \in \mathbb{Z}, \\ +\infty & \text{otherwise.} \end{cases}$$

In addition, also vector-valued controls can be considered [5]. Since $u = \mathrm{prox}_{g/\alpha}(v)$ if and only if $0 \in \alpha(u - v) + \partial g(u)$, implicit function theorems [21] can be used to prove the semismoothness of proximal maps.

7 VARIATIONAL DISCRETIZATION

For computations, one has to choose finite-dimensional approximations of the infinite-dimensional problem (1.1). The application of the semismooth Newton method to discretized problems is straightforward. In this section, we will focus on the variational discretization of [15]. We will consider the following optimal control problem: Minimize

$$\frac{1}{2}\|y - z\|_{L^2(\Omega)}^2 + \frac{\alpha}{2}\|u\|_{L^2(\Omega)}^2 + \int_\Omega \tilde{g}(u(x)) \, \mathrm{d}x$$

over all $(y, u) \in H_0^1(\Omega) \times L^2(\Omega)$ subject to

$$\int_\Omega \nabla y \cdot \nabla v \, \mathrm{d}x = \int_\Omega uv \, \mathrm{d}x \quad \forall v \in H_0^1(\Omega),$$

where $\Omega \subseteq \mathbb{R}^d$ is a bounded domain, $z \in L^2(\Omega)$ is given, and $\tilde{g} : \mathbb{R} \to \bar{\mathbb{R}}$ is as in Section 6. The state space is discretized using piecewise linear and continuous finite element functions: let $V_h \subseteq H_0^1(\Omega)$ be the corresponding finite element subspace. Given $u \in L^2(\Omega)$, let $S_h u := y_h \in V_h$ denote the solution of

$$(7.1) \qquad \int_\Omega \nabla y_h \cdot \nabla v_h \, \mathrm{d}x = \int_\Omega uv_h \, \mathrm{d}x \quad \forall v_h \in V_h,$$

which gives rise to a linear and bounded operator $S_h : L^2(\Omega) \to V_h$. Note that the control space $U = L^2(\Omega)$ is not discretized.

The dual of the discretized problem is now posed in V_h: find $\xi_h \in V_h$ such that

$$(7.2) \qquad \xi - z + S_h \operatorname{prox}_{g/\alpha}(S_h^* \xi/\alpha) = 0.$$

The crucial observation of [15] is the following: in this equation S_h is applied to elements of the type $\operatorname{prox}_{g/\alpha}(q_h)$, where q_h is a piecewise linear finite element function. That is, if $\operatorname{prox}_{g/\alpha}$ has a favorable structure, then $S_h \operatorname{prox}_{g/\alpha}(q_h)$ can be evaluated numerically even if $\operatorname{prox}_{g/\alpha}(q_h)$ does not belong to a finite element subspace. This is the case for the first four examples discussed in Section 6.2. For instance, if g models box constraints on the control, then $\operatorname{prox}_{g/\alpha}(q_h)$ is piecewise linear. This piecewise linear function has kinks in the interior of elements in general, nevertheless the right-hand side for $u = \operatorname{prox}_{g/\alpha}(q_h)$ in (7.1) can be computed exactly [15, 16].

In addition, when applying the semismooth Newton method to the dual equation (7.2), these kinks do not accumulate. This is not the case when applying the semismooth Newton to the primal equation as in Remark 5.15. To overcome this difficulty, [17] suggested to apply the semismooth Newton method to the normal map equation as in Remark 5.16. The variational discretization is not limited to piecewise linear finite elements, in [36] the application to finite elements of polynomial degree 2 is discussed.

8 NUMERICAL EXPERIMENTS

We will consider the following optimal control problem: Minimize

$$(8.1) \qquad J(y, u) := \frac{1}{2}\|y - z\|_{L^2(\Omega)}^2 + \frac{\alpha}{2}\|u\|_{L^2(\Omega)}^2 + \int_\Omega \tilde{g}(u(x))\,dx$$

over all $(y, u) \in H_0^1(\Omega) \times L^2(\Omega)$ subject to

$$(8.2) \qquad \int_\Omega \nabla y \cdot \nabla v\,dx = \int_\Omega uv\,dx \quad \forall v \in H_0^1(\Omega),$$

where $\Omega = (0,1)^2$, $z \in L^2(\Omega)$ is given, and $\tilde{g} : \mathbb{R} \to \bar{\mathbb{R}}$ is as in Section 6. Here, (8.2) is the weak formulation of $-\Delta y = u$ with homogeneous Dirichlet boundary conditions.

We set $U := L^2(\Omega)$ and $Y := L^2(\Omega)$. We define $Su := y$, where $y \in H_0^1(\Omega)$ is the uniquely determined solution of (8.2). Then $S \in \mathcal{L}(Y, U)$ is bounded, moreover, S^* is linear and continuous from $L^2(\Omega) = Y$ to $L^\infty(\Omega)$.

For discretization we used standard piecewise linear and conforming finite elements on a triangulation of Ω to discretize states (y) and dual quantities (ξ). Controls (u) were discretized with piecewise constant functions. To evaluate $\operatorname{prox}_{g/\alpha}(S^* \xi/\alpha)$, first the piecewise linear function $S^* \xi$ is projected onto the subspace of piecewise constant functions, so that the proximal map is computed for this projection. All implementations were done in Matlab.

8.1 EXAMPLE 1: BOX CONSTRAINTS AND L^1-COST

Here, we work with

$$\tilde{g}(u) = \beta|u| + \begin{cases} +\infty & \text{if } |u| > R, \\ 0 & \text{if } |u| \leq R \end{cases}$$

with

$$\beta = 10^{-2}, \ R = 1000.$$

The parameter α is set to $\alpha = 10^{-5}$ if not mentioned otherwise. The (local) semismooth Newton method for this problem was analyzed in [37]. As argued in Section 6.2, g satisfies Assumption 2. Due to the

smoothing properties of S^*, the assumptions of Lemma 6.5 are satisfied, so that Assumption 1 holds for this example. As desired state we used

$$z(x_1, x_2) = 10x_1 \sin(5x_1)\cos(7x_2).$$

We always start the iterations with $\xi_0 = -z$, which corresponds to the initial guess $u_0 = 0$ for the control. For this problem, the unglobalized semismooth Newton method does not converge, hence globalization is necessary.

In the implementation, we used the following values for the parameters of Algorithm 2:

$$\sigma = 0.1, \ \beta = 0.5, \ \tau = 1, \ \delta_{\text{tol}} = 10^{-12}.$$

In our implementation we used the following inexactness criterion

$$\|M_k d_k + \nabla\Phi(\xi_k)\|_Y \le \min\left(10^{-4}, \ 0.1\|\nabla\Phi(\xi_k)\|_Y, \ \|\nabla\Phi(\xi_k)\|_Y^2\right),$$

which is slightly stronger than the one which is used in the convergence analysis, and is inspired by the choices taken in [25]. In addition, we stopped the iteration if the threshold in the linesearch is smaller than the machine epsilon, i.e.,

$$(8.3) \qquad\qquad |\langle d_k, \nabla\Phi(\xi_k)\rangle| \le \texttt{eps(dual)},$$

where the Matlab expression $\texttt{eps(dual)}$ returns the positive distance from $\texttt{dual}$ to the next larger floating-point number, and $\texttt{dual}$ is the current value of the dual objective. This is motivated by the following observation: if the step size t_k leads to a decrease of Φ we get the chain of inequalities

$$\Phi(\xi_k) + t_k\langle d_k, \nabla\Phi(\xi_k)\rangle \le \Phi(\xi_k + t_k d_k) \le \Phi(\xi_k),$$

where the first inequality is due to convexity. Hence, $\Phi(\xi_k + t_k d_k)$ lies in an interval of length $t_k\,|\langle d_k, \nabla\Phi(\xi_k)\rangle|$. If the condition (8.3) is satisfied, then the only floating point number in this interval is $\Phi(\xi_k)$, hence it is impossible to decrease the dual functional value any further. In addition, Corollary 5.6 implies that $\|\nabla\Phi(\xi_k)\|_Y$ is small in this situation.

The optimal control $u = \text{prox}_{g/\alpha}(S^*\xi/\alpha)$ computed on a mesh consisting of 20000 triangles and with mesh-size $h = 0.014$ is depicted in Figure 1. The right-hand plot in Figure 1 shows $\partial\,\text{prox}_{g/\alpha}(S^*\xi_k/\alpha)$, where ξ_k is the final iterate of the dual variable. In this plot, dark regions correspond to $\partial\,\text{prox}_{g/\alpha} = 0$, which means that the control is either zero or at the bounds $\pm R$, while light regions correspond to $\partial\,\text{prox}_{g/\alpha} = 0$, which are regions, where the control is between zero and the bounds.

The results for computations on different meshes can be found in Table 1. The mesh size is denoted by h. The finest mesh consists of $2\cdot 10^6$ triangles. On all meshes the globalized semismooth Newton method needed 8 iterations. The final value of the dual objective Φ together with the value of the final residual and primal-dual gap can be found in Table 1 as well. In all cases the iteration stopped because the condition (8.3) was satisfied. Since the primal-dual gap is small compared to Φ, the optimal primal objective value equals $-\Phi$. In addition, the number of iterations of the globalized semismooth Newton method (it) as well as the total number of iterations of the inner conjugate gradient method (cg) almost stay constant for the different discretizations.

We also performed experiments with varying parameter α on a fixed discretization with $h = 2.83\cdot 10^{-3}$. The results can be found in Table 2. As one expects, the iteration number increases with decreasing α. The third column of Table 2 reports the value $\|\partial\,\text{prox}_{g/\alpha}(S^*\xi/\alpha)\|_{L^1(\Omega)}$, which corresponds to the size of the inactive set, i.e., the size of the set, where $\text{prox}_{g/\alpha}(S^*\xi/\alpha)$ takes values not in $\{-R, 0, +R\}$.

In addition, we conducted experiments using the variational discretization of [15] as described in Section 7. We studied again varying α on a fixed mesh with $h = 2.83\cdot 10^{-3}$. The results can be found in Table 3. While the results are comparable to those in Table 2 it is surprising that the iteration numbers grow only moderately with decreasing α. So it seems that the variational discretization should be favored to the piecewise constant discretization at least for small values of α. This might be due to the fact that the variational discretization can better resolve the boundary of active sets.
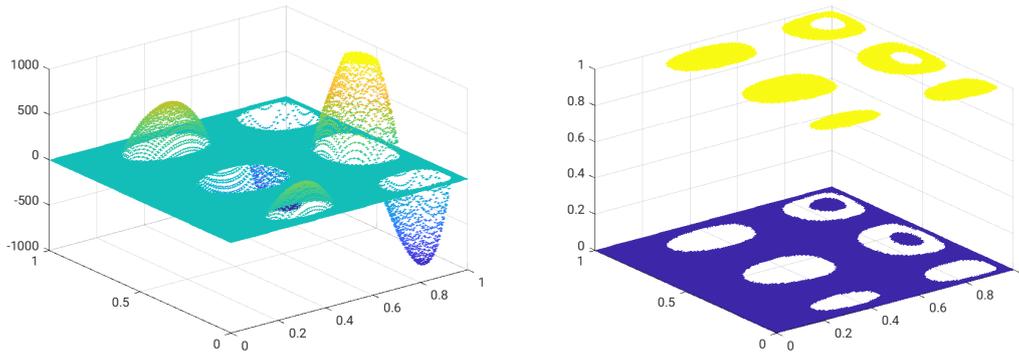
Figure 1: Example 1: optimal control (left), $\partial \operatorname{prox}_{g/\alpha}$ (right)

h	it	cg	Φ	gap	residual
$4.42 \cdot 10^{-2}$	8	98	-3.06	$4.44 \cdot 10^{-16}$	$1.47 \cdot 10^{-10}$
$2.21 \cdot 10^{-2}$	8	98	-3.41	$2.66 \cdot 10^{-15}$	$4.53 \cdot 10^{-9}$
$1.13 \cdot 10^{-2}$	8	99	-3.60	$5.33 \cdot 10^{-15}$	$1.28 \cdot 10^{-9}$
$5.66 \cdot 10^{-3}$	8	99	-3.70	$4.44 \cdot 10^{-16}$	$6.32 \cdot 10^{-10}$
$2.83 \cdot 10^{-3}$	8	100	-3.76	$8.88 \cdot 10^{-16}$	$1.83 \cdot 10^{-9}$
$1.41 \cdot 10^{-3}$	8	99	-3.78	$4.44 \cdot 10^{-16}$	$3.18 \cdot 10^{-9}$

Table 1: Example 1, results for different meshes

8.2 EXAMPLE 2: SINGULAR CONTROL PROBLEM

Here, we work with box constraints, i.e.,

$$\tilde{g}(u) = \begin{cases} +\infty & \text{if } |u| > R, \\ 0 & \text{if } |u| \leq R \end{cases}$$

with $R = 1$. The desired state was set to $z := Sf$ with

$$f(x_1, x_2) = \chi_{[0,0.2]}(x_1) \cdot 5 \cdot \sin(\pi x_2).$$

Hence, the control is used to counter the unwanted perturbation f in the equation. This example is adapted from an example in [35, Section 4.4]. It is challenging to solve for small values of α, as for $\alpha \searrow 0$ the corresponding solutions are not of bang-bang type.

We used the same discretization and setup as in the previous experiment. The solution for two different values of $\alpha \in \{10^{-8}, 10^{-10}\}$ and for a discretization with $h = 2.83 \cdot 10^{-3}$ can be found in Figure 2. The control is zero on a large part of the domain. On the interface of the regions between $u = 0$ and $u \neq 0$ the control exhibits scattering behavior: it oscillates between upper and lower bounds, where the oscillations increase with decreasing mesh size and parameter α. The results of computations for different values of α are reported in Table 4. As one can see, the iteration number grows much stronger with decreasing α than in the previous experiment. Most of these iterations are spent in damping oscillations of the control variable in the region, where the control eventually is zero. In addition, the inner CG iteration also took considerably more steps than for the previous experiment.

α	it	cg	$\|\partial\operatorname{prox}_{g/\alpha}\|_{L^1}$	Φ	gap	residual
$1.00\cdot10^{-4}$	5	41	$4.57\cdot10^{-1}$	-4.70	$2.66\cdot10^{-15}$	$1.57\cdot10^{-10}$
$1.00\cdot10^{-5}$	8	100	$2.70\cdot10^{-1}$	-3.76	$8.88\cdot10^{-16}$	$1.84\cdot10^{-9}$
$1.00\cdot10^{-6}$	11	153	$8.75\cdot10^{-2}$	-3.35	$7.11\cdot10^{-15}$	$1.30\cdot10^{-7}$
$1.00\cdot10^{-7}$	19	290	$2.23\cdot10^{-2}$	-3.29	$5.33\cdot10^{-14}$	$3.58\cdot10^{-9}$
$1.00\cdot10^{-8}$	25	377	$4.15\cdot10^{-3}$	-3.28	$1.71\cdot10^{-13}$	$9.07\cdot10^{-9}$
$1.00\cdot10^{-9}$	29	427	$6.06\cdot10^{-4}$	-3.28	$2.12\cdot10^{-12}$	$1.71\cdot10^{-8}$
$1.00\cdot10^{-10}$	72	723	$1.06\cdot10^{-4}$	-3.28	$2.09\cdot10^{-12}$	$4.24\cdot10^{-9}$

Table 2: Example 1, results for different α

α	it	cg	$\|\partial\operatorname{prox}_{g/\alpha}\|_{L^1}$	Φ	residual
$1.00\cdot10^{-4}$	5	41	$4.57\cdot10^{-1}$	-4.70	$1.61\cdot10^{-10}$
$1.00\cdot10^{-5}$	8	99	$2.70\cdot10^{-1}$	-3.76	$3.74\cdot10^{-9}$
$1.00\cdot10^{-6}$	11	153	$8.74\cdot10^{-2}$	-3.35	$1.40\cdot10^{-7}$
$1.00\cdot10^{-7}$	15	239	$2.22\cdot10^{-2}$	-3.29	$1.27\cdot10^{-9}$
$1.00\cdot10^{-8}$	20	302	$4.23\cdot10^{-3}$	-3.28	$2.11\cdot10^{-9}$
$1.00\cdot10^{-9}$	21	328	$6.02\cdot10^{-4}$	-3.28	$1.82\cdot10^{-10}$
$1.00\cdot10^{-10}$	20	294	$6.01\cdot10^{-5}$	-3.28	$1.47\cdot10^{-6}$

Table 3: Example 1, results for different α, variational discretization

That lead us to consider a nested iteration approach: here the iteration for some value $\alpha = 10^{-(k+1)}$ is started with the result for the value $\alpha = 10^{-k}$. The results can be found in Table 5. There the column 'iterations' denotes the number of iterations when started with the solution of the previous iterations. For this warm-started semismooth Newton the number of iterations per value of α stays constant. In addition, the cumulated iteration numbers are still less than those for the experiment in Table 4, and the inner CG iteration numbers do not grow as severely as in that setting.

For this example, the usage of the variational discretization did not give any advantage like in the previous experiment, so we chose not to report the corresponding numbers.

REFERENCES

[1] A. Alphonse, C. Christof, M. Hintermüller, and I. P. A. Papadopoulos, A globalized inexact semismooth Newton method for nonsmooth fixed-point equations involving variational inequalities, *Comput. Optim. Appl.* 93 (2026), 1–55, doi:10.1007/s10589-025-00722-8, https://doi.org/10.1007/s10589-025-00722-8.

[2] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second edition edition, 2017, doi:10.1007/978-3-319-48311-5. With a foreword by Hédy Attouch.

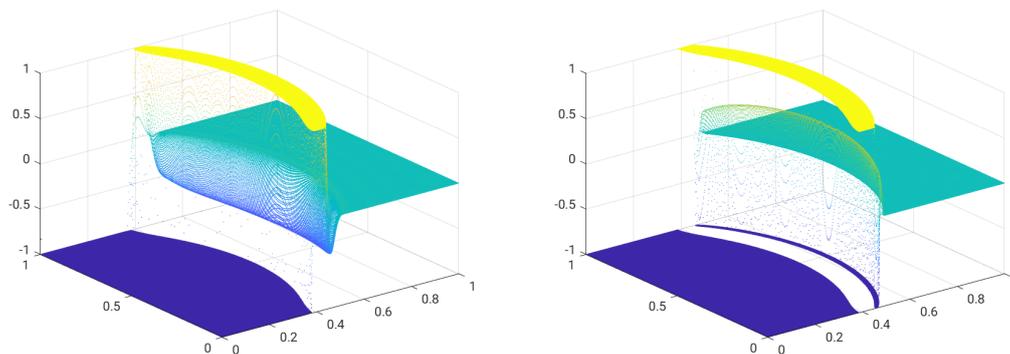[3] M. Bergounioux, K. Ito, and K. Kunisch, Primal-dual strategy for constrained optimal control

Figure 2: Example 2: optimal controls for $\alpha = 10^{-8}$ (left), $\alpha = 10^{-10}$ (right)

α	it	cg	$\|\partial\operatorname{prox}_{g/\alpha}\|_{L^1}$	Φ	residual
$1.00 \cdot 10^{-4}$	9	45	$7.75 \cdot 10^{-1}$	$-4.61 \cdot 10^{-5}$	$6.99 \cdot 10^{-18}$
$1.00 \cdot 10^{-5}$	17	122	$6.16 \cdot 10^{-1}$	$-3.10 \cdot 10^{-5}$	$1.97 \cdot 10^{-17}$
$1.00 \cdot 10^{-6}$	16	265	$5.48 \cdot 10^{-1}$	$-2.90 \cdot 10^{-5}$	$1.48 \cdot 10^{-11}$
$1.00 \cdot 10^{-7}$	24	650	$4.65 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$2.90 \cdot 10^{-14}$
$1.00 \cdot 10^{-8}$	61	4506	$4.43 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$7.48 \cdot 10^{-11}$
$1.00 \cdot 10^{-9}$	52	8768	$4.22 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$4.79 \cdot 10^{-14}$
$1.00 \cdot 10^{-10}$	101	59713	$4.13 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$2.63 \cdot 10^{-6}$

Table 4: Example 2, results for different α

problems, *SIAM J. Control Optim.* 37 (1999), 1176–1194, doi:10.1137/s0363012997328609, https://doi.org/10.1137/S0363012997328609.

[4] C. Clason and K. Kunisch, Multi-bang control of elliptic systems, *Ann. Inst. H. Poincaré C Anal. Non Linéaire* 31 (2014), 1109–1130, doi:10.1016/j.anihpc.2013.08.005, https://doi.org/10.1016/j.anihpc.2013.08.005.

[5] C. Clason, C. Tameling, and B. Wirth, Convex relaxation of discrete vector-valued optimization problems, *SIAM Rev.* 63 (2021), 783–821, doi:10.1137/21m1426237, https://doi.org/10.1137/21M1426237.

[6] J. W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, *SIAM J. Numer. Anal.* 4 (1967), 10–26, doi:10.1137/0704002, https://doi.org/10.1137/0704002.

[7] F. Facchinei, Minimization of SC^1 functions and the Maratos effect, *Oper. Res. Lett.* 17 (1995), 131–137, doi:10.1016/0167-6377(94)00059-f, https://doi.org/10.1016/0167-6377(94)00059-F.

[8] M. Gerdts, S. Horn, and S. J. Kimmerle, Line search globalization of a semismooth Newton method for operator equations in Hilbert spaces with applications in optimal control, *J. Ind. Manag. Optim.* 13 (2017), 47–62, doi:10.3934/jimo.2016003, https://doi.org/10.3934/jimo.2016003.

[9] H. Gfrerer, On a globally convergent semismooth* Newton method in nonsmooth nonconvex

α	it	cg	$\|\partial \operatorname{prox}_{g/\alpha}\|_{L^1}$	Φ	residual
$1.00 \cdot 10^{-4}$	9	45	$7.75 \cdot 10^{-1}$	$-4.61 \cdot 10^{-5}$	$6.99 \cdot 10^{-18}$
$1.00 \cdot 10^{-5}$	7	118	$6.16 \cdot 10^{-1}$	$-3.10 \cdot 10^{-5}$	$7.64 \cdot 10^{-18}$
$1.00 \cdot 10^{-6}$	6	294	$5.48 \cdot 10^{-1}$	$-2.90 \cdot 10^{-5}$	$7.67 \cdot 10^{-12}$
$1.00 \cdot 10^{-7}$	7	631	$4.65 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$4.02 \cdot 10^{-15}$
$1.00 \cdot 10^{-8}$	7	1738	$4.43 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$3.86 \cdot 10^{-9}$
$1.00 \cdot 10^{-9}$	7	3857	$4.22 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$9.13 \cdot 10^{-14}$
$1.00 \cdot 10^{-10}$	8	11190	$4.13 \cdot 10^{-1}$	$-2.87 \cdot 10^{-5}$	$7.06 \cdot 10^{-6}$

Table 5: Example 2, results for different α, nested iterations

optimization, *Comput. Optim. Appl.* 91 (2025), 67–124, doi:10.1007/s10589-025-00658-z, https://doi.org/10.1007/s10589-025-00658-z.

[10] C. Gräser, Globalization of nonsmooth Newton methods for optimal control problems, in *Numerical mathematics and advanced applications*, Springer, Berlin, 2008, 605–612.

[11] R. Griesse, T. Grund, and D. Wachsmuth, Update strategies for perturbed nonsmooth equations, *Optim. Methods Softw.* 23 (2008), 321–343, doi:10.1080/10556780701523551, https://doi.org/10.1080/10556780701523551.

[12] E. Hans and T. Raasch, Global convergence of damped semismooth Newton methods for ℓ_1 Tikhonov regularization, *Inverse Problems* 31 (2015), 025005, 31, doi:10.1088/0266-5611/31/2/025005, https://doi.org/10.1088/0266-5611/31/2/025005.

[13] R. Herzog, G. Stadler, and G. Wachsmuth, Directional sparsity in optimal control of partial differential equations, *SIAM J. Control Optim.* 50 (2012), 943–963, doi:10.1137/100815037, https://doi.org/10.1137/100815037.

[14] M. Hintermüller, K. Ito, and K. Kunisch, The primal-dual active set strategy as a semismooth Newton method, *SIAM J. Optim.* 13 (2002), 865–888, doi:10.1137/s1052623401383558, https://doi.org/10.1137/S1052623401383558.

[15] M. Hinze, A variational discretization concept in control constrained optimization: the linear-quadratic case, *Comput. Optim. Appl.* 30 (2005), 45–61, doi:10.1007/s10589-005-4559-5, https://doi.org/10.1007/s10589-005-4559-5.

[16] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*, volume 23 of Mathematical Modelling: Theory and Applications, Springer, New York, 2009.

[17] M. Hinze and M. Vierling, A globalized semi-smooth Newton method for variational discretization of control constrained elliptic optimal control problems, in *Constrained optimization and optimal control for partial differential equations*, volume 160 of Internat. Ser. Numer. Math., Birkhäuser, Basel, 2012, 171–182, doi:10.1007/978-3-0348-0133-1_9, https://doi.org/10.1007/978-3-0348-0133-1_9.

[18] M. Hinze and M. Vierling, The semi-smooth Newton method for variationally discretized control constrained elliptic optimal control problems; implementation, convergence and globalization, *Optim. Methods Softw.* 27 (2012), 933–950, doi:10.1080/10556788.2012.676046, https://doi.org/10.1080/10556788.2012.676046.

[19] K. Ito and K. Kunisch, *Lagrange multiplier approach to variational problems and applications*, volume 15 of Advances in Design and Control, SIAM, Philadelphia, PA, 2008, doi:10.1137/1.9780898718614.

[20] C. T. Kelley and E. W. Sachs, A trust region method for parabolic boundary control problems, *SIAM J. Optim.* 9 (1999), 1064–1081, doi:10.1137/s1052623496308965, https://doi.org/10.1137/S1052623496308965.

[21] F. Kruse, Semismooth implicit functions, *J. Convex Anal.* 25 (2018), 595–622.

[22] K. Kunisch and A. Rösch, Primal-dual active set strategy for a general class of constrained optimal control problems, *SIAM J. Optim.* 13 (2002), 321–334, doi:10.1137/s1052623499358008, https://doi.org/10.1137/S1052623499358008.

[23] K. Kunisch and D. Wachsmuth, On time optimal control of the wave equation and its numerical realization as parametric optimization problem, *SIAM J. Control Optim.* 51 (2013), 1232–1262, doi:10.1137/120877520, https://doi.org/10.1137/120877520.

[24] K. Malanowski, Remarks on differentiability of metric projections onto cones of nonnegative functions, *J. Convex Anal.* 10 (2003), 285–294.

[25] A. Milzarek, F. Schaipp, and M. Ulbrich, A semismooth Newton stochastic proximal point algorithm with variance reduction, *SIAM J. Optim.* 34 (2024), 1157–1185, doi:10.1137/22m1488181, https://doi.org/10.1137/22M1488181.

[26] W. Ouyang and A. Milzarek, A trust region-type normal map-based semismooth Newton method for nonsmooth nonconvex composite optimization, *Mathematical Programming* (2024), doi:10.1007/s10107-024-02110-2, https://doi.org/10.1007/s10107-024-02110-2.

[27] K. Pieper, *Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems*, PhD thesis, Technische Universität München, 2015, https://mediatum.ub.tum.de/1241413.

[28] F. Pörner and D. Wachsmuth, An iterative Bregman regularization method for optimal control problems with inequality constraints, *Optimization* 65 (2016), 2195–2215, doi:10.1080/02331934.2016.1238082, https://doi.org/10.1080/02331934.2016.1238082.

[29] B. Pötzl, A. Schiela, and P. Jaap, Second order semi-smooth proximal Newton methods in Hilbert spaces, *Comput. Optim. Appl.* 82 (2022), 465–498, doi:10.1007/s10589-022-00369-9, https://doi.org/10.1007/s10589-022-00369-9.

[30] B. Pötzl, A. Schiela, and P. Jaap, Inexact proximal Newton methods in Hilbert spaces, *Comput. Optim. Appl.* 87 (2024), 1–37, doi:10.1007/s10589-023-00515-x, https://doi.org/10.1007/s10589-023-00515-x.

[31] L. Qi and D. Sun, A survey of some nonsmooth equations and smoothing Newton methods, in *Progress in optimization*, volume 30 of Appl. Optim., Kluwer Acad. Publ., Dordrecht, 1999, 121–146, doi:10.1007/978-1-4613-3285-5_7, https://doi.org/10.1007/978-1-4613-3285-5_7.

[32] S. M. Robinson, Local structure of feasible sets in nonlinear programming. III. Stability and sensitivity, in *Nonlinear Analysis and Optimization*, B. Cornet, V. H. Nguyen, and J. P. Vial (eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 1987, 45–66, doi:10.1007/bfb0121154, https://doi.org/10.1007/bfb0121154.

[33] S. M. Robinson, Normal maps induced by linear transformations, *Math. Oper. Res.* 17 (1992), 691–714, doi:10.1287/moor.17.3.691, https://doi.org/10.1287/moor.17.3.691.

[34] R. T. Rockafellar, Integral functionals, normal integrands and measurable selections, in *Nonlinear operators and the calculus of variations (Summer School, Univ. Libre Bruxelles, Brussels, 1975)*, Lecture Notes in Math., Vol. 543, Springer, Berlin-New York, 1976, 157–207.

[35] S. Rotin, *Konvergenz des Proximal-Punkt-Verfahrens für inkorrekt gestellte Optimalsteuerprobleme mit partiellen Differentialgleichungen*, PhD thesis, Universität Trier, 1999, doi:10.25353/ubtr-xxxx-901d-57ad.

[36] D. Sevilla and D. Wachsmuth, Polynomial integration on regions defined by a triangle and a conic, in *ISSAC 2010—Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, ACM, New York, 2010, 163–170, doi:10.1145/1837934.1837968, https://doi.org/10.1145/1837934.1837968.

[37] G. Stadler, Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices, *Comput. Optim. Appl.* 44 (2009), 159–181, doi:10.1007/s10589-007-9150-9, https://doi.org/10.1007/s10589-007-9150-9.

[38] L. N. Trefethen and D. Bau, III, *Numerical linear algebra*, SIAM, Philadelphia, PA, 1997, doi:10.1137/1.9780898719574.

[39] M. Ulbrich, Semismooth Newton methods for operator equations in function spaces, *SIAM J. Optim.* 13 (2002), 805–842, doi:10.1137/s1052623400371569, https://doi.org/10.1137/S1052623400371569.

[40] M. Ulbrich, *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, volume 11 of MOS-SIAM Series on Optimization, SIAM, Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2011, doi:10.1137/1.9781611970692.

[41] R. Winther, Some superlinear convergence results for the conjugate gradient method, *SIAM J. Numer. Anal.* 17 (1980), 14–17, doi:10.1137/0717002, https://doi.org/10.1137/0717002.

[42] X. Y. Zhao, D. Sun, and K. C. Toh, A Newton-CG augmented Lagrangian method for semidefinite programming, *SIAM J. Optim.* 20 (2010), 1737–1765, doi:10.1137/080718206, https://doi.org/10.1137/080718206.