Numerische Mathematik (2021) 149:305–340
https://doi.org/10.1007/s00211-021-01222-7

Numerische Mathematik

Numerical approximation of control problems of non-monotone and non-coercive semilinear elliptic equations

Eduardo Casas[1] · Mariano Mateos[2] · Arnd Rösch[3]

Received: 28 January 2020 / Revised: 23 February 2021 / Accepted: 26 July 2021 /
Published online: 12 September 2021
© The Author(s) 2021

Abstract

We analyze the numerical approximation of a control problem governed by a non-monotone and non-coercive semilinear elliptic equation. The lack of monotonicity and coercivity is due to the presence of a convection term. First, we study the finite element approximation of the partial differential equation. While we can prove existence of a solution for the discrete equation when the discretization parameter is small enough, the uniqueness is an open problem for us if the nonlinearity is not globally Lipschitz. Nevertheless, we prove the existence and uniqueness of a sequence of solutions bounded in $L^\infty(\Omega)$ and converging to the solution of the continuous problem. Error estimates for these solutions are obtained. Next, we discretize the control problem. Existence of discrete optimal controls is proved, as well as their convergence to solutions of the continuous problem. The analysis of error estimates is quite involved due to the possible non-uniqueness of the discrete state for a given control. To overcome this difficulty we define an appropriate discrete control-to-state mapping in a neighbourhood of a strict solution of the continuous control problem. This allows us to introduce a reduced functional and obtain first order optimality conditions as well as

The first two authors were partially supported by the Spanish Ministerio de Economía, Industria y Competitividad under project MTM2017-83185-P.

✉ Mariano Mateos
mmateos@uniovi.es

Eduardo Casas
eduardo.casas@unican.es

Arnd Rösch
arnd.roesch@uni-due.de

[1] Departamento de Matemática Aplicada y Ciencias de la Computación, E.T.S.I. Industriales y de Telecomunicación, Universidad de Cantabria, 39005 Santander, Spain

[2] Departamento de Matemáticas, Campus de Gijón, Universidad de Oviedo, 33203 Gijón, Spain

[3] Fakultät für Mathematik, Universtät Duisburg-Essen, 45127 Essen, Germany

error estimates. Some numerical experiments are included to illustrate the theoretical results.

Mathematics Subject Classification Primary 49K20 · 35J61 · 65M15 · 49M25

1 Introduction

In this paper, we consider the numerical approximation of the optimal control problem

$$\text{(P)} \quad \min_{u \in U_{\text{ad}}} J(u) := \frac{1}{2}\int_\Omega (y_u(x) - y_d(x))^2\,dx + \frac{\nu}{2}\int_\Omega u^2(x)\,dx,$$

where $\Omega \subset \mathbb{R}^n$, $n = 2$ or $n = 3$, is a convex domain with boundary Γ, y_u is the solution of the following state equation

$$\begin{cases} Ay + b(x) \cdot \nabla y + f(x, y) = u \text{ in } \Omega, \\ y = 0 \text{ on } \Gamma, \end{cases} \tag{1.1}$$

A is an elliptic operator, $b : \Omega \longrightarrow \mathbb{R}^n$ is a given function, and $f : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is non-decreasing monotone in the second variable. Moreover, $y_d \in L^2(\Omega)$ is a given function, $\nu > 0$, and

$$U_{\text{ad}} = \{u \in L^2(\Omega) : \alpha \le u(x) \le \beta \text{ for a.e. } x \in \Omega\}$$

with $-\infty \le \alpha < \beta \le +\infty$. This problem is studied in [12], where existence and uniqueness results for the equation, as well as existence of optimal controls and optimality conditions are obtained. For the convenience of the reader, these results are summarized in Sect. 2. In this work, we will discretize the problem and obtain approximation results. The reader is also referred to [8,15] for a similar control problem associated to a non-monotone quasilinear elliptic equation. The main difference with respect to the above equation is that the operators considered in [8,15] are coercive, while our equation is neither monotonone nor coercive.

In Sect. 3 we study the approximation of the state equation by finite elements. The reader is referred to [32] for the linear case or [8,16,22] for the case of non-monotone but coercive quasilinear equations. In the quasilinear case, the discrete equation has at least one solution for every h, which easily follows from an application of Brouwer's fixed point theorem and the coercivity of the operator. However, there is not a uniqueness result. In the linear case, we only can prove existence of solution if h is small enough, but we have a unique solution for each of these values of h. In the semilinear discrete case corresponding to (1.1), the existence of a discrete solution requires, as in the linear case, a parameter h small. But, as in the quasilinear case, the uniqueness of discrete solutions is an open issue. We prove existence and uniqueness of a discrete solution if the equation is linear or if the non-monotone term is bounded. In the general case we can prove the existence and uniqueness of a bounded sequence of solutions as h tends to 0, but we cannot rule out the possible existence of a divergent sequence of solutions in the $L^\infty(\Omega)$-norm. Error estimates are provided for the bounded approximations of the solutions of the state equation.

In Sect. 4 we discretize the control problem, using either piecewise constant or continuous piecewise linear approximations of the control. We prove the existence of a number $h_0 > 0$ such that the discrete optimal control problem has at least one solution $(\bar{y}_h, \bar{u}_h)$ for every discretization parameter $h < h_0$. We also prove the boundedness of these solutions in $H_0^1(\Omega) \times L^2(\Omega)$. Moreover, every limit in the $H_0^1(\Omega) \times L^2(\Omega)$ weak topology when $h \to 0$ of a sequence of discrete solutions is a solution of the continuous optimal control problem. In addition, the converge is not only weak, it is strong in the $H_0^1(\Omega) \times L^2(\Omega)$ topology. Next, we define a discrete control-to-state mapping in a neighborhood of a strict solution of the continuous problem, as well as an associated reduced functional and we state first order optimality conditions. In Sect. 5 we obtain error estimates and in the last section we include a numerical experiment.

To finish this introduction let us mention some papers concerning error estimates for the numerical approximation of non-linear elliptic control problems. Early references for the numerical analysis of linear quadratic control problems are the papers [23] and [24]. The first reference we are aware of dealing with the numerical approximation of optimal control problems governed by a semilinear elliptic equation is [3]; state constraints were included in the analysis in [29]. Different aspects of Neumann boundary optimal control problems have been treated in [9,13] or [26]. The case of Dirichlet boundary control was first treated in [14]. In all these references, the equations were coercive and monotone. Optimal control problems governed by quasi-linear elliptic equations have been studied in [8,16,17,20]. In these works, the equations were coercive but not monotone. It is also worth mentioning the works [2,30]. In the first one, the authors investigate under which conditions discrete local minima are indeed global. In the second one the authors study how to numerically verify second order optimality conditions, which are very important for the study of local minima for non-convex optimal control problems.

NOTATION: Along the paper we will consider the following operators

$$Ay = -\sum_{i,j=1}^{n} \partial_{x_j}(a_{ij}(x)\partial_{x_i} y) \text{ and } A^*\varphi = -\sum_{i,j=1}^{n} \partial_{x_i}(a_{ij}(x)\partial_{x_j}\varphi), \tag{1.2}$$

$$\mathcal{A}y = Ay + b(x)\cdot\nabla y \text{ and } \mathcal{A}^*\varphi = A^*\varphi - \operatorname{div}[b(x)\varphi]. \tag{1.3}$$

As usual we will denote $C(\bar{\Omega})$ the space of continuous functions in $\bar{\Omega}$, the closure of Ω. $C^{0,\delta}(\bar{\Omega})$ is the space of Hölder functions in $\bar{\Omega}$ if $0 < \delta < 1$ and of Lipschitz functions if $\delta = 1$. For $p \in [1, +\infty]$, $s \geq 0$, we will denote $L^p(\Omega)$ and $W^{s,p}(\Omega)$ respectively the Lebesgue and Sobolev spaces. We also abbreviate $H^s(\Omega) = W^{s,2}(\Omega)$. $H_0^1(\Omega)$ is the space of elements in $H^1(\Omega)$ with null trace on Γ. $H^{-1}(\Omega)$ is the dual of $H_0^1(\Omega)$. See [1] for definitions and further properties of these spaces. In the Sobolev space $H_0^1(\Omega)$ we take the norm

$$\|y\|_{H_0^1(\Omega)} = \left(\int_\Omega |\nabla y(x)|^2\,dx\right)^{\frac{1}{2}}.$$

According to the Poincaré inequality, there exists a constant C_Ω such that

$$\|y\|_{L^2(\Omega)} \leq C_\Omega \|y\|_{H_0^1(\Omega)} \quad \forall y \in H_0^1(\Omega). \tag{1.4}$$

From this inequality and Sobolev's embedding theorem, we also know that there exists a constant K_Ω such that

$$\|y\|_{L^4(\Omega)} \leq K_\Omega \|y\|_{H_0^1(\Omega)} \quad \forall y \in H_0^1(\Omega). \tag{1.5}$$

We will denote $Y = H_0^1(\Omega) \cap C(\bar{\Omega})$, which is a Banach space when endowed with the norm

$$\|y\|_Y = \|y\|_{H_0^1(\Omega)} + \|y\|_{C(\bar{\Omega})}.$$

2 Preliminary results

Assumption 1 We assume that Ω is a convex domain in $\mathbb{R}^n$ with $n = 2$ or 3. We also suppose that Ω is polygonal if $n = 2$ or polyhedral if $n = 3$. Γ denotes its boundary, which is Lipschitz. The following conditions are satisfied by the coefficients of the operator $\mathcal{A}$:

$$\begin{cases} a_{ij} \in C^{0,1}(\bar{\Omega}) \text{ for } i, j = 1, \ldots, n, \\ \exists \Lambda > 0 \text{ such that } \displaystyle\sum_{i,j=1}^n a_{ij}(x)\xi_i\xi_j \geq \Lambda|\xi|^2 \ \ \forall \xi \in \mathbb{R}^n \text{ and for a.e. } x \in \Omega, \end{cases} \tag{2.1}$$

$$b \in L^{\bar{p}}(\Omega) \text{ with } \bar{p} > 2 \text{ if } n = 2 \text{ and } \bar{p} \geq 3 \text{ if } n = 3, \text{ and } \operatorname{div} b \in L^2(\Omega). \tag{2.2}$$

The following properties on the operators $\mathcal{A}$ and $\mathcal{A}^*$ were proved in Theorem 2.2, Corollary 2.4, Theorem 2.5 and Corollary 2.6 of [12]. The proofs of these results make use of [25, Theorems 2.2.2.3 and 3.2.1.2].

Theorem 2.1 *Under Assumption 1, both operators $\mathcal{A}$ and $\mathcal{A}^*$ define isomorphisms between the spaces $H_0^1(\Omega)$ and $H^{-1}(\Omega)$, and $H^2(\Omega) \cap H_0^1(\Omega)$ and $L^2(\Omega)$.*

Regarding the Eq. (1.1) we make the following assumption on the non-linear function f.

Assumption 2 Function $f : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function, monotone non-decreasing with respect to the second variable, and satisfying

$$\begin{cases} f(\cdot, 0) \in L^2(\Omega) \text{ and } \forall M > 0 \ \exists \phi_M \in L^2(\Omega) \text{ such that} \\ |f(x, y_2) - f(x, y_1)| \leq \phi_M(x)|y_2 - y_1| \text{ for a.e. } x \in \Omega \text{ and } |y_i| \leq M, \ i = 1, 2. \end{cases} \tag{2.3}$$

The following result concerning existence, uniqueness and regularity of a solution of (1.1) follows from Theorems 2.6 and 2.8 of [12].

Theorem 2.2 *Under Assumptions 1 and 2, for every $u \in L^2(\Omega)$ the Eq. (1.1) has a unique solution $y_u \in H^2(\Omega) \cap H_0^1(\Omega)$. Moreover, the estimate*

$$\|y_u\|_{H^2(\Omega)} \leq C_{A,f}\Big(\|u\|_{L^2(\Omega)} + 1\Big) \ \forall u \in L^2(\Omega) \tag{2.4}$$

holds for some constant $C_{A,f}$ depending on A and f.

Additional regularity assumptions on f are necessary to consider the differentiability of the functional J.

Assumption 3 We suppose that $f : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function of class C^2 with respect to the second variable satisfying:

$$f(\cdot, 0) \in L^2(\Omega) \text{ and } \frac{\partial f}{\partial y}(x, y) \geq 0 \ \text{ for a.e. } x \in \Omega \text{ and } \forall y \in \mathbb{R}. \tag{2.5}$$

For every $M > 0$ there exists a constant $C_{f,M} > 0$ such that

$$\left|\frac{\partial f}{\partial y}(x, y)\right| + \left|\frac{\partial^2 f}{\partial y^2}(x, y)\right| \leq C_{f,M} \text{ for a.e. } x \in \Omega \text{ and for all } |y| \leq M. \tag{2.6}$$

For every $M > 0$ and $\varepsilon > 0$ there exists $\delta > 0$, depending on M and ε, such that

$$\left|\frac{\partial^2 f}{\partial y^2}(x, y_2) - \frac{\partial^2 f}{\partial y^2}(x, y_1)\right| < \varepsilon \ \text{ if } |y_1|, |y_2| \leq M, \ |y_2 - y_1| \leq \delta, \ \text{ for a.e. } x \in \Omega. \tag{2.7}$$

It is easy to check that Assumption 3 implies Assumption 2. Typical examples of functions satisfying these assumptions are $f(x, y) = a_0(x)y^{2n+1}$ or $f(x, y) = a_0(x)\exp(y)$, where $a_0 \in L^\infty(\Omega)$, $a_0(x) \geq 0$, and n is a positive integer.

Concerning the differentiability of J we have the following result [12, Theorems 2.12 and 3.2 and Corollary 2.6].

Theorem 2.3 *Let us suppose that Assumptions 1 and 3 hold. Then, the mapping $G : L^2(\Omega) \longrightarrow Y$ given by $G(u) = y_u$ is well defined and of class C^2. Moreover, given $u, v \in L^2(\Omega)$, $z_v = DG(u)v$ is the solution of*

$$\begin{cases} Az + b(x) \cdot \nabla z + \dfrac{\partial f}{\partial y}(x, y_u)z = v \textit{ in } \Omega, \\ z = 0 \textit{ on } \Gamma. \end{cases} \tag{2.8}$$

The functional J is of class C^2. Moreover, given $u, v, v_1, v_2 \in L^2(\Omega)$ we have

$$J'(u)v = \int_\Omega (\varphi_u + \nu u)v\,dx, \tag{2.9}$$

$$J''(u)(v_1, v_2) = \int_\Omega \Big[1 - \varphi_u \frac{\partial^2 f}{\partial y^2}(x, y_u)\Big] z_{v_1} z_{v_2}\,dx + \nu \int_\Omega v_1 v_2\,dx, \tag{2.10}$$

where $z_{v_i} = DG(u)v_i$, $i = 1, 2$, and $\varphi_u \in H^2(\Omega) \cap H_0^1(\Omega)$ is the unique solution of the adjoint equation

$$\begin{cases} A^*\varphi - \operatorname{div}[b(x)\varphi] + \dfrac{\partial f}{\partial y}(x, y_u)\varphi = y_u - y_d \text{ in } \Omega, \\ \varphi = 0 \text{ on } \Gamma. \end{cases} \tag{2.11}$$

Since (P) is not a convex problem, we distinguish between local and global solutions. We say that $\bar{u}$ is a local solution of (P) if there exists $\varepsilon > 0$ such that

$$J(\bar{u}) \le J(u) \quad \forall u \in U_{\mathrm{ad}} : \|u - \bar{u}\|_{L^2(\Omega)} \le \varepsilon.$$

As usual, we say that $\bar{u}$ is a strict local solution if the above inequality is strict whenever $u \ne \bar{u}$. The reader is referred to [12, Definition 3.3 and Lemma 3.4] for a discussion of different notions of local solutions.

Theorem 2.4 *Under Assumptions 1 and 3, (P) has at least one solution. Moreover, if $\bar{u}$ is a local solution of (P), then there exist two unique elements $\bar{y}, \bar{\varphi} \in H^2(\Omega) \cap H_0^1(\Omega)$ such that*

$$\begin{cases} A\bar{y} + b(x) \cdot \nabla \bar{y} + f(x, \bar{y}) = \bar{u} \text{ in } \Omega, \\ \bar{y} = 0 \text{ on } \Gamma, \end{cases} \tag{2.12}$$

$$\begin{cases} A^*\bar{\varphi} - \operatorname{div}[b(x)\bar{\varphi}] + \dfrac{\partial f}{\partial y}(x, \bar{y})\bar{\varphi} = \bar{y} - y_d \text{ in } \Omega, \\ \bar{\varphi} = 0 \text{ on } \Gamma, \end{cases} \tag{2.13}$$

$$\int_\Omega (\bar{\varphi} + \nu \bar{u})(u - \bar{u})\,dx \ge 0 \quad \forall u \in U_{\mathrm{ad}}. \tag{2.14}$$

Further, the regularity $\bar{u} \in H^1(\Omega) \cap C(\bar{\Omega})$ holds. In addition, if $U_{\mathrm{ad}} = L^2(\Omega)$, then we have $\bar{u} \in H^2(\Omega) \cap H_0^1(\Omega)$.

This theorem follows from [12, Theorems 3.1 and 3.6, and Corollary 3.7].

We finish this section by establishing the second order optimality conditions. To this end, we define the cone of critical directions as follows:

$$C_{\bar{u}} = \{v \in L^2(\Omega) : J'(\bar{u})v = 0 \text{ and (2.15) holds}\}$$

$$v(x, t) \begin{cases} \ge 0 & \text{if } \bar{u}(x, t) = \alpha, \\ \le 0 & \text{if } \bar{u}(x, t) = \beta. \end{cases} \tag{2.15}$$

Let us observe that (2.14) implies that

$$\bar{\varphi}(x) + \nu\bar{u}(x) \begin{cases} \geq 0 & \text{if } \bar{u}(x) = \alpha, \\ \leq 0 & \text{if } \bar{u}(x) = \beta. \end{cases}$$

Therefore, if $v \in L^2(\Omega)$ satisfies (2.15), then $J'(\bar{u})v \geq 0$ holds, and $J'(\bar{u})v = 0$ if and only if $v(x) = 0$ whenever $\bar{\varphi}(x) + \nu\bar{u}(x) \neq 0$.

In the case where there are not control constraints, namely $U_{\mathrm{ad}} = L^2(\Omega)$, then $J'(\bar{u}) = 0$ and $C_{\bar{u}} = L^2(\Omega)$.

Theorem 2.5 *Under Assumptions 1 and 3, if $\bar{u}$ is a local solution of (P), then $J''(\bar{u})v^2 \geq 0$ $\forall v \in C_{\bar{u}}$. Conversely, if $\bar{u} \in U_{\mathrm{ad}}$ satisfies (2.12)–(2.14) along with $(\bar{y}, \bar{\varphi})$ and*

$$J''(\bar{u})v^2 > 0 \quad \forall v \in C_{\bar{u}} \setminus \{0\}, \tag{2.16}$$

then there exist $\varepsilon > 0$ and $\kappa > 0$ such that

$$J(\bar{u}) + \frac{\kappa}{2}\|u - \bar{u}\|^2_{L^2(\Omega)} \leq J(u) \ \ \forall u \in U_{\mathrm{ad}} : \|u - \bar{u}\|_{L^2(\Omega)} \leq \varepsilon. \tag{2.17}$$

This result was established in [12, Corollary 3.9].

3 Approximation of the state equation

In this section we consider the finite element discretization of the Eq. (1.1). The goal is to prove the existence of solution for the discrete problems and to derive some error estimates. We proceed in three steps. First we study the linear equation; see Lemma 3.1. Next we replace the local Lipschitz condition stated in Assumption 2 by the more restrictive global condition (3.12). Using this condition, we prove the existence of a unique discrete solution in Theorem 3.5 and error estimates in Theorem 3.6. Finally, we remove assumption (3.12) to obtain the main result of this section, Theorem 3.7. It will be assumed, without express mention, that Assumption 1 holds.

From Theorem 2.2 we know that, under the Assumptions 1 and 2, given $u \in L^2(\Omega)$, (1.1) has a unique solution $y \in H^2(\Omega) \cap H^1_0(\Omega)$. In the rest of the section u denotes a fixed element of $L^2(\Omega)$ and y the corresponding solution of (1.1).

To formulate a discrete version of (1.1) we introduce a quasi-uniform family of triangulations $\{\mathcal{T}_h\}_{h>0}$ of $\bar{\Omega}$; cf. [5, Definition (4.4.13)]. We denote N_h the number of interior nodes of $\mathcal{T}_h$. Associated with these triangulations we consider the finite dimensional spaces

$$Y_h = \{y_h \in C(\bar{\Omega}) : \ y_{h|T} \in P_1(T) \ \forall T \in \mathcal{T}_h \text{ and } y_h \equiv 0 \text{ on } \Gamma\},$$

where $P_1(T)$ denotes the space of polynomials in T of degree less than or equal to one. Now, we introduce the discrete version of (1.1) as follows

$$\begin{cases} \text{Find } y_h \in Y_h \text{ such that} \\ a(y_h, z_h) + \displaystyle\int_\Omega f(x, y_h(x)) z_h(x)\,dx = \int_\Omega u(x) z_h(x)\,dx \ \ \forall z_h \in Y_h. \end{cases} \tag{3.1}$$

Above $a : H^1(\Omega) \times H^1(\Omega) \longrightarrow \mathbb{R}$ denotes the bilinear form associated to the operator $\mathcal{A}$

$$\begin{aligned} a(y_1, y_2) &= \langle \mathcal{A} y_1, y_2 \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ &= \int_\Omega \Big(\sum_{i,j=1}^n a_{ij}(x) \partial_{x_i} y_1 \partial_{x_j} y_2 + [b(x) \cdot \nabla y_1] y_2 \Big)\,dx. \end{aligned}$$

From Theorem 2.1 we have

$$|a(y_1, y_2)| \le \|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))} \|y_1\|_{H_0^1(\Omega)} \|y_2\|_{H_0^1(\Omega)}.$$

Due to the presence of b in the definition of the bilinear form a, it is not necessarily coercive. However, we can prove, see [12, Lemma 2.1] that it satisfies Gårding's inequality. There exists $C_{\Lambda,b}$ such that

$$a(z, z) \ge \frac{\Lambda}{4} \|z\|^2_{H_0^1(\Omega)} - C_{\Lambda,b} \|z\|^2_{L^2(\Omega)} \quad \forall z \in H_0^1(\Omega). \tag{3.2}$$

From [12, Corollary 2.6] we know that $\mathcal{A}^* : H^2(\Omega) \cap H_0^1(\Omega) \longrightarrow L^2(\Omega)$ is an isomorphism. Then, we argue similarly to [32] to deduce the the following result.

Lemma 3.1 *Let $a_0 \in L^2(\Omega)$ be a non-negative function. There exists $h_{\mathcal{A},a_0} > 0$ depending on $\mathcal{A}$ and $\|a_0\|_{L^2(\Omega)}$ such that the variational problem*

$$\begin{cases} \text{Find } y_h \in Y_h \text{ such that} \\ a(y_h, z_h) + \displaystyle\int_\Omega a_0(x) y_h(x) z_h(x)\,dx = \int_\Omega u(x) z_h(x)\,dx \ \ \forall z_h \in Y_h \end{cases} \tag{3.3}$$

has a unique solution for every $h \le h_{\mathcal{A},a_0}$ and for every $u \in L^2(\Omega)$. Moreover, there exists a constant $C_{\mathcal{A},a_0}$ depending on $\mathcal{A}$ and a_0 such that

$$\|y_h\|_{H_0^1(\Omega)} \le C_{\mathcal{A},a_0} \|\mathcal{A}^{-1} u\|_{H_0^1(\Omega)} \quad \forall h \le h_{\mathcal{A},a_0}. \tag{3.4}$$

Proof Because of the linearity of the system (3.3), it is enough to show the existence of $h_{\mathcal{A},a_0}$ such that the only solution of the homogeneous problem is $y_h = 0$. Therefore, let us assume that $y_h \in Y_h$ satisfies

$$a(y_h, z_h) + \int_\Omega a_0(x) y_h(x) z_h(x)\, dx = 0 \quad \forall z_h \in Y_h. \tag{3.5}$$

Then, from Gårding's inequality (3.2) and the fact that $a_0 \geq 0$ we get

$$0 = a(y_h, y_h) + \int_\Omega a_0(x) y_h^2(x)\, dx \geq \frac{\Lambda}{4} \|y_h\|^2_{H_0^1(\Omega)} - C_{\Lambda,b} \|y_h\|^2_{L^2(\Omega)},$$

hence

$$\|y_h\|_{H_0^1(\Omega)} \leq 2\sqrt{\frac{C_{\Lambda,b}}{\Lambda}} \|y_h\|_{L^2(\Omega)}. \tag{3.6}$$

Now, let us take $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$ satisfying of $\mathcal{A}^*\psi + a_0\psi = y_h$ in Ω. We have

$$\|\psi\|_{H^2(\Omega)} \leq K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)} + 1\Big)\|y_h\|_{L^2(\Omega)};$$

see Lemma 3.2 below. Let us denote by $\hat{\psi}_h \in Y_h$ the $H_0^1(\Omega)$ projection of ψ on Y_h, i.e.:

$$\int_\Omega \nabla\hat{\psi}_h \cdot \nabla z_h\, dx = \int_\Omega \nabla\psi \cdot \nabla z_h\, dx \quad \forall z_h \in Y_h. \tag{3.7}$$

Then, there exist constants $\hat{c}_2$ and $\hat{c}_\infty$ such that

$$\|\psi - \hat{\psi}_h\|_{H_0^1(\Omega)} \leq \hat{c}_2 h \|\psi\|_{H^2(\Omega)} \leq \hat{c}_2 K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)} + 1\Big) h \|y_h\|_{L^2(\Omega)}, \tag{3.8}$$

$$\|\psi - \hat{\psi}_h\|_{L^\infty(\Omega)} \leq \hat{c}_\infty h^{2-\frac{n}{2}} \|\psi\|_{H^2(\Omega)} \leq \hat{c}_\infty K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)} + 1\Big) h^{2-\frac{n}{2}} \|y_h\|_{L^2(\Omega)}; \tag{3.9}$$

see Theorems 18.1 and 19.3 of [18].

Now, from (3.5) we get

$$\begin{aligned}\|y_h\|^2_{L^2(\Omega)} &= \int_\Omega (\mathcal{A}^*\psi + a_0\psi) y_h\, dx = a(y_h, \psi) + \int_\Omega a_0 \psi y_h\, dx \\ &= a(y_h, \psi - \hat{\psi}_h) + \int_\Omega a_0(\psi - \hat{\psi}_h) y_h\, dx \\ &\leq \|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))} \|y_h\|_{H_0^1(\Omega)} \|\psi - \hat{\psi}_h\|_{H_0^1(\Omega)} \\ &\quad + \|a_0\|_{L^2(\Omega)} \|y_h\|_{L^2(\Omega)} \|\psi - \hat{\psi}_h\|_{L^\infty(\Omega)}.\end{aligned}$$

From the estimates (3.8) and (3.9) we get

$$\|y_h\|^2_{L^2(\Omega)} \le \|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\|y_h\|_{H^1_0(\Omega)}\hat{c}_2 K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)h\|y_h\|_{L^2(\Omega)} + \|a_0\|_{L^2(\Omega)}\hat{c}_\infty K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)h^{2-\frac{n}{2}}\|y_h\|^2_{L^2(\Omega)}.$$

Taking $h_1 > 0$ such that

$$\|a_0\|_{L^2(\Omega)}\hat{c}_\infty K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)h_1^{2-\frac{n}{2}} = \frac{1}{2}, \tag{3.10}$$

we deduce for $h \le h_1$

$$\|y_h\|_{L^2(\Omega)} \le 2\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\hat{c}_2 K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)h\|y_h\|_{H^1_0(\Omega)}.$$

Now, we select h_2 as follows

$$4\sqrt{\frac{C_{\Lambda,b}}{\Lambda}}\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\hat{c}_2 K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)h_2 = 1. \tag{3.11}$$

Finally, we infer from (3.6) that $y_h = 0$ if $h < \min\{h_1, h_2\}$.

Let us conclude the demonstration by proving the estimate (3.4). To this end we set $h_{\mathcal{A},a_0} = \min\{h_1, h_2\}/2$. Let $y \in Y$ be the solution of $\mathcal{A}y = u$ in Ω and let $y_h \in Y_h$ be the solution of (3.3) for $h \le h_{\mathcal{A},a_0}$. Then, using again ψ and $\hat{\psi}_h$, and arguing similarly as we did above, we get

$$\begin{aligned}
\|y_h\|^2_{L^2(\Omega)} &= \int_\Omega (\mathcal{A}^*\psi + a_0\psi)y_h\,dx = a(y_h,\psi) + \int_\Omega a_0\psi y_h\,dx \\
&= a(y_h, \psi - \hat{\psi}_h) + \int_\Omega a_0(\psi - \hat{\psi}_h)y_h\,dx + \int_\Omega u\hat{\psi}_h\,dx \\
&= a(y_h, \psi - \hat{\psi}_h) + \int_\Omega a_0(\psi - \hat{\psi}_h)y_h\,dx + a(y, \hat{\psi}_h) \\
&\le \|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\|y_h\|_{H^1_0(\Omega)}\|\psi - \hat{\psi}_h\|_{H^1_0(\Omega)} \\
&\quad + \|a_0\|_{L^2(\Omega)}\|y_h\|_{L^2(\Omega)}\|\psi - \hat{\psi}_h\|_{L^\infty(\Omega)} \\
&\quad + \|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\|y\|_{H^1_0(\Omega)}\|\hat{\psi}_h\|_{H^1_0(\Omega)}.
\end{aligned}$$

Then, using that $h \le h_1$, and taking into account that $\|\hat{\psi}_h\|_{H^1_0(\Omega)} \le \|\psi\|_{H^1_0(\Omega)} \le \|\psi\|_{H^2(\Omega)} \le K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)\|y_h\|_{L^2(\Omega)}$, see Lemma 3.2 below, and arguing as above we deduce

$$\|y_h\|_{L^2(\Omega)} \le 2\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\hat{c}_2 K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)h\|y_h\|_{H^1_0(\Omega)} + 2\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)\|y\|_{H^1_0(\Omega)}.$$

Moreover, since $h \le h_2/2$ we obtain

$$\|y_h\|_{L^2(\Omega)} \le \frac{1}{4}\sqrt{\frac{\Lambda}{C_{\Lambda,b}}}\|y_h\|_{H_0^1(\Omega)} \\ + 2\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)\|y\|_{H_0^1(\Omega)},$$

and

$$\|y_h\|^2_{L^2(\Omega)} \le \frac{1}{8}\frac{\Lambda}{C_{\Lambda,b}}\|y_h\|^2_{H_0^1(\Omega)} \\ + 8\|\mathcal{A}\|^2_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}K^2_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)^2\|y\|^2_{H_0^1(\Omega)}.$$

Now, from (3.2) we obtain

$$\frac{\Lambda}{4}\|y_h\|^2_{H_0^1(\Omega)} - C_{\Lambda,b}\|y_h\|^2_{L^2(\Omega)} \le a(y_h, y_h) + \int_\Omega a_0(x)y_h^2(x)\,dx \\ = \int_\Omega u y_h\,dx = a(y, y_h) \le \|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}\|y\|_{H_0^1(\Omega)}\|y_h\|_{H_0^1(\Omega)}.$$

Then, from the last inequalities we infer

$$\frac{\Lambda}{8}\|y_h\|^2_{H_0^1(\Omega)} \le \|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}\|y\|_{H_0^1(\Omega)}\|y_h\|_{H_0^1(\Omega)} \\ + C_{\Lambda,b}8\|\mathcal{A}\|^2_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}K^2_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)^2\|y\|^2_{H_0^1(\Omega)}.$$

Young's inequality implies

$$\frac{\Lambda}{16}\|y_h\|^2_{H_0^1(\Omega)} \le \|\mathcal{A}\|^2_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}\left(\frac{4}{\Lambda} + 8C_{\Lambda,b}K^2_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)^2\right)\|y\|^2_{H_0^1(\Omega)}.$$

This implies (3.4). Indeed, it is enough to observe that $y = \mathcal{A}^{-1}u$. □

Lemma 3.2 *Let $a_0 \in L^2(\Omega)$ be a non-negative function. For every $u \in L^2(\Omega)$ there exists a unique solution $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$ of*

$$\begin{cases} \mathcal{A}^*\psi + a_0\psi = u \text{ in } \Omega, \\ \psi = 0 \text{ on } \Gamma. \end{cases}$$

Moreover, there exists a constant $K_{\mathcal{A}^*}$ only depending on $\mathcal{A}^*$ such that

$$\|\psi\|_{H^2(\Omega)} \le K_{\mathcal{A}^*}\Big(\|a_0\|_{L^2(\Omega)}+1\Big)\|u\|_{L^2(\Omega)} \quad \forall u \in L^2(\Omega).$$

Proof The existence and uniqueness of a solution $\psi \in H^2(\Omega) \cap H^1_0(\Omega)$ follows from [12, Corollary 2.6]. Let us prove the estimate. To this end we set $u = u^+ - u^-$, where $u^+ = \max\{u, 0\}$ and $u^- = -\min\{u, 0\}$. Now, we take ψ_1, ψ_2, ϕ_1 and ϕ_2 in $H^2(\Omega) \cap H^1_0(\Omega)$ satisfying

$$\mathcal{A}^*\psi_1 + a_0\psi_1 = u^+ \text{ and } \mathcal{A}^*\psi_2 + a_0\psi_2 = u^-,$$
$$\mathcal{A}^*\phi_1 = u^+ \text{ and } \mathcal{A}^*\phi_2 = u^-.$$

Then, the identity $\psi = \psi_1 - \psi_2$ holds. From the comparison principle proven in [12, Lemma 2.8] we know that all the above functions are non-negative. Using the same Lemma and the fact that $\mathcal{A}^*(\psi_1 - \phi_1) = -a_0\psi_1 \le 0$ and $\mathcal{A}^*(\psi_2 - \phi_2) = -a_0\psi_2 \le 0$, we infer that $0 \le \psi_1 \le \phi_1$ and $0 \le \psi_2 \le \phi_2$. Hence, we have

$$\|\psi\|_{L^\infty(\Omega)} \le \|\psi_1\|_{L^\infty(\Omega)} + \|\psi_2\|_{L^\infty(\Omega)} \le \|\phi_1\|_{L^\infty(\Omega)} + \|\phi_2\|_{L^\infty(\Omega)}.$$

Now, from Theorem 2.1 we obtain

$$\|\phi_1\|_{L^\infty(\Omega)} \le C\|u^+\|_{L^2(\Omega)} \text{ and } \|\phi_2\|_{L^\infty(\Omega)} \le C\|u^-\|_{L^2(\Omega)},$$

where C depends on $\mathcal{A}^*$. Combining the above inequalities it follows

$$\|\psi\|_{L^\infty(\Omega)} \le C\Big(\|u^+\|_{L^2(\Omega)} + \|u^-\|_{L^2(\Omega)}\Big) \le 2C\|u\|_{L^2(\Omega)}.$$

Using the above estimate and applying again Theorem 2.1 to the equation $\mathcal{A}^*\psi = u - a_0\psi$ we infer

$$\begin{aligned}\|\psi\|_{H^2(\Omega)} &\le C\Big(\|u\|_{L^2(\Omega)} + \|a_0\|_{L^2(\Omega)}\|\psi\|_{L^\infty(\Omega)}\Big)\\ &\le C\max\{1, 2C\}\Big(1 + \|a_0\|_{L^2(\Omega)}\Big)\|u\|_{L^2(\Omega)},\end{aligned}$$

which proves the lemma. □

Remark 3.3 Notice that in the proof, we have also stated the existence of a constant $K^\infty_{\mathcal{A}^*}$ which does not depend on a_0 such that

$$\|\psi\|_{L^\infty(\Omega)} \le K^\infty_{\mathcal{A}^*}\|u\|_{L^2(\Omega)}.$$

Since the non-linear discrete Eq. (3.1) is neither monotone nor coercive, the proof of existence or uniqueness of solution is not obvious. We will establish the existence for h small enough. In a first step, we make the following assumption

Assumption 2′ $f : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function, monotone non-decreasing with respect to the second variable, and satisfying

$$\begin{cases} \exists\phi_f \in L^2(\Omega) \text{ such that } |f(x, y)| \le \phi_f(x)\ \forall y \in \mathbb{R} \text{ and}\\ |f(x, y_2) - f(x, y_1)| \le \phi_f(x)|y_2 - y_1| \text{ for a.e. } x \in \Omega \text{ and } \forall y_1, y_2 \in \mathbb{R}. \end{cases} \tag{3.12}$$

Remark 3.4 Let us observe that under Assumption 2, given $M > 0$ and setting $f_M(x, y) = f(x, \mathrm{Proj}_{[-M,+M]}(y))$, we have that f_M satisfies Assumption 2'. Indeed, we have

$$\begin{cases} |f_M(x, y)| \le |f(x, 0)| + \phi_M(x)|\,\mathrm{Proj}_{[-M,+M]}(y)| \le |f(x, 0)| + \phi_M(x)M, \\ |f_M(x, y_2) - f_M(x, y_1)| \le \phi_M(x)|y_2 - y_1|. \end{cases} \tag{3.13}$$

Hence, (3.12) holds with

$$\phi_f(x) = |f(x, 0)| + \phi_M(x)\max(M, 1).$$

This property will be used later to remove the Assumption 2'.

It is obvious that (3.12) is more restrictive than (2.3). Indeed, Assumption 2 obviously follows from the above hypothesis. Later we will get rid of this assumption. Now, we address the existence of a solution of (3.1). In the next theorem we apply Lemma 3.1 with $a_0 = 0$, and we set $h_{\mathcal{A}} = h_{\mathcal{A},0}$ and $C_{\mathcal{A}} = C_{\mathcal{A},0}$.

Theorem 3.5 *If Assumptions 1 and 2' hold, then there exists $0 < h_{\mathcal{A},f} \le h_{\mathcal{A}}$ independent of u such that (3.1) has a unique solution.*

Proof. Let us take $h \in (0, h_{\mathcal{A}}]$. We define the function $F : Y_h \longrightarrow Y_h$ where $F_h(w_h) = y_h(w_h)$ satisfies

$$a(y_h(w_h), z_h) = \int_\Omega [u(x) - f(x, w_h(x))]z_h(x)\,dx \quad \forall z_h \in Y_h$$

From Lemma 3.1 we know that $y_h(w_h)$ is well defined and

$$\|y_h(w_h)\|_{H_0^1(\Omega)} \le C_{\mathcal{A}}\|\mathcal{A}^{-1}(u - f(\cdot, w_h))\|_{H_0^1(\Omega)}.$$

From this inequality and (3.12) we deduce that $\|y_h(w_h)\|_{H_0^1(\Omega)} \le r\ \forall w_h \in Y_h$ with

$$\begin{aligned} r &= C_{\mathcal{A}} \sup_{w_h \in Y_h} \|\mathcal{A}^{-1}(u - f(\cdot, w_h))\|_{H_0^1(\Omega)} \\ &\le C_{\mathcal{A}}\|\mathcal{A}^{-1}\|_{\mathcal{L}(H^{-1}(\Omega), H_0^1(\Omega))} C_\Omega(\|u\|_{L^2(\Omega)} + \|\phi_f\|_{L^2(\Omega)}). \end{aligned}$$

From this estimate, the continuity of F_h and Brouwer's fixed point theorem we obtain the existence of at least one fixed point y_h. Obviously, y_h is solution of (3.1). It remains to prove the uniqueness of a solution for h small enough. Let us assume that $y_{h,1}, y_{h,2} \in Y_h$ are two solutions of (3.1). Then subtracting the equations satisfied by these solutions we get

$$a(y_{h,2} - y_{h,1}, y_{h,2} - y_{h,1}) + \int_\Omega [f(x, y_{h,2}) - f(x, y_{h,1})](y_{h,2} - y_{h,1})\,dx = 0.$$

Using (3.2) along with the monotonicity of f we get

$$\frac{\Lambda}{4}\|y_{h,2}-y_{h,1}\|^2_{H^1_0(\Omega)} - C_{\Lambda,b}\|y_{h,2}-y_{h,1}\|^2_{L^2(\Omega)} \le 0. \tag{3.14}$$

We define

$$a_0(x) = \begin{cases} \dfrac{f(x,y_{h,2}(x)) - f(x,y_{h,1}(x))}{y_{h,2}(x)-y_{h,1}(x)} & \text{if } \ y_{h,2}(x) \neq y_{h,1}(x), \\ 0 & \text{otherwise.} \end{cases}$$

From (3.12) we get that $a_0 \in L^2(\Omega)$ and $\|a_0\|_{L^2(\Omega)} \le \|\phi_f\|_{L^2(\Omega)}$. Now we take $\psi \in H^2(\Omega) \cap H^1_0(\Omega)$ such that $\mathcal{A}^*\psi + a_0\psi = y_{h,2} - y_{h,1}$ in Ω. According to Lemma 3.2 we have that $\|\psi\|_{H^2(\Omega)} \le K_{\mathcal{A}^*}\big(1 + \|\phi_f\|_{L^2(\Omega)}\big)\|y_{h,2}-y_{h,1}\|_{L^2(\Omega)}$. We denote by $\hat{\psi}_h \in Y_h$ the $H^1_0(\Omega)$-projection of ψ in Y_h; see (3.7). Then, from (3.14), the definition of a_0, the choice of ψ and $\hat{\psi}_h \in Y_h$, and using (1.4), (1.5), (3.8) and that $y_{h,1}$ and $y_{h,2}$ are solutions of (3.1) we infer

$$\begin{aligned}
&\frac{\Lambda}{4C_{\Lambda,b}}\|y_{h,2}-y_{h,1}\|^2_{H^1_0(\Omega)} \le \|y_{h,2}-y_{h,1}\|^2_{L^2(\Omega)} = \int_\Omega [\mathcal{A}^*\psi + a_0\psi](y_{h,2}-y_{h,1})\,dx \\
&\quad = a(y_{h,2}-y_{h,1},\psi) + \int_\Omega [f(x,y_{h,2}) - f(x,y_{h,1})]\psi\,dx \\
&\quad = a(y_{h,2}-y_{h,1},\psi-\hat{\psi}_h) + \int_\Omega [f(x,y_{h,2}) - f(x,y_{h,1})](\psi-\hat{\psi}_h)\,dx \\
&\quad \le \|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))}\|y_{h,2}-y_{h,1}\|_{H^1_0(\Omega)}\|\psi-\hat{\psi}_h\|_{H^1_0(\Omega)} \\
&\qquad + \|\phi_f\|_{L^2(\Omega)}\|y_{h,2}-y_{h,1}\|_{L^4(\Omega)}\|\psi-\hat{\psi}_h\|_{L^4(\Omega)} \\
&\quad \le \Big(\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K^2_\Omega\Big)\|\psi-\hat{\psi}_h\|_{H^1_0(\Omega)}\|y_{h,2}-y_{h,1}\|_{H^1_0(\Omega)} \\
&\quad \le \Big(\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K^2_\Omega\Big)\hat{c}_2 h\|\psi\|_{H^2(\Omega)}\|y_{h,2}-y_{h,1}\|_{H^1_0(\Omega)} \\
&\quad \le \Big(\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K^2_\Omega\Big)K_{\mathcal{A}^*}\big(1+\|\phi_f\|_{L^2(\Omega)}\big)\hat{c}_2\, h \\
&\qquad \|y_{h,2}-y_{h,1}\|_{L^2(\Omega)}\|y_{h,2}-y_{h,1}\|_{H^1_0(\Omega)} \\
&\quad \le \Big(\|\mathcal{A}\|_{\mathcal{L}(H^1_0(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K^2_\Omega\Big)C_\Omega K_{\mathcal{A}^*}\big(1+\|\phi_f\|_{L^2(\Omega)}\big)\hat{c}_2\, h \\
&\qquad \|y_{h,2}-y_{h,1}\|^2_{H^1_0(\Omega)}.
\end{aligned}$$

From this inequality we obtain that $y_{h,2} - y_{h,1} = 0$ if

$$h < h_{\mathcal{A},f} = \min\left\{h_{\mathcal{A}}, \frac{\Lambda}{4C_{\Lambda,b}\tilde{C}}\right\},$$

where

$$\tilde{C} = \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K_\Omega^2\Big)C_\Omega K_{\mathcal{A}^*}\big(1+\|\phi_f\|_{L^2(\Omega)}\big)\hat{c}_2. \quad \square$$

Next we prove some error estimates for $y - y_h$.

Theorem 3.6 *If Assumptions 1 and 2' hold, then there exists $h_0 \le h_{\mathcal{A},f}$ and constants $M_{\mathcal{A},f}$ and $M_{\infty,\mathcal{A},f}$ independent of u such that for every $h < h_0$*

$$\|y-y_h\|_{L^2(\Omega)} + h\|y-y_h\|_{H_0^1(\Omega)} \le M_{\mathcal{A},f}\Big(\|u\|_{L^2(\Omega)}+1\Big)h^2, \tag{3.15}$$

$$\|y-y_h\|_{L^\infty(\Omega)} \le M_{\infty,\mathcal{A},f}\Big(\|u\|_{L^2(\Omega)}+1\Big)h^{2-\frac{n}{2}}, \tag{3.16}$$

where y and y_h denote the solutions of (1.1) and (3.1).

Proof The proof is divided into three steps.

Step 1. $\|y-y_h\|_{L^2(\Omega)} \le K_1 h\|y-y_h\|_{H_0^1(\Omega)}$.

We proceed similarly as we did in the proof of the above theorem. We define

$$a_0(x) = \begin{cases} \dfrac{f(x,y(x)) - f(x,y_h(x))}{y(x)-y_h(x)} & \text{if } y(x) \ne y_h(x), \\ 0 & \text{otherwise.} \end{cases}$$

Now we take $\psi \in H^2(\Omega)\cap H_0^1(\Omega)$ such that $\mathcal{A}^*\psi + a_0\psi = y - y_h$ in Ω, and $\hat{\psi}_h \in Y_h$ as the projection of ψ in Y_h. Then, subtracting the equations satisfied by y and y_h, using the estimate of Lemma 3.2 and taking $\hat{c}_2$ as in (3.8), we obtain

$$\begin{aligned}
\|y-y_h\|_{L^2(\Omega)}^2 &= \int_\Omega (\mathcal{A}^*\psi + a_0\psi)(y-y_h)\,dx \\
&= a(y-y_h,\psi) + \int_\Omega [f(x,y)-f(x,y_h)]\psi\,dx \\
&= a(y-y_h,\psi-\hat{\psi}_h) + \int_\Omega [f(x,y)-f(x,y_h)](\psi-\hat{\psi}_h)\,dx \\
&\le \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K_\Omega^2\Big)\|y-y_h\|_{H_0^1(\Omega)}\|\psi-\hat{\psi}_h\|_{H_0^1(\Omega)} \\
&\le \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K_\Omega^2\Big)\|y-y_h\|_{H_0^1(\Omega)}\hat{c}_2 h\|\psi\|_{H^2(\Omega)} \\
&\le K_1\|y-y_h\|_{H_0^1(\Omega)}\|y-y_h\|_{L^2(\Omega)}h,
\end{aligned}$$

which proves the desired estimate with a constant

$$K_1 = \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)}K_\Omega^2\Big)\hat{c}_2 K_{\mathcal{A}^*}\Big(1+\|\phi_f\|_{L^2(\Omega)}\Big)$$

independent of u.

Step 2. $\|y-y_h\|_{H_0^1(\Omega)} \le K_2\Big(\|u\|_{L^2(\Omega)}+1\Big)h$.

Let us denote $\hat{y}_h \in Y_h$ the projection of y in Y_h, so that

$$\int_\Omega \nabla \hat{y}_h \nabla z_h \, dx = \int_\Omega \nabla y \nabla z_h \, dx \quad \forall z_h \in Y_h.$$

Hence, we have with (2.4)

$$\|y - \hat{y}_h\|_{H_0^1(\Omega)} \le \hat{c}_2 C_{A,f} \Big(\|u\|_{L^2(\Omega)} + 1 \Big) h. \tag{3.17}$$

From the estimate proved in Step 1, (3.2) along with the monotonicity of f, and (3.17) we get for

$$h < h_0 = \min \left\{ h_{\mathcal{A},f}, \sqrt{\frac{\Lambda}{20 K_1^2 C_{\Lambda,b}}} \right\}$$

the following estimate

$$\begin{aligned}
\frac{\Lambda}{5} \|y - y_h\|^2_{H_0^1(\Omega)} &\le a(y - y_h, y - y_h) + \int_\Omega [f(x, y) - f(x, y_h)](y - y_h) \, dx \\
&= a(y - y_h, y - \hat{y}_h) + \int_\Omega [f(x, y) - f(x, y_h)](y - \hat{y}_h) \, dx \\
&\le \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)} K_\Omega^2 \Big) \|y - y_h\|_{H_0^1(\Omega)} \|y - \hat{y}_h\|_{H_0^1(\Omega)} \\
&\le \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)} K_\Omega^2 \Big) \|y - y_h\|_{H_0^1(\Omega)} \hat{c}_2 C_{A,f} \Big(\|u\|_{L^2(\Omega)} + 1 \Big) h,
\end{aligned}$$

which proves Step 2 with

$$K_2 = \Big(\|\mathcal{A}\|_{\mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))} + \|\phi_f\|_{L^2(\Omega)} K_\Omega^2 \Big) \hat{c}_2 C_{A,f},$$

and (3.15) follows for $M_{\mathcal{A},f} = K_2 \max\{1, K_1\}$.

Step 3. Proof of (3.16). The proof of (3.16) follows from (3.15) and the inverse inequality

$$\|z_h\|_{L^\infty(\Omega)} \le c_{\infty,2} \|z_h\|_{L^2(\Omega)} h^{-\frac{n}{2}} \; \forall z_h \in Y_h,$$

where $c_{\infty,2}$ in independent of h; see [18, Theorem 17.2]. Though the proof is quite classical we include it here for convenience of the reader. Taking $\tilde{y}_h \in Y_h$ as the function interpolating y at the nodes of the triangulation, we know, see [18, Theorem 16.2 and Theorem 17.1], that

$$\|y - \tilde{y}_h\|_{L^\infty(\Omega)} \le \tilde{c}_\infty h^{2-\frac{n}{2}} \|y\|_{H^2(\Omega)} \text{ and } \|y - \tilde{y}_h\|_{L^2(\Omega)} \le \tilde{c}_2 h^2 \|y\|_{H^2(\Omega)}.$$

Hence, we have with (2.4)

$$\begin{aligned}\|y-y_h\|_{L^\infty(\Omega)} &\leq \|y-\tilde{y}_h\|_{L^\infty(\Omega)} + \|\tilde{y}_h - y_h\|_{L^\infty(\Omega)}\\ &\leq \|y-\tilde{y}_h\|_{L^\infty(\Omega)} + c_{\infty,2}\|\tilde{y}_h - y_h\|_{L^2(\Omega)}h^{-\frac{n}{2}}\\ &\leq \|y-\tilde{y}_h\|_{L^\infty(\Omega)} + c_{\infty,2}\Big(\|\tilde{y}_h - y\|_{L^2(\Omega)} + \|y-y_h\|_{L^2(\Omega)}\Big)h^{-\frac{n}{2}}\\ &\leq (\tilde{c}_\infty C_{A,f} + c_{\infty,2}\tilde{c}_2 + c_{\infty,2}M_{\mathcal{A},f})\Big(\|u\|_{L^2(\Omega)}+1\Big)h^{2-\frac{n}{2}},\end{aligned}$$

which implies (3.16). □

Now, we replace Assumption 2' by the weaker Assumption 2'.

Theorem 3.7 *Under Assumptions 1 and 2, for all $M \geq 1 + \|y\|_{C(\bar{\Omega})}$ there exists $h_M > 0$ such that for every $h < h_M$ (3.1) has a unique solution y_h satisfying $\|y_h\|_{C(\bar{\Omega})} \leq M$. Moreover, there exist constants K_M and $K_{\infty,M}$ independent of u such that*

$$\|y-y_h\|_{L^2(\Omega)} + h\|y-y_h\|_{H_0^1(\Omega)} \leq K_M\Big(\|u\|_{L^2(\Omega)}+1\Big)h^2, \tag{3.18}$$

$$\|y-y_h\|_{L^\infty(\Omega)} \leq K_{\infty,M}\Big(\|u\|_{L^2(\Omega)}+1\Big)h^{2-\frac{n}{2}}. \tag{3.19}$$

Further, if there exist other solutions $\{\tilde{y}_h\}_{h<h_M}$ of (3.1) with $y_h \neq \tilde{y}_h$ for all h, then $\lim_{h\to 0}\|\tilde{y}_h\|_{C(\bar{\Omega})} = \infty$.

Proof Given M, we define the function $f_M : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ by

$$f_M(x,y) = f(x, \mathrm{Proj}_{[-M,+M]}(y)).$$

Then, according to Remark 3.4, f_M satisfies the conditions (3.12). Moreover, if y is the solution of (1.1), then $f_M(x,y(x)) = f(x,y(x))$ in Ω, thus y also satisfies

$$\begin{cases} Ay + b(x)\cdot\nabla y + f_M(x,y) = u \text{ in } \Omega,\\ y = 0 \text{ on } \Gamma.\end{cases}$$

According to Theorem 3.5, there exists $\tilde{h}_M = h_{\mathcal{A},f_M}$, which depends on $\mathcal{A}$ and $\|f(x,0)\|_{L^2(\Omega)} + \max(M,1)\|\phi_M\|_{L^2(\Omega)}$ but is independent of u, such that the variational problem

$$\begin{cases}\text{Find } y_h \in Y_h \text{ such that}\\ a(y_h,z_h) + \displaystyle\int_\Omega f_M(x,y_h(x))z_h(x)\,dx = \int_\Omega u(x)z_h(x)\,dx \;\; \forall z_h \in Y_h.\end{cases} \tag{3.20}$$

has a unique solution for every $h < \tilde{h}_M$. Moreover, from Theorem 3.6 we have the estimate

$$\|y-y_h\|_{L^\infty(\Omega)} \leq M_{\infty,\mathcal{A},f_M}\Big(\|u\|_{L^2(\Omega)}+1\Big)h^{2-\frac{n}{2}}.$$

Taking h_M such that

$$0 < h_M \le \tilde{h}_M \text{ and } M_{\infty,\mathcal{A},f_M}\Big(\|u\|_{L^2(\Omega)} + 1\Big)h_M^{2-\frac{n}{2}} \le 1, \tag{3.21}$$

we have

$$\|y_h\|_{C(\bar{\Omega})} \le \|y - y_h\|_{C(\bar{\Omega})} + \|y\|_{C(\bar{\Omega})} \le M.$$

Hence, $f_M(x, y_h(x)) = f(x, y_h(x))$ in Ω for all $h \le h_M$. Consequently, y_h is a solution of (3.1). Moreover, if $\hat{y}_h$ is another solution of (3.1) such that $\|\hat{y}_h\|_{C(\bar{\Omega})} \le M$, then $\hat{y}_h$ also solves (3.20). Hence, Theorem 3.5 implies that $y_h = \hat{y}_h$. Moreover, the estimates (3.18) and (3.19) follow from (3.15) and (3.16).

Finally, let us assume that $\{\tilde{y}_h\}_{h<h_M}$ are solutions of (3.1) with $y_h \ne \tilde{y}_h$. We argue by contradiction and assume that there exists a constant C_∞ such that $\|\tilde{y}_h\|_{C(\bar{\Omega})} \le C_\infty$ for all $h < h_M$. We take $\tilde{M} = \max\{1 + \|y\|_{C(\bar{\Omega})}, C_\infty\}$. Then, y_h and $\tilde{y}_h$ are two different solutions of (3.20) for every $h < h_{\tilde{M}}$, which contradicts the uniqueness already established. □

Using the previous theorem, we are going to establish a well defined local mapping $u_h \to y_h$ by ignoring those solutions of (3.1) with big $C(\bar{\Omega})$-norms.

Theorem 3.8 *Suppose that Assumptions 1 and 2 hold. Let $\bar{y} \in Y$ be the solution of (1.1) corresponding to the control $\bar{u} \in L^2(\Omega)$. Given $\rho > 0$ arbitrary, there exist $\rho^* > 0$ and $h_0 > 0$ such that (3.1) has a unique solution $y_h(u) \in \bar{B}^Y_{\rho^*}(\bar{y})$ for every $u \in \bar{B}_\rho(\bar{u}) \subset L^2(\Omega)$ and for all $h < h_0$, where*

$$\bar{B}_\rho(\bar{u}) = \{u \in L^2(\Omega) : \|u - \bar{u}\|_{L^2(\Omega)} \le \rho\} \textit{ and } \bar{B}^Y_{\rho^*}(\bar{y}) = \{y \in Y : \|y - \bar{y}\|_Y \le \rho^*\}.$$

Furthermore, there exist constants K_2 and K_∞ such that

$$\|y_u - y_h(u)\|_{L^2(\Omega)} + h\|y_u - y_h(u)\|_{H_0^1(\Omega)} \le K_2\Big(\|\bar{u}\|_{L^2(\Omega)} + \rho + 1\Big)h^2, \tag{3.22}$$

$$\|y_u - y_h(u)\|_{L^\infty(\Omega)} \le K_\infty\Big(\|\bar{u}\|_{L^2(\Omega)} + \rho + 1\Big)h^{2-\frac{n}{2}} \quad \forall u \in \bar{B}_\rho(\bar{u}), \tag{3.23}$$

where y_u and $y_h(u)$ are the solutions of (1.1) and (3.1), respectively, associated with the control u.

Proof Let us fix $\rho > 0$. In [12, Lemma 3.5], it was proved the existence of a constant M_ρ such that

$$\|y_u - \bar{y}\|_Y \le M_\rho\|u - \bar{u}\|_{L^2(\Omega)} \le M_\rho\rho \quad \forall u \in \bar{B}_\rho(\bar{u}). \tag{3.24}$$

Hence, we have

$$\|y_u\|_{C(\bar{\Omega})} \le \|\bar{y}\|_{C(\bar{\Omega})} + M_\rho\rho \quad \forall u \in \bar{B}_\rho(\bar{u}).$$

Now, we set $M = 1 + \|\bar{y}\|_{C(\bar{\Omega})} + M_\rho \rho$. According to Theorem 3.7 and (3.21), there exists $h_M > 0$ such that for all $h < h_M$ and for every $u \in \bar{B}_\rho(\bar{u})$ (3.1) has a unique solution $y_h(u)$ satisfying $\|y_h(u)\|_{C(\bar{\Omega})} \le M$. Moreover, the estimates (3.18) and (3.19) hold for $y_u - y_h(u)$. Further, it is enough to observe that $\|u\|_{L^2(\Omega)} \le \|\bar{u}\|_{L^2(\Omega)} + \rho$ holds to deduce (3.22) and (3.23). Finally, we define $\rho^* = 1 + M_\rho \rho$ and take $h_0 \in (0, h_M]$ such that

$$\Big(\|\bar{u}\|_{L^2(\Omega)} + \rho + 1\Big)(K_2 h_0 + K_\infty h_0^{2-\frac{n}{2}}) \le 1. \tag{3.25}$$

Then, for every $u \in \bar{B}_\rho(\bar{u})$, (3.24) implies that $y_u \in B^Y_{\rho^*}(\bar{y})$ holds. Furthermore, (3.22), (3.23) and (3.25) imply

$$\|y_h(u) - \bar{y}\|_Y \le \|y_h(u) - y_u\|_Y + \|y_u - \bar{y}\|_Y < 1 + M_\rho \rho = \rho^* \quad \forall h < h_0.$$

Hence, $y_h(u) \in B^Y_{\rho^*}(\bar{u})$ holds. Thus, we have proved that (3.1) has a solution in $B^Y_{\rho^*}(\bar{u}) \cap Y_h$ for every $u \in \bar{B}_\rho(\bar{u})$. It remains to prove the uniqueness. This follows from the fact that $h_0 \le h_M$ and, thanks to (3.25), any element $y_h \in \bar{B}^Y_{\rho^*}(\bar{u})$ satisfies

$$\|y_h\|_{C(\bar{\Omega})} \le \|y_h - \bar{y}\|_{C(\bar{\Omega})} + \|\bar{y}\|_{C(\bar{\Omega})} \le \rho^* + \|\bar{y}\|_{C(\bar{\Omega})} = M. \qquad \square$$

□

4 Approximation of the control problem (P)

In this section, we discretize the control problem (P) and study the convergence of the discretizations. To this end, we suppose without express mention that Assumptions 1 and 2 hold.

Let us consider the functional $\mathcal{J} : L^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$ given by

$$\mathcal{J}(y, u) = \frac{1}{2}\int_\Omega (y(x) - y_d(x))^2\, dx + \frac{\nu}{2}\int_\Omega u^2\, dx.$$

Let us denote by $\mathcal{U}_h$ one of the following two spaces:

$$\mathcal{U}_h = \mathcal{U}_h^0 := \{u_h \in L^2(\Omega) : \ u_{h|T} \in P_0(T)\ \forall T \in \mathcal{T}_h\},$$
$$\mathcal{U}_h = \mathcal{U}_h^1 := \{u_h \in C(\bar{\Omega}) : \ u_{h|T} \in P_1(T)\ \forall T \in \mathcal{T}_h\},$$

where $P_0(T)$ and $P_1(T)$ denote the space of polynomials in T of degree 0 and ≤ 1, respectively. We also set $U_{\mathrm{ad},h} = \mathcal{U}_h \cap U_{\mathrm{ad}}$. If $\mathcal{U}_h = \mathcal{U}_h^0$, then we will denote $\Pi_h : L^2(\Omega) \longrightarrow \mathcal{U}_h^0$ the $L^2(\Omega)$ linear projection. If $\mathcal{U}_h = \mathcal{U}_h^1$, then $\Pi_h : L^2(\Omega) \longrightarrow \mathcal{U}_h^1$ will denote Cartensen's quasi-interpolation operator. In both cases it is known that $\Pi_h u$ converges to u in $L^2(\Omega)$ as h tends to 0 for all $u \in L^2(\Omega)$, and $\Pi_h u \in U_{\mathrm{ad},h}$ for all $u \in U_{\mathrm{ad}}$.

We will approximate Problem (P) by the problem

$$(\mathcal{P}_h) \qquad \min\{\mathcal{J}(y_h, u_h) : (y_h, u_h) \in Y_h \times U_{\mathrm{ad},h} \text{ satisfies (4.1)}\}.$$

where

$$a(y_h, z_h) + \int_\Omega f(x, y_h(x)) z_h(x)\, dx = \int_\Omega u_h(x) z_h(x)\, dx \ \forall z_h \in Y_h. \tag{4.1}$$

Theorem 4.1 *There exists $h_0 > 0$ such that problem $(\mathcal{P}_h)$ has at least one solution $(\bar{y}_h, \bar{u}_h)$ for all $h < h_0$. Moreover, if $\{(\bar{y}_h, \bar{u}_h)\}_{h<h_0}$ is a sequence of solutions of problems $(\mathcal{P}_h)$, then it is bounded in $H_0^1(\Omega) \times L^2(\Omega)$ and there exist subsequences converging weakly in $H_0^1(\Omega) \times L^2(\Omega)$. In addition, if a subsequence, denoted in the same way, satisfies that $(\bar{y}_h, \bar{u}_h) \rightharpoonup (\bar{y}, \bar{u})$ in $H_0^1(\Omega) \times L^2(\Omega)$ as $h \to 0$, then $(\bar{y}, \bar{u}) \in Y \times U_{\mathrm{ad}}$, $\bar{u}$ is a solution of (P) with associated stated $\bar{y}$, and $(\bar{y}_h, \bar{u}_h) \to (\bar{y}, \bar{u})$ strongly in $H_0^1(\Omega) \times L^2(\Omega)$.*

Proof *Claim 1—Existence of discrete solutions:* Let us prove the existence of a solution of $(\mathcal{P}_h)$ for every h small enough. Let $F_h : Y_h \times U_{\mathrm{ad},h} \longrightarrow Y_h^*$ be the mapping defined by

$$\langle F_h(y_h, u_h), w_h \rangle_{Y_h^*, Y_h} = a(y_h, w_h) + \int_\Omega [f(x, y_h) - u_h] w_h\, dx \quad \forall w_h \in Y_h.$$

Since F_h is continuous and $U_{\mathrm{ad},h}$ is closed, then the set of feasible points (y_h, u_h) for $(\mathcal{P}_h)$, which is $F_h^{-1}(\{0\})$, is closed in $Y_h \times \mathcal{U}_h$. Moreover, $\mathcal{J}$ is continuous and coercive. Hence, it is enough to prove the existence of feasible points for $(\mathcal{P}_h)$. We choose a constant $u \in U_{\mathrm{ad},h}$ to guarantee $u \in U_{\mathrm{ad},h}$ for every $h > 0$. This can be done by $u \equiv \alpha$ if $\alpha > -\infty$, or $u \equiv \beta$ if $\beta < \infty$, or $u \equiv 0$ otherwise. According to Theorem 3.7, there exists $h_0 > 0$ such that (4.1) has a solution $y_h(u) \in Y_h$ for every $h < h_0$ satisfying $y_h(u) \to y_u$ in Y. Therefore, $(y_h(u), u)$ is a feasible point for $(\mathcal{P}_h)$ for every $h < h_0$.

Claim 2—Uniform boundedness of discrete solutions in $H_0^1(\Omega) \times L^2(\Omega)$ and weak convergence: Let us denote by $\{(\bar{y}_h, \bar{u}_h)\}_{h<h_0}$ a sequence of solutions for problems $(\mathcal{P}_h)$. We prove the boundedness of this sequence in $H_0^1(\Omega) \times L^2(\Omega)$. Since

$$\mathcal{J}(\bar{y}_h, \bar{u}_h) \le \mathcal{J}(y_h(u), u) \to \mathcal{J}(y_u, u),$$

we infer that $\{(\bar{y}_h, \bar{u}_h)\}_{h<h_0}$ is bounded in $L^2(\Omega) \times L^2(\Omega)$. Moreover, since $(\bar{y}_h, \bar{u}_h)$ satisfies (4.1), taking $z_h = \bar{y}_h$ in (4.1) we deduce from (3.2) and the monotonicity of f

$$\begin{aligned}\frac{\Lambda}{4} \|\bar{y}_h\|^2_{H_0^1(\Omega)} - C_{\Lambda,b} \|\bar{y}_h\|^2_{L^2(\Omega)} &\le a(\bar{y}_h, \bar{y}_h) + \int_\Omega [f(x, \bar{y}_h) - f(x, 0)] \bar{y}_h\, dx \\ &= \int_\Omega [\bar{u}_h - f(x, 0)] \bar{y}_h\, dx.\end{aligned}$$

From here we obtain $\forall h < h_0$.

$$\frac{\Lambda}{4}\|\bar{y}_h\|^2_{H_0^1(\Omega)} \leq C_{\Lambda,b}\|\bar{y}_h\|^2_{L^2(\Omega)} + \Big(\|\bar{u}_h\|_{L^2(\Omega)} + \|f(\cdot,0)\|_{L^2(\Omega)}\Big)\|\bar{y}_h\|_{L^2(\Omega)}.$$

This inequality along with the boundedness of $\{(\bar{y}_h, \bar{u}_h)\}_{h<h_0}$ in $L^2(\Omega) \times L^2(\Omega)$ implies that $\{\bar{y}_h\}_{h<h_0}$ is bounded in $H_0^1(\Omega)$ as well.

Since $\{(\bar{y}_h, \bar{u}_h)\}_{h<h_0}$ is bounded in $H_0^1(\Omega) \times L^2(\Omega)$, it has weakly convergent subsequences in this topology. Now, we take a subsequence, denoted in the same way, such that

$$(\bar{y}_h, \bar{u}_h) \overset{h\to 0}{\rightharpoonup} (\bar{y}, \bar{u}) \text{ in } H_0^1(\Omega) \times L^2(\Omega),$$
$$\bar{y}_h \xrightarrow{h\to 0} \bar{y} \text{ in } L^2(\Omega) \text{ and } \bar{y}_h(x) \to \bar{y}(x) \text{ a.e. in } \Omega.$$

Claim 3—Validity of the state equation for the limit element: The proof of this claim is split into three main steps: First, we prove that $f(\cdot, \bar{y}_h) \to f(\cdot, \bar{y})$ strongly in $L^1(\Omega)$. Next, we use this to prove (4.3), which is a weak version of (1.1) for bounded test functions. Finally, we prove that $\bar{y} \in L^\infty(\Omega)$ to conclude this part of the proof.

To prove that $f(\cdot, \bar{y}_h) \to f(\cdot, \bar{y})$ strongly in $L^1(\Omega)$, we show that $\{f(x, \bar{y}_h(x)) - f(x, 0)\}_{h<h_0}$ is uniformly integrable. Then, the convergence will follow from Vitali's convergence theorem and the pointwise convergence of the sequence $f(x, \bar{y}_h(x)) \to f(x, \bar{y}(x))$ in Ω; see [4, Volume I, Theorem 4.5.4] or [31, Chapter 6, Exercise 10].

We get from (4.1) and the boundedness of $\{(\bar{y}_h, \bar{u}_h)\}_{h<h_0}$ in $H_0^1(\Omega) \times L^2(\Omega)$ the existence of a constant C such that

$$\int_\Omega [f(x, \bar{y}_h) - f(x, 0)]\bar{y}_h \, dx$$
$$= \int_\Omega [\bar{u}_h - f(x, 0)]\bar{y}_h \, dx - a(\bar{y}_h, \bar{y}_h) \leq C \quad \forall h < h_0. \tag{4.2}$$

Let $\varepsilon > 0$ be arbitrarily small. Using (2.3) with $M = 2C/\varepsilon$, we deduce the existence of a function $\phi_\varepsilon \in L^2(\Omega)$ such that

$$|f(x, y) - f(x, 0)| \leq \phi_\varepsilon(x)|y| \leq \phi_\varepsilon(x)\frac{2C}{\varepsilon} \quad \text{if} \quad |y| \leq \frac{2C}{\varepsilon}.$$

From the integrability of ϕ_ε we infer the existence of $\lambda_0 > 0$ such that

$$\int_{\{x:\phi_\varepsilon(x)\geq\lambda_0\}} \phi_\varepsilon(x)\, dx \leq \frac{\varepsilon^2}{4C}.$$

Let us set $\lambda = \frac{2C\lambda_0}{\varepsilon}$ and $\Omega_{h,\lambda} = \{x \in \Omega : |f(x, \bar{y}_h(x)) - f(x, 0)| > \lambda\}$. We notice the following properties:

- If $x \in \Omega_{h,\lambda}$ and $|\bar{y}_h(x)| > \frac{2C}{\varepsilon}$, then

$$|f(x,\bar{y}_h(x)) - f(x,0)| \leq \frac{\varepsilon}{2C}[f(x,\bar{y}_h(x)) - f(x,0)]y_h(x).$$

- If $x \in \Omega_{h,\lambda}$ and $|\bar{y}_h(x)| \leq \frac{2C}{\varepsilon}$ then

$$|f(x,\bar{y}_h(x)) - f(x,0)| \leq \phi_\varepsilon(x)\frac{2C}{\varepsilon} \text{ and } \phi_\varepsilon(x) \geq \lambda_0.$$

From here, and using (4.2), we infer

$$\begin{aligned}\int_{\Omega_{h,\lambda}} |f(x,\bar{y}_h(x)) - f(x,0)|\,dx &\leq \frac{2C}{\varepsilon}\int_{\{x:\phi_\varepsilon(x)\geq\lambda_0\}} \phi_\varepsilon(x)\,dx \\ &+ \frac{\varepsilon}{2C}\int_{\Omega_{h,\lambda}} [f(x,\bar{y}_h(x)) - f(x,0)]\bar{y}_h(x)\,dx \leq \varepsilon.\end{aligned}$$

Since λ was chosen independently of h, the uniform integrability follows and $f(\cdot,\bar{y}_h) \to f(\cdot,\bar{y})$ strongly in $L^1(\Omega)$.

Next, given $z \in H_0^1(\Omega) \cap L^\infty(\bar{\Omega})$, we can take a sequence $\{z_h\}_{h<h_0}$ with $z_h \in Y_h$ for every h such that $z_h \to z$ in $H_0^1(\Omega)$ and $\|z_h\|_{L^\infty(\Omega)} \leq \|z\|_{L^\infty(\Omega)}$. For instance, we can take z_h Carstensen's quasi-interpolation of z; see [7]. Hence, using Lebesgue's dominated convergence theorem, we can pass to the limit in (4.1) and deduce that

$$a(\bar{y},z) + \int_\Omega f(x,\bar{y})z\,dx = \int_\Omega \bar{u}z\,dx \;\forall z \in H_0^1(\Omega) \cap L^\infty(\Omega). \tag{4.3}$$

Finally, we prove that $\bar{y} \in L^\infty(\Omega)$, and consequently, by a truncation argument, it will follow that (4.3) holds for all $z \in H_0^1(\Omega)$ Let us set

$$\tilde{a}(y,z) = a(y,z) + C_{\Lambda,b}\int_\Omega yz\,dx,$$

where $C_{\Lambda,b}$ is given by (3.2). Then we have that $\tilde{a}$ is coercive in $H_0^1(\Omega)$ and

$$\tilde{a}(\bar{y},z) + \int_\Omega [f(x,\bar{y}) - f(x,0)]z\,dx = \int_\Omega [\bar{u} + C_{\Lambda,b}\bar{y} - f(x,0)]z\,dx \;\; \forall z \in Y. \tag{4.4}$$

The above identity holds, in particular, for $y_k = \text{Proj}_{[-k,+k]}(\bar{y})$ for every $k \geq 1$:

$$\begin{aligned}&\tilde{a}(\bar{y},y_k) + \int_\Omega [f(x,\bar{y}) - f(x,0)]y_k\,dx \\ &\quad = \int_\Omega [\bar{u} + C_{\Lambda,b}\bar{y} - f(x,0)]y_k\,dx \;\; \forall k \geq 1.\end{aligned} \tag{4.5}$$

Moreover, from Fatou's Lemma, (4.5), denoting $g = \bar{u} + C_{\Lambda,b}\bar{y} - f(x,0) \in L^2(\Omega)$ and taking into account that $\tilde{a}(\bar{y}, y_k) \geq \tilde{a}(y_k, y_k) \geq 0$, we have

$$\begin{aligned}\int_\Omega [f(x,\bar{y}(x)) - f(x,0)]\bar{y}(x)\,dx &\leq \liminf_{k\to\infty} \int_\Omega [f(x,\bar{y}(x)) - f(x,0)]y_k(x)\,dx \\ &\leq \liminf_{k\to\infty} \int_\Omega g(x)\bar{y}_k(x)dx \leq \int_\Omega g(x)\bar{y}(x)dx.\end{aligned} \tag{4.6}$$

Hence, $[f(\cdot,\bar{y}) - f(\cdot,0)]\bar{y} \in L^1(\Omega)$ holds. Since $0 \leq [f(x,\bar{y}(x)) - f(x,0)]y_k(x) \leq [f(x,\bar{y}(x)) - f(x,0)]\bar{y}(x)$, we can apply the Lebesgue's dominated convergence theorem to pass to the limit in (4.5):

$$\tilde{a}(\bar{y},\bar{y}) + \int_\Omega [f(x,\bar{y}) - f(x,0)]\bar{y}\,dx = \int_\Omega [\bar{u} + C_{\Lambda,b}\bar{y} - f(x,0)]\bar{y}\,dx.$$

Then, combining this identity and (4.5), we get for $y^k = \bar{y} - y_k$:

$$\tilde{a}(\bar{y}, y^k) + \int_\Omega [f(x,\bar{y}) - f(x,0)]y^k\,dx = \int_\Omega g y^k\,dx \ \ \forall k \geq 1.$$

From the monotonicity of f and the definition of y^k we get

$$\frac{\Lambda}{4}\|y^k\|^2_{H_0^1(\Omega)} \leq \tilde{a}(y^k, y^k) \leq \tilde{a}(\bar{y}, y^k) \leq \int_\Omega g y^k\,dx \ \ \forall k \geq 1.$$

Then, we can proceed as in [34, Theorem 4.1] or [35, §7.2] to infer the existence of k_0 such that $y^k = 0$ for $k \geq k_0$. Hence, $\bar{y} \in L^\infty(\Omega)$ holds. Moreover, from (2.5) and (2.6) we deduce that $f(\cdot,\bar{y}) \in L^2(\Omega)$. Therefore, we have that $\mathcal{A}\bar{y} \in L^2(\Omega)$ and, consequently, $\bar{y} \in C(\bar{\Omega})$; see [12, Corollary 2.2]. Thus, $\bar{y} \in Y$ and (4.3) implies that $\bar{y}$ is the solution of (1.1) associated with $\bar{u}$.

Claim 4—Optimality of $\bar{u}$: Let us prove that $\bar{u}$ is a solution of (P). First, notice that it follows from the inclusion $U_{\text{ad},h} \subset U_{\text{ad}}$ that $\bar{u} \in U_{\text{ad}}$. To prove the optimality, we take an arbitrary element $u \in U_{\text{ad}}$ and set $u_h = \Pi_h u \in U_{\text{ad},h}$. Moreover, Theorem 3.7 implies that there exists $y_h(u_h) \in Y_h$ solution of (4.1) for every h small enough and such that $y_h(u_h) \to y_u$ in Y. Hence, we deduce from the optimality of $(\bar{y}_h, \bar{u}_h)$

$$\begin{aligned}J(\bar{u}) = \mathcal{J}(\bar{y},\bar{u}) &\leq \liminf_{h\to 0} \mathcal{J}(\bar{y}_h,\bar{u}_h) \leq \limsup_{h\to 0} \mathcal{J}(\bar{y}_h,\bar{u}_h) \\ &\leq \limsup_{h\to 0} \mathcal{J}(\bar{y}_h(u_h), u_h) = \mathcal{J}(y_u, u) = J(u).\end{aligned}$$

Since u was taken arbitrary in U_{ad}, the above inequalities prove that $\bar{u}$ is a solution of (P).

Claim 5—Strong convergence in $H_0^1(\Omega) \times L^2(\Omega)$. If we take above $u = \bar{u}$ we deduce from the previous inequalities and the strong convergence $\bar{y}_h \to \bar{y}$ in $L^2(\Omega)$

that $\|\bar u_h\|_{L^2(\Omega)} \to \|\bar u\|_{L^2(\Omega)}$ and, hence, $\bar u_h \to \bar u$ strongly in $L^2(\Omega)$. Finally, we prove the strong convergence $\bar y_h \to \bar y$ in $H_0^1(\Omega)$ as follows

$$\begin{aligned} a(\bar y, \bar y) &\le \liminf_{h\to 0} a(\bar y_h, \bar y_h) \le \limsup_{h\to 0} a(\bar y_h, \bar y_h)\\ &\le \limsup_{h\to 0} \int_\Omega \big[\bar u_h - f(x,0)\big]\bar y_h\,dx - \liminf_{h\to 0}\int_\Omega [f(x,\bar y_h) - f(x,0)]\bar y_h\,dx\\ &= \int_\Omega \big[\bar u - f(x,0)\big]\bar y\,dx - \int_\Omega [f(x,\bar y) - f(x,0)]\bar y\,dx = a(\bar y,\bar y), \end{aligned}$$

where we have used (4.6). The above inequalities imply that $a(\bar y_h, \bar y_h) \to a(\bar y, \bar y)$. Hence, from (3.2) and the weak convergence $\bar y_h \rightharpoonup \bar y$ in $H_0^1(\Omega)$ we infer that $\bar y_h \to \bar y$ strongly in $H_0^1(\Omega)$. □

Next, we prove a kind of converse theorem. More precisely, we assume that $\bar u \in U_{\rm ad}$ is a strict local minimum of (P) with associated state $\bar y$. This means that there exists $\rho > 0$ such that

$$J(\bar u) < J(u) \quad \forall u \in (\bar B_\rho(\bar u) \cap U_{\rm ad}) \setminus \{\bar u\}. \tag{4.7}$$

Under assumptions of Theorem 4.1 there exists $\rho^* > 0$ and $h_0 > 0$ such that for every $u \in \bar B_\rho(\bar u)$ there exists a unique solution of (4.1) $y_h(u) \in \bar B^Y_{\rho^*}(\bar y)$; see Theorem 3.8. Then, for every $h < h_0$ we have a well defined mapping $G_h : \bar B_\rho(\bar u) \longrightarrow \bar B^Y_{\rho^*}(\bar y) \cap Y_h$ given by $G_h(u) = y_h(u)$. Now we define the functional $J_h : \bar B_\rho(\bar u) \longrightarrow \mathbb{R}$ by

$$J_h(u) = \mathcal{J}(y_h(u), u) = \frac{1}{2}\int_\Omega (y_h(u) - y_d)^2\,dx + \frac{\nu}{2}\int_\Omega u^2\,dx.$$

Associated with J_h we define the discrete control problem

$$(\mathrm{P}_h^\rho) \qquad \min_{u_h \in \bar B_\rho(\bar u)\cap U_{{\rm ad},h}} J_h(u_h).$$

Theorem 4.2 *Under Assumptions 1 and 2, and with the above notations, there exists $h_\rho \in (0, h_0]$ such that (P_h^ρ) has at least one solution $\bar u_h$ for every $h \le h_\rho$. Moreover, the convergence $\bar u_h \stackrel{h\to 0}{\longrightarrow} \bar u$ in $L^2(\Omega)$ holds.*

Proof Since $\Pi_h \bar u \stackrel{h\to 0}{\longrightarrow} \bar u$, then there exists $h_\rho \in (0, h_0]$ such that $\Pi_h \bar u \in \bar B_\rho(\bar u) \cap U_{{\rm ad},h}$ for every $h \le h_\rho$. Hence, $\bar B_\rho(\bar u) \cap U_{{\rm ad},h}$ is a compact non-empty subset in $\mathcal{U}_h$ for every $h \le h_\rho$. Let us prove that J_h is continuous. Let $\{u_{hk}\}_{k=1}^\infty \subset \bar B_\rho(\bar u)$ be a sequence converging to u_h in $\mathcal{U}_h$. Let $\{y_h(u_{hk})\}_{k=1}^\infty \subset \bar B^Y_{\rho^*}(\bar y) \cap Y_h$ be the associated discrete states. From the boundedness of this sequence in Y_h, we deduce the existence of a subsequence, denoted in the same way, such that $y_h(u_{hk}) \stackrel{k\to\infty}{\longrightarrow} y_h$ for some $y_h \in \bar B^Y_{\rho^*}(\bar y) \cap Y_h$. Now, it is immediate to pass to the limit in the Eq. (4.1) satisfied by (y_{hk}, u_{hk}) and to deduce that $y_h = y_h(u_h)$. Since every subsequence of $\{y_h(u_{hk})\}_{k=1}^\infty$

converges to the same limit $y_h(u_h)$, it follows that the whole sequence converges to $y_h(u_h)$. This proves the continuity of G_h and, consequently, the continuity of J_h. Therefore, (P_h^ρ) consists of the minimization of a continuous function on a non-empty compact set, which implies the existence of a solution $\bar{u}_h$.

It remains to prove that $\{\bar{u}_h\}_{h\le h_\rho}$ converges to $\bar{u}$ strongly in $L^2(\Omega)$. First, from the boundedness of $\{\bar{u}_h\}_{h\le h_\rho} \subset \bar{B}_\rho(\bar{u})$ and the inclusions $U_{\mathrm{ad},h} \subset U_{\mathrm{ad}}$ we deduce the existence of a subsequence, denoted in the same way, and an element $\tilde{u} \in \bar{B}_\rho(\bar{u}) \cap U_{\mathrm{ad}}$ such that $\bar{u}_h \rightharpoonup \tilde{u}$ in $L^2(\Omega)$. This implies that $y_{\bar{u}_h} \to y_{\tilde{u}}$ strongly in Y; see [12, Theorem 2.9]. Therefore, using (3.22) and (3.23) we infer

$$\|y_h(\bar{u}_h) - y_{\tilde{u}}\|_Y \le \|y_h(\bar{u}_h) - y_{\bar{u}_h}\|_Y + \|y_{\bar{u}_h} - y_{\tilde{u}}\|_Y \longrightarrow 0.$$

This convergence and the optimality of $\bar{u}_h$ imply

$$J(\tilde{u}) \le \liminf_{h\to 0} J_h(\bar{u}_h) \le \limsup_{h\to 0} J_h(\bar{u}_h) \le \limsup_{h\to 0} J_h(\Pi_h\bar{u}) = J(\bar{u}). \tag{4.8}$$

This inequality and (4.7) lead to the identity $\bar{u} = \tilde{u}$. Moreover, (4.8) implies that $\bar{u}_h \to \bar{u}$ strongly in $L^2(\Omega)$. This property is satisfied by every weakly convergent subsequence of $\{\bar{u}_h\}_{h\le h_\rho}$, hence the whole sequence converges strongly to $\bar{u}$. □

Remark 4.3 By selecting h_ρ sufficiently small, we have that the solutions $\bar{u}_h$ of (P_h^ρ) belong to the open ball $B_\rho(\bar{u})$. Indeed, this is an obvious consequence of the strong convergence $\bar{u}_h \to \bar{u}$ in $L^2(\Omega)$. From now on, we will assume that h_ρ has been chosen so that $\bar{u}_h$ is included in the open ball. From Theorem 3.8 we deduce that $(y_h(\bar{u}_h), \bar{u}_h)$ is a local solution of $(\mathcal{P}_h)$. Thus, Theorem 4.2 proves that strict local solutions of (P) can be approximated by local solutions of $(\mathcal{P}_h)$.

The next goal is to derive the optimality conditions satisfied by a solution of (P_h^ρ). To this end, we firstly analyze the differentiability of the mapping G_h and the functional J_h.

Theorem 4.4 *Suppose that Assumptions 1 and 3 hold. Then, there exists $\bar{h}_\rho \le h_0$ such that for every $h < \bar{h}_\rho$ the mapping $G_h : B_\rho(\bar{u}) \longrightarrow Y_h$ is of class C^2. Moreover, if $z_h = G_h'(u)v$ for $u \in B_\rho(\bar{u})$ and $v \in L^2(\Omega)$, then z_h is the unique solution of the variational problem*

$$\begin{cases} \textit{Find } z_h \in Y_h \textit{ such that} \\ a(z_h, w_h) + \displaystyle\int_\Omega \frac{\partial f}{\partial y}(x, y_h(u))z_h(x)w_h(x)\,dx = \int_\Omega v(x)w_h(x)\,dx \ \ \forall w_h \in Y_h. \end{cases} \tag{4.9}$$

Proof For every $h < h_0$ let $\mathcal{F}_h : Y_h \times B_\rho(\bar{u}) \longrightarrow Y_h^*$ be the mapping defined by

$$\langle \mathcal{F}_h(y_h, u), w_h\rangle_{Y_h^*,Y_h} = a(y_h, w_h) + \int_\Omega [f(x, y_h) - u]w_h\,dx \quad \forall w_h \in Y_h.$$

It is clear that $\mathcal{F}_h$ is of class C^2 and $\mathcal{F}_h(G_h(u), u)) = 0\ \forall u \in B_\rho(\bar{u})$. Hence, the differentiability of G_h is a consequence of the implicit function theorem applied to $\mathcal{F}_h$. We only need to prove that

$$\frac{\partial \mathcal{F}_h}{\partial y}(y_h(u), u) : Y_h \longrightarrow Y_h^*$$
$$\left\langle \frac{\partial \mathcal{F}_h}{\partial y}(y_h(u), u) z_h, w_h \right\rangle_{Y_h^*, Y_h} = a(z_h, w_h) + \int_\Omega \frac{\partial f}{\partial y}(x, y_h(u)) z_h w_h \, dx$$

is an isomorphism. This is equivalent to prove that (4.9) has a unique solution $z_h \in Y_h$ for every $u \in L^2(\Omega)$. We prove this. First we observe that $y_h(u) \in \bar{B}^Y_{\rho^*}(\bar{y})\ \forall u \in B_\rho(\bar{u})$. Therefore, $\|y_h(u)\|_{C(\bar{\Omega})} \le \rho^*\ \forall u \in B_\rho(\bar{u})$ holds. From (2.6) we infer the existence of a constant C_{f,ρ^*} such that

$$\left|\frac{\partial f}{\partial y}(x, y_h(u))\right| \le C_{f,\rho^*} \quad \forall u \in B_\rho(\bar{u}).$$

Then, Lemma 3.1 implies the existence of $\bar{h}_\rho \le h_0$ depending on C_{f,ρ^*} and $\mathcal{A}$ such that (4.9) has a unique solution for every $h \le \bar{h}_\rho$ and for all $(u, v) \in B_\rho(\bar{u}) \times L^2(\Omega)$. □

As an immediate corollary of the above theorem we get the differentiability of the objective functional J_h.

Corollary 4.5 *Under the assumptions of Theorem 4.4, the mapping $J_h : B_\rho(\bar{u}) \longrightarrow \mathbb{R}$ is of class $C^2\ \forall h < \bar{h}_\rho$, and*

$$J_h'(u)v = \int_\Omega (\varphi_h(u) + \nu u) v \, dx \quad \forall (u, v) \in B_\rho(\bar{u}) \times L^2(\Omega), \tag{4.10}$$

where $\varphi_h(u) \in Y_h$ is the adjoint state, i.e. it is the solution of the variational problem

$$\begin{cases} \textit{Find } \varphi_h \in Y_h \textit{ such that} \\ a(w_h, \varphi_h) + \displaystyle\int_\Omega \frac{\partial f}{\partial y}(x, y_h(u)) \varphi_h w_h \, dx = \int_\Omega (y_h(u) - y_d) w_h \, dx \quad \forall w_h \in Y_h. \end{cases} \tag{4.11}$$

Let us observe that (4.11) is a linear system of equations, adjoint to the one defined by (4.9). Therefore, the existence and uniqueness of a solution of (4.11) is a consequence of the same property for (4.9).

Now, we can formulate the first order optimality conditions satisfied by a solution of (P_h^ρ).

Theorem 4.6 *Assume that $h < \bar{h} = \min\{h_\rho, \bar{h}_\rho\}$ with h_ρ and $\bar{h}_\rho$ given by Theorems 4.2 and 4.4. Then, (P_h^ρ) has a solution $\bar{u}_h$ in the open ball $B_\rho(\bar{u})$ for every $h < \bar{h}$.*

Moreover, for any of these solutions there exist two unique functions $\bar{y}_h, \bar{\varphi}_h \in Y_h$ *satisfying*

$$a(\bar{y}_h, w_h) + \int_\Omega f(x, \bar{y}_h) w_h \, dx = \int_\Omega \bar{u}_h w_h \, dx \quad \forall w_h \in Y_h, \tag{4.12}$$

$$a(w_h, \bar{\varphi}_h) + \int_\Omega \frac{\partial f}{\partial y}(x, \bar{y}_h) \bar{\varphi}_h w_h \, dx = \int_\Omega (\bar{y}_h - y_d) w_h \, dx \quad \forall w_h \in Y_h, \tag{4.13}$$

$$\int_\Omega (\bar{\varphi}_h + \nu \bar{u}_h)(u_h - \bar{u}_h) \, dx \geq 0 \quad \forall u_h \in U_{\mathrm{ad},h}. \tag{4.14}$$

Proof The existence of a solution of (P_h^ρ) in the open ball follows from Remark 4.3. Then, the inequality $J_h'(\bar{u}_h)(u_h - \bar{u}_h) \geq 0 \; \forall u_h \in U_{\mathrm{ad},h}$ holds for every $h < \bar{h}$. This along with (4.10) leads straightforward to (4.12)–(4.14). □

5 Error estimates

In this section, we suppose that Assumptions 1 and 3 hold. In the whole section $\bar{u}$ will denote a strict local solution of (P) with associated state $\bar{y}$ and adjoint state $\bar{\varphi}$. Following Theorem 4.6, in the sequel we will assume that $h < \bar{h}$, and we consider the discrete problems (P_h^ρ) having solutions $\bar{u}_h \in B_\rho(\bar{u}) \cap U_{\mathrm{ad},h}$ and satisfying the optimality conditions (4.12)–(4.14). We know that $\bar{u}_h \to \bar{u}$ strongly in $L^2(\Omega)$. The goal is to provide some error estimates for the difference $\bar{u} - \bar{u}_h$. We will distinguish two cases depending on the set U_{ad}. Firstly we analyze the case where $U_{\mathrm{ad}} \subsetneq L^2(\Omega)$, next we treat the case where $U_{\mathrm{ad}} = L^2(\Omega)$. Let us prove a preliminary result that we will use later.

Theorem 5.1 *Let* $u \in B_\rho(\bar{u})$ *be arbitrary. Let* $\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$ *and* $\varphi_h \in Y_h$ *denote the solutions of* (2.11) *and* (4.11)*, respectively. Then, there exist constants* k_2 *and* k_∞ *such that*

$$\|\varphi - \varphi_h\|_{L^2(\Omega)} + h \|\varphi - \varphi_h\|_{H_0^1(\Omega)} \leq k_2 h^2 \quad \forall u \in B_\rho(\bar{u}). \tag{5.1}$$

$$\|\varphi - \varphi_h\|_{L^\infty(\Omega)} \leq k_\infty h^{2 - \frac{n}{2}} \quad \forall u \in B_\rho(\bar{u}). \tag{5.2}$$

Proof First, we recall that $\|y_h(u)\|_{C(\bar{\Omega})} \leq \rho^* \; \forall u \in B_\rho(\bar{u})$. Hence, with (2.6) we get

$$\left|\frac{\partial f}{\partial y}(x, y_h(u))\right| + \left|\frac{\partial^2 f}{\partial y^2}(x, y_h(u))\right| \leq C_{f,\rho^*} \quad \forall u \in B_\rho(\bar{u}). \tag{5.3}$$

Now we introduce the function $\varphi^h \in H^2(\Omega) \cap H_0^1(\Omega)$, the unique solution of

$$\begin{cases} A^* \varphi^h - \mathrm{div}[b(x) \varphi^h] + \dfrac{\partial f}{\partial y}(x, y_h(u)) \varphi^h = y_h(u) - y_d \text{ in } \Omega, \\ \varphi = 0 \text{ on } \Gamma. \end{cases} \tag{5.4}$$

From Lemma 3.2 we deduce the existence of a constant C_1 such that $\|\varphi^h\|_{H^2(\Omega)} \leq C_1$ $\forall u \in B_\rho(\bar{u})$. From the continuous embedding $H^2(\Omega) \subset C(\bar{\Omega})$ we get that $\|\varphi^h\|_{C(\bar{\Omega})} \leq C_2$ $\forall u \in B_\rho(\bar{u})$. On the other side, from (3.18) we obtain for some constant K_2

$$\|y_u - y_h(u)\|_{L^2(\Omega)} \leq K_2 h^2 \quad \forall u \in B_\rho(\bar{u}). \tag{5.5}$$

Now, we write $\varphi - \varphi_h = (\varphi - \varphi^h) + (\varphi^h - \varphi_h(u))$ and we estimate both summands. For the first summand we subtract the equations (2.11) and (5.4):

$$\begin{aligned} &A^*(\varphi - \varphi^h) - \operatorname{div}(b(x)(\varphi - \varphi^h)) + \frac{\partial f}{\partial y}(x, y_u)(\varphi - \varphi^h) \\ &\quad = (y_u - y_h(u)) + \left[\frac{\partial f}{\partial y}(x, y_h(u)) - \frac{\partial f}{\partial y}(x, y_u)\right]\varphi^h. \end{aligned}$$

Using the mean value theorem, we have that there exists a measurable function $0 < \theta(x) < 1$ such that, if we name $\hat{y} = y_h(u) + \theta(y_u - y_h(u))$, then

$$\frac{\partial f}{\partial y}(x, y_h(u)) - \frac{\partial f}{\partial y}(x, y_u) = \frac{\partial^2 f}{\partial y^2}(x, \hat{y})(y_h(u) - y_u).$$

Using Lemma 3.2, (5.3), the boundedness of $\{\varphi^h\}_h$ in $C(\bar{\Omega})$, and (5.5) we infer

$$\begin{aligned} &\|\varphi - \varphi^h\|_{H^2(\Omega)} \\ &\quad \leq C_3\Big(\|y_u - y_h(u)\|_{L^2(\Omega)} + \Big\|\frac{\partial f}{\partial y}(x, y_h(u)) - \frac{\partial f}{\partial y}(x, y_u)\Big\|_{L^2(\Omega)}\|\varphi^h\|_{C(\bar{\Omega})}\Big) \\ &\quad \leq C_3\Big(1 + \Big\|\frac{\partial^2 f}{\partial y^2}(x, \hat{y})\Big\|_{L^\infty(\Omega)}\|\varphi^h\|_{C(\bar{\Omega})}\Big)\|y_u - y_h(u)\|_{L^2(\Omega)} \\ &\quad \leq C_3(1 + C_{f,\rho^*}C_2)\|y_u - y_h(u)\|_{L^2(\Omega)} \leq C_3(1 + C_{f,\rho^*}C_2)Kh^2 \;\; \forall u \in B_\rho(\bar{u}). \end{aligned} \tag{5.6}$$

To estimate the term $\varphi^h - \varphi_h(u)$ we observe that the equation satified by $\varphi_h(u)$ is the corresponding discretization of the equation satisfied by φ^h. Both equations are linear. Hence, we can use [32] to deduce that

$$\|\varphi^h - \varphi_h(u)\|_{L^2(\Omega)} + h\|\varphi^h - \varphi_h(u)\|_{H_0^1(\Omega)} \leq C_4 h^2 \quad \forall u \in B_\rho(\bar{u}).$$

Using the estimate in $L^2(\Omega)$, interpolation error estimates, and an inverse inequality we obtain

$$\|\varphi^h - \varphi_h(u)\|_{L^\infty(\Omega)} \leq C_5 h^{2-\frac{n}{2}} \quad \forall u \in B_\rho(\bar{u}). \tag{5.7}$$

The reader can also consider to apply Theorem 3.6 with a function f that is linear and change the equation by its adjoint. Finally, (5.1) and (5.2) follow from the above estimates and (5.6). □

Error estimates can be deduced from the abstract error estimate of [17, Theorem 2.14].

Lemma 5.2 *Let $\bar{u}$ be a local minimizer of (P) with associated state $\bar{y}$ and satisfying (2.16). Let $\{(\bar{y}_h, \bar{u}_h)\}$ be a sequence of local minimizers of the problems $(\mathcal{P}_h)$ converging strongly to $(\bar{y}, \bar{u})$ in $H_0^1(\Omega) \times L^2(\Omega)$ (see Theorem 4.2 and Remark 4.3). Then, there exists $h_0 > 0$ such that for every $h < h_0$*

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)} \le C\big[h^4 + \|\bar{u} - u_h\|^2_{L^2(\Omega)} + J'(\bar{u})(u_h - \bar{u})\big]^{1/2} \; \forall u_h \in U_{\text{ad},h}.$$

Proof We use [17, Theorem 2.14]. To this end, we have to check the Assumptions (A2), (A3) and (A7) of [17]. First we observe that there exist positive constants r, M_1 and M_2 such that for all $v, v_1, v_2 \in L^2(\Omega)$ and all $u \in U_{\text{ad}}$ such that $\|\bar{u}-u\|_{L^2(\Omega)} < r$

$$|J'(u)v| \le M_1\|v\|_{L^2(\Omega)}, \; |J''(u)(v_1, v_2)| \le M_2\|v_1\|_{L^2(\Omega)}\|v_2\|_{L^2(\Omega)}.$$

Moreover, for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $u_1, u_2 \in L^\infty(\Omega)$ with $\|u_i - \bar{u}\|_{L^\infty(\Omega)} < r$, $i = 1, 2$, and $\forall v \in L^2(\Omega)$

$$\|u_1 - u_1\|_{L^\infty(\Omega)} < \delta \Rightarrow \begin{cases} |(J'(u_1) - J'(u_2))v| \le \varepsilon\|v\|_{L^2(\Omega)} \\ |(J''(u_1) - J''(u_2))v^2| \le \varepsilon\|v\|^2_{L^2(\Omega)}. \end{cases}$$

Hence, (A2) holds. Assumption (A3) says that for any element $u \in U_{\text{ad}}$ there exists a family $\{u_h\}_{h>0}$ with $u_h \in U_{\text{ad},h}$ such that $\|u - u_h\|_{L^2(\Omega)} \to 0$ when $h \to 0$, which is well known to be satisfied for our choices of $U_{\text{ad},h}$. Finally, estimate (5.1) implies

$$\begin{aligned} |(J_h'(u) - J'(u))(u_h - \bar{u})| &= \int_\Omega (\varphi_h(u) - \varphi_u)(u_h - \bar{u})dx \\ &\le \|\varphi_h(u) - \varphi_u\|_{L^2(\Omega)}\|u_h - \bar{u}\|_{L^2(\Omega)} \\ &\le k_2h^2\|u_h - \bar{u}\|_{L^2(\Omega)}. \end{aligned}$$

Therefore, Assumption (A7) holds with $\varepsilon_h = h^2$. Then, [17, Theorem 2.14] claims the existence of a constant C independent of h such that

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)} \le C\big[h^4 + \|\bar{u} - u_h\|^2_{L^2(\Omega)} + J'(\bar{u})(u_h - \bar{u})\big]^{1/2} \; \forall u_h \in U_{\text{ad},h} \; \forall h < h_0,$$

and the result follows. □

Next, we obtain error estimates for unconstrained problems.

Theorem 5.3 *Suppose $U_{\rm ad} = L^2(\Omega)$ and set $\mathcal{U}_h = \mathcal{U}_h^i$, $i = 0, 1$. Let $\bar u$ be a local minimizer of (P) with associated state $\bar y$ and satisfying* (2.16). *Let $\{(\bar y_h, \bar u_h)\}$ be a sequence of local minimizers of the problems $(\mathcal{P}_h)$ converging strongly to $(\bar y, \bar u)$ in $H_0^1(\Omega) \times L^2(\Omega)$. Then there exists $h_0 > 0$ such that*

$$\|\bar u - \bar u_h\|_{L^2(\Omega)} \le Ch^{1+i} \quad \forall h < h_0.$$

Proof We apply Lemma 5.2. In this case $J'(\bar u) = 0$. For $i = 0$ we take $u_h = \Pi_h \bar u$ and for $i = 1$, we take $u_h = I_h \bar u$, the nodal interpolation of $\bar u$ and the result follows from the approximation properties of the projection in the $L^2(\Omega)$ sense and the nodal interpolation respectively. □

In the following result, we obtain error estimates for constrained problems.

Theorem 5.4 *Suppose $-\infty < \alpha$ or $\beta < \infty$ and set $\mathcal{U}_h = \mathcal{U}_h^i$, $i = 0, 1$. Let $\bar u$ be a local minimizer of (P) with associated state $\bar y$ and satisfying* (2.16). *Let $\{(\bar y_h, \bar u_h)\}$ be a sequence of local minimizers of the problems $(\mathcal{P}_h)$ converging strongly to $(\bar y, \bar u)$ in $H_0^1(\Omega) \times L^2(\Omega)$. Then there exists $h_0 > 0$ such that*

$$\|\bar u - \bar u_h\|_{L^2(\Omega)} \le Ch \quad \forall h < h_0.$$

Proof We apply again Lemma 5.2 with $u_h = \Pi_h \bar u \in U_{{\rm ad},h}$, where we recall that Π_h is either the linear projection in the $L^2(\Omega)$ sense onto $\mathcal{U}_h^0$ or Carstensen's quasi-interpolation operator, depending on the approximation space for the controls. In both cases we have that $\|\bar u - u_h\|_{L^2(\Omega)} \le Ch$; see [21, Lemma 4.3] for Carstenen's quasi-interpolation operator. For the last term we have

$$\begin{aligned} J'(\bar u)(u_h - \bar u) &= \int_\Omega (\bar\varphi + \nu \bar u)(u_h - \bar u)dx \le C\|\bar\varphi + \nu\bar u\|_{H_0^1(\Omega)}\|u_h - \bar u\|_{H^{-1}(\Omega)} \\ &\le Ch^2, \end{aligned}$$

where the estimate $\|u_h - \bar u\|_{H^{-1}(\Omega)} \le Ch^2$ follows by duality for the $L^2(\Omega)$ projection and is proved in [21, Lemma 4.4] for Carstenen's quasi-interpolation operator. □

Finally, we deduce error estimates in the norm of $L^\infty(\Omega)$. We start with a result for the adjoint state.

Corollary 5.5 *Let us suppose that the assumptions of Theorem 5.3 or Theorem 5.4 are satisfied and let $\bar\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$ and $\bar\varphi_h \in Y_h$ be the solutions of* (2.13) *and* (4.13). *Then there exists $h_0 > 0$ and a constant C independent of h such that*

$$\|\bar\varphi - \bar\varphi_h\|_{L^\infty(\Omega)} \le Ch^{2-n/2} \quad \forall h < h_0.$$

Proof By the triangle inequality

$$\|\bar\varphi - \bar\varphi_h\|_{L^\infty(\Omega)} \le \|\bar\varphi - \varphi_{\bar u_h}\|_{L^\infty(\Omega)} + \|\varphi_{\bar u_h} - \bar\varphi_h\|_{L^\infty(\Omega)}.$$

Using either Theorem 5.3 or Theorem 5.4, we have that there exists some $h_0 > 0$ such that $\bar{u}_h \in \bar{B}_\rho(\bar{u})$ for all $h < h_0$. Therefore, we can use (5.2) to obtain

$$\|\varphi_{\bar{u}_h} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \leq k_\infty h^{2-n/2}$$

Using the same technique as in the proof of Theorem 5.1 and the Sobolev embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$, which is valid for $n \leq 3$, we have that

$$\|\bar{\varphi} - \varphi_{\bar{u}_h}\|_{L^\infty(\Omega)} \leq C_3(1 + C_{f,\rho^*} C_2)\|\bar{y} - y_{\bar{u}_h}\|_{L^2(\Omega)}.$$

Next, we use [12, Lemma 3.5] and either the estimate proved in Theorem 5.4 or the ones proved in Theorem 5.3, depending on wether we have control constraints or not. Since $\bar{u}_h \in \bar{B}_\rho(\bar{u})$ for all $h < h_0$, we know that there is a constant $M_{\bar{B}_\rho(\bar{u})}$ such that

$$\|\bar{y} - y_{\bar{u}_h}\|_{L^2(\Omega)} \leq M_{\bar{B}_\rho(\bar{u})}\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)} \leq Ch.$$

The result follows from the previous estimates, just taking into account that $2 - n/2 \leq 1$. □

To deduce error estimates for the control variable in $L^\infty(\Omega)$, we replace Assumption (2.2) and the assumption on the target y_d by the following one, which is not very restrictive in practice:

$$b \in L^{\bar{p}}(\Omega) \text{ with } \bar{p} > n,\ \operatorname{div} b,\ y_d \in L^q(\Omega) \text{ with } q > 2, \tag{5.8}$$

Using that Ω is convex, we know that there exists some $2 < p \leq \min\{\bar{p}, q\}$ such that $\bar{\varphi} \in W^{2,p}(\Omega)$; see, e.g., [25] for $n = 2$ and [19, Corollary 3.12] for $n = 3$.

Corollary 5.6 *Let $\bar{u}$ be a local minimizer of (P) with associated state $\bar{y}$ and satisfying (2.16). Let $\{(\bar{y}_h, \bar{u}_h)\}$ be a sequence of local minimizers of the problems $(\mathcal{P}_h)$ converging strongly to $(\bar{y}, \bar{u})$ in $H_0^1(\Omega) \times L^2(\Omega)$. Suppose further that one of the following conditions is satisfied:*

1. *$\mathcal{U}_h = \mathcal{U}_h^1$ and $U_{\text{ad}} = L^2(\Omega)$;*
2. *$\mathcal{U}_h = \mathcal{U}_h^0$ and (5.8) holds;*
3. *$\mathcal{U}_h = \mathcal{U}_h^1$, $n = 2$ and (5.8) holds.*

Then there exists $h_0 > 0$ and a constant C independent of h such that

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} \leq Ch^{2-n/2} \quad \forall h < h_0.$$

Proof **Case 1**—*$\mathcal{U}_h = \mathcal{U}_h^1$ and $U_{\text{ad}} = L^2(\Omega)$*. The optimality conditions (2.14) and (4.14) and Corollary 5.5 lead straightforward to

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} = \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \leq Ch^{2-n/2} \quad \forall h < h_0.$$

Case 2—*$\mathcal{U}_h = \mathcal{U}_h^0$ and* (5.8) *holds*. In this case (2.14) and (4.14) lead to

$$\bar{u}(x) = \text{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}(x)\right), \ \bar{u}_T = \text{Proj}_{[\alpha,\beta]}\left(-\frac{1}{|T|\nu}\int_T \bar{\varphi}_h(x)dx\right) \ \forall T \in \mathcal{T}_h,$$

where $\text{Proj}_{[\alpha,\beta]}(s) = \max(\alpha, \min(\beta, s))$ and $\bar{u}_T$ is the constant value of $\bar{u}_h$ at the triangle T. From the mean value theorem, for every element $T \in \mathcal{T}_h$, we deduce the existence of some $x_T \in T$ such that

$$\int_T \bar{\varphi}_h(x)dx = |T|\bar{\varphi}_h(x_T).$$

Since $\text{Proj}_{[\alpha,\beta]}(s)$ is a contraction, we have that for every $T \in \mathcal{T}_h$ and almost every $x \in T$,

$$|\bar{u}(x) - \bar{u}_h(x)| = |\bar{u}(x) - u_T| \le \frac{1}{\nu}|\bar{\varphi}(x) - \bar{\varphi}_h(x_T)|. \tag{5.9}$$

Since $\bar{\varphi} \in W^{2,p}(\Omega)$ for some $p > 2$, by the Sobolev imbedding theorem, also $\bar{\varphi} \in C^{0,\delta}(\bar{\Omega})$ for $\delta = 1$ if $n = 2$ and some $1/2 < \delta \le 1$ depending on p if $n = 3$. Therefore, there exists a constant $\Lambda_{\bar{\varphi}} > 0$ such that

$$\begin{aligned} |\bar{\varphi}(x) - \bar{\varphi}_h(x_T)| &\le |\bar{\varphi}(x) - \bar{\varphi}(x_T)| + |\bar{\varphi}(x_T) - \bar{\varphi}_h(x_T)| \\ &\le \Lambda_{\bar{\varphi}} h^\delta + \|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)}. \end{aligned} \tag{5.10}$$

From (5.9), (5.10), Corollary 5.5 and the fact that $2 - n/2 \le \delta$, we have that

$$|\bar{u}(x) - \bar{u}_h(x)| \le Ch^{2-n/2} \ \text{ for a.e. } x \in \Omega,$$

and the result follows.

Case 3—*$\mathcal{U}_h = \mathcal{U}_h^1$, $n = 2$ and* (5.8) *holds*. If there are no control constraints, we are in the situation of Case 1, so we assume that $-\infty < \alpha$ or $\beta < +\infty$. In this case, (4.14) implies that $\bar{u}_h$ is the projection in the $L^2(\Omega)$-sense of $-\frac{1}{\nu}\bar{\varphi}_h$ onto $U_{\text{ad},h}$, but we do not have a pointwise projection formula.

The estimate follows from the results of [28, Sections 3,4]. Notice that, although that reference is about linear equations, the proof only requires $L^2(\Omega)$-error estimates for the control, which we have in Theorem 5.4, $L^\infty(\Omega)$-error estimates and Lipschitz regularity for the adjoint state, which we have from Corollary 5.5 and assumption (5.8) and the fact that the discrete optimal control is a projection in the $L^2(\Omega)$-sense of $-\frac{1}{\nu}\bar{\varphi}_h$. Notice also that the technique of proof cannot be translated to $n = 3$, since the analogous of [28, Lemma 3.5] for $n = 3$ does not hold. □

Remark 5.7 Under additional regularity conditions, higher orders of convergence can be proved. Indeed, let us suppose that $\varphi_u \in W^{2,p}(\Omega)$ for some $p > n$ if $u \in L^\infty(\Omega)$. For $n = 2$, condition (5.8) is sufficient for this regularity, while for $n = 3$ we have to assume that b, $\text{div}\, b$, $y_d \in L^{\bar{p}}(\Omega)$ for some $\bar{p} > 3$ and also that the internal angles of

Ω are small enough; see [19]. Using the same technique as in the proof of Theorem 5.1 together with [33, Theorem 2.1], [6, Theorem 2.2], and the fact that $\{\bar{u}_h\}$ is bounded in $L^\infty(\Omega)$, we obtain

$$\|\varphi_{\bar{u}_h} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \le Ch^{2-\frac{n}{p}} |\log h| \|\varphi_{\bar{u}_h}\|_{W^{2,p}(\Omega)} \le Cp|\log h| h^{2-\frac{n}{p}}.$$

From this estimate, it can proved as in Corollaries 5.5 and 5.6, that

$$\begin{aligned} \|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} &= \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \le Cp|\log h| h^{2-\frac{n}{p}} && \text{if } \mathcal{U}_h = \mathcal{U}_h^1 \text{ and } U_{\text{ad}} = L^2(\Omega) \\ \|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} &\le \Lambda_{\bar{\varphi}} h + \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \le Ch && \text{if } \mathcal{U}_h = \mathcal{U}_h^0, \end{aligned}$$

where $\Lambda_{\bar{\varphi}}$ is the Lipschitz constant of $\bar{\varphi}$.

Assume that we have that $\varphi_u \in W^{2,p}(\Omega)$ for all $p < +\infty$ if $u \in L^\infty(\Omega)$. If further $\mathcal{U}_h = \mathcal{U}_h^1$ and $U_{\text{ad}} = L^2(\Omega)$, then we obtain by setting $p = |\log h|$ in the above inequality

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} = \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \le C|\log h|^2 h^2.$$

See [11, Lemma 3] for the proof of a similar result. This high regularity can be achieved, for instance, if b, $\operatorname{div} b$, $y_d \in L^{\bar{p}}(\Omega)$ for all $\bar{p} < +\infty$ and Ω is a rectangle or a rectangular parallelepiped or its boundary Γ is of class $C^{1,1}$.

Also, when $\mathcal{U}_h = \mathcal{U}_h^1$ and $U_{\text{ad}} \subsetneq L^2(\Omega)$, the order of convergence usually observed in experiments for the $L^2(\Omega)$-error of the control is $O(h^{3/2})$. A detailed explanation of this phenomenon can be found in [10, Section 10]. In our case, this order is achieved if $p > n$. The proof is based on the assumption that the measure of set $\cup\{T \in \mathcal{T}_h : \bar{u} \notin H^2(T)\}$ is of order h. This assumption is not restrictive and is usually satisfied in practice; see [27].

6 Numerical experiments

We are going to build an example with an explicitly known local solution satisfying the second order sufficient optimality condition (2.16).

Let us consider $\Omega = (0, 1)^n$, $Ay = -\Delta y$, $f(x, y) = \exp(y)$, $\nu = 1$ and $b(x) = (B(x_1), 0)$ if $n = 2$ or $b(x) = (B(x_1), 0, 0)$ if $n = 3$, where $B(x) = 5x^{3/4}(1 - 2x)$. With these choices, Assumptions 1, 2 and 3 are satisfied, but notice that Assumption 2' does not hold. The lower control constraint will be $\alpha = -\infty$ and we will investigate both the upper unconstrained case $\beta = \infty$ and the constrained case $\beta = 2^{-2n-1}$. To define the target state y_d, we first define $\bar{\varphi}(x) = -\Pi_{i=1}^n x_i(1 - x_i)$ and $\bar{u}(x) = \operatorname{proj}_{[\alpha,\beta]}(-\bar{\varphi}(x)/\nu)$. Next, we take $\bar{y} \in H^2(\Omega) \cap H_0^1(\Omega)$ solution of the state equation and set $y_d(x) = \Delta\bar{\varphi}(x) + \operatorname{div}(b(x)\bar{\varphi}(x)) + \bar{y}(x) - \frac{\partial f}{\partial y}(x, \bar{y}(x))\bar{\varphi}(x)$. (In practice, we do not have $\bar{y}$, but we can use $y_h(\bar{u})$ to compute a good approximation of y_d).

Table 1 Experimental order of convergence for the control error

n	Piecewise constant controls				Continuous piecewise linear controls			
	Unconstrained		Constrained		Unconstrained		constrained	
	L^2	L^∞	L^2	L^∞	L^2	L^∞	L^2	L^∞
2	1.0	1.0	1.0	1.0	2.0	1.9	1.5	1.0
3	1.0	1.0	1.0	1.0	2.0	2.0	1.8	1.1

Table 2 Experimental order of convergence for the state error

n	Piecewise constant controls				Continuous piecewise linear controls			
	Unconstrained		Constrained		Unconstrained		Constrained	
	L^2	L^∞	L^2	L^∞	L^2	L^∞	L^2	L^∞
2	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.9
3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.1

With these choices, it is clear that $(\bar u, \bar y, \bar\varphi)$ satisfies first order optimality conditions (2.12)–(2.14). From (2.10), we have

$$J''(\bar u)v^2 = \int_\Omega \left(1 - \bar\varphi(x)e^{\bar y(x)}\right) z_v^2(x)dx + \int_\Omega v^2(x)dx \ \forall v \in L^2(\Omega).$$

Since $\bar\varphi(x) < 0$ for all $x \in \Omega$, the condition (2.16) holds and hence $\bar u$ is a local solution of (P).

The problem is discretized using the finite element method. To solve the discrete problems, we use a semi-smooth Newton method as described in [10, Section 14]. The success of the conjugate gradient method used to solve the unconstrained quadratic programs arising in the optimization process is an indication that the solutions of the finite dimensional problems are strict local minima.

The mesh of size $h_i = 2^{-i}$ is obtained splitting Ω into 2^{in} congruent cells obtained by translation of $(0, h_i)^n$ and dividing each cell into $n!$ $n-$simplices. In this family of meshes, the experimental order of convergence for the error of the variable $z \in \{u, y, \varphi\}$ measured in the norm of $X = L^2(\Omega)$ or $L^\infty(\Omega)$ can be computed as

$$EOC_i = \log_2(\|\bar z_{h_{i-1}} - \bar z\|_X) - \log_2(\|\bar z_{h_i} - \bar z\|_X).$$

We report on the $L^2(\Omega)$ and $L^\infty(\Omega)$ experimental order of convergence of the error for the control, the state, and the adjoint state for $i = 8$ if $n = 2$, and $i = 5$ if $n = 3$. We summarize the results Tables 1, 2 and 3.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence,

Table 3 Experimental order of convergence for the adjoint state error

n	Piecewise constant controls				Continuous piecewise linear controls			
	Unconstrained		Constrained		Unconstrained		Constrained	
	L^2	L^∞	L^2	L^∞	L^2	L^∞	L^2	L^∞
2	2.0	1.9	2.0	1.9	2.0	1.9	2.0	1.9
3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces, Pure and Applied Mathematics (Amsterdam), vol. 140, 2nd edn. Elsevier, Amsterdam (2003)
2. Ali, A.A., Deckelnick, K., Hinze, M.: Global minima for semilinear optimal control problems. Comput. Optim. Appl. **65**(1), 261–288 (2016). https://doi.org/10.1007/s10589-016-9833-1
3. Arada, N., Casas, E., Tröltzsch, F.: Error estimates for the numerical approximation of a semilinear elliptic control problem. Comput. Optim. Appl. **23**(2), 201–229 (2002). https://doi.org/10.1023/A:1020576801966
4. Bogachev, V.I.: Measure Theory, vol. I, II. Springer, Berlin (2007). https://doi.org/10.1007/978-3-540-34514-5
5. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods, Texts in Applied Mathematics, vol. 15, 3rd edn. Springer, New York (2008)
6. Cameron, A.W.: Estimates for solutions of elliptic partial differential equations with explicit constants and aspects of the finite element method for second-order equations. Ph.D. thesis, Cornell University. https://hdl.handle.net/1813/17599 (2010)
7. Carstensen, C.: Quasi-interpolation and a posteriori error analysis in finite element methods. M2AN Math. Model. Numer. Anal. **33**(6), 1187–1202 (1999). https://doi.org/10.1051/m2an:1999140
8. Casas, E., Dhamo, V.: Error estimates for the numerical approximation of a quasilinear Neumann problem under minimal regularity of the data. Numer. Math. **117**(1), 115–145 (2011). https://doi.org/10.1007/s00211-010-0344-1
9. Casas, E., Mateos, M.: Error estimates for the numerical approximation of Neumann control problems. Comput. Optim. Appl. **39**(3), 265–295 (2008). https://doi.org/10.1007/s10589-007-9056-6
10. Casas, E., Mateos, M.: Optimal control of partial differential equations. In: Computational Mathematics, Numerical Analysis and Applications. SEMA SIMAI Springer Ser., vol. 13, pp. 3–59. Springer, Cham (2017)
11. Casas, E., Mateos, M.: State error estimates for the numerical approximation of sparse distributed control problems in the absence of Tikhonov regularization. Vietnam J. Math. **49**, 713–738 (2021). https://doi.org/10.1007/s10013-021-00491-x
12. Casas, E., Mateos, M., Rösch, A.: Analysis of control problems of nonmontone semilinear elliptic equations. ESAIM Control Optim. Calc. Var. **26**, Paper No. 80, 21 (2020). https://doi.org/10.1051/cocv/2020032
13. Casas, E., Mateos, M., Tröltzsch, F.: Error estimates for the numerical approximation of boundary semilinear elliptic control problems. Comput. Optim. Appl. **31**(2), 193–219 (2005). https://doi.org/10.1007/s10589-005-2180-2
14. Casas, E., Raymond, J.P.: Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations. SIAM J. Control Optim. **45**(5), 1586–1611 (2006). https://doi.org/10.1137/050626600

15. Casas, E., Tröltzsch, F.: First- and second-order optimality conditions for a class of optimal control problems with quasilinear elliptic equations. SIAM J. Control Optim. **48**(2), 688–718 (2009). https://doi.org/10.1137/080720048
16. Casas, E., Tröltzsch, F.: Numerical analysis of some optimal control problems governed by a class of quasilinear elliptic equations. ESAIM Control Optim. Calc. Var. **17**(3), 771–800 (2011). https://doi.org/10.1051/cocv/2010025
17. Casas, E., Tröltzsch, F.: A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. Comput. Optim. Appl. **53**(1), 173–206 (2012). https://doi.org/10.1007/s10589-011-9453-8
18. Ciarlet, P.G.: Basic error estimates for elliptic problems. In: Handbook of Numerical Analysis, Handb. Numer. Anal., Vol. II, pp. 17–351. North-Holland, Amsterdam (1991)
19. Dauge, M.: Neumann and mixed problems on curvilinear polyhedra. Integral Equ. Oper. Theory **15**(2), 227–261 (1992). https://doi.org/10.1007/BF01204238
20. de los Reyes, J.C., Dhamo, V.: Error estimates for optimal control problems of a class of quasilinear equations arising in variable viscosity fluid flow. Numer. Math. **132**(4), 691–720 (2016). https://doi.org/10.1007/s00211-015-0737-2
21. de los Reyes, J.C., Meyer, C., Vexler, B.: Finite element error analysis for state-constrained optimal control of the Stokes equations. Control Cybern. **37**(2), 251–284 (2008)
22. Douglas, J., Jr., Dupont, T.: A Galerkin method for a nonlinear Dirichlet problem. Math. Comput. **29**, 689–696 (1975). https://doi.org/10.2307/2005280
23. Falk, R.S.: Approximation of a class of optimal control problems with order of convergence estimates. J. Math. Anal. Appl. **44**, 28–47 (1973). https://doi.org/10.1016/0022-247X(73)90022-X
24. Geveci, T.: On the approximation of the solution of an optimal control problem governed by an elliptic equation. RAIRO Anal. Numér. **13**(4), 313–328 (1979). https://doi.org/10.1051/m2an/1979130403131
25. Grisvard, P.: Elliptic Problems in Nonsmooth Domains, Monographs and Studies in Mathematics, vol. 24. Pitman (Advanced Publishing Program), Boston (1985)
26. Krumbiegel, K., Pfefferer, J.: Superconvergence for Neumann boundary control problems governed by semilinear elliptic equations. Comput. Optim. Appl. **61**, 373–408 (2015)
27. Meyer, C., Rösch, A.: Superconvergence properties of optimal control problems. SIAM J. Control Optim. **43**(3), 970–985 (2004). https://doi.org/10.1137/S0363012903431608
28. Meyer, C., Rösch, A.: L^∞-estimates for approximated optimal control problems. SIAM J. Control Optim. **44**(5), 1636–1649 (2005). https://doi.org/10.1137/040614621
29. Neitzel, I., Pfefferer, J., Rösch, A.: Finite element discretization of state-constrained elliptic optimal control problems with semilinear state equation. SIAM J. Control Optim. **53**(2), 874–904 (2015). https://doi.org/10.1137/140960645
30. Rösch, A., Wachsmuth, D.: Numerical verification of optimality conditions. SIAM J. Control Optim. **47**(5), 2557–2581 (2008). https://doi.org/10.1137/060663714
31. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill Book Co., New York (1987)
32. Schatz, A.H.: An observation concerning Ritz-Galerkin methods with indefinite bilinear forms. Math. Comput. **28**, 959–962 (1974). https://doi.org/10.2307/2005357
33. Schatz, A.H.: Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids. I. Global estimates. Math. Comput. **67**(223), 877–899 (1998). https://doi.org/10.1090/S0025-5718-98-00959-4
34. Stampacchia, G.: Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. Ann. Inst. Fourier (Grenoble) **15**(fasc. 1), 189–258 (1965)
35. Tröltzsch, F.: Optimal control of partial differential equations. In: Graduate Studies in Mathematics, vol. 112. American Mathematical Society, Providence (2010). https://doi.org/10.1090/gsm/112. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.