Scientific
Research
Publishing

# Proximal Methods for Elliptic Optimal Control Problems with Sparsity Cost Functional

## Andreas Schindele, Alfio Borzì

Institut für Mathematik, Universität Würzburg, Würzburg, Germany
Email: andreas.schindele@mathematik.uni-wuerzburg.de, alfio.borzi@mathematik.uni-wuerzburg.de

---

## Abstract

**First-order proximal methods that solve linear and bilinear elliptic optimal control problems with a sparsity cost functional are discussed. In particular, fast convergence of these methods is proved. For benchmarking purposes, inexact proximal schemes are compared to an inexact semismooth Newton method. Results of numerical experiments are presented to demonstrate the computational effectiveness of proximal schemes applied to infinite-dimensional elliptic optimal control problems and to validate the theoretical estimates.**

## Keywords

**Optimal Control, Elliptic PDE, Nonsmooth Optimization, Proximal Method, Semismooth Newton Method**

---

## 1. Introduction

In recent years, a great research effort has been made to solve optimization problems governed by Partial Differential Equations (PDEs); see, e.g., [1]-[3] and references therein. In many cases, this research has focused on objective functionals with differentiable $L^2$ terms and non-smoothness resulted from the presence of control and state constraints. However, more recently, the investigation of $L^1$ cost functionals has become a central topic in PDE-based optimization [4]-[6], because they give rise to sparse controls that are advantageous in many applications like optimal actuator placement [4] or impulse control [7].

A representative formulation of optimal control problems with $L^1$ control costs is the following

$$\min_{(y,u)\in H_0^1(\Omega)\times L^2(\Omega)} \quad \frac{1}{2}\|y-z\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L^2}^2 + \beta\|u\|_{L^1}$$
$$\text{s.t.} \quad c(y,u)=0, \quad u_a \leq u \leq u_b \text{ a.e. in } \Omega, \tag{1.1}$$

---

where $c(y,u) = 0$ represents a PDE for the state $y$ including the control $u$. This problem has been discussed in [4] [5] for the case where $c(y,u)$ represents a linear elliptic operator. Nonlinear PDE constraints have been considered in [6]. However, in these references a linear control mechanism is discussed. Concerning the optimization methodology for (1.1), the semi-smooth Newton method has been the solver of choice in [4]-[6].

On the other hand, in the field of signal acquisition and reconstruction, $l^1$-based optimization and sparsity have been exploited to successfully recover "functions" from few samples; see, e.g., [8]-[10].

In this framework, it was shown [11] that $l^1$-based inverse problems in signal recovery can be very efficiently solved by proximal methods. Nowadays, these iterative schemes are the method of choice in magnetic resonance imaging and a special proximal method called "Fast Iterative Soft Thresholding Algorithm" (FISTA) [12] is considered the state-of-the-art method for solving finite-dimensional optimization problems of the following form

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \beta \|x\|_1,$$

where the rectangular matrix $A$ represents a blur operator.

We remark that the research and successful application of proximal schemes are attracting attention of many scientists and practitioners, which result in many new developments in this field. We refer to, e.g., [13] for recent results and additional references.

The purpose of our work is to contribute to the field of PDE-based optimization with $L^1$ control costs by investigating proximal methods in this infinite-dimensional setting. In particular, we aim at implementing and analysing proximal schemes for solving (1.1) that exploit first-order optimality conditions. Our investigation is motivated by the fact that proximal methods may have a computational performance that is comparable to that of semismooth Newton methods. However, in contrast to the latter, proximal schemes do not require the construction of second-order derivatives and the implementation of, e.g., a Krylov solver.

For our investigation, we consider (1.1) with elliptic operators and linear and bilinear control mechanisms. Notice that the latter case has been a much less investigated problem. One of our main contributions is to prove convergence for all variants of the proximal schemes that we discuss in this paper. In particular, we prove an $\mathcal{O}(1/k^2)$ convergence rate of the value of reduced cost functional, where $k$ is the number of proximal iterations. This notion of convergence is used in $l^1$-based optimization and in some application fields [14].

We remark that many arguments in our analysis are similar to those presented in the finite-dimensional case. However, some additional arguments are necessary in infinite dimensions, especially regarding the structure of our differential constraints and the discussion of our inexact proximal schemes. We refer to [13] for further results concerning the formulation of proximal schemes for infinite-dimensional optimization problems from a different perspective.

In the next section, we discuss linear and bilinear elliptic optimal control problems, where for completeness, some conditions for the existence of a unique control-to-state operators are considered. Section 3 is devoted to optimal control problems with sparsity costs and governed by elliptic equations with linear and bilinear control mechanisms. We discuss conditions for convexity of the bilinear problem and state the optimality conditions. In Section 4, we present a Fast Inexact Proximal method (FIP) that represents an infinite-dimensional extension of the FISTA method. In Section 5, the convergence rate of this method is proven to be $\mathcal{O}(1/k^2)$. In Section 6, an inexact semismooth Newton method in function spaces is presented as the state of the art method for comparison purposes. For completeness, the theory of this method is extended to the case of elliptic bilinear control problems. A numerical comparison of the FIP and Semismooth-Newton methods is presented in Section 7. A section of conclusion completes this work.

## 2. Elliptic Models with Linear and Bilinear Control Mechanisms

In this section, we discuss elliptic PDE models with linear and bilinear control structures. Notice that these models are already discussed in many references; see, e.g., [2] [15]-[17]. However, in this section, we report the main results required for our analysis of convergence of the proposed proximal methods.

Consider the following boundary value problem

$$Ay + u = f \quad \text{in } \Omega \tag{2.1}$$

$$y = 0 \quad \text{on } \partial\Omega, \tag{2.2}$$

where $\Omega \subset \mathbb{R}^n$, with $n \leq 3$, is a bounded domain and $f \in L^2(\Omega)$. The operator $A : H_0^1(\Omega) \to H^{-1}(\Omega)$

represents a second-order linear elliptic differential operator of the following form

$$Ay = -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{ij} \frac{\partial}{\partial x_j} y\right) + a_0 y,$$

such that $a_{i,j}, a_0 \in L^{\infty}(\Omega)$, and $a_{i,j}$ satisfies the coercivity condition $\sum_{i,j=1}^{n} a_{ij}(x)\xi_1\xi_2 \geq \theta \sum_{j=1}^{n} \xi_j^2$ a.e. in $\Omega$ for some $\theta > 0$ and $a_0 \geq 0$. For the existence and uniqueness of solutions to (2.1) see ([15], Section 6).

Further, we consider the following bilinear elliptic control problem

$$Ay + uy = f \quad \text{in } \Omega \tag{2.3}$$

$$y = 0 \quad \text{on } \partial\Omega. \tag{2.4}$$

In both linear and bilinear control settings, we require $u \in U_{ad}$, with the following set of feasible controls

$$U_{ad} := \left\{ u \in L^2(\Omega) : u_a \leq u \leq u_b \text{ a.e. in } \Omega \right\}, \tag{2.5}$$

where $u_a \leq 0 \leq u_b$, $u_a < u_b$.

Now, we discuss the existence of a unique weak solution to (2.3)-(2.4). For this purpose, we need the Poincaré-Friedrichs lemma and denote with $c_\Omega$ the Poincaré-Friedrichs constant; see, e.g., [15].

We denote $\|\cdot\| := \|\cdot\|_{L^2(\Omega)}$ induced by the inner product $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_{L^2(\Omega)}$.

**Theorem 2.1.** *Let* $u \in U_{ad}$, *where*

$$u_a > -a_0 - \frac{\theta}{c_\Omega}. \tag{2.6}$$

*Then, there exists a unique weak solution* $y \in H_0^1(\Omega)$ *to the bilinear elliptic problem (2.3)-(2.4) and the following property holds*

$$\|y\|_{H^1(\Omega)} \leq C_1 \|f\|. \tag{2.7}$$

*Proof.* The proof is immediate using the Lemma of Lax-Milgram and the following result

$$a(y,y) := \langle Ay + uy, y\rangle = \int_{\Omega}\left(\sum_{i,j=1}^{n} a_{ij}\frac{\partial y}{\partial x_j}\frac{\partial y}{\partial x_i} + (a_0 + u)y^2\right)dx$$

$$\geq \int_{\Omega}\left(\theta|\nabla y|^2 + (a_0 + u_a)y^2\right)dx$$

$$= \kappa c_\Omega \int_{\Omega}|\nabla y|^2\, dx + (\theta - \kappa c_\Omega)\int_{\Omega}|\nabla y|^2\, dx + \int_{\Omega}(a_0 + u_a)y^2 dx$$

$$\geq \kappa c_\Omega \int_{\Omega}|\nabla y|^2\, dx + \left(\frac{\theta}{c_\Omega} - \kappa\right)\int_{\Omega}y^2 dx + \int_{\Omega}(a_0 + u_a)y^2 dx$$

$$\geq \min\left(\kappa c_\Omega, u_a + a_0 + \frac{\theta}{c_\Omega} - \kappa\right)\|y\|_{H^1}^2 = C_0\|y\|_{H^1}^2.$$

With $\kappa := \dfrac{u_a + a_0 + \theta/c_\Omega}{c_\Omega + 1} > 0$, we have that $\kappa c_\Omega = u_a + a_0 + \theta/c_\Omega - \kappa$ and therefore $C_0 := \kappa c_\Omega$. In the forth line the Poincaré-Friedrichs was used. Hence, $C_1 := C_0^{-1}$. $\square$

**Remark 2.1.** *In the case of* $\Omega = (0,1)^n$, $n \leq 3$, *and* $A = -\Delta$, *including homogeneous Dirichlet boundary conditions, we have* $a_0 = 0$, $\theta = 1$ *and* $c_\Omega = 1/2n$ *such that we can ensure invertibility for* $u_a > -2n$.

**Remark 2.2.** *In order to ensure a unique solution, we require condition (2.6) for the choice of* $u_a$ *in the bilinear case.*

Next, we recall the following theorem stating higher regularity of solutions to (2.3)-(2.4); see ([18], Theorem 4.3.1.4).

**Theorem 2.2.** *Let* $\Omega \subset \mathbb{R}^n$, $n \leq 3$, *be a convex and bounded polygonal or polyhedral domain. If in addition to the assumptions of Theorem 2.1, we have that* $a_{ij} \in C^1(\bar{\Omega})$, *then* $y \in H_0^1(\Omega) \cap H^2(\Omega)$ *and the following*

*holds*

$$\|y\|_{H^2(\Omega)} \le \tilde{C}\left(\|f\| + \|y\|\right) \le C\|f\|, \tag{2.8}$$

for some appropriate constants $\tilde{C}, C > 0$ that only depend on $\Omega$.

**Remark 2.3.** *Because* $H^2(\Omega)$ *can be embedded in* $L^\infty(\Omega)$ *for* $n \le 3$ [19] *and using* (2.8), *we obtain*

$$\|y\|_{L^\infty(\Omega)} \le C\|f\|. \tag{2.9}$$

Theorem 2.1 and Theorem 2.2 ensure the existence of a unique control-to-state operator

$$S : U_{ad} \to H_0^1(\Omega) \cap H^2(\Omega), \qquad u \mapsto S(u), \tag{2.10}$$

where in the linear case $S_l(u) = A^{-1}(f - u)$ represents the unique solution to (2.1) and in the bilinear case $S_b(u) = (A + u)^{-1} f$ is the unique solution to (2.3). In the following, we use the expression $S(u)$ when it is valid for both the linear and the bilinear systems.

**Remark 2.4.** *The control-to-state operator* $S_b(u)$ *is not Fréchet-differentiable in the* $L^2$ *topology since for every* $\varepsilon > 0$ *there is always an* $h \in L^2(\Omega)$ *with* $\|h\| \le \varepsilon$ *such that* $u + h \notin U_{ad}$ *and therefore it is not necessarily defined. However, we only need the following weaker form of differentiability, which is a directional differentiability in all* $v \in U_{ad}$ *in the directions* $(u - v)$ *for* $u \in U_{ad}$. *This is called Q-differentiability; see* [16].

**Definition 2.1. (Q-differentiability).** *Let* $U \subset X$ *be a convex set and* $T : U \to Y$. *Then* $T$ *is called Q-differentiable in* $v \in U$, *if there exists a mapping* $T'_U(v) \in \mathcal{L}(X,Y)$, *such that for all* $u \in U$ *the following holds*

$$\frac{\|T(v + u - v) - T(v) - T'_U(v)(u - v)\|_Y}{\|u - v\|_X} \to 0 \quad \text{if } \|u - v\|_X \to 0.$$

In the following, we omit the index $U$ and write $T' = T'_U$.

The $Q$-derivatives of $S_b(u)$ have the following properties.

**Lemma 2.3.** *The control-to-state-operator* $S_b$ *is at least two times Q-differentiable in* $U_{ad}$ *with* $X, Y = L^2(\Omega)$ *and its derivatives have the following properties for all directions* $\phi_1, \phi_2 \in L^2(\Omega)$:

i) $S_b'(u)(\phi_1) \in H_0^1(\Omega) \cap H^2(\Omega)$ *is the solution* $y'$ *of*

$$Ay' + uy' = -\phi_1 S_b(u). \tag{2.11}$$

ii) $S_b''(u)(\phi_1, \phi_2) \in H_0^1(\Omega) \cap H^2(\Omega)$ *is the solution* $y''$ *of*

$$Ay'' + uy'' = -\phi_2 S_b'(u)(\phi_1) - \phi_1 S_b'(u)(\phi_2). \tag{2.12}$$

iii) *The following inequalities hold*

$$\|S_b'(u)(\phi_1)\| \le C_2 \|\phi_1\| \|f\|, \tag{2.13}$$

$$\|S_b''(u)(\phi_1, \phi_2)\| \le C_3 \|\phi_1\| \|\phi_2\| \|f\|. \tag{2.14}$$

*Proof.* Part (i) and (ii) can be shown by direct calculation (see ([16], Lemma 2.9). So part (iii) is left to be proved. If $y' := S_b'(u)(\phi_1) \in H_0^1(\Omega) \cap H^2(\Omega)$ is a solution to

$$Ay' + uy' = -\varphi_1 S_b(u),$$

for $u \in U_{ad}$ and $f \in L^2(\Omega)$, by using (2.9), we obtain

$$\|y'\| \le C\|y'\|_{L^\infty} \le \tilde{C}\|\phi_1 S_b(u)\| \le \tilde{C}\|\phi_1\| \|S_b(u)\|_{L^\infty(\Omega)} \le C_2 \|\phi_1\| \|f\|, \tag{2.15}$$

where the constants depend on the measure of $\Omega$ and not on $y$. Therefore we obtain (2.13) and $y' \in L^\infty(\Omega)$. conclude Furthermore, let $y'' := S_b''(u)(\phi_1, \phi_2) \in H_0^1(\Omega) \cap H^2(\Omega)$ be a solution to the following problem

$$Ay'' + uy'' = -\phi_2 S_b'(u)(\phi_1) - \phi_1 S_b'(u)(\phi_2),$$

for $u \in U_{ad}$ and $f \in L^2(\Omega)$. With the same arguments as above and using (2.15), we obtain

$$\|y''\| \le C\|\phi_2 S_b'(u)(\phi_1) + \phi_1 S_b'(u)(\phi_2)\| \le C_3 \|\phi_1\| \|\phi_2\| \|f\|.$$

Therefore, we obtain (2.14), which completes the proof. □

## 3. Elliptic Optimal Control Problems with Sparsity Cost Functional

In this section, we discuss optimal control problems governed by the linear- and bilinear-control elliptic systems discussed in the previous section. We consider the following cost functional

$$J(y,u) := \frac{1}{2}\|y - z\|^2 + \frac{\alpha}{2}\|u\|^2 + \beta\|u\|_{L^1}, \tag{3.1}$$

where $z \in L^2(\Omega)$, $y \in H_0^1(\Omega)$, $u \in U_{ad}$ and $\alpha, \beta > 0$. This functional is made of a Fréchet-differentiable classical tracking type cost with $L^2$-regularization and a nondifferentiable $L^1$-control cost. Using the control-to-state operator (2.10), we have the following reduced optimal control problem

$$\min_{u \in U_{ad}} \hat{J}(u) := \frac{1}{2}\|S(u) - z\|^2 + \frac{\alpha}{2}\|u\|^2 + \beta\|u\|_{L^1}. \tag{3.2}$$

The nondifferentiable part $\hat{J}_1(u) := \beta\|u\|_{L^1}$ is convex. Therefore, in order to state convexity of the reduced functional $\hat{J}(u)$, we investigate the second-derivative of the differentiable part $\hat{J}_2(u) := \frac{1}{2}\|S(u) - z\|^2 + \frac{\alpha}{2}\|u\|^2$.

We have

$$\hat{J}_2''(\overline{u})(v,w) = \langle S'(\overline{u})(v), S'(\overline{u})(w)\rangle + \langle S(\overline{u}) - z, S''(\overline{u})(v,w)\rangle + \alpha\langle v,w\rangle.$$

In particular, in the linear case, we have

$$\hat{J}_2''(\overline{u})(v,v) = \|A^{-1}v\|^2 + \alpha\|v\|^2 > 0, \text{ for all } v \in L^2(\Omega), \|v\| \neq 0. \tag{3.3}$$

We conclude that the reduced functional is strictly convex in the linear case.

In the bilinear case, we have a non-convex optimization problem. However, local convexity can be guaranteed under some conditions. To be specific, we chose the sufficient condition stated in the following theorem.

**Lemma 3.1.** *Let* $C''(u) := \sup_{\|v\| \leq 1}\|S_b''(u)(v,v)\|$, *if the following inequality holds*

$$C''(u)\|S_b(u) - z\| < \alpha, \tag{3.4}$$

then the reduced functional $\hat{J}(u)$ is strictly convex in a neighborhood of $u \in U_{ad}$.

*Proof.* Since $\hat{J}_1(u) := \beta\|u\|_{L^1}$ is convex, we have to prove that $\hat{J}_2(u) := \frac{1}{2}\|S_b(u) - z\|^2 + \frac{\alpha}{2}\|u\|^2$ is strictly convex in $u$. Therefore we show that the reduced Hessian is positive definite in $U_{ad}$ as follows

$$\hat{J}_2''(u)(v,v) = \langle S_b'(u)(v), S_b'(u)(v)\rangle + \langle S_b(u) - z, S_b''(u)(v,v)\rangle + \alpha\langle v,v\rangle$$
$$\geq (\alpha - C''(u)\|S_b(u) - z\|)\|v\|^2,$$

and thus $\hat{J}(u)$ is strictly convex in $u$. □

We remark that the result of Lemma 3.1 is well known. It expresses local convexity of the reduced objective when the state function is sufficiently close to the target and the weight of the quadratic $L^2$ cost of the control is sufficiently large. Indeed, local convexity may result with much weaker assumptions. However, since our focus is the investigation of proximal schemes, we make the following strong assumption.

**Assumption 1.** *We assume that* (3.4) *holds for all* $u \in U_{ad}$.

Because of Lemma 2.3, this assumption holds if the regularization parameter $\alpha > C_3\|f\|(C\|f\| + \|z\|)$.

In the next step, the first-order optimality conditions for (3.2) are derived. First, we need the definition of the subdifferential.

**Definition 3.1.** *Let* $H$ *be a Hilbert space and* $F : H \to \mathbb{R}$ *be convex. We call the set-valued mapping* $\partial F : H \rightrightarrows H^*$ *given by*

$$\partial F(u) := \{g \in H^* : \langle g, v - u\rangle \leq F(v) - F(u) \text{ for all } v \in H\}.$$

the subdifferential of $F$ in $u$.

From ([20], Remark 3.2), we obtain that $\bar{u}$ is a solution of (3.2), if and only if there exists a $\bar{\lambda} \in \partial \hat{J}_1(\bar{u})$ such that

$$\left\langle S'(\bar{u})^*\left(S(\bar{u})-z\right)+\alpha\bar{u}+\bar{\lambda}, u-\bar{u}\right\rangle \geq 0, \quad \text{for all } u \in U_{ad}, \tag{3.5}$$

where $^*$ denotes the adjoint operator. From (3.5), one can derive the optimality system by using the Lagrange multipliers $\bar{\lambda}_a, \bar{\lambda}_b \in L^2(\Omega)$. We have the following theorem (see [4], Theorem 2.1).

**Theorem 3.2.** *The optimal solution $\bar{u}$ of (3.2) is characterized by the existence of* $\left(\bar{\lambda}, \bar{\lambda}_a, \bar{\lambda}_b\right) \in L^2(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ *such that*

$$S'(\bar{u})^*\left(S(\bar{u})-z\right)+\alpha\bar{u}+\bar{\lambda}+\bar{\lambda}_b-\bar{\lambda}_a = 0, \tag{3.6}$$

$$\bar{\lambda}_b \geq 0, \quad u_b - \bar{u} \geq 0, \quad \bar{\lambda}_b\left(u_b - \bar{u}\right) = 0, \tag{3.7}$$

$$\bar{\lambda}_a \geq 0, \quad \bar{u} - u_a \geq 0, \quad \bar{\lambda}_a\left(\bar{u} - u_a\right) = 0, \tag{3.8}$$

$$\bar{\lambda} = \beta \quad \text{a.e. on } \{x \in \Omega : \bar{u} > 0\}, \tag{3.9}$$

$$|\bar{\lambda}| \leq \beta \quad \text{a.e. on } \{x \in \Omega : \bar{u} = 0\}, \tag{3.10}$$

$$\bar{\lambda} = -\beta \quad \text{a.e. on } \{x \in \Omega : \bar{u} < 0\}. \tag{3.11}$$

If one introduces the parameter $\bar{\mu} := \bar{\lambda} + \bar{\lambda}_b - \bar{\lambda}_a$, it is shown in [4] that conditions (3.7)-(3.11) are equivalent to

$$B(\bar{u}, \bar{\mu}) = 0, \tag{3.12}$$

where

$$B(\bar{u}, \bar{\mu}) := \bar{u} - \max\left\{0, \bar{u} + c\left(\bar{\mu} - \beta\right)\right\} - \min\left\{0, \bar{u} + c\left(\bar{\mu} + \beta\right)\right\}$$
$$+ \max\left\{0, \bar{u} - u_b + c\left(\bar{\mu} - \beta\right)\right\} + \min\left\{0, \bar{u} - u_a + c\left(\bar{\mu} + \beta\right)\right\},$$

where $c > 0$ is arbitrary. With this setting, the optimality system (3.6)-(3.11) reduces to the following

$$S'(\bar{u})^*\left(S(\bar{u})-z\right)+\alpha\bar{u}+\bar{\mu} = 0, \tag{3.13}$$

$$B(\bar{u}, \bar{\mu}) = 0. \tag{3.14}$$

In the linear-control case, the Equation (3.13) becomes the following

$$-A^{-*}\left(A^{-1}\left(f - \bar{u}\right) - z\right)+\alpha\bar{u}+\bar{\mu} = 0, \tag{3.15}$$

where $A^{-*} = \left(A^*\right)^{-1}$. By setting $\bar{y} = A^{-1}\left(f - \bar{u}\right)$ and $\bar{p} := -A^{-*}\left(\bar{y} - z\right)$, (3.15) becomes $\bar{p} + \alpha\bar{u} + \bar{\mu} = 0$.

We summarize the previous considerations into the following theorem.

**Theorem 3.3.** (*Linear optimality conditions*) *The optimal solution $(\bar{y}, \bar{u}) \in H_0^1(\Omega) \times L^2(\Omega)$ to (3.2) in the linear control case is characterized by the existence of the dual pair $(\bar{p}, \bar{\mu}) \in H_0^1(\Omega) \times L^2(\Omega)$ such that*

$$A\bar{y} + \bar{u} - f = 0 \tag{3.16}$$

$$A^*\bar{p} + \bar{y} - z = 0 \tag{3.17}$$

$$\bar{p} + \alpha\bar{u} + \bar{\mu} = 0 \tag{3.18}$$

$$B(\bar{u}, \bar{\mu}) = 0. \tag{3.19}$$

Furthermore, the reduced gradient and the reduced Hessian of $\hat{J}_2(u)$ are given by

$$\nabla \hat{J}_2(u) = \alpha u + p, \text{ and } \hat{J}_2''(u) = \alpha I + A^{-*}A^{-1}. \tag{3.20}$$

Notice that with an abuse of notation, we denote the reduced Hessian with $\hat{J}_2''(u)$, which is also used to denote the second derivative operator.

For the bilinear-control system, we have $S'(u)(\phi_1) = -(A+u)^{-1}\left[\phi_1(A+u)^{-1}f\right]$ and therefore $S'(u)^*(\phi_1) = -(A+u)^{-1}f(A+u)^{-*}\phi_1$ such that (3.13) becomes the following

$$-(A+\overline{u})^{-1}f(A+\overline{u})^{-*}\left((A+\overline{u})^{-1}f - z\right) + \alpha\overline{u} + \overline{\mu} = 0. \tag{3.21}$$

By setting $\overline{y} = (A+\overline{u})^{-1}f$ and $\overline{p} := -(A+\overline{u})^{-*}(\overline{y}-z)$ this can be written as follows, $\overline{y}\,\overline{p} + \alpha\overline{u} + \overline{\mu} = 0$. We summarize the previous considerations into the following theorem.

**Theorem 3.4.** (*Bilinear optimality system*) *The optimal solution* $(\overline{y},\overline{u}) \in H_0^1(\Omega) \times L^2(\Omega)$ *to (3.2) in the bilinear control case is characterized by the existence of the dual pair* $(\overline{p},\overline{\mu}) \in H_0^1(\Omega) \times L^2(\Omega)$ *such that*

$$\begin{aligned}
A\overline{y} + \overline{u}\,\overline{y} - f &= 0 \\
A^*\overline{p} + \overline{y} + \overline{u}\,\overline{p} - z &= 0 \\
\overline{y}\,\overline{p} + \alpha\overline{u} + \overline{\mu} &= 0 \\
B(\overline{u},\overline{\mu}) &= 0.
\end{aligned} \tag{3.22}$$

Furthermore, the explicit reduced gradient and the reduced Hessian of $\hat{J}_2(u)$ are given by

$$\nabla\hat{J}_2(u) = \alpha u + py, \tag{3.23}$$

and

$$\begin{aligned}
\hat{J}_2''(u)(v_1,v_2) &= \left\langle v_1, y(A+u)^{-*}(A+u)^{-1}(yv_2) - y(A+u)^{-*}(pv_2) \right\rangle \\
&\quad - \left\langle v_1, p(A+u)^{-1}(yv_2) + \alpha v_2 \right\rangle.
\end{aligned}$$

## 4. Proximal Methods for Elliptic Control Problems

In this section, we discuss first-order proximal methods to solve our linear and bilinear optimal control problems. The starting point to discuss proximal methods consists of identifying a smooth and a nonsmooth part in the reduced objective $\hat{J}(u)$. That is, we consider the following optimization problem

$$\min_{u \in U_{ad}} \hat{J}(u) := \hat{J}_1(u) + \hat{J}_2(u), \tag{4.1}$$

where, we assume

$$\hat{J}_1(u) \text{ is continuous, convex and nondifferentiable} \tag{4.2}$$

$$\begin{aligned}
&\hat{J}_2(u) \text{ is Q-differentiable with respect to } U_{ad}, \text{ convex,} \\
&\text{and has Lipschitz-continuous gradient:}
\end{aligned} \tag{4.3}$$

$$\left\|\nabla\hat{J}_2(u) - \nabla\hat{J}_2(v)\right\| \le L(\hat{J}_2)\|u-v\|, \quad \forall u,v \in U_{ad}, \tag{4.4}$$

where $L(\hat{J}_2) > 0$. Notice that our optimal control problem (3.2) has this additive structure where (4.2) holds for $\hat{J}_1(u) = \beta\|u\|_{L^1}$ and $\hat{J}_2(u) = \frac{1}{2}\|S(u)-z\|^2 + \frac{\alpha}{2}\|u\|^2$ is at least two times Q-differentiable, it is convex under appropriate conditions discussed in the previous section, and it has Lipschitz-continuous gradient that we prove in the following lemma.

**Lemma 4.1.** *The functional* $\hat{J}_2(u) = \frac{1}{2}\|S(u)-z\|^2 + \frac{\alpha}{2}\|u\|^2$ *has a Lipschitz-continuous gradient for* $S(u) = A^{-1}(f-u)$ *(linear-control case) and for* $S(u) = (A+u)^{-1}f$ *(bilinear-control case).*

*Proof.* For the linear-control case, we have

$$\begin{aligned}
\left\|\nabla\hat{J}_2(u) - \nabla\hat{J}_2(v)\right\| &= \left\|\alpha(u-v) + A^{-*}A^{-1}(u-v)\right\| \\
&\le \alpha\|u-v\| + \left\|A^{-*}A^{-1}\right\|\|u-v\| = \left(\alpha + \left\|A^{-*}A^{-1}\right\|_{L^2,L^2}\right)\|u-v\|,
\end{aligned}$$

such that we have the Lipschitz-constant $L(\hat{J}_2) = (\alpha + \|A^{-*}A^{-1}\|_{L^2,L^2})$.

For the bilinear-control case, we use the mean value theorem. There exists a $\xi \in U_{ad}$ such that

$$\left\|\nabla\hat{J}_2(u) - \nabla\hat{J}_2(v)\right\| \leq \sup_{h\in L^2(\Omega),\|h\|\leq 1}\left|\hat{J}_2'(u)(h) - \hat{J}_2'(v)(h)\right| = \sup_{h\in L^2(\Omega),\|h\|\leq 1}\left|\hat{J}_2''(\xi)(h,u-v)\right|$$

$$= \sup_{h\in L^2(\Omega),\|h\|\leq 1}\left|\left\langle S_b'(\xi)(u-v), S_b'(\xi)(h)\right\rangle + \left\langle S_b''(\xi)(u-v,h), S_b(\xi)-z\right\rangle + \alpha\left\langle u-v,h\right\rangle\right|$$

$$\leq \left(C_2^2\|f\|^2 + C_1C_3\|f\|^2 - C_3\|f\|\|z\| + \alpha\right)\|u-v\|,$$

for the last inequality, we use (2.7), (2.13), (2.14), which completes the proof. $\square$

The following lemma is essential in the formulation of proximal methods.

**Lemma 4.2.** *Let $J_2$ be Q-differentiable with respect to $U_{ad}$, and it has a Lipschitz continuous gradient with Lipschitz constant $L(J_2)$. Then for all $L \geq L(J_2) > 0$, the following holds*

$$\hat{J}_2(u) \leq \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(v), u-v\right\rangle + \frac{L}{2}\|u-v\|^2, \quad \forall u,v \in U_{ad}. \tag{4.5}$$

*Proof.*

$$\hat{J}_2(u) = \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(v), u-v\right\rangle + \int_0^1\left\langle\nabla\hat{J}_2(v+t(u-v)) - \nabla\hat{J}_2(v), u-v\right\rangle dt$$

$$\leq \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(v), u-v\right\rangle + \int_0^1\left\|\nabla\hat{J}_2(v+t(u-v)) - \nabla\hat{J}_2(v)\right\|\|u-v\|dt$$

$$\leq \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(v), u-v\right\rangle + \int_0^1 Lt\|u-v\|^2 dt$$

$$\leq \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(v), u-v\right\rangle + \frac{L}{2}\|u-v\|^2.$$

$\square$

Notice that $L := L(\hat{J}_2)$ represents the smallest value of $L$ such that (4.5) is satisfied. We remark that the discussion that follows is valid for $L \geq L(\hat{J}_2)$ as in Lemma 4.5. However, as we discuss below, the efficiency of our proximal schemes depends on how close is the chosen $L$ to the minimal and optimal value $L(\hat{J}_2)$. Now, since this value is usually not available analytically, we discuss and implement below some numerical strategies for determining a sufficiently accurate approximation of $L(\hat{J}_2)$. In particular, we consider a power iteration [21], and the backtracking approach discussed in Remark 5.1.

Further, notice that also in the case of choosing $L \gg L(\hat{J}_2)$, our proximal scheme still converges with rate $1/k$ (resp. $1/k^2$) times a convergence constant. However, this convergence constant grows considerably as $L$ becomes larger and therefore the convergence of the proximal method appears recognizably slower. On the other hand, if $L$ is chosen smaller than the Lipschitz constant, then convergence cannot be guaranteed.

The strategy of the proximal scheme is to minimize an upper bound of the objective functional at each iteration, instead of minimizing the functional directly. Lemma 4.2 gives us the following upper bound for all $v \in U_{ad}$. We have

$$\min_{u\in U_{ad}}\left\{\hat{J}_1(u) + \hat{J}_2(u)\right\} \leq \min_{u\in U_{ad}}\left\{\hat{J}_1(u) + \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(y), u-v\right\rangle + \frac{L}{2}\|u-v\|^2\right\},$$

where, we have equality if $u = v$. Furthermore, we have the following equation

$$\mathrm{argmin}_{u\in U_{ad}}\left\{\hat{J}_1(u) + \hat{J}_2(v) + \left\langle\nabla\hat{J}_2(v), u-v\right\rangle + \frac{L}{2}\|u-v\|^2\right\}$$

$$= \mathrm{argmin}_{u\in U_{ad}}\left\{\hat{J}_1(u) + \frac{L}{2}\left\|u - \left(v - \frac{1}{L}\nabla\hat{J}_2(v)\right)\right\|^2\right\}. \tag{4.6}$$

Now, consider (4.6) and recall that $\hat{J}_1(u) = \beta\|u\|_{L^1}$. We have the following lemma.

**Lemma 4.3.** *The following equation holds*

$$\text{argmin}_{u \in U_{ad}} \left\{ \tau \|u\|_{L^1} + \frac{1}{2} \|u - v\|^2 \right\} = \mathbb{S}_\tau^{U_{ad}}(v) \quad \text{for any } v \in L^2(\Omega),$$

*where the projected soft thresholding function is defined as follows*

$$\mathbb{S}_\tau^{U_{ad}}(v) := \begin{cases} \min\{v - \tau, u_b\} & \text{on } \{x \in \Omega : v(x) > \tau\} \\ 0 & \text{on } \{x \in \Omega : |v(x)| \le \tau\} \\ \max\{v + \tau, u_a\} & \text{on } \{x \in \Omega : v(x) < -\tau\} \end{cases}$$

*Proof.* There exists a $\gamma(\bar{u}) \in \partial \|\bar{u}\|_{L^1}$, the subdifferential of $\|\cdot\|_{L_1}$ such that the solution $\bar{u} := \text{argmin}_{u \in U_{ad}} \left\{ \tau \|u\|_{L^1} + \frac{1}{2} \|u - v\|^2 \right\}$ fulfills the following variational inequality; see, e.g., ([20], Remark 3.2);

$$\langle \bar{u} - v + \tau \gamma(\bar{u}), u - \bar{u} \rangle \ge 0, \quad \forall u \in U_{ad}. \tag{4.7}$$

Now, we show that $\hat{u} := \mathbb{S}_\tau^{U_{ad}}(v)$ fulfills (4.7). The following investigation of the different cases is meant to be pointwise. We have

- $v - \tau > u_b \ge 0$:

  It follows that $\hat{u} = u_b$ and therefore $\gamma(\hat{u}) = 1$ such that

  $$(u_b - v + \tau)(u - u_b) \ge 0, \forall u \in U_{ad}.$$

- $0 < v - \tau < u_b$:

  It follows that $\hat{u} = v - \tau > 0$ and $\gamma(\hat{u}) = 1$ such that

  $$(\hat{u} - v + \tau)(u - u_b) = 0, \forall u \in U_{ad}.$$

- $|v| \le \tau$:

  It follows that $\hat{u} = 0$ and $\gamma(\hat{u}) = \frac{v}{\tau} \in B_1(0)$ such that

  $$\left(\hat{u} - v + \tau\left(\frac{v}{\tau}\right)\right)(u - \hat{u}) = 0, \forall u \in U_{ad}.$$

- $u_a < v + \tau < 0$:

  It follows that $\hat{u} = v + \tau < 0$ and therefore $\gamma(\hat{u}) = -1$ such that

  $$(\hat{u} - v - \tau)(u - \hat{u}) = 0, \forall u \in U_{ad}$$

- $v + \tau < u_a \le 0$:

  It follows that $\hat{u} = u_a$ and therefore $\gamma(\hat{u}) = -1$ such that

  $$(u_a - v - \tau)(u - u_a) \ge 0, \forall u \in U_{ad}.$$

$\square$

Based on this lemma, we conclude that the solution to (4.6) is given by

$$\text{argmin}_{u \in U_{ad}} \left\{ \hat{J}_1(u) + \frac{L}{2} \left\| u - \left( v - \frac{1}{L} \nabla \hat{J}_2(v) \right) \right\|^2 \right\} = \mathbb{S}_{\frac{\beta}{L}}^{U_{ad}} \left( v - \frac{1}{L} \nabla \hat{J}_2(v) \right),$$

thus obtaining an approximation to the optimal $u$ sought. Therefore we can use this result to define an iterative scheme as follows

$$u_k \leftarrow \mathbb{S}_{\frac{\beta}{L}}^{U_{ad}} \left( u_{k-1} - \frac{1}{L} \nabla \hat{J}_2(u_{k-1}) \right),$$

starting from a given $u_0$. The resulting algorithm implements a proximal scheme as follows

---

**Algorithm 1** (Proximal (P) method)

---

**Require:** $\beta$, $\hat{J}_2$, $u_0$, $U_{ad}$, $TOL$

  **Initialize:** $L = L(\hat{J}_2)$; $v_0 = u_0$; $t_0 = 1$; $B_0 = 1$, $k = 1$;

  **while** $\|B_{k-1}\| > TOL$ **do**

    1. $u_k = \mathbb{S}^{U_{ad}}_{\frac{\beta}{L}} \left( u_{k-1} - \frac{1}{L} \nabla \hat{J}_2 (u_{k-1}) \right)$

    2. $\mu_k = -\alpha u_k - S'(u_k)^* (S(u_k) - z)$   (3.13)

    3. $B_k = B(u_k, \mu_k)$

    4. $k = k + 1$

  **end while**

---

This scheme is discussed in [9] [12] for the case of finite-dimensional optimization problems. Notice that the iterated thresholding scheme discussed in [9] coincides with Algorithm 1 for the special case $L = 1$. The convergence results for Algorithm 1 presented in [12] can be extended to our elliptic control problems, using the theoretical results presented above. Therefore we can state the following theorem.

**Theorem 4.4.** *Let* $\{u_k\}$ *be a sequence generated by Algorithm* 1 *and* $u^*$ *be the solution to* (3.2) *with linear- or bilinear-control elliptic equality constraints. Then, for every* $k \geq 1$ *the following holds*

$$\hat{J}(u_k) - \hat{J}(u^*) \leq \frac{L(\hat{J}_2) \|u_0 - u^*\|^2}{2k}.$$

In [22], an acceleration strategy for proximal methods applied to convex optimization problems fulfilling (4.4) is formulated, which improves the rate of convergence of these schemes from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$. Specifically, one defines the sequence $\{t_k, v_k\}$ with

$$t_0 = 1, \qquad t_k := 1 + \sqrt{1 + 4t_{k-1}^2}\big/2, \tag{4.8}$$

and

$$v_0 := u_0, \qquad v_k := u_k + \frac{(t_{k-1} - 1)}{t_k}(u_k - u_{k-1}). \tag{4.9}$$

Correspondingly, the optimization variable $u_k$ is updated by the following

$$u_k \leftarrow \mathbb{S}^{U_{ad}}_{\frac{\beta}{L}} \left( v_{k-1} - \frac{1}{L} \nabla \hat{J}_2 (v_{k-1}) \right).$$

This procedures is summarized in the following algorithm

---

**Algorithm 2** (Fast proximal (FP) method)

---

**Require:** $\beta$, $\hat{J}_2$, $u_0$, $U_{ad}$, $TOL$

  **Initialize:** $L = L(\hat{J}_2)$; $v_0 = u_0$; $t_0 = 1$; $B_0 = 1$, $k = 1$;

  **while** $\|B_{k-1}\| > TOL$ **do**

    1. $u_k = \mathbb{S}^{U_{ad}}_{\frac{\beta}{L}} \left( v_{k-1} - \frac{1}{L} \nabla \hat{J}_2 (v_{k-1}) \right)$

    2. $\mu_k = -\alpha u_k - S'(u_k)^* (S(u_k) - z)$   (3.13)

    3. $B_k = B(u_k, \mu_k)$

    4. $t_k = \dfrac{1 + \sqrt{1 + 4t_{k-1}^2}}{2}$

    5. $v_k = u_k + \left( \dfrac{t_{k-1} - 1}{t_k} \right)(u_k - u_{k-1})$

    6. $k = k + 1$

  **end while**

---

The following convergence result represents an extension of ([12], Theorem 4.4). We have

**Theorem 4.5.** *Let* $\{u_k\}$ *be a sequence generated by Algorithm 2 and* $u^*$ *be the solution of* (3.2) *with linear- or bilinear-control elliptic equality constraints. Then, for every* $k \geq 1$ *the following holds*

$$\hat{J}(u_k) - \hat{J}(u^*) \leq \frac{2L(\hat{J}_2)\|u_0 - u^*\|^2}{(k+1)^2}.$$

Algorithm 1 and Algorithm 2 require the calculation of

$$\nabla \hat{J}_2(u) = -A^{-*}\left(A^{-1}(f - u) - z\right) + \alpha u \quad \text{(linear)}$$

$$\text{resp.} \quad \nabla \hat{J}_2(u) = -(A+u)^{-*}\left((A+u)^{-1}f - z\right)(A+u)^{-1}f + \alpha u \quad \text{(bilinear)}.$$

However, the exact inversion of a discretized elliptic differential operator $A$ may become too expensive. Therefore one has to use iterative methods; e.g., the conjugate gradient method [23]. For this reason, we discuss an inexact version of the proximal scheme, where the equality constraints and the corresponding adjoint equations are solved up to a given tolerance quantified by $\varepsilon > 0$. In the following, we denote with $\nabla_\varepsilon \hat{J}_2(u)$ the inexact gradient that corresponds to an inexact inversion of the equation $Ay = f - u$, resp. $(A+u)y = f$, that results in an approximated state variable $y^\varepsilon$, resp. $p^\varepsilon$, in the following sense

$$\|Ay^\varepsilon - f + u\| \leq \varepsilon, \qquad \text{resp.} \quad \|(A+u)y^\varepsilon - f\| \leq \varepsilon.$$

Hence, there exists an $\tilde{e} \in L^2(\Omega)$ with $\|\tilde{e}\| < \varepsilon$ such that

$$y^\varepsilon = A^{-1}(f - u + \tilde{e}), \qquad \text{resp.} \quad y^\varepsilon = (A+u)^{-1}(f + \tilde{e}). \tag{4.10}$$

We denote the inexact inversion method for the problem $\mathcal{A}y = g$, with an error $\|\mathcal{A}y^\varepsilon - g\| \leq \varepsilon$, with $inv(\mathcal{A}, g, \varepsilon)$. With this notation, the inexact gradient computation is illustrated in Algorithm 3.

---

**Algorithm 3** (Calculation of the inexact gradient $\nabla_\varepsilon \hat{J}_2(u)$)

**Require:** $\beta$, $\hat{J}_2$, $u_0$, $U_{ad}$, $TOL$

   1. $y^\varepsilon = inv(A, f - u, \varepsilon)$, resp. $y^\varepsilon = inv(A + u, f, \varepsilon)$

   2. $p^\varepsilon = inv(A^*, z - y^\varepsilon, \varepsilon)$, resp. $p^\varepsilon = inv(A^* + u, z - y^\varepsilon, \varepsilon)$

   3. $\nabla_\varepsilon \hat{J}_2(u) = p^\varepsilon + \alpha u$, resp. $\nabla_\varepsilon \hat{J}_2(u) = p^\varepsilon y^\varepsilon + \alpha u$

---

With this preparation, we formulate our inexact proximal (IP) scheme with Algorithm 4.

---

**Algorithm 4** (Inexact proximal (IP) method)

**Require:** $\beta$, $\hat{J}_2$, $u_0$, $U_{ad}$, $TOL$, $\varepsilon_0$

  **Initialize:** $L = L(\hat{J}_2)$; $v_0 = u_0$; $t_0 = 1$; $B_0 = 1$; $k = 1$

  **while** $\|B_{k-1}\| > TOL$ **do**

    1. $\varepsilon_k := \dfrac{\varepsilon_0}{k}$

    2. $u_k = \mathbb{S}^{U_{ad}}_{\frac{\beta}{L}}\left(u_{k-1} - \dfrac{1}{L}\nabla_{\varepsilon_k}\hat{J}_2(u_{k-1})\right)$

    3. $\mu_k = -\alpha u_k - S'(u_k)^*(S(u_k) - z)$   (3.13)

    4. $B_k = B(u_k, \mu_k)$

    5. $k = k + 1$

  **end while**

---

We also formulate the accelerated (fast) version of our IP scheme in Algorithm 5. We refer to it as the FIP

method.

---

**Algorithm 5** (Fast inexact proximal (FIP) method)

---

**Require:** $\beta$, $\hat{J}_2$, $u_0$, $U_{ad}$, $TOL$, $\varepsilon_0$
  **Initialize:** $L = L(\hat{J}_2)$; $v_0 = u_0$; $t_0 = 1$; $B_0 = 1$, $k = 1$;
  **while** $\|B_{k-1}\| > TOL$ **do**

  1. $\varepsilon_k := \dfrac{\varepsilon_0}{(k+1)^3}$

  2. $u_k = \mathbb{S}^{U_{ad}}_{\frac{\beta}{L}}\left(v_{k-1} - \dfrac{1}{L}\nabla_{\varepsilon_k}\hat{J}_2(v_{k-1})\right)$

  3. $\mu_k = -\alpha u_k - S'(u_k)^*(S(u_k)-z)$  (3.13)

  4. $B_k = B(u_k,\mu_k)$

  5. $t_k = \dfrac{1+\sqrt{1+4t_{k-1}^2}}{2}$

  6. $v_k = u_k + \left(\dfrac{t_{k-1}-1}{t_k}\right)(u_k - u_{k-1})$

  7. $k = k+1$
  **end while**

---

# 5. Convergence Analysis of Inexact Proximal Methods

In this section, we investigate the convergence of our IP and FIP schemes. Notice that our analysis differs from that presented in [12] where finite-dimensional problems and exact inversion are considered. We start investigating the error of the inexact gradient $\hat{J}_2^{\prime\varepsilon}(u)$.

**Lemma 5.1.** *The following estimate holds*

$$\left\|\nabla_\varepsilon \hat{J}_2(u) - \nabla \hat{J}_2(u)\right\| \le c\varepsilon,$$

*for some $c > 0$.*

  *Proof.* In the linear case, we have $\nabla \hat{J}_2(u) = -A^{-*}\left(A^{-1}(f-u)-z\right)+\alpha u$. Using (4.10) there exist the errors $\tilde{e}_1, \tilde{e}_2 \in L^2(\Omega)$ with $\|\tilde{e}_1\|, \|\tilde{e}_2\| < \varepsilon$ such that

$$\begin{aligned}
\left\|\nabla_\varepsilon \hat{J}_2(u) - \nabla \hat{J}_2(u)\right\| &= \left\|-A^{-*}\left(A^{-1}(f-u+\tilde{e}_1)-z+\tilde{e}_2\right)+A^{-*}\left(A^{-1}(f-u)-z\right)\right\| \\
&= \left\|-A^{-*}A^{-1}\tilde{e}_1 + A^{-*}\tilde{e}_2\right\| < c\varepsilon,
\end{aligned}$$

(5.1)

where

$$c = \left\|A^{-*}A^{-1}\right\| + \left\|A^{-*}\right\|.$$

(5.2)

  In the bilinear case, we have $\nabla \hat{J}_2(u) = -(A+u)^{-*}\left((A+u)^{-1}f-z\right)(A+u)^{-1}f + \alpha u$. Furthermore, Theorem 2.1 implies that the solution $\tilde{y} := (A+u)^{-1}g$ of the equation $A\tilde{y}+u\tilde{y}=g$ has the following property

$$\left\|(A+u)^{-1}g\right\| \le C_1\|g\|, \quad \text{for all } g \in L^2(\Omega).$$

(5.3)

  We also have $\left\|(A+u)^{-*}g\right\| = \left\|(A^*+u)^{-1}g\right\| \le C_1\|g\|$.

  Using (4.10) there exist the errors $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3 \in L^2(\Omega)$ with $\|\tilde{e}_1\|, \|\tilde{e}_2\|, \|\tilde{e}_3\| < \varepsilon$ such that

$$\left\| \nabla_\varepsilon \hat{J}_2(u) - \nabla \hat{J}_2(u) \right\|$$

$$= \left\| -(A+u)^{-*} \left( (A+u)^{-1}(f+\tilde{e}_1) - z + \tilde{e}_2 \right)(A+u)^{-1}(f+\tilde{e}_3) \right.$$
$$\left. + (A+u)^{-*} \left( (A+u)^{-1} f - z \right)(A+u)^{-1} f \right\|$$

$$= \left\| -(A+u)^{-*} \left( (A+u)^{-1} \tilde{e}_1 + \tilde{e}_2 \right)(A+u)^{-1}(f+\tilde{e}_3) \right.$$
$$\left. -(A+u)^{-*} \left( (A+u)^{-1} f - z \right)(A+u)^{-1} \tilde{e}_3 \right\|$$

$$\leq \left\| (A+u)^{-*} \left( (A+u)^{-1} \tilde{e}_1 + \tilde{e}_2 \right) \right\| \left( \|y\| + \left\| (A+u)^{-1} \tilde{e}_3 \right\| \right)$$
$$+ \left\| (A+u)^{-*} \left( (A+u)^{-1} f - z \right) \right\| \left\| (A+u)^{-1} \tilde{e}_3 \right\|$$

$$\leq C_1 \left\| (A+u)^{-1} \tilde{e}_1 + \tilde{e}_2 \right\| \left( C_1 \|f\| + C_1 \|\tilde{e}_3\| \right)$$
$$+ C_1 \left( \|y\| + \|z\| \right) C_1 \|\tilde{e}_3\|$$

$$\leq C_1^2 \left[ (C_1 \varepsilon + \varepsilon)(\|f\| + \varepsilon) + (C_1 \|f\| + \|z\|)\varepsilon \right] \leq c\varepsilon,$$

where

$$c = C_1^2 \left[ 2C_1 \|f\| + \|f\| + C_1 + 1 + \|z\| \right]. \tag{5.4}$$

For the three last inequalities, we use (5.3), (2.7), and $\varepsilon < 1$. □

We refer to the estimation error of the inexact gradient in step $k$ as follows

$$e_k := \nabla_\varepsilon \hat{J}_2(u_k) - \nabla \hat{J}_2(u_k), \quad \text{where } \|e_k\| \leq c\varepsilon_k.$$

Now, we define

$$Q_L(u,v) := \beta \|u\|_{L^1} + \hat{J}_2(v) + \left\langle \nabla \hat{J}_2(v), u - v \right\rangle + \frac{L}{2} \|u - v\|^2,$$

$$Q_L^e(u,v) := \beta \|u\|_{L^1} + \hat{J}_2(v) + \left\langle \nabla \hat{J}_2(v), u - v \right\rangle + \frac{L}{2} \|u - v\|^2 + \left\langle e, u - v \right\rangle,$$

and

$$P_L^e(v) := \arg\min_{u \in U_{ad}} \left\{ Q_L^e(u,v) \right\}, \tag{5.5}$$

such that one step of Algorithm 4, resp. Algorithm 5, can be written as follows

$$u_k = P_L^{e_{k-1}}(u_{k-1}), \quad \text{resp.} \quad u_k = P_L^{e_{k-1}}(v_{k-1}).$$

In order to prove the convergence of the IP method, we need the following two lemmas.

**Lemma 5.2.** *For any* $v \in U_{ad}$, *one has* $w = P_L^e(v)$ *iff there exists* $\gamma(v) \in \partial \|w\|_{L^1}$, *the subdifferential of* $\|\cdot\|_{L^1}$, *such that*

$$\left\langle \nabla \hat{J}_2(v) + L(w - v) + \beta \gamma(v) + e, u - w \right\rangle \geq 0, \quad \forall u \in U_{ad}. \tag{5.6}$$

*Proof.* This is immediate from the variational inequality of (5.5). For a proof see, e.g., [20]. □

**Lemma 5.3.** *Let* $v \in U_{ad}$ *and* $L > L(\hat{J}_2)$, *then for any* $u \in U_{ad}$, *we have*

$$\hat{J}(u) - \hat{J}\left( P_L^e(v) \right) \geq \frac{L}{2} \left\| P_L^e(v) - v \right\|^2 + L \left\langle v - u, P_L^e(v) - v \right\rangle + \left\langle P_L^e(v) - u, e \right\rangle.$$

*Proof.* From (4.5), we have

$$\hat{J}\left( P_L^e(v) \right) \leq Q_L\left( P_L^e(v), v \right),$$

and therefore

$$\hat{J}(u) - \hat{J}\left(P_L^e(v)\right) \geq \hat{J}(u) - Q_L\left(P_L^e(v), v\right). \tag{5.7}$$

Now since $\beta\|\cdot\|_{L^1}$ and $\hat{J}_2$ are convex, we have

$$\beta\|u\|_{L^1} \geq \beta\left\|P_L^e(v)\right\|_{L^1} + \left\langle u - P_L^e(v), \beta\gamma(v)\right\rangle \text{ and } \hat{J}_2(u) \geq \hat{J}_2(v) + \left\langle u - v, \nabla\hat{J}_2(v)\right\rangle.$$

Summing the above inequalities gives

$$\hat{J}(u) \geq \beta\left\|P_L^e(v)\right\|_{L^1} + \left\langle u - P_L^e(v), \beta\gamma(v)\right\rangle + \hat{J}_2(v) + \left\langle u - v, \nabla\hat{J}_2(v)\right\rangle, \tag{5.8}$$

so using (5.6), (5.8), and the definition of $Q_L$ in (5.7) gives the following

$$\begin{aligned}
\hat{J}(u) - \hat{J}\left(P_L^e(v)\right) &\geq -\frac{L}{2}\left\|P_L^e(v) - v\right\|^2 + \left\langle u - P_L^e(v), \nabla\hat{J}_2(v) + \beta\gamma(v)\right\rangle \\
&\geq -\frac{L}{2}\left\|P_L^e(v) - v\right\|^2 + L\left\langle u - P_L^e(v), v - P_L^e(v)\right\rangle + \left\langle P_L^e(v) - u, e\right\rangle \\
&= \frac{L}{2}\left\|P_L^e(v) - v\right\|^2 + L\left\langle v - u, P_L^e(v) - v\right\rangle + \left\langle P_L^e(v) - u, e\right\rangle.
\end{aligned}$$

$$\square$$

Now, we prove a $\mathcal{O}$ convergence rate for Algorithm 4 (IP scheme).

**Theorem 5.4.** *Let* $(u_k)$ *be the sequence generated by Algorithm* 4 *and* $u^*$ *be the solution of* (3.2) *with linear or bilinear elliptic equality constraints; let c be determined by* (5.2) *resp.* (5.4). *Then for any* $k \geq 1$, *we have*

$$\hat{J}(u_k) - \hat{J}(u^*) \leq \frac{L\left(\hat{J}_2\right)\left\|u_0 - u^*\right\|^2 + 2c\left\|u_b - u_a\right\|\varepsilon_0}{2k}. \tag{5.9}$$

*Proof.* Using Lemma 5.3 with $u = u^*$, $v = u_n$ and $L = L\left(\hat{J}_2\right)$ we obtain

$$\begin{aligned}
\frac{2}{L}\left(J\left(u^*\right) - J\left(u_{n+1}\right)\right) &\geq \left\|u_{n+1} - u_n\right\|^2 + 2\left\langle u_n - u^*, u_{n+1} - u_n\right\rangle + \frac{2}{L}\left\langle u_{n+1} - u^*, e_k\right\rangle \\
&= \left\|u^* - u_{n+1}\right\|^2 - \left\|u^* - u_n\right\|^2 + \frac{2}{L}\left\langle u_{n+1} - u^*, e_k\right\rangle.
\end{aligned}$$

Summing this inequality over $n = 0, \cdots, k - 1$ gives

$$\frac{2}{L}\left(kJ\left(u^*\right) - \sum_{n=0}^{k-1}J\left(u_{n+1}\right)\right) \geq \left\|u^* - u_k\right\|^2 + \left\|u^* - u_0\right\|^2 + \frac{2}{L}\sum_{n=0}^{k-1}\left\langle u_{n+1}, e_k\right\rangle - \frac{2}{L}k\left\langle u^*, e_k\right\rangle. \tag{5.10}$$

Using again Lemma 5.3 with $u = v = u_n$, we obtain

$$\frac{2}{L}\left(J\left(u_n\right) - J\left(u_{n+1}\right)\right) \geq \left\|u_n - u_{n+1}\right\|^2 + \frac{2}{L}\left\langle u_{n+1} - u_n, e_k\right\rangle.$$

Multiplying this inequality by $n$ and summing again over $n = 0, \cdots, k - 1$ gives

$$\begin{aligned}
&\frac{2}{L}\sum_{n=0}^{k-1}\left(nJ\left(u_n\right) - (n+1)J\left(u_{n+1}\right) + J\left(u_{n+1}\right)\right) \\
&\geq \sum_{n=0}^{k-1}n\left\|u_n - u_{n+1}\right\|^2 + \frac{2}{L}\sum_{n=0}^{k-1}\left(-n\left\langle u_n, e_k\right\rangle + (n+1)\left\langle u_{n+1}, e_k\right\rangle - \left\langle u_{n+1}, e_k\right\rangle\right),
\end{aligned}$$

which simplifies to the following

$$\frac{2}{L}\left(-kJ\left(u_k\right) + \sum_{n=0}^{k-1}J\left(u_{n+1}\right)\right) \geq \sum_{n=0}^{k-1}n\left\|u_n - u_{n+1}\right\|^2 + \frac{2}{L}k\left\langle u_k, e_k\right\rangle - \frac{2}{L}\sum_{n=0}^{k-1}\left\langle u_{n+1}, e_k\right\rangle. \tag{5.11}$$

Adding (5.10) and (5.11) together, we get

$$\frac{2k}{L}\left(J\left(u^*\right)-J\left(u_k\right)\right)\geq\left\|u^*-u_k\right\|^2+\sum_{n=0}^{k-1}n\left\|u_n-u_{n+1}\right\|^2-\left\|u^*-u_0\right\|^2+\frac{2}{L}k\left\langle u_k-u^*,e_k\right\rangle,$$

and hence with $\varepsilon_k=\dfrac{\varepsilon_0}{k}$ and $u_a\leq u_k,u^*\leq u_b$, it follows that

$$J\left(u_k\right)-J\left(u^*\right)\leq\frac{L\left\|u_0-u^*\right\|^2}{2k}+L\left\langle u^*-u_k,e\right\rangle\leq\frac{L\left\|u_0-u^*\right\|^2}{2k}+c\left\|u^*-u_k\right\|\varepsilon_k$$

$$\leq\frac{L\left(\hat{J}_2\right)\left\|u_0-u^*\right\|^2+2c\left\|u^*-u_k\right\|\varepsilon_0}{2k}\leq\frac{L\left(\hat{J}_2\right)\left\|u_0-u^*\right\|^2+2c\left\|u_b-u_a\right\|\varepsilon_0}{2k}.$$

$\square$

Next, we present a convergence result for the FIP method. For this purpose, we need the following lemma.

**Lemma 5.5.** *Let* $\left(u_k\right)$, $\left(v_k\right)$ *and* $\left(t_k\right)$ *be the sequences generated by Algorithm 5, let* $e_k$ *be the error of the inexact gradient, and let* $u^*$ *be the solution to (3.2), then for any* $k\geq 1$, *we have*

$$\frac{2}{L}t_{k-1}^2w_k-\frac{2}{L}t_kw_{k+1}\geq\left\|r_{k+1}\right\|^2-\left\|r_k\right\|^2+\frac{2}{L}t_k\left\langle r_{k+1},e_k\right\rangle,$$

with $w_k:=J\left(u_k\right)-J\left(u^*\right)$, $r_k:=t_{k-1}u_k-\left(t_{k-1}-1\right)u_{k-1}-u^*$.

*Proof.* We apply Lemma 4.2 at the points $\left(u:=u_k,v:=v_k\right)$ and likewise at the points $\left(u:=u^*,v:=v_k\right)$. We obtain the following

$$2L^{-1}\left(w_k-w_{k-1}\right)\geq\left\|u_{k+1}-v_k\right\|^2+2\left\langle u_{k+1}-v_k,v_k-u_k\right\rangle+2L^{-1}\left\langle u_{k+1}-u_k,e_k\right\rangle,$$

$$-2L^{-1}w_{k-1}\geq\left\|u_{k+1}-v_k\right\|^2+2\left\langle u_{k+1}-v_k,v_k-u^*\right\rangle+2L^{-1}\left\langle u_{k+1}-u^*,e_k\right\rangle,$$

where we used the fact that $u_{k+1}=P_L^e\left(v_k\right)$. Now, we multiply the first inequality above by $\left(t_k-1\right)$ and add it to the second inequality to obtain the following

$$\frac{2}{L}\left(\left(t_k-1\right)w_k-t_kw_{k+1}\right)\geq t_k\left\|u_{k+1}-v_k\right\|^2+2\left\langle u_{k+1}-v_k,t_kv_k-\left(t_k-1\right)u_k-u^*\right\rangle$$

$$+\frac{2}{L}t_k\left\langle u_{k+1}-u_k,e_k\right\rangle+\frac{2}{L}\left\langle u_k-u^*,e_k\right\rangle.$$

Multiplying this inequality by $t_k$ and using $t_{k-1}^2=t_k^2-t_k$, which holds due to (4.8), we obtain

$$\frac{2}{L}\left(t_{k-1}^2w_k-t_k^2w_{k+1}\right)\geq\left\|t_k\left(u_{k+1}-v_k\right)\right\|^2+2t_k\left\langle u_{k+1}-v_k,t_kv_k-\left(t_k-1\right)u_k-u^*\right\rangle$$

$$+\frac{2}{L}t_k\left\langle t_ku_{k+1}-\left(t_k-1\right)u_k-u^*,e_k\right\rangle.$$

Applying the Pythagoras relation $\left\|a-b\right\|^2+2\left\langle b-a,a-c\right\rangle=\left\|b-c\right\|^2-\left\|a-c\right\|^2$, to the right-hand side of the last inequality with $a:=t_kv_k,b:=t_ku_{k+1},c:=\left(t_k-1\right)u_k+u^*$, we obtain

$$\frac{2}{L}\left(t_{k-1}^2w_k-t_k^2w_{k+1}\right)\geq\left\|t_ku_{k+1}-\left(t_k-1\right)u_k-u^*\right\|^2-\left\|t_kv_k-\left(t_k-1\right)u_k-u^*\right\|^2$$

$$+\frac{2}{L}t_k\left\langle t_ku_{k+1}-\left(t_k-1\right)u_k-u^*,e_k\right\rangle.$$

Therefore, with $v_k$ (see (4.9)) and $r_k$ defined as

$$t_kv_k=t_ku_k+\left(t_{k-1}-1\right)\left(u_k-u_{k-1}\right),\quad r_k:=t_{k-1}u_k-\left(t_{k-1}-1\right)u_{k-1}-u^*,$$

it follows that

$$\frac{2}{L}t_{k-1}^2w_k-\frac{2}{L}t_kw_{k+1}\geq\left\|r_{k+1}\right\|^2-\left\|r_k\right\|^2+\frac{2}{L}t_k\left\langle r_{k+1},e_k\right\rangle.$$

$\square$

We also have the following lemmas.

**Lemma 5.6.** *The positive sequence* $(t_k)$ *generated by the FIP scheme via* (4.8) *with* $t_0 = 1$ *satisfies* $(k+2)/2 \le t_k \le k+1$ *for all* $k \ge 0$.

*Proof.* The proof is immediate by mathematical induction. $\qquad\square$

**Lemma 5.7.** *Let* $(a_k)$ *and* $(b_k)$ *be positive sequences of reals and* $(c_k)$ *be a sequence of reals satisfying*

$$a_k + b_k \ge a_{k+1} + b_{k+1} + c_{k+1} \quad \forall k \ge 1 \text{ and } a_1 + b_1 + c_1 \le d, d > 0.$$

*Then,* $a_k \le d - \sum_{n=1}^{k} c_n$.

*Proof.* The proof is immediate by mathematical induction. $\qquad\square$

Now, we can prove a convergence rate of $\mathcal{O}(1/k^2)$ for Algorithm 5 (FIP scheme).

**Theorem 5.8.** *Let* $(u_k)$ *be the sequence generated by Algorithm 5, let* $u^*$ *be the solution to* (3.2) *with linear or bilinear elliptic equality constraints; let c be determined by* (5.2) *resp.* (5.4). *Then for any* $k \ge 0$, *the following holds*

$$J(u_k) - J(u^*) \le \frac{2L(\hat{J}_2)\|u_0 - u^*\|^2 + 2c\|u_b - 2u_a\|\varepsilon_0}{(k+1)^2}. \tag{5.12}$$

*Proof.* Let us define the quantities

$$a_k := \frac{2}{L}t_{k-1}^2 w_k, \quad b_k := \|r_k\|^2, \quad c_k := \frac{2}{L}t_{k-1}\langle r_k, e_k \rangle, \quad d := \|u_0 - u^*\|^2.$$

As in Lemma 5.5, we define $w_k := J(u_k) - J(u^*)$. Then, by Lemma 5.5, the following holds for every $k \ge 1$

$$a_k - a_{k+1} \ge b_{k+1} - b_k + c_{k+1} \Leftrightarrow a_k + b_k \ge a_{k+1} + b_{k+1} + c_{k+1},$$

and hence assuming that $a_1 + b_1 + c_1 \le d$ holds true, invoking Lemma 5.7, we obtain

$$\frac{2}{L}t_{k-1}^2 w_k \le \|x_0 - x^*\|^2 + \frac{2}{L}t_{k-1}k \max_{n=1,\cdots,k}(\|r_n\|)\|e_n\|,$$

which combined with $t_{k-1} \ge (k+1)/2$ (Lemma 5.6) gives the following

$$w_k \le \frac{2L\|u_0 - u^*\|^2}{(k+1)^2} + 2c \max_{n=1,\cdots,k}(\|r_n\|)\varepsilon_k. \tag{5.13}$$

Furthermore with Lemma 5.6 and $u_a \le u^*, u_k \le u_b$, we have that

$$\|r_n\| = \|t_{n-1}u_n - (t_{n-1}-1)u_{n-1} - u^*\| \le \left\|nu_b - \left(\frac{n-1}{2}u_a + u_a\right)\right\| \le n\|u_b - 2u_a\|,$$

which combined with (5.13) and $\varepsilon_k = \frac{\varepsilon_0}{(k+1)^3}$ gives the following

$$w_k \le \frac{2L(\hat{J}_2)\|u_0 - u^*\|^2 + 2c\|u_b - 2u_a\|\varepsilon_0}{(k+1)^2}.$$

What remains to be proved is the validity of the relation $a_1 + b_1 + c_1 \le d$. Since $t_0 = 1$, we have

$$a_1 = \frac{2}{L}t_0^2 w_1 = \frac{2}{L}w_1, \quad b_1 = \|r_1\|^2 = \|u_1 - u^*\|^2, \quad c_1 = 2\langle u_1 - u^*, e_1 \rangle.$$

Applying Lemma 4.2 to the points $u := u^*$ and $v = v_0 = u_0$, we get

$$\frac{2}{L}\left(J(u^*) - J(u_1)\right) \ge \|u_1 - v_0\|^2 + 2\langle v_0 - u^*, u_1 - v_0 \rangle + \frac{2}{L}\langle u_1 - u^*, e_1 \rangle$$

$$= \|u_1 - u^*\|^2 - \|v_0 - u^*\|^2 + \frac{2}{L}\langle u_1 - u^*, e_1 \rangle,$$

that is $-a_1 \geq b_1 - d + c_1 \Leftrightarrow a_1 + b_1 + c_1 \leq d$ holds true. $\qquad \square$

**Remark 5.1.** *The IP and FIP methods converge also replacing L with an upper bound of it. In particular, we can prove $\mathcal{O}\left(1/k^2\right)$ convergence of the FIP method using a backtracking stepsize rule for the Lipschitz constant (Step 1 in Algorithm 5) as in* [12].

We complete this section formulating a fast inexact proximal scheme where the Lipschitz constant $L$ is obtained by forward tracking, (nevertheless we call it backtracking as in [12]), thus avoiding any need to compute the reduced Hessian. Our fast inexact proximal backtracking (FIPB) method is presented in Algorithm 6.

---

**Algorithm 6** (Fast inexact proximal backtracking (FIPB) method)

---

**Require:** $\beta$, $\hat{J}_2$, $u_0$, $U_{ad}$, $TOL$, $\varepsilon_0$, $\eta > 1$, $L_0 > 0$
  **Initialize:** $v_0 = u_0$; $t_0 = 1$; $B_0 = 1$, $k = 1$;
  **while** $\|B_{k-1}\| > TOL$ **do**

    1. Backtracking: Find the smallest nonnegative integer $i$ such that with
    $\tilde{L} = \eta^i L_{k-1}$

$$\hat{J}_2\left(\tilde{v}\right) \leq \hat{J}_2\left(v_{k-1}\right) + \left\langle \nabla \hat{J}_2\left(v_{k-1}\right), \tilde{v} - v_{k-1}\right\rangle + \frac{\tilde{L}}{2}\left\|\tilde{v} - v_{k-1}\right\|^2$$

    where $\quad \tilde{v} = \mathbb{S}_{\frac{\beta}{\tilde{L}}}^{U_{ad}}\left(v_{k-1} - \frac{1}{\tilde{L}}\nabla_\varepsilon \hat{J}_2\left(v_{k-1}\right)\right)$

    2. Set $\quad L_k = \tilde{L}$

    3. $\varepsilon_k := \dfrac{\varepsilon_0}{\left(k+1\right)^3}$

    4. $u_k = \mathbb{S}_{\frac{\beta}{L_k}}^{U_{ad}}\left(v_{k-1} - \dfrac{1}{L_k}\nabla_\varepsilon \hat{J}_2\left(v_{k-1}\right)\right)$

    5. $\mu_k = -\alpha u_k - S'\left(u_k\right)^*\left(S\left(u_k\right) - z\right)$   (3.13)

    6. $B_k = B\left(u_k, \mu_k\right)$

    7. $t_k = \dfrac{1 + \sqrt{1 + 4t_{k-1}^2}}{2}$

    8. $v_k = u_k + \left(\dfrac{t_{k-1} - 1}{t_k}\right)\left(u_k - u_{k-1}\right)$

    7. $k = k + 1$
  **end while**

---

# 6. The Inexact Semismooth Newton Method

We consider the semismooth Newton method as a benchmark scheme for solving elliptic non-smooth optimal control problems; see, e.g., [4]-[6]. This method is proven to be equivalent to the primal-dual active set method in [24]. The inexact semismooth Newton (ISSN) method is presented in [25] for finite-dimensional problems. In this section, we discuss the ISSN method for infinite-dimensional optimization problems and use it for comparison with our inexact proximal schemes. To support our use of the ISSN scheme to solve bilinear control problems, we extend two theoretical results in [3] [4]. For the analysis that follows, we need the following definition.

**Definition 6.1.** *Let X and Y be Banach spaces, $D \subset X$ be open and $\mathcal{F} : D \to Y$ be a nonlinear mapping. We say that $\mathcal{F}$ is generalized differentiable in an open subset $U \subset D$ if there exists a set-valued mapping $\partial^* \mathcal{F} : D \rightrightarrows \mathcal{L}\left(X, Y\right)$ with $\partial^* \mathcal{F}\left(x\right) \neq \varnothing$ for all $x \in D$ such that*

$$\lim_{h \to 0} \frac{1}{\|h\|_X} \|\mathcal{F}(x+h) - \mathcal{F}(x) - \mathcal{G}(x+h)h\|_Y = 0, \tag{6.1}$$

for every $\mathcal{G} \in \partial^* \mathcal{F}$ and for every $x \in U$. We call $\partial^* \mathcal{F}$ the generalized differential and every $\mathcal{G} \in \partial^* \mathcal{F}$ a generalized derivative.

This definition is similar to the semismoothness stated in [3] and also known under the name "slant differentiability"; see, e.g., [24]. Now, we discuss the solution of the following nonlinear equation $\mathcal{F}(x) = 0$. We have the following theorem.

**Theorem 6.1.** ([24], *Theorem* 1.1) *Suppose that* $x^*$ *is a solution to* $\mathcal{F}(x) = 0$ *and that* $\mathcal{F}$ *is generalized differentiable in an open neighborhood U containing* $x^*$ *with a generalized derivative* $\mathcal{G}$. *If* $\mathcal{G}(x)$ *is invertible for all* $x \in U$ *and* $\left\{ \left\| \mathcal{G}(x)^{-1} \right\|_{Y,X} : x \in U \right\}$ *is bounded, then the semismooth Newton* (*SSN*) *iteration*

$$x^{k+1} = x^k - \mathcal{G}(x^k)^{-1} \mathcal{F}(x^k),$$

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

An inexact version of the SSN scheme discussed in this theorem is formulated in ([3], Algorithm 3.19), where the update $d_k$ to $x^k$ is obtained as follows. Choose a boundedly invertible operator $B_k \in \mathcal{L}(X, Y)$ and compute $d_k = -B_k^{-1} \mathcal{F}(x_k)$. For this scheme, superlinear convergence is proven in ([3], Theorem 3.20), provided that there exists a $\mathcal{G}_k \in \partial^* \mathcal{F}(x_k)$ such that

$$\lim_{\|d_k\|_X \to 0} \frac{\|(B_k - \mathcal{G}_k)d_k\|_Y}{\|d_k\|_X} = 0.$$

However, this procedure is difficult to realize in practice. For this reason, in our ISSN scheme, the "exact" update step $x^{k+1} = x^k + d_k$ with $d_k = -\mathcal{G}(x^k)^{-1} \mathcal{F}(x^k)$, as discussed in [24], is replaced by $x^{k+1} = x^k + d_k$ with $d_k$ satisfying the following inequality

$$\|\mathcal{G}(x_k)d_k + \mathcal{F}(x_k)\|_Y \leq \eta_k \|\mathcal{F}(x_k)\|_Y. \tag{6.2}$$

Our ISSN scheme is given in Algorithm 7.

---

**Algorithm 7** (Inexact semismooth Newton (ISSN) method)

---

**Require:** $\mathcal{F}$, $x_0 \in D$
  **Initialize:** $k = 0$
  **while** $\mathcal{F}(x_k) = 0$ **do**
      1. Calculate the direction $d_k$ such that
$$\|\mathcal{G}(x_k)d_k + \mathcal{F}(x_k)\|_Y \leq \eta_k \|\mathcal{F}(x_k)\|_Y \quad (6.3)$$
      with $\eta_k < 1$ and $\eta_k \to 0$
      2. $x_{k+1} = x_k + d_k$
      3. $k = k + 1$
  **end while**

---

On the basis of the proof of Theorem 3.20 in [3], we prove the following theorem that states convergence of Algorithm 7. We have

**Theorem 6.2.** *Suppose that* $x^*$ *is a solution to* $\mathcal{F}(x) = 0$ *and that* $\mathcal{F}$ *is generalized differentiable and Lipschitz continuous in an open neighborhood U containing* $x^*$ *with a generalized derivative* $\mathcal{G}$. *If* $\mathcal{G}(x)$ *is invertible for all* $x \in U$ *and* $\left\{ \left\| \mathcal{G}(x)^{-1} \right\|_{Y,X} : x \in U \right\}$ *is bounded, then Algorithm 7 converges superlinearly to*

$x^*$, *provided that* $\left\| x_0 - x^* \right\|_X$ *is sufficiently small.*

*Proof.* Let $r_k := \mathcal{G}(x_k) d_k + \mathcal{F}(x_k)$ and $v_k := x_k - x^*$. Furthermore, let $\delta > 0$ be so small that $\left\| x_0 - x^* \right\|_X < \delta$ and $\mathcal{F}$ is Lipschitz continuous in $x^* + \delta B_X \subset U$ with $L > 0$. Now, we show inductively that $\left\| x_{k+1} - x^* \right\| < \delta$ for all $k$. So we assume that $\left\| x_k - x^* \right\| < \delta$ for some $k \geq 0$. Then there holds $\left\| \mathcal{F}(x_k) \right\|_Y \leq L \left\| v_k \right\|_X$. We estimate the $Y$-norm of $r_k$ as

$$\left\| r_k \right\|_Y \leq \eta_k \left\| \mathcal{F}(x_k) \right\|_Y \leq L \eta_k \left\| v_k \right\|_X. \tag{6.4}$$

Next, using $\mathcal{F}(x^*) = 0$, we obtain

$$\begin{aligned} \mathcal{G}(x_k) v_{k+1} &= \mathcal{G}(x_k)(d_k + v_k) = r_k - \mathcal{F}(x_k) + \mathcal{G}(x_k) v_k \\ &= r_k - \left[ \mathcal{F}(x^* + v_k) - \mathcal{F}(x^*) - \mathcal{G}(x^* + v_k) v_k \right]. \end{aligned} \tag{6.5}$$

This result, the generalized differentiability of $\mathcal{F}$ at $x^*$, and (6.4) give the following

$$\left\| \mathcal{G}(x_k) v_{k+1} \right\|_Y = o\left( \left\| v_k \right\|_X \right) \text{ as } \left\| v_k \right\|_Y \to 0. \tag{6.6}$$

Hence, for sufficiently small $\delta > 0$, we have $\left\| \mathcal{G}(x_k) v_{k+1} \right\|_Y \leq \dfrac{1}{2 C_{\mathcal{G}^{-1}}} \left\| v_k \right\|_X$, with

$C_{\mathcal{G}^{-1}} = \sup \left\{ \left\| \mathcal{G}(x)^{-1} \right\|_{Y,X} : x \in U \right\}$, and thus

$$\left\| v_{k+1} \right\|_X \leq \left\| \mathcal{G}(x_k)^{-1} \right\|_{Y,X} \left\| \mathcal{G}(x_k) v_{k+1} \right\|_Y \leq \frac{1}{2} \left\| v_k \right\|_X.$$

This gives $x_{k+1} \in x^* + \dfrac{\left\| v_k \right\|_X}{2} \overline{B}_X \subset x^* + \dfrac{\delta}{2} B_X \subset U$, which inductively proves that $x_k \to x^*$ in $Y$. We conclude from (6.6) the following $\left\| v_{k+1} \right\|_X \leq C_{\mathcal{G}^{-1}} \left\| \mathcal{G}(x_k) v_{k+1} \right\|_Y = o\left( \left\| v_k \right\|_X \right)$, which completes the proof. $\qquad\square$

Our purpose is to solve the nonlinear and nonsmooth equation system (3.13)-(3.14) by the semismooth Newton iteration. We introduce the operator

$$\mathcal{T} : L^2(\Omega) \to L^s(\Omega), \quad \mathcal{T}(u) := \mathcal{I}\left( S'(u)^* (z - S(u)) \right),$$

where $\mathcal{I}$ is the Sobolev embedding (see [4] and [19], Theorem 5.4]) of $H_0^1(\Omega)$ into $L^s(\Omega)$ with $s > 2$. This embedding is necessary to show that the function $\mathcal{F}$ defined in (6.7) is generalized differentiable. Now, by using $\overline{\mu} = -\alpha \overline{u} + \mathcal{T}(\overline{u})$ from (3.13) and choosing $c := \alpha^{-1}$, Equation (3.14) becomes $\mathcal{F}(\overline{u}) = 0$, where

$$\begin{aligned} \mathcal{F}(u) := {}& u - \alpha^{-1} \max\{0, \mathcal{T}(u) - \beta\} - \alpha^{-1} \min\{0, \mathcal{T}(u) + \beta\} \\ &+ \alpha^{-1} \max\{0, \mathcal{T}(u) - \beta - \alpha u_b\} + \alpha^{-1} \min\{0, \mathcal{T}(u) + \beta - \alpha u_a\}. \end{aligned} \tag{6.7}$$

The function $\mathcal{F}$ is generalized differentiable (see [4], Theorem 4.2 for the linear case, analogue for the bilinear case) and a generalized derivative is given by

$$\mathcal{G}(u)(v) = v - \alpha^{-1} \chi_{(\mathcal{I}_- \cup \mathcal{I}_+)} \left( \mathcal{T}'(u)(v) \right), \tag{6.8}$$

where $\mathcal{I}_- := \left\{ x \in \Omega : \alpha u_a \leq \mathcal{T}(u) + \beta \leq 0 \text{ a.e. in } \Omega \right\}$ and $\mathcal{I}_+ := \left\{ x \in \Omega : 0 \leq \mathcal{T}(u) - \beta \leq \alpha u_b \text{ a.e. in } \Omega \right\}$.

Using Theorem 6.2, the following theorem guarantees the superlinear convergence of the semismooth Newton method applied to our problems. To prove this we extend the proof of Theorem 4.3 in [4].

**Theorem 6.3.** *If* (3.4) *is fulfilled, then* $\mathcal{G}(u)$ *is invertible for all* $u \in U_{ad}$ *and* $\left\{ \left\| \mathcal{G}(x)^{-1} \right\|_{L^2, L^2} : u \in U_{ad} \right\}$ *is bounded.*

*Proof.* The linear-control case is investigated in [4], so we focus on the bilinear case. Define $J := \mathcal{I}_- \cup \mathcal{I}_+$, and for $\mathcal{D} \subset \Omega$ and $v \in L^2(\Omega)$ the restriction operator $E_{\mathcal{D}} : L^2(\Omega) \to L^2(\mathcal{D})$ by $E_{\mathcal{D}}(v) := v|_{\mathcal{D}}$. The corresponding adjoint operator is the extension-by-zero operator $E_{\mathcal{D}}^* : L^2(\mathcal{D}) \to L^2(\Omega)$. We assume that $\mathcal{G}(u)(v) = w$.

From (6.8) we obtain that $E_{\Omega\setminus J}v = E_{\Omega\setminus J}w$. Thus, $v_J := E_J v \in L^2(J)$ satisfies

$$v_J - \alpha^{-1}E_J \mathcal{T}'(u)\left(E_J^* v_J\right) = E_J w + \alpha^{-1}E_J \mathcal{T}'(u)\left(E_{\Omega\setminus J}^* E_{\Omega\setminus J}w\right). \tag{6.9}$$

Now, we define $g(\varphi) := \left\langle E_J w + \alpha^{-1}E_J \mathcal{T}'(u)\left(E_{\Omega\setminus J}^* E_{\Omega\setminus J}w\right), \varphi \right\rangle_{L^2(J)}$ and

$$a(v_1, v_2) := \left\langle v_1, v_2 \right\rangle_{L^2(J)} + \alpha^{-1}\left[\left\langle S(u) - z, S''(u)\left(E_J^* v_1, E_J^* v_2\right)\right\rangle_{L^2(\Omega)} + \left\langle S'(u)\left(E_J^* v_1\right), S'(u)\left(E_J^* v_2\right)\right\rangle_{L^2(\Omega)}\right],$$

for $\varphi, v_1, v_2 \in L^2(J)$. We use

$$\left\langle \mathcal{T}'(u)(w_1), w_2\right\rangle_{L^2(\Omega)} = \left\langle z - S(u), S''(u)(w_1, w_2)\right\rangle_{L^2(\Omega)} - \left\langle S'(u)(w_1), S'(u)(w_2)\right\rangle_{L^2(\Omega)}$$

to see that (6.9) is equivalent to $a(v_J, \varphi) = g(\varphi)$, for all $\varphi \in L^2(J)$.

Using $\left\langle v_1, v_2\right\rangle_{L^2(J)} = \left\langle E_J^* v_1, E_J^* v_2\right\rangle_{L^2(\Omega)}$ and (3.4) we have coercivity of $a$ for $u \in U_{ad}$ and therefore the Lax-Milgram-Lemma can be applied to show that (6.9) admits a unique solution $v_J \in L^2(J)$. Moreover, this solution satisfies $\|v_J\|_{L^2(J)} \leq \tilde{C}\|g\|_{L^2(J)} \leq C\|w\|_{L^2(\Omega)}$, with a constant $C$ independent of $u$. For the last inequality, we use the fact that $\mathcal{T}'(u)$ is bounded due to the boundedness of $S(u)$, $S'(u)$ and $S''(u)$ as shown in (2.7), (2.13), and (2.14). □

## 7. Numerical Experiments

In this section, we present results of numerical experiments to validate the computational performance of our inexact proximal methods and to demonstrate the convergence rate of $\mathcal{O}(1/k^2)$ proved in Theorem 5.8. In the following procedures, for validation purposes, we formulate control problems for which we know the exact solution. We have

**Procedure 1.** (*Linear case*)
1) Choose $\hat{y} \in H_0^1(\Omega)$ and $\hat{p} \in H_0^1(\Omega)$ arbitrary.

2) Set $\hat{u} := \begin{cases} \max\left\{\dfrac{-\hat{p}+\beta}{\alpha}, u_a\right\} & \text{on } \{x \in \Omega : \hat{p}(x) > \beta\} \\[2mm] \min\left\{\dfrac{-\hat{p}-\beta}{\alpha}, u_b\right\} & \text{on } \{x \in \Omega : \hat{p}(x) < -\beta\} \\[2mm] 0 & \text{elsewhere} \end{cases}$

3) Set $\mu := -\hat{p} - \alpha\hat{u}$, $f := A\hat{y} + \hat{u}$, and $z := A^*\hat{p} + \hat{y}$.

**Lemma 7.1.** *Procedure* 1 *provides a solution* $(\hat{y}, \hat{u})$ *of the optimal control problem* (3.2) *with linear-control elliptic equality constraints.*

*Proof.* We show that the optimality conditions (3.16)-(3.19) in Theorem 3.3 are fulfilled. (3.16)-(3.18) are obviously fulfilled because of 3) in Procedure 1. Now, we consider different cases to show (3.19):

- $|\hat{p}| \leq \beta$: From 2) we have $\hat{u} = 0$ and from 3) $\mu = -\hat{p}$ and therefore

$$B(\hat{u}, \mu) = 0 - \max\left\{0, c(-\hat{p}-\beta)\right\} - \min\left\{0, c(-\hat{p}+\beta)\right\}$$
$$+ \max\left\{0, -u_b + c(-\hat{p}-\beta)\right\} + \min\left\{0, -u_a + c(-\hat{p}+\beta)\right\} = 0.$$

- $\hat{p} > \beta$:
  * $u_a \leq \dfrac{-\hat{p}+\beta}{\alpha} \leq u_b$: From 2) we have $\hat{u} = \dfrac{-\hat{p}+\beta}{\alpha} < 0$ and from 3) we have $\mu = -\hat{p} - \alpha\hat{u} = -\beta$, therefore

$$B(\hat{u}, \mu) = \hat{u} - \max\left\{0, \hat{u} - c(2\beta)\right\} - \min\left\{(0, \hat{u})\right\}$$
$$+ \max\left\{0, \hat{u} - u_b - c(2\beta)\right\} + \min\left\{0, \hat{u} - u_a\right\} = \hat{u} - 0 - \hat{u} + 0 + 0 = 0.$$

* $\dfrac{-\hat{p}+\beta}{\alpha} \le u_a$ : From 2) we have $\hat{u} = u_a$ and from 3) we have $\mu = -\hat{p} - \alpha u_a$, therefore

$$B(\hat{u},\mu) = u_a - \max\left\{0, u_a + c(-\hat{p} - \alpha u_a - \beta)\right\} - \min\left\{0, u_a + c(-\hat{p} - \alpha u_a + \beta)\right\}$$
$$+ \max\left\{0, u_a - u_b + c(-\hat{p} - \alpha u_a - \beta)\right\} + \min\left\{0, u_a - u_a + c(-\hat{p} - \alpha u_a + \beta)\right\}$$
$$= u_a - 0 - \left(u_a + c(-\hat{p} - \alpha u_a + \beta)\right) + 0 + c(-\hat{p} - \alpha u_b + \beta) = 0.$$

- $\hat{p} < -\beta$

  * $u_a \le \dfrac{-\hat{p}-\beta}{\alpha} \le u_b$ : From 2) we have $\hat{u} = \dfrac{-\hat{p}-\beta}{\alpha} > 0$ and from 3) we have $\mu = -\hat{p} - \alpha\hat{u} = \beta$. So

$$B(\hat{u},\mu) = \hat{u} - \max\left\{0,\hat{u}\right\} - \min\left\{0,\hat{u} + c(2\beta)\right\}$$
$$+ \max\left\{0,\hat{u} - u_b\right\} + \min\left\{0,\hat{u} - u_a + c(2\beta)\right\}$$
$$= \hat{u} - \hat{u} - 0 + 0 + 0 = 0.$$

  * $\dfrac{-\hat{p}-\beta}{\alpha} \ge u_b$ : From 2) we have $\hat{u} = u_b$ and from 3) we have $\mu = -\hat{p} - \alpha u_b$, therefore

$$B(\hat{u},\mu) = u_b - \max\left\{0, u_b + c(-\hat{p} - \alpha u_b - \beta)\right\} - \min\left\{0, u_b + c(-\hat{p} - \alpha u_b + \beta)\right\}$$
$$+ \max\left\{0, u_b - u_b + c(-\hat{p} - \alpha u_b - \beta)\right\} + \min\left\{0, u_b - u_a + c(-\hat{p} - \alpha u_b + \beta)\right\}$$
$$= u_b - \left(u_b + c(-\hat{p} - \alpha u_b - \beta)\right) - 0 + c(-\hat{p} - \alpha u_b - \beta) + 0 = 0.$$

$\square$

**Procedure 2.** (*Bilinear case*)
1) Choose $\hat{y} \in H_0^1(\Omega)$ and $\hat{p} \in H_0^1(\Omega)$ arbitrary.

2) Set $\hat{u} := \begin{cases} \max\left\{\dfrac{-\hat{p}\hat{y}+\beta}{\alpha}, u_a\right\} & \text{on } \left\{x \in \Omega : \hat{p}(x)\hat{y}(x) > \beta\right\} \\[2ex] \min\left\{\dfrac{-\hat{p}\hat{y}-\beta}{\alpha}, u_b\right\} & \text{on } \left\{x \in \Omega : \hat{p}(x)\hat{y}(x) < -\beta\right\}. \\[2ex] 0 & \text{elsewhere} \end{cases}$

3) Set $\mu := -\hat{p}\hat{y} - \alpha\hat{u}$, $f := A\hat{y} + \hat{u}\hat{y}$, and $z := A^*\hat{p} + \hat{y} + \hat{u}\hat{p}$.

**Lemma 7.2.** *Procedure* 2 *provides a solution* $(\hat{y},\hat{u})$ *to the optimal control problem* (3.2) *with bilinear-control elliptic equality constraints.*

*Proof.* The proof is similar to the one of the linear case. $\square$

Next, we specify the elliptic operator, the domain of computation, the choice of $\hat{y}$ and $\hat{p}$, and the optimization and numerical parameters. We consider the following examples.

**Case 1.** (1 *dimensional*) $\Omega = (0,1)$, $A = -\Delta$, $u_a \equiv -1$, $\hat{y} = \sin(\pi x)$ and $\hat{p} = 2\beta\sin(2\pi x)$. We discretize $\Omega$ with gridsize $h = 1/1024$ $A$ is discretized by second-order finite differences. Then we have $c_\Omega = \dfrac{1}{2}$, $a_0 = 0$ and $\theta = 1$ such that (2.6) holds. The results are shown in **Table 1**.

**Case 2.** (2 *dimensional*) $\Omega = (0,1)^2$, $A = -\Delta$, $u_a \equiv -1$, $\hat{y} = \sin(\pi x_1)\sin(\pi x_1)$ and $\hat{p} = 4\beta\sin(2\pi x_1)\sin(\pi x_2)$. We discretize $\Omega$ with gridsize $h = 1/256$. $A$ is discretized by second-order finite differences. Then we have $c_\Omega = \dfrac{1}{4}$, $a_0 = 0$ and $\theta = 1$ such that (2.6) holds. The results are shown in **Table 2**.

We compare the FIP, FIPB and ISSN schemes in terms of computational time. In the FIP method, we estimate an approximation to the Lipschitz constant $L(\hat{J}_2) := \|J_2''\|_{L^2,L^2}$ with a power iteration. This power iteration is stopped if the difference between two iterates of the norm $\|J_2''\|_{L^2,L^2}$ is less or equal than a tolerance of $10^{-5}$. For the FIPB method, we use backtracking with $\eta = 1.5$ and $L_0 = 0.001$. All algorithms are stopped if $\|B(u_k,\mu_k)\| < 10^{-8}$. We can see in **Table 1** and **Table 2** that the computational performance of the FIP and FIPB methods is comparable to that of the ISSN method.

**Table 1.** Example 1: Comparison of the FIP, FIPB and ISSN methods.

|  |  | linear case ( $u_b \equiv 15\beta$ ) | | | bilinear case ( $u_b \equiv 7\beta$ ) | | |
|---|---|---|---|---|---|---|---|
| $\alpha$ | $\beta$ | FIP | FIPB | ISSN | FIP | FIPB | ISSN |
| 0.5 | 0.1 | 0.441s | 3.86s | 0.591s | 2.89s | 8.62s | 4.11s |
|  | 0.01 | 0.333s | 8.26s | 0.587s | 2.07s | 9.57s | 2.75s |
| 0.05 | 0.1 | 2.33s | 8.74 | 2.56s | 6.94s | 17.8 | 6.62s |
|  | 0.01 | 1.82s | 7.78s | 1.26s | 3.11s | 19.42s | 4.37s |
| 0.005 | 0.1 | 6.48s | 51.7 | 2.49s | 15.0s | 7.2s | 7.9s |
|  | 0.01 | 6.48s | 5.50s | 2.68s | 8.04s | 7.15s | 6.61s |

**Table 2.** Example 2: Comparison of the FIP, FIPB, and ISSN methods.

|  |  | linear case ( $u_b \equiv 15\beta$ ) | | | bilinear case ( $u_b \equiv 7\beta$ ) | | |
|---|---|---|---|---|---|---|---|
| $\alpha$ | $\beta$ | FIP | FIPB | ISSN | FIP | FIPB | ISSN |
| 0.5 | 0.1 | 6.55s | 34.0s | 6.83s | 58.3s | 156s | 123s |
|  | 0.01 | 5.27s | 28.6s | 6.46s | 44.9s | 105s | 75.3s |
| 0.05 | 0.1 | 21.8s | 42.3s | 39.3s | 77.7s | 118s | 117s |
|  | 0.01 | 15.4s | 38.9s | 14.6s | 55.8s | 95.8s | 112s |
| 0.005 | 0.1 | 34.1s | 47.8s | 38.9s | 268s | 90.5s | 172s |
|  | 0.01 | 40.8s | 59.5s | 45.0s | 104s | 63.6s | 139s |

In order to validate the convergence rate of $\mathcal{O}\left(1/k^2\right)$, the theoretical upper bound of Theorem 5.20 and the actual error of the functional in correspondence to Example 1 and Example 2 with $\beta = 0.1$ and $\alpha = 0.005$, are plotted in **Figure 1**. We see that the observed convergence may be faster than the theoretical prediction.

We conclude this section considering challenging linear- and a bilinear-control cases. However, the exact solutions are not known. In these cases, the target function is not attainable. We have

**Case 3.** (*Linear case*) $\Omega = (0,1)^2$, $A = -\Delta$, $u_a \equiv -20$, $u_b \equiv 20$, $z = 1 + \sin(2\pi x)\sin(2\pi y) \notin H_0^1(\Omega)$ and $f \equiv 1$. *We discretize* $\Omega$ *with gridsize* $h = 1/256$. *A is discretized by second-order finite differences.*

**Case 4.** (*Bilinear case*) $\Omega = (0,1)^2$, $A = -\Delta$, $u_a \equiv -10$, $u_b \equiv 10$, $z = 1 + \sin(2\pi x)\sin(2\pi y) \notin H_0^1(\Omega)$ and $f \equiv 1$. *We discretize* $\Omega$ *with gridsize* $h = 1/256$. *A is discretized by second-order finite differences.*

In **Figure 2**, we present the optimal controls obtained for the Examples 3 and 4, respectively. Notice that the controls obtained with the FIP, FIPB, and ISSN schemes are indistinguishable. We observe that in the case of a small $\alpha$ there is an abrupt change between $u = 0$ and $u = u_b$, whereas for bigger $\alpha$ the change is continuous. We also see that by increasing $\beta$ the support of u decreases, as expected. The different computational times of the FIP, FIPB, and ISSN schemes are also given in the figure. We see that the FIPB scheme may outperform the ISSN scheme and vice versa. We also have a case where the ISSN scheme has difficulty to converge; see **Figure 2**, test case (d). Notice that very similar results are also obtained using a globalized version [7] of the ISSN scheme. These results and further results of numerical experiments demonstrate that fast inexact proximal scheme represent an effective alternative to semi-smooth Newton methods.

## 8. Conclusion

Inexact proximal schemes for solving linear- and bilinear elliptic optimal control problems were discussed. A complete analysis of these methods was presented and a convergence rate of $\mathcal{O}\left(1/k^2\right)$ was proven. For benchmarking purposes, the proposed inexact proximal schemes were compared to an inexact semismooth Newton method. Results of numerical experiments demonstrated the computational effectiveness of inexact proximal schemes and successfully validated the theoretical estimates.
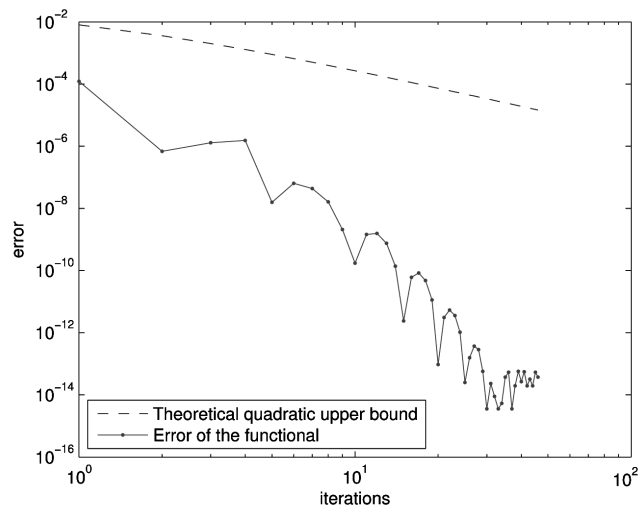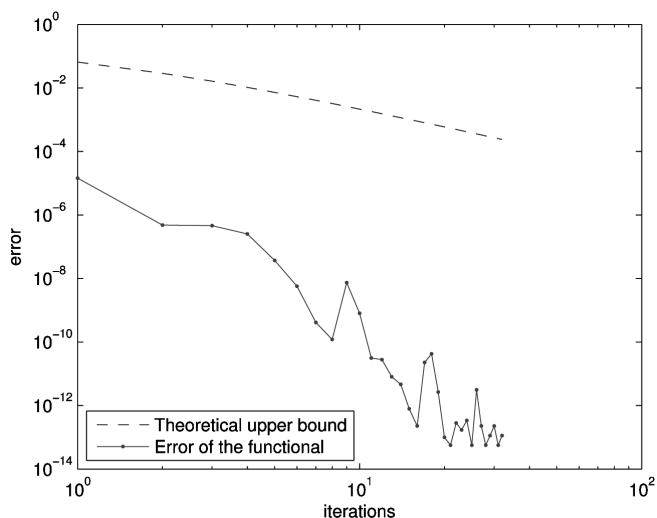
(a) Example 1 - Linear problem
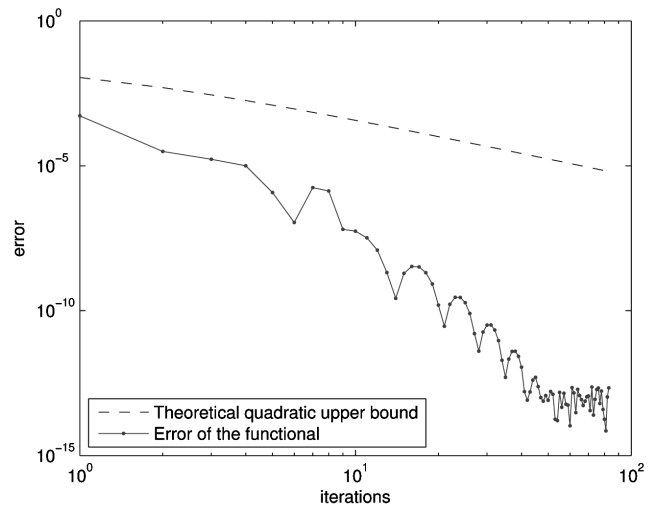


(b) Example 1 - Bilinear problem
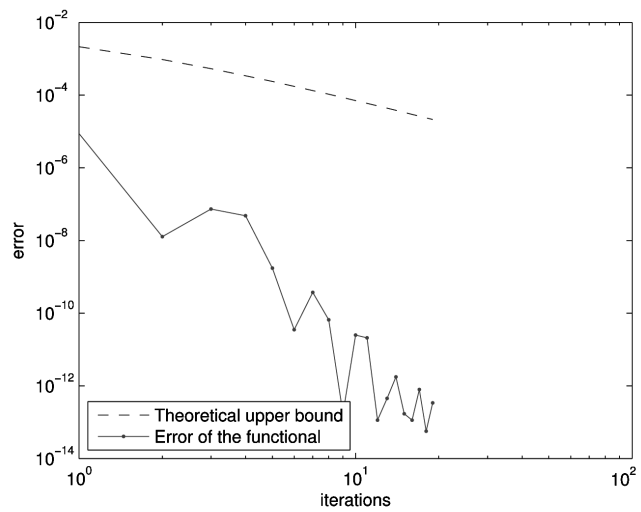


(c) Example 2 - Linear problem

(d) Example 2 - Bilinear problem

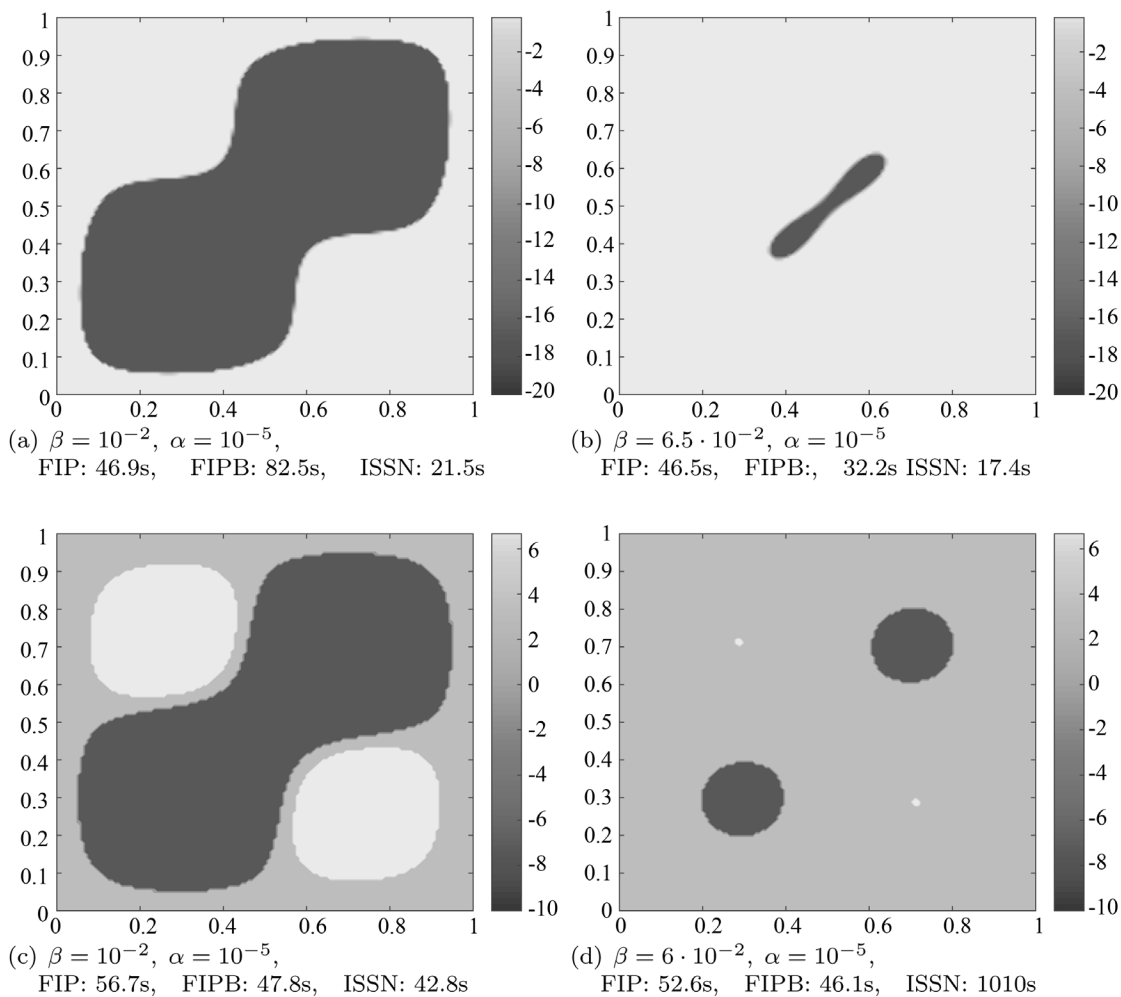**Figure 1.** Validation of the theoretical upper bound (Theorem 58).



(a) $\beta = 10^{-2}$, $\alpha = 10^{-5}$,
FIP: 46.9s,    FIPB: 82.5s,    ISSN: 21.5s

(b) $\beta = 6.5 \cdot 10^{-2}$, $\alpha = 10^{-5}$
FIP: 46.5s,    FIPB:,    32.2s ISSN: 17.4s

(c) $\beta = 10^{-2}$, $\alpha = 10^{-5}$,
FIP: 56.7s,    FIPB: 47.8s,    ISSN: 42.8s

(d) $\beta = 6 \cdot 10^{-2}$, $\alpha = 10^{-5}$,
FIP: 52.6s,    FIPB: 46.1s,    ISSN: 1010s

**Figure 2.** Optimal controls $u$ for the Case 3 (top) and Case 4 (bottom).

## Acknowledgements

## References

[1] Borzì, A. and Schulz, V. (2011) Computational Optimization of Systems Governed by Partial Differential Equations. SIAM, Philadelphia. http://dx.doi.org/10.1137/1.9781611972054

[2] Tröltzsch, F. (2009) Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen. Vieweg. http://dx.doi.org/10.1007/978-3-8348-9357-4

[3] Ulbrich, M. (2011) Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. SIAM, Philadelphia. http://dx.doi.org/10.1137/1.9781611970692

[4] Stadler, G. (2009) Elliptic Optimal Control Problems with $L^1$-Control Cost and Applications for the Placement of Control Devices. *Computational Optimization and Applications*, **44**, 159-181. http://dx.doi.org/10.1007/s10589-007-9150-9

[5] Wachsmuth, G. and Wachsmuth, D. (2010) Convergence and Regularization Results for Optimal Control Problems with Sparsity Function. *ESAIM*: *Control Optimisation and Calculus of Variations*, **17**, 858-886. http://dx.doi.org/10.1051/cocv/2010027

[6] Casas, E., Herzog, R. and Wachsmuth, G. (2012) Optimality Conditions and Error Analysis of Semilinear Elliptic Control Problems with $L^1$ Cost Functional. *SIAM Journal on Optimization*, **22**, 795-820. http://dx.doi.org/10.1137/110834366

[7] Ciaramella, G. and Borzì, A. (2016) A LONE Code for the Sparse Control of Quantum Systems. *Computer Physics Communications*, **200**, 312-323. http://dx.doi.org/10.1016/j.cpc.2015.10.028

[8] Candes, E., Romberg, J. and Tao, T. (2006) Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Communications on Pure and Applied Mathematics*, **59**, 1207-1223. http://dx.doi.org/10.1002/cpa.20124

[9] Defrise, M., Daubechies, I. and Mol, C.D. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics*, **57**, 1413-1457. http://dx.doi.org/10.1002/cpa.20042

[10] Donoho, D.L. and Elad, M. (2003) Maximal Sparsity Representation via $\ell_1$ Minimization. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 2197-2202. http://dx.doi.org/10.1073/pnas.0437847100

[11] Combettes, P.L. and Wajs, V.R. (2005) Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Modeling & Simulation*, **4**, 1168-1200. http://dx.doi.org/10.1137/050626090

[12] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. http://dx.doi.org/10.1137/080716542

[13] Lorenz, A., Bredies, K. and Reiterer, S. (2015) Minimization of Non-Smooth, Non-Convex Functionals by Iterative Thresholding. *Journal of Optimization Theory and Applications*, **165**, 78-112. http://dx.doi.org/10.1007/s10957-014-0614-7

[14] Wech, T., Seiberlich, N., Schindele, A., Grau, V., Diffley, L., Gyngell, M.L., Borzì, A., Köstler, H. and Schneider, J.E. (2016) Development of Real-Time Magnetic Resonance Imaging of Mouse Hearts at 9.4 Tesla—Simulations and First Application. *IEEE Transactions on Medical Imaging*, **35**, 912-920. http://dx.doi.org/10.1109/TMI.2015.2501832

[15] Evans, L.C. (2010) Partial Differential Equations. 2nd Edition, Department of Mathematics, University of California, Berkeley, American Mathematical Society. http://dx.doi.org/10.1090/gsm/019

[16] Kröner, A. and Vexler, B. (2009) A Priori Error Estimates for Elliptic Optimal Control Problems with a Bilinear State Equation. *Journal of Computational and Applied Mathematics*, **230**, 781-802. http://dx.doi.org/10.1016/j.cam.2009.01.023

[17] Lions, J. (1971) Optimal Control of Systems Governed by Partial Differential Equations. Springer, Berlin.

[18] Grisvard, P. (1985) Elliptic Problems in Nonsmooth Domains. Pitman Publishing, Boston.

[19] Adams, R.A. (1975) Sobolev Spaces. Pure and Applied Mathematics. Academic Press, Inc., Cambridge.

[20] Ekeland, I. and Témam, R. (1999) Convex Analysis and Variational Problems. Society for Industrial and Applied Mathematics, Philadelphia. http://dx.doi.org/10.1137/1.9781611971088

[21] Wilkinson, J.H. (1988) The Algebraic Eigenvalue Problem. Oxford University Press, Oxford.

[22] Nesterov, Y.E. (1983) A Method for Solving the Convex Programming Problem with Covergence Rate $O(1/k^2)$. *Doklady Akademii Nauk SSSR*, **269**, 543-547.

[23] Hestenes, M.R. and Stiefel, E. (1952) Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, **49**, 409-436. http://dx.doi.org/10.6028/jres.049.044

[24] Hintermüller, M., Ito, K. and Kunisch, K. (2002) The Primal-Dual Active Set Strategy as a Semismooth Newton Method. *SIAM Journal on Optimization*, **13**, 865-888. http://dx.doi.org/10.1137/S1052623401383558

[25] Martínez, J. and Qi, L. (1995) Inexact Newton Methods for Solving Nonsmooth Equations. *Journal of Computational and Applied Mathematics*, **60**, 127-145.