ESAIM: COCV 20 (2014) 803–822
DOI: 10.1051/cocv/2013084

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

NEW REGULARITY RESULTS AND IMPROVED ERROR ESTIMATES FOR OPTIMAL CONTROL PROBLEMS WITH STATE CONSTRAINTS *

Eduardo Casas[1], Mariano Mateos[2] and Boris Vexler[3]

Abstract. In this paper we are concerned with a distributed optimal control problem governed by an elliptic partial differential equation. State constraints of box type are considered. We show that the Lagrange multiplier associated with the state constraints, which is known to be a measure, is indeed more regular under quite general assumptions. We discretize the problem by continuous piecewise linear finite elements and we are able to prove that, for the case of a linear equation, the order of convergence for the error in $L^2(\Omega)$ of the control variable is $h|\log h|$ in dimensions 2 and 3.

Mathematics Subject Classification. 49K20, 49M05, 49M25, 65N30, 65N15.

Received May 31, 2013. Revised November 25, 2013.
Published online June 5, 2014.

1. Introduction

This paper deals with the following optimal control problem

$$(\mathrm{P}) \qquad \min_{u\in U_{\mathrm{ad}}} J(u),$$

where

$$J(u) = \frac{1}{2}\int_\Omega (y_u(x) - y_d(x))^2\,\mathrm{d}x + \frac{\nu}{2}\int_\Omega u^2(x)\,\mathrm{d}x,$$
$$U_{\mathrm{ad}} = \{u \in L^2(\Omega) : a(x) \le y_u(x) \le b(x) \text{ for all } x \in \Omega\},$$

and y_u is the solution of the Dirichlet problem

$$\begin{cases} Ay = u & \text{in } \Omega, \\ y = 0 & \text{on } \Gamma, \end{cases} \tag{1.1}$$

A being an elliptic operator.

Keywords and phrases. Optimal control, state constraints, elliptic equations, Borel measures, error estimates.

* *The first two authors were partially supported by the Spanish Ministerio de Economía y Competitividad under project MTM2011-22711.*

[1] Departmento de Matemática Aplicada y Ciencias de la Computación, E.T.S.I. Industriales y de Telecomunicación, Universidad de Cantabria, 39005 Santander, Spain. eduardo.casas@unican.es

[2] Departmento de Matemáticas, E.P.I. Gijón, Universidad de Oviedo, Campus de Gijón, 33203 Gijón, Spain. mmateos@uniovi.es

[3] Center for Mathematical Sciences, Technische Universität München, Bolzmannstrasse 3, 85748 Garching b. München, Germany. vexler@ma.tum.de

Article published by EDP Sciences

The inherent difficulty of these control problems is the fact that the Lagrange multiplier associated with the state constraints is a Borel measure $\bar{\mu}$ in $\Omega \subset \mathbb{R}^n$. This leads to an adjoint state $\bar{\varphi} \in W_0^{1,s}(\Omega)$ for every $1 \le s < \frac{n}{n-1}$. In this paper, under mild assumptions, we prove that the adjoint state $\bar{\varphi}$ belongs to $H^1(\Omega) \cap L^\infty(\Omega)$ and $\bar{\mu} \in H^{-1}(\Omega)$. This implies, in particular, that Dirac measures are excluded as Lagrange multipliers for $n > 1$. As a consequence we also obtain the $H^1(\Omega) \cap L^\infty(\Omega)$ regularity for the optimal control $\bar{u}$. As far as we know, $H^1(\Omega)$ regularity results for $\bar{u}$ were proved in [11], where the presence of pointwise control constraints played a crucial role in the proof. However, no additional regularity for the adjoint state $\bar{\varphi}$ and the Lagrange multiplier $\bar{\mu}$ was obtained there. Our regularity result has been inspired by a recent result by Pieper and Vexler [27] for a sparse control problem with controls in a measure space.

We use this new regularity result to improve the error estimates for the finite element discretization of the control problem. For the error between the optimal control $\bar{u}$ and its discrete counterpart $\bar{u}_h$ we prove an estimate of order $\mathcal{O}(h|\log h|)$ for both the two and three dimensional case.

Error estimates for optimal control problems with state constraints governed by elliptic equations are derived in several publications. In [9, 10, 22] error estimates are given for optimal control problems with finitely many state constraints. In Deckelnick and Hinze [15,16] error estimates of order $h^{1-\varepsilon}$ in 2d and $h^{\frac{1}{2}-\varepsilon}$ in 3d are derived for a problem with pointwise state constraints, see also Meyer [26] for a proof of similar results with a different technique. These estimates are obtained for domains with smooth boundary Γ. For a convex polygonal domain, Meyer [26] obtains an order of $O(h^\lambda)$ where $\lambda \in (1/2, 1)$ depends on the biggest interior angle of Ω. For the three dimensional case an improvement to $h^{\frac{3}{4}}$ is achieved in Rösch and Steinig [28] and an estimate of order $h|\log h|$ is shown in [21] for a problem with control and state constraints. For the proof of the later result the presence of control constraints ensuring the uniform boundedness of $\bar{u}$ and $\bar{u}_h$ plays a crucial role. Liu, Gong and Yan [23] and Gong and Yan [19] treat problem (P) by transforming it into a biharmonic obstacle problem. They deal with the linear-quadratic case and the Laplace operator in dimension 2. For the error analysis, they suppose that Ω is polygonal, the obstacle is in $H^4(\Omega)$ and, following [3], some extra assumptions on the active set are needed to have the adjoint state in $H_0^1(\Omega)$. Discretization is made with the nonconforming Morley finite element in the first reference and with Lagrange P_1 finite elements in the second one. They prove $O(h)$ convergence for the Morley finite element. For P_1 elements they obtain again $O(h)$ in superconvergent meshes, and $O(h^{1/2})$ in quasi-uniform meshes. These results are valid if the optimal state is in $H^3(\Omega)$.

This means that our result improves the known estimate of almost order $\mathcal{O}(h^{\frac{1}{2}})$ to $\mathcal{O}(h|\log h|)$ for a purely state-constrained problem in the three dimensional case and also for plane polygonal convex domains, where in general the optimal state is only in $H^{2+\alpha}(\Omega)$, for some $\alpha \in (0, 1)$ depending on the biggest interior angle of the domain (see, *e.g.* [20], Thm. 5.1.1.4, or [4] for a general regularity result about the biharmonic operator).

The plan of this paper is as follows. In Section 2 we recall the optimality system and we discuss some consequences of it. In particular, the structure of the Lagrange multiplier $\bar{\mu}$ is studied. Section 3 is devoted to the proof of our main regularity result. In Section 4 we consider some extensions of this result to more general control problems. In particular, the case of semilinear state equations is considered. The error estimates for the finite element approximations are proved in Section 5 and some numerical examples confirming these estimates are given in Section 6.

2. Assumptions and preliminary results

Concerning the Dirichlet problem (1.1) we make the following hypotheses.

Assumption 2.1. Ω denotes an open bounded subset of $\mathbb{R}^n$, $n = 2$ or 3, with a Lipschitz boundary Γ. A is the linear operator

$$Ay = -\sum_{i,j=1}^{n} \partial_{x_j}[a_{ij}(x)\,\partial_{x_i} y] + a_0(x)y,$$

with $a_{ij}, a_0 \in L^\infty(\Omega)$, $a_0(x) \geq 0$ for almost all $x \in \Omega$. Furthermore, there exists some $\Lambda > 0$ such that

$$\sum_{i,j=1}^{n} a_{ij}(x)\, \xi_i\, \xi_j \geq \Lambda\, |\xi|^2 \text{ for a.a. } x \in \Omega \quad \text{and} \quad \forall \xi \in \mathbb{R}^n.$$

Some additional regularity will be required for Ω, Γ and the coefficients a_{ij} in the section devoted to the numerical approximation.

Under this assumption, it is well known that for every $u \in H^{-1}(\Omega)$ equation (1.1) has a unique solution in the Sobolev space $H_0^1(\Omega)$. This solution will be denoted by y_u. Moreover, if $u \in W^{-1,p}(\Omega)$ for some $p > n$, then y_u belongs to the Hölder space $C^\theta(\bar{\Omega})$ for some $0 < \theta < 1$ depending on p; see [18], Theorem 8.29. We also have the estimate

$$\begin{cases} \|y_u\|_{H_0^1(\Omega)} \leq M_0 \|u\|_{H^{-1}(\Omega)}, \\ \|y_u\|_{C^\theta(\bar{\Omega})} \leq M_p \|u\|_{W^{-1,p}(\Omega)}. \end{cases} \tag{2.1}$$

Since $n = 2$ or 3, we have that $L^2(\Omega) \subset W^{-1,p}(\Omega)$, with continuous embedding, for any $p < \infty$ if $n = 2$ and any $p \leq 6$ if $n = 3$. Hence, the above estimates remain valid replacing the norm of u in the corresponding Sobolev space by the $L^2(\Omega)$-norm, with the obvious changes of the constants M_0 and M_p.

Assumption 2.2. Along this paper y_d is an element of $L^2(\Omega)$ and $\nu > 0$. We also assume the following hypotheses on the functions a and b.

$$a, b \in C(\bar{\Omega}). \tag{2.2a}$$

$$a(x) < b(x) \;\forall x \in \bar{\Omega}. \tag{2.2b}$$

$$a(x) < 0 < b(x) \;\forall x \in \Gamma. \tag{2.2c}$$

$$Aa, Ab \in L^\infty(\Omega). \tag{2.2d}$$

Associated with these data we define the set

$$\mathcal{Y}_{\text{ab}} = \{y \in C_0(\Omega) : a(x) \leq y(x) \leq b(x) \quad \forall x \in \Omega\},$$

where $C_0(\Omega)$ is the space of continuous functions in $\bar{\Omega}$ vanishing on Γ. Then, the admissible control set can be rewritten as follows

$$U_{\text{ad}} = \{u \in L^2(\Omega) : y_u \in \mathcal{Y}_{\text{ab}}\}.$$

Notice that the assumptions (2.2a)–(2.2d) hold if a and b are constants satisfying $a < 0 < b$. This is the case for the typical state constraint $|y_u(x)| \leq \delta$. The assumptions (2.2a)–(2.2c) are natural and similar to assumptions usually required for optimal control problems with state constraints. The additional regularity assumption (2.2d) is crucial for our main regularity result, see Theorem 3.1, as well as for the error estimate, see Corollary 5.7.

It is obvious that the control problem (P) is strictly convex. Hence, it has at most one solution. The existence of a solution can be proved by standard arguments. Hereafter, $\bar{u}$ will denote the solution of (P) and $\bar{y}$ the corresponding state. Before establishing the first-order optimality conditions fulfilled by the optimal control $\bar{u}$, we introduce some notation. We denote by $\mathcal{M}(\Omega)$ the space of real and regular Borel measures in Ω, which is identified with the dual of $C_0(\Omega)$. This is a Banach space for the norm

$$\|\mu\|_{\mathcal{M}(\Omega)} = |\mu|(\Omega) = \sup_{y \in C_0(\Omega), \|y\|_\infty \leq 1} \int_\Omega y \,\mathrm{d}\mu.$$

Above $|\mu|$ denotes the total variation measure corresponding to μ. We also consider the Jordan decomposition $\mu = \mu^+ - \mu^-$. Then, we know that $|\mu| = \mu^+ + \mu^-$; see, for instance, Rudin ([29], Chap. 6), for details.

Theorem 2.3. *Under the Assumptions 2.1 and (2.2a)–(2.2c), there exist a measure $\bar\mu \in \mathcal{M}(\Omega)$ and an element $\bar\varphi \in W_0^{1,s}(\Omega)$, for every $1 \le s < \frac{n}{n-1}$, such that*

$$\begin{cases} A\bar y = \bar u & \text{in } \Omega, \\ \bar y = 0 & \text{on } \Gamma, \end{cases} \tag{2.3a}$$

$$\begin{cases} A^*\bar\varphi = \bar y - y_d + \bar\mu & \text{in } \Omega, \\ \bar\varphi = 0 & \text{on } \Gamma, \end{cases} \tag{2.3b}$$

$$\int_\Omega (y - \bar y)\,\mathrm{d}\bar\mu \le 0 \quad \forall y \in \mathcal{Y}_{\mathrm{ab}}, \tag{2.3c}$$

$$\bar\varphi + \nu\bar u = 0, \tag{2.3d}$$

where A^* is the adjoint operator of A, given by the expression

$$A^*\varphi = -\sum_{i,j=1}^n \partial_{x_j}[a_{ji}(x)\,\partial_{x_i}\varphi] + a_0(x)\varphi.$$

Moreover, the Lagrange multiplier $\bar\mu$ and the adjoint state $\bar\varphi$ are unique.

Proof. First, let us prove the uniqueness of $\bar\varphi$ and $\bar\mu$. The uniqueness of $\bar\varphi$ follows from the uniqueness of $\bar u$ and (2.3d). Hence, (2.3b) implies the uniqueness of $\bar\mu$ as a distribution, which is equivalent to the uniqueness of $\bar\mu$ as a measure.

The existence of $\bar\mu$ and $\bar\varphi$ satisfying (2.3b)–(2.3d) is well known under the assumption of the Slater condition, see, for instance, [6]. Let us check that the Slater condition is fulfilled:

$$\exists u_0 \in L^2(\Omega) \text{ such that } a(x) < y_{u_0}(x) < b(x)\ \forall x \in \bar\Omega.$$

Define

$$\rho_1 = \min_{x\in\bar\Omega}(b(x) - a(x)),$$

Due to (2.2a) and (2.2b), we have that $\rho_1 > 0$. Moreover, using (2.2c) we can find $\varepsilon > 0$ and $\rho_2 > 0$ such that

$$a(x) < -\rho_2 < \rho_2 < b(x) \quad \forall x \in \bar\Omega \ \text{ such that } \ d(x,\Gamma) < \varepsilon. \tag{2.4}$$

Set $\rho = \min\{\rho_1, \rho_2\}$. Now, we use Uryshon's lemma to obtain a function $\phi \in C_0(\Omega)$ such that

$$0 \le \phi \le 1 \ \text{ and } \ \phi(x) = \begin{cases} 0 \text{ if } d(x,\Gamma) \le \dfrac{\varepsilon}{2}, \\ 1 \text{ if } d(x,\Gamma) \ge \varepsilon. \end{cases}$$

Setting $a_\phi = \phi a$ and $b_\phi = \phi b$, we have that $a_\phi, b_\phi \in C_0(\Omega)$ and

$$a(x) \le a_\phi(x) \le b_\phi(x) \le b(x) \quad \forall x \in \bar\Omega.$$

We know that the space $C_0^\infty(\Omega)$ of smooth functions having a compact support in Ω is dense in $C_0(\Omega)$. Therefore, we can select one of these functions, denoted by y, such that

$$\left\| y - \frac{1}{2}(a_\phi + b_\phi)\right\|_{L^\infty(\Omega)} < \frac{\rho}{4}.$$

It is obvious that $Ay \in W^{-1,p}(\Omega)$ for every $p \ge 1$. Fix some $p \in (n, +\infty)$ and take $u_0 \in L^p(\Omega)$ such that

$$\|u_0 - Ay\|_{W^{-1,p}(\Omega)} < \frac{\rho}{4M_p},$$

where M_p is given by (2.1). Let us prove that u_0 satisfies the Slater assumption. First, we observe

$$\begin{aligned}\left\| y_{u_0} - \frac{1}{2}(a_\phi + b_\phi)\right\|_{L^\infty(\Omega)} &\le \|y_{u_0} - y\|_{L^\infty(\Omega)} + \left\| y - \frac{1}{2}(a_\phi + b_\phi)\right\|_{L^\infty(\Omega)} \\ &< M_p \|u_0 - Ay\|_{W^{-1,p}(\Omega)} + \frac{\rho}{4} < \frac{\rho}{2}. \end{aligned} \tag{2.5}$$

To prove that $a(x) < y_{u_0}(x) < b(x)$ in $\bar{\Omega}$ we distinguish three cases.

Case I. $d(x,\Gamma) \le \frac{\varepsilon}{2}$. In this case, $a_\phi(x) = b_\phi(x) = 0$, therefore (2.4) and (2.5) lead to

$$a(x) < -\rho_2 \le -\rho < y_{u_0}(x) < \rho \le \rho_2 < b(x).$$

Case II. $d(x,\Gamma) \ge \varepsilon$. For this x we have that $a_\phi(x) = a(x)$ and $b_\phi(x) = b(x)$. From the definition of ρ_1 and ρ we get

$$-\rho \ge -\rho_1 \ge a(x) - b(x) \quad \text{and} \quad \rho \le \rho_1 \le b(x) - a(x).$$

Thus, we obtain with (2.5)

$$\begin{aligned} a(x) &\le \frac{1}{2}(a(x)+b(x)) - \frac{\rho_1}{2} \le \frac{1}{2}(a_\phi(x)+b_\phi(x)) - \frac{\rho}{2} < y_{u_0}(x) \\ &< \frac{1}{2}(a_\phi(x)+b_\phi(x)) + \frac{\rho}{2} \le \frac{1}{2}(a(x)+b(x)) + \frac{\rho_1}{2} \le b(x). \end{aligned}$$

Case III. $\frac{\varepsilon}{2} < d(x,\Gamma) < \varepsilon$. Using again (2.4) and (2.5), it follows

$$\begin{aligned} a(x) &\le \frac{1}{2}a_\phi(x) + \frac{1}{2}a(x) \le \frac{1}{2}(a_\phi(x)+b_\phi(x)) - \frac{\rho_2}{2} \\ &\le \frac{1}{2}(a_\phi(x)+b_\phi(x)) - \frac{\rho}{2} < y_{u_0}(x) < \frac{1}{2}(a_\phi(x)+b_\phi(x)) + \frac{\rho}{2} \\ &\le \frac{1}{2}(a_\phi(x)+b_\phi(x)) + \frac{\rho_2}{2} \le \frac{1}{2}b_\phi(x) + \frac{1}{2}b(x) \le b(x). \end{aligned}$$

□

Remark 2.4. Let us notice that the assumption (2.2d) was not necessary for the proof of the Slater condition. We only used the assumptions (2.2a)–(2.2c). Moreover, the proof can be simplified if we assume that the coefficients a_{ij} of the operator A are Lipschitz continuous in $\bar{\Omega}$. Indeed, under this assumption, if we take ϕ of class C^2 in Ω and satisfying the conditions of the proof, then $u_0 = A[\frac{1}{2}(a_\phi + b_\phi)] \in L^2(\Omega)$ satisfies the Slater condition. Furthermore, if a and b are constants, with $a < 0 < b$, then $u_0 = 0$ satisfies the Slater condition.

Regarding the adjoint state equation (2.3b), some explanation is necessary. Following Stampacchia [31], given a measure $\mu \in \mathcal{M}(\Omega)$, we say that an element $\varphi \in L^1(\Omega)$ is a solution of the Dirichlet problem

$$\begin{cases} A^*\varphi = \mu & \text{in } \Omega, \\ \quad \varphi = 0 & \text{on } \Gamma, \end{cases} \tag{2.6}$$

if

$$\int_\Omega \varphi A z \,\mathrm{d}x = \int_\Omega z \,\mathrm{d}\mu \quad \forall z \in \mathcal{Z},$$

with

$$\mathcal{Z} = \{z \in H_0^1(\Omega) : Az \in C_0(\Omega)\}.$$

Using again ([18], Thm. 8.29), we deduce that $\mathcal{Z} \subset C_0(\Omega)$, hence the above integrals are well defined. With this definition, we know that there exists a unique solution that additionally belongs to $W_0^{1,s}(\Omega)$ for every $s < \frac{n}{n-1}$.

Moreover, if E denotes the support of μ, then $\varphi \in H^1_{loc}(\Omega \setminus E) \cap C(\bar{\Omega} \setminus E)$, and for any compact $K \subset \Omega \setminus E$ the following estimate holds

$$\|\varphi\|_{C(K)} \leq C_K \|\mu\|_{\mathcal{M}(\Omega)}, \tag{2.7}$$

see ([31], Thm. 9.3). Of course, all these properties are enjoyed by $\bar{\varphi}$.

The following complementarity result is well known, but we have not found any proof in the literature for the case of non-constant bounds a and b. For the sake of completeness we give a proof, see also *e.g.* [6] for the proof in the case of constant bounds.

Proposition 2.5. *Let a, b be two functions satisfying (2.2a)–(2.2c) and let $\bar{y} \in C_0(\Omega)$ and $\bar{\mu} \in \mathcal{M}(\Omega)$ satisfy (2.3c). Then, the following embeddings hold*

$$\begin{cases} \operatorname{supp} \bar{\mu}^+ \subset \{x \in \Omega : \bar{y}(x) = b(x)\}, \\ \operatorname{supp} \bar{\mu}^- \subset \{x \in \Omega : \bar{y}(x) = a(x)\}, \end{cases} \tag{2.8}$$

where $\bar{\mu} = \bar{\mu}^+ - \bar{\mu}^-$ is the Jordan decomposition of $\bar{\mu}$.

Proof. Let us denote

$$K_a = \{x \in \Omega : \bar{y}(x) = a(x)\} \quad \text{and} \quad K_b = \{x \in \Omega : \bar{y}(x) = b(x)\}.$$

For every integer $k \geq 1$, we set

$$\Omega_k = \left\{x \in \Omega : a(x) + \frac{1}{k} < \bar{y}(x) < b(x) - \frac{1}{k}\right\}.$$

From the assumptions (2.2a)−(2.2c) we infer that Ω_k is a nonempty open set for every k sufficiently large. Let us take an arbitrary element $y \in C_0(\Omega_k)$ such that $\|y\|_\infty \leq 1$. We extend y by zero to Ω and denote this extension again by y. Then, $y \in C_0(\Omega)$ and $y_k = \bar{y} + \frac{1}{k} y \in \mathcal{Y}_{\text{ab}}$, hence (2.3c) implies

$$\int_{\Omega_k} y \,\mathrm{d}\bar{\mu} = k \int_\Omega (y_k - \bar{y}) \,\mathrm{d}\bar{\mu} \leq 0.$$

This implies that $|\bar{\mu}|(\Omega_k) = 0$ for every k, therefore

$$|\bar{\mu}|(\Omega \setminus (K_a \cup K_b)) = \lim_{k \to \infty} |\bar{\mu}|(\Omega_k) = 0.$$

This means that the support of $\bar{\mu}$ is contained in $K_a \cup K_b$. It is enough to prove that $\bar{\mu}$ is nonnegative on K_b and nonpositive on K_a to conclude (2.8). We show that $\bar{\mu}$ is nonpositive on K_a; the proof of the nonnegativity of $\bar{\mu}$ on K_b is analogous. Take a number ρ satisfying

$$0 < \rho < \frac{1}{2} \inf_{x \in \bar{\Omega}} (b(x) - a(x)).$$

This choice is possible thanks to (2.2a)−(2.2c). Now, we define the sets

$$\Omega_a = \{x \in \Omega : \bar{y}(x) < a(x) + \rho\} \quad \text{and} \quad \Omega_b = \{x \in \Omega : \bar{y}(x) > b(x) - \rho\}.$$

Since $\bar{y} \in C_0(\Omega)$, we get with (2.2a)−(2.2c) that Ω_a and Ω_b are open sets and

$$K_a \subset \Omega_a, \quad K_b \subset \Omega_b \quad \text{and} \quad \bar{\Omega}_a \cap \bar{\Omega}_b = \emptyset.$$

Let $y \in C(K_a) \setminus \{0\}$ be a nonnegative function. Using Tietze's extension theorem, we can extend y to Ω, denoted y again, in such a way that $\operatorname{supp} y \subset \Omega_a$. Moreover, taking $\max\{y(x), 0\}$, we can assume that $y \geq 0$ in Ω. Then, we have

$$y_\rho = \bar{y} + \frac{\rho}{\|y\|_\infty} y \in \mathcal{Y}_{\text{ab}}.$$

Indeed, if $x \notin \Omega_a$, then $y_\rho = \bar{y}$ and $a(x) \le y_\rho(x) \le b(x)$. If $x \in \Omega_a$, the definition of ρ and Ω_a implies

$$a(x) \le \bar{y}(x) \le y_\rho(x) \le \bar{y}(x) + \rho < a(x) + 2\rho < b(x).$$

Finally, from $\operatorname{supp} y \subset \Omega_a$, $\operatorname{supp} \bar{\mu} \subset K_a \cup K_b$ and (2.3c) it follows

$$\int_{K_a} y \,\mathrm{d}\bar{\mu} = \int_\Omega y \,\mathrm{d}\bar{\mu} = \frac{\|y\|_\infty}{\rho} \int_\Omega (y_\rho - \bar{y}) \,\mathrm{d}\bar{\mu} \le 0.$$

Since this inequality holds for every nonnegative function $y \in C(K_a)$, we conclude that $\bar{\mu}$ is nonpositive on K_a, as desired. □

The classical regularity result for $\bar{u}$ is deduced from the equality (2.3d): $\bar{u} \in W_0^{1,s}(\Omega)$ for every $1 \le s < \frac{n}{n-1}$. In the next section, we will show that $\bar{u} \in H_0^1(\Omega)$.

3. A regularity result for the adjoint state $\bar{\varphi}$ and the Lagrange multiplier $\bar{\mu}$

The goal of this section is to prove the following theorem.

Theorem 3.1. *Under the Assumptions 2.1 and 2.2, the following regularity result holds:*

$$\bar{\varphi}, \bar{u} \in H_0^1(\Omega) \cap L^\infty(\Omega) \quad \textit{and} \quad \bar{\mu} \in \mathcal{M}(\Omega) \cap H^{-1}(\Omega).$$

We state two auxiliary lemmas before proving the theorem.

Lemma 3.2. *Let $\mu \in \mathcal{M}(\Omega)$ be a positive measure with a compact support in Ω and $\varphi \in W_0^{1,s}(\Omega)$ be the solution of (2.6). Define*

$$\varphi^*(x) := \int_\Omega g_A(x,\xi) \,\mathrm{d}\mu(\xi) \quad \forall x \in \Omega,$$

where g_A is Green's function corresponding to the Dirichlet problem (2.6). Then, we have that

$$\varphi \in L^\infty(\Omega) \Leftrightarrow \sup_{x \in \operatorname{supp} \mu} \varphi^*(x) < \infty.$$

Proof. For the case $A = -\Delta$ this lemma is proven in [27]. For the general case, let us consider the solution $z \in W_0^{1,s}(\Omega)$ of the Dirichlet problem

$$\begin{cases} -\Delta z = \mu & \text{in } \Omega, \\ z = 0 & \text{on } \Gamma. \end{cases}$$

Observe that the positivity of μ implies that φ and z are nonnegative almost everywhere in Ω. Moreover, since $A^*\varphi = \Delta z = 0$ in the open set $\Omega \setminus \operatorname{supp} \mu$ and $\varphi = z = 0$ on Γ, we deduce that $\varphi, z \in C(\bar{\Omega} \setminus \operatorname{supp} \mu)$. Therefore, given $\varepsilon > 0$ we can choose a compact set K_ε such that $\operatorname{supp} \mu \subset K_\varepsilon \subset \Omega$ and

$$\varphi(x) + z(x) < \varepsilon \quad \text{for a.a. } x \in \Omega \setminus K_\varepsilon. \tag{3.1}$$

We know from [31] that there exists a positive number C_ε such that

$$\frac{1}{C_\varepsilon} g(x,\xi) \le g_A(x,\xi) \le C_\varepsilon \, g(x,\xi) \quad \forall x, \xi \in K_\varepsilon,$$

where g denotes the Green's function associated with the Dirichlet problem corresponding to $-\Delta$. Analogously to φ^* we define

$$z^*(x) = \int_\Omega g(x,\xi) \,\mathrm{d}\mu(\xi).$$

Integration with respect to μ in the above inequalities and taking into account that $\mu \geq 0$ and $\operatorname{supp}\mu \subset K_\varepsilon$ yields for all $x \in K_\varepsilon$

$$\begin{aligned}\frac{1}{C_\varepsilon} z^*(x) = \frac{1}{C_\varepsilon}\int_\Omega g(x,\xi)\,\mathrm{d}\mu(\xi) = \frac{1}{C_\varepsilon}\int_{K_\varepsilon} g(x,\xi)\mathrm{d}\mu(\xi) \leq \int_{K_\varepsilon} g_A(x,\xi)\,\mathrm{d}\mu(\xi)\\ = \int_\Omega g_A(x,\xi)\,\mathrm{d}\mu(\xi) = \varphi^*(x) \leq C_\varepsilon \int_{K_\varepsilon} g(x,\xi)\,\mathrm{d}\mu(\xi) = C_\varepsilon\, z^*(x).\end{aligned}$$

Since $\varphi = \varphi^*$ almost everywhere in Ω, these inequalities imply that

$$\varphi \in L^\infty(\Omega) \Leftrightarrow \varphi^* \in L^\infty(\Omega) \Leftrightarrow z^* \in L^\infty(\Omega) \Leftrightarrow \sup_{x\in\operatorname{supp}\mu} z^*(x) < \infty,$$

where the last equivalence is due to a result by Pieper and Vexler [27].

Finally, the inequalities

$$\frac{1}{C_\varepsilon} z^*(x) \leq \varphi^*(x) \leq C_\varepsilon z^*(x) \quad \forall x \in \operatorname{supp}\mu \subset K_\varepsilon$$

and the above equivalences imply that

$$\sup_{x\in\operatorname{supp}\mu} \varphi^*(x) < \infty \Leftrightarrow \sup_{x\in\operatorname{supp}\mu} z^*(x) < \infty \Leftrightarrow \varphi \in L^\infty(\Omega).$$ □

For any function φ and $\alpha \leq \beta$, we denote $\operatorname{Proj}_{[\alpha,\beta]}(\varphi)(x) = \min\{\max\{\alpha, \varphi(x)\}, \beta\}$.

Lemma 3.3. *Let $\mu \in \mathcal{M}(\Omega)$ and let $\varphi \in W_0^{1,s}(\Omega)$ for all $s < n/(n-1)$ be the solution of (2.6). Then, $\operatorname{Proj}_{[-M,M]}(\varphi) \in H_0^1(\Omega)$ for every $M > 0$.*

Proof. This result can be deduced from ([14], Thm. 10.1 and Eq. (2.22)) or ([17], Eq. (7)). For convenience, we provide the reader with a simple proof.

Let us take a sequence of functions $\{\mu_k\}_k \subset L^2(\Omega)$, such that $\mu_k \overset{*}{\rightharpoonup} \mu$ in $\mathcal{M}(\Omega)$ and $\|\mu_k\|_{L^1(\Omega)} \leq \|\mu\|_{\mathcal{M}(\Omega)}$. Let $\varphi_k \in H_0^1(\Omega)$ be the unique solution of

$$\begin{cases} A^*\varphi_k = \mu_k & \text{in } \Omega,\\ \varphi_k = 0 & \text{on } \Gamma.\end{cases}$$

Due to the compact embedding of $\mathcal{M}(\Omega)$ into $W^{-1,s}(\Omega)$ for all $s < \frac{n}{n-1}$, we have that $\varphi_k \to \varphi$ strongly in $W^{1,s}(\Omega)$. Take $M > 0$ fixed and define $\varphi_k^M = \operatorname{Proj}_{[-M,M]}(\varphi_k)$. By the continuity of $\operatorname{Proj}_{[-M,M]}\colon W^{1,s}(\Omega) \to W^{1,s}(\Omega)$ we also have

$$\lim_k \varphi_k^M = \operatorname{Proj}_{[-M,M]}(\varphi) \text{ strongly in } W_0^{1,s}(\Omega)\ \forall\, 1 \leq s < \frac{n}{n-1}\cdot \tag{3.2}$$

On the other hand, using the uniform ellipticity of the operator A^*, we have

$$\begin{aligned}\Lambda\|\nabla\varphi_k^M\|_{L^2(\Omega)}^2 &\leq \sum_{i,j=1}^n \int_\Omega a_{ij}(x)\partial_{x_j}\varphi_k^M \partial_{x_i}\varphi_k^M\,\mathrm{d}x + \int_\Omega a_0(x)\varphi_k^M\varphi_k^M\,\mathrm{d}x\\ &\leq \sum_{i,j=1}^n \int_\Omega a_{ij}(x)\partial_{x_j}\varphi_k \partial_{x_i}\varphi_k^M\,\mathrm{d}x + \int_\Omega a_0(x)\varphi_k\varphi_k^M\,\mathrm{d}x\\ &= \int_\Omega \mu_k\varphi_k^M\,\mathrm{d}x \leq \|\varphi_k^M\|_{L^\infty(\Omega)}\|\mu_k\|_{L^1(\Omega)} \leq M\|\mu\|_{\mathcal{M}(\Omega)}.\end{aligned}$$

Therefore the sequence $\{\varphi_k^M\}_k$ is bounded in $H_0^1(\Omega)$ and there exist $\varphi_M \in H_0^1(\Omega)$ and a subsequence of $\{\varphi_k^M\}_k$, denoted in the same way, such that $\varphi_k^M \rightharpoonup \varphi_M$ weakly in $H_0^1(\Omega)$. Using this fact and (3.2) we readily have that $\operatorname{Proj}_{[-M,M]}(\varphi) = \varphi_M \in H_0^1(\Omega)$. □

Proof of Theorem 3.1. Let us consider the Jordan decomposition of $\bar\mu$: $\bar\mu = \bar\mu^+ - \bar\mu^-$. We also decompose

$$\bar\varphi = \varphi_0 + \varphi_+ - \varphi_-$$

where φ_0, φ_+ and φ_- are the solutions of (2.6) with right hand side equal to $\bar y - y_d$, $\bar\mu^+$ and $\bar\mu^-$, respectively. Since $\bar y - y_d \in L^2(\Omega)$, then we know that $\varphi_0 \in H_0^1(\Omega) \cap C(\bar\Omega)$. We will prove that $\varphi_+ \in L^\infty(\Omega)$, the proof of the boundedness of φ_- being analogous, which implies the boundedness of $\bar\varphi$. Then, the $H_0^1(\Omega)$-regularity of $\bar\varphi$ follows from Lemma 3.3. The proof of the theorem is concluded by (2.3b) and (2.3d): $\bar\mu = A^*\bar\varphi - \bar y + y_d \in H^{-1}(\Omega)$ and $\bar u = -\frac{1}{\nu}\bar\varphi \in H_0^1(\Omega) \cap L^\infty(\Omega)$.

Let us prove the boundedness of φ_+. First, we observe that φ_+ and φ_- are nonnegative functions and $\varphi_- \in C(\bar\Omega \setminus \operatorname{supp}\bar\mu^-)$. By Lemma 3.2, we have that $\varphi_+ \in L^\infty(\Omega)$ if and only if φ_+^* is upper bounded in $\operatorname{supp}\bar\mu^+$, where φ_+^* is defined as in Lemma 3.2 by

$$\varphi_+^*(x) := \int_\Omega g_A(x,\xi)\,\mathrm{d}\mu^+(\xi) \quad \forall x \in \Omega.$$

We argue by contradiction and we assume that φ_+^* is not bounded in $\operatorname{supp}\bar\mu^+$. Take $0 < \rho < 1$ such that $a(x) < \rho b(x)$ for every $x \in \bar\Omega$. The existence of ρ follows from (2.2a) and (2.2c). We define the compact set

$$K = \{x \in \Omega : \bar y(x) \geq \rho b(x)\}.$$

By (2.7) and (2.8) we have

$$\|\varphi_-\|_{C(K)} \leq C_K \|\bar\mu^-\|_{\mathcal{M}(\Omega)}.$$

Let us set

$$M = \|\varphi_0\|_{C(K)} + C_K\|\bar\mu^-\|_{\mathcal{M}(\Omega)} + \nu\left(\|a_0\|_{L^\infty(\Omega)}\|b-a\|_{L^\infty(\Omega)} + \|Ab\|_{L^\infty(\Omega)}\right).$$

Since φ_+^* is not bounded in $\operatorname{supp}\bar\mu^+$, there exists an element $x^0 \in \operatorname{supp}\bar\mu^+$ such that $\varphi_+^*(x^0) > M$. From the positivity of Green's function g_A we deduce, with Fatou's Lemma, that φ_+^* is lower semicontinuous. Hence, the set $\{x \in \Omega : \varphi_+^*(x) > M\}$ is open. Let us take $\varepsilon > 0$ such that $\varphi_+^*(x) > M$ $\forall x \in B_\varepsilon(x^0)$ and $B_\varepsilon(x^0) \subset K$. Therefore, we have for almost every $x \in B_\varepsilon(x^0)$

$$\begin{aligned}\bar\varphi(x) &= \varphi_0(x) + \varphi_+^*(x) - \varphi_-^*(x)\\ &> M - \|\varphi_0\|_{C(K)} - \|\varphi_-^*\|_{C(K)}\\ &\geq \nu\left(\|a_0\|_{L^\infty(\Omega)}\|b-a\|_{L^\infty(\Omega)} + \|Ab\|_{L^\infty(\Omega)}\right).\end{aligned}$$

We consider the difference $\tilde y = \bar y - b$. There holds $\tilde y \leq 0$ and due to the fact that $x_0 \in \operatorname{supp}\bar\mu^+$ we have by Proposition 2.5 that $\tilde y(x_0) = \bar y(x_0) - b(x_0) = 0$ takes its maximum at $x = x_0$. Now, (2.2a), (2.2d) and the above inequality imply

$$\begin{aligned}-\sum_{i,j=1}^n \partial_{x_j}[a_{ij}(x)\,\partial_{x_i}(\tilde y(x))] &= \bar u(x) - a_0(x)(\bar y(x) - b(x)) - Ab\\ &= -\frac{1}{\nu}\bar\varphi(x) - a_0(x)(\bar y(x) - b(x)) - Ab\\ &< -\|a_0\|_{L^\infty(\Omega)}\|b-a\|_{L^\infty(\Omega)} - a_0(x)(\bar y(x) - b(x)) \leq 0\end{aligned}$$

for almost all $x \in B_\varepsilon(x^0)$. Hence, the maximum principle implies that $\tilde y$ is constant in the ball $B_\varepsilon(x^0)$, which contradicts the above inequality. □

Remark 3.4. The assumptions (2.2a)–(2.2c) are quite natural for the study of pointwise state constraints. However, the assumption (2.2d) is unusual, but it is necessary to prove the regularity results of Theorem 3.1.

Indeed, Theorem 3.1 can fail if it does not hold. There are several examples in the literature where the multipliers $\bar\mu$ are Dirac measures; see *e.g.* ([24], Sect. 8.1), ([25], Sect. 6.2) or [12]. In these examples, $Ab \notin L^2(\Omega)$. We provide an example of a linear quadratic control problem such that $Ab \in L^p(\Omega)$ for all $p < +\infty$, and however the Lagrange multiplier is a Dirac measure.

Consider $\Omega = B_1(0) \subset \mathbb{R}^2$ the unit ball, $y_d \equiv 1$, $\nu = 1$ and $A = -\Delta$. Let $\tilde u$ be the solution of the unconstrained linear quadratic control problem

$$(\tilde P) \ \min_{u\in L^2(\Omega)} J(u) = \frac{1}{2}\|y_u - y_d\|_{L^2(\Omega)} + \frac{1}{2}\|u\|^2_{L^2(\Omega)}.$$

Let $\tilde y = y_{\tilde u}$. From the optimality system

$$-\Delta\tilde y = -\tilde\varphi,\ -\Delta\tilde\varphi = \tilde y - 1 \text{ in } \Omega,\ \tilde y = \tilde\varphi = 0 \text{ on } \Gamma,\ \tilde u = -\tilde\varphi \text{ in } \Omega,$$

we deduce that $\tilde y \not\equiv 0$. Now, consider the point $x_0 = (0,0)$ and take some $b < \tilde y(x_0)$. We define the problem with a state constraint only at the point x_0

$$(P_0)\ \min_{u\in U_{\rm ad}} J(u), \quad \text{with } U_{\rm ad} = \{u \in L^2(\Omega),\ y_u(x_0) \le b\}.$$

This problem has a unique solution $\bar u$ with related state $\bar y$. Moreover, the state constraint at x_0 is attained: $\bar y(x_0) = b$. Indeed, if $\bar y(x_0) < b$, then $\bar\mu = 0$ and $\bar u$ would satisfy the optimality system for problem $(\tilde P)$. Therefore, $\bar u$ would be a solution of the unconstrained problem $(\tilde P)$, and by uniqueness $\bar u = \tilde u$. This would imply $\bar y(x_0) > b$, which is a contradiction. The optimality system for (P_0) reads like

$$-\Delta\bar y = -\bar\varphi,\ -\Delta\bar\varphi = \bar y - 1 + \bar\lambda\delta_{x_0} \text{ in } \Omega,\ \bar y = \bar\varphi = 0 \text{ on } \Gamma, \bar u = -\bar\varphi \text{ in } \Omega,$$

and $\bar\lambda \in \mathbb{R}$, $\bar\lambda > 0$. Again, if $\bar\lambda = 0$ we would have $\bar u = \tilde u$ leading to a contradiction. Recall that $n = 2$; then we have that $\bar\varphi \in W^{1,s}(\Omega)$ for all $s < 2$ and hence $\bar\varphi \in L^p(\Omega)$ for all $p < \infty$, but $\bar\varphi \notin L^\infty(\Omega)$. Therefore, we conclude that $-\Delta\bar y \in L^p(\Omega)$ for all $p < \infty$, but $-\Delta\bar y \notin L^\infty(\Omega)$.

Consider $b(x) = \bar y(x) + |x|^2$, where $|x|$ denotes the Euclidean norm of x in $\mathbb{R}^2$ and the problem

$$(\mathrm{P})\ \min_{u\in U_{\rm ad}} J(u), \quad \text{with } U_{\rm ad} = \{u \in L^2(\Omega),\ y_u(x) \le b(x)\ \forall x \in \Omega\}.$$

The function b satisfies the assumptions (2.2a)–(2.2c), but not (2.2d) because $-\Delta b = -\Delta\bar y - 4$. We have that $\bar y(x) \le b$ in $\bar\Omega$ since $b - \bar y = |x|^2$ and $\bar u$, $\bar y$, $\bar\varphi$ and $\bar\mu = \bar\lambda\delta_{x_0}$ satisfy the optimality system for (P). Since the problem is convex, necessary conditions are also sufficient and hence we have found that the solution does not satisfy the claims of Theorem 3.1.

We finish this section establishing a corollary of Theorem 3.1 that will be useful to prove the error estimates in Section 5.

Corollary 3.5. *Let us suppose that the Assumptions 2.1 and 2.2 hold, and $a_{ij} \in C^{0,1}(\Omega)$ for $1 \le i,j \le n$. Then, for every open set $\Omega' \subset \bar\Omega' \subset \Omega$ and any $1 \le p < \infty$, there exists a constant $C > 0$ independent of $\bar u$ and p such that*

$$\|\bar y\|_{W^{2,p}(\Omega')} \le Cp\|\bar u\|_{L^\infty(\Omega)}. \tag{3.3}$$

The fact that $\bar y \in W^{2,p}(\Omega')$ follows by elliptic regularity. The exact dependence of the constant on p can be traced for example from Theorem 9.9 in [18].

4. Some extensions

More general formulations for state-constrained optimal control problems of elliptic equations can be found in the literature, see for instance [7] or [11]. The formulation of the problem (P) was given in a simple way for an easier reading of the paper and the technique used in the Proof of Theorem 3.1. In this section, we point out under which assumptions Theorem 3.1 can be extended to more general formulations.

4.1. A general cost functional and control constraints

Instead of the quadratic cost functional considered in the formulation of (P), a more general functional can be treated

$$J(u) = \int_\Omega L(x, y_u(x), u(x))\,\mathrm{d}x,$$

with $L\colon \Omega \times \mathbb{R}^2 \to \mathbb{R}$ a Carathédory function. Some differentiability hypotheses on L must be assumed to get the corresponding optimality system (2.3a)−(2.3d). We make the following assumption: L is of class C^1 with respect to the second and third variables, $L(\cdot, 0, 0) \in L^1(\Omega)$, and for all $M > 0$ there is a function $\psi_M \in L^2(\Omega)$ such that

$$\left|\frac{\partial L}{\partial u}(x,y,u)\right| + \left|\frac{\partial L}{\partial y}(x,y,u)\right| \le \psi_M(x), \text{ for a.a. } x \in \Omega \text{ and } |y|, |u| \le M.$$

Additionally, we can consider control constraints $u \in \mathcal{U}_{\alpha\beta}$, with

$$\mathcal{U}_{\alpha\beta} = \{u \in L^\infty(\Omega) : -\infty < \alpha \le u(x) \le \beta < +\infty \quad \text{for a.a. } x \in \Omega\}$$

with $\alpha, \beta \in \mathbb{R}$ and $\alpha < \beta$. In this case, the control problem is in general not convex and we can have local and global solutions. To prove the existence of a solution of the control problem, besides the Assumptions 2.1 and 2.2, we need the convexity of L with respect to u. Assuming that the Slater condition is fulfilled, if $\bar{u}$ is a local solution then the optimality system (2.3a)−(2.3d) holds with the following changes. Instead of (2.3b) and (2.3d), we have

$$\begin{cases} A^*\bar{\varphi} = \dfrac{\partial L}{\partial y}(x, \bar{y}, \bar{u}) + \bar{\mu} & \text{in } \Omega, \\ \quad\bar{\varphi} = 0 & \text{on } \Gamma, \end{cases} \tag{4.1}$$

$$\int_\Omega \left(\bar{\varphi} + \frac{\partial L}{\partial u}(x, \bar{y}, \bar{u})\right)(u - \bar{u})\,\mathrm{d}x \ge 0 \quad \forall u \in \mathcal{U}_{\alpha\beta}. \tag{4.2}$$

Relations (2.3b) and (2.3d) played an important role in the Proof of Theorem 3.1. Relations (4.1) and (4.2) can be used in a similar way to get the same regularity results. If L is independent of u, then we get from from (4.2) that $\bar{u}(x) = \alpha$ for a.a. $x \in B_\varepsilon(x^0)$. If L depends on u, then we additionally assume that L is of class C^2 with respect to u and

$$\exists \kappa > 0 \text{ such that } \frac{\partial^2 L}{\partial u^2}(x,y,u) \ge \kappa \text{ for a.a. } \ x \in \Omega \ \text{ and } \ \forall y \in \mathbb{R}.$$

Then, (4.2) leads to the formula

$$\bar{u}(x) = \mathrm{Proj}_{[\alpha,\beta]}(\bar{s}(x)),$$

where $\bar{s}(x)$ is (uniquely) defined through the equation

$$\bar{\varphi}(x) + \frac{\partial L}{\partial u}(x, \bar{y}(x), \bar{s}(x)) = 0 \quad \text{for a.a. } x \in \Omega;$$

see [1] for more details. The above equality and the assumption $\frac{\partial^2 L}{\partial u^2}(x,y,u) \ge \kappa$ implies that $\bar{s}(x) < 0$ is as small as needed assuming that $\bar{\varphi}(x) > M$ for M sufficiently large. To argue as in the Proof of Theorem 3.1, the assumption (2.2d) is not enough. We need new one involving α, β, a and b: $\alpha < Ab$ and $\beta > Aa$ in Ω. If a and b are constants satisfying $a < 0 < b$, then the precedent conditions are simplified: $\alpha < 0 < \beta$.

Remark 4.1. The conditions relating α with b and β with a cannot be removed. We provide a model example where b is constant and $\alpha \equiv 0$ and the multiplier is not an element in $H^{-1}(\Omega)$. Consider problem $(\tilde{P})$ as in Remark 3.4, but with $y_d(x) = |x|^4 - 4|x|^2 + 67$. The unique solution of this problem is given by $\tilde{u}(x) = 16(1-|x|^2)$, whose associate state is $\tilde{y}(x) = y_d(x) - 64$. Define $\alpha(x) \equiv 0$, fix some $0 < \epsilon < 1$, define $b = \epsilon\tilde{y}(0)$ and consider the problem

$$(P^0)\ \min J(u) = \frac{1}{2}\|y_u - y_d\|^2_{L^2(\Omega)} + \frac{1}{2}\|u\|^2_{L^2(\Omega)},\ u(x) \geq 0 \text{ for all } x \in \bar{\Omega},\ y(x_0) \leq b$$

Since $u = \frac{1}{2}\epsilon\tilde{u}$ is admissible for (P^0), it is clear that this problem has a unique solution $\bar{u}$. The optimality system reads like:

$$-\Delta\bar{y} = \bar{u} \text{ in } \Omega,\ y = 0 \text{ on } \Gamma, -\Delta\bar{\varphi} = \bar{y} - y_d + \bar{\lambda}\delta x_0 \text{ in } \Omega,\ \bar{\varphi} = 0 \text{ on } \Gamma,\ \bar{u}(x) = \max(-\bar{\varphi}(x), 0) \text{ in } \bar{\Omega}$$

Notice that the state constraint must be attained and the Lagrange multiplier $\bar{\lambda} \in \mathbb{R}$ associated to the state constraint must be strictly positive (in other case, the solution of the optimality system would be $\tilde{u}$, which is impossible). Notice also that due to the radial symmetry of the problem and the fact that $-\Delta\bar{y} = \bar{u} \geq 0$ in Ω, we gather that $\bar{y}$ attains its absolute maximum at x_0, with value b. Therefore $\bar{u}$ is the solution of problem

$$(P)\ \min J(u) = \frac{1}{2}\|y_u - y_d\|^2_{L^2(\Omega)} + \frac{1}{2}\|u\|^2_{L^2(\Omega)},\ u(x) \geq 0 \text{ for all } x \in \bar{\Omega},\ y(x) \leq b \text{ for all } x \in \bar{\Omega}$$

and the associated multiplier to the state constraint is $\bar{\mu} = \bar{\lambda}\delta_{x_0}$, which does not satisfy the claims of Theorem 3.1.

4.2. A semilinear elliptic equation

We can extend our results to semilinear equations replacing (1.1)

$$\begin{cases} Ay + a_0(x,y) = u & \text{in } \Omega, \\ y = 0 & \text{on } \Gamma, \end{cases}$$

with

$$Ay = -\sum_{i,j=1}^{n} \partial_{x_j}[a_{ij}(x)\,\partial_{x_i}y].$$

Here, $a_0 \colon \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function of class C^1 with respect to the second variable, with $a(\cdot, 0) \in L^p(\Omega)$ for some $p > \frac{n}{2}$, and satisfying

$$\begin{cases} \dfrac{\partial a_0}{\partial y}(x,y) \geq 0 \ \text{ for a.a. } \ x \in \Omega \ \text{ and } \ \forall y \in \mathbb{R}, \\ \forall M > 0\ \exists C_M > 0 \text{ s.t. } \dfrac{\partial a_0}{\partial y}(x,y) \leq C_M \ \text{ for a.a. } x \in \Omega \ \text{ and } \forall |y| \leq M. \end{cases}$$

In this situation, the regularity results of Theorem 3.1 still hold under the assumption (2.2d) and assuming that the functions $a_0(x, a(x))$ and $a_0(x, b(x))$ belong to $L^\infty(\Omega)$. The proof follows the same steps of the one of Theorem 3.1. It is enough to take into account the monotonicity of a_0 with respect to the second variable and to take

$$M = \|\varphi_0\|_{C(K)} + C_K\|\bar{\mu}^-\|_{\mathcal{M}(\Omega)} + \nu\left(\|Ab\|_{L^\infty(\Omega)} + \|a_0(x, a(x))\|_{L^\infty(\Omega)}\right).$$

Then, we have with $\tilde{y} = \bar{y} - b$

$$\begin{aligned} A\tilde{y} &= \bar{u}(x) - a_0(x, \bar{y}(x)) - Ab \\ &\leq -\frac{1}{\nu}\bar{\varphi}(x) - a_0(x, a(x)) - Ab < 0 \ \text{ in } B_\varepsilon\left(x^0\right), \end{aligned}$$

and we can use again the maximum principle to get the contradiction.

5. Numerical approximation

In this section, we make some additional assumptions on Ω as well as on the coefficients a_{ij} and on the bounds a, b, which avoid some technicalities and which are necessary to assure a higher regularity of the optimal state and adjoint state. This extra regularity is needed to improve the error estimates.

Assumption 5.1. Hereafter we will suppose that Ω is convex and $a_{ij} \in C^{1,\theta}(\bar{\Omega})$, for some $0 < \theta \leq 1$.

Under Assumptions 2.1 and 5.1, the solution y_u of (1.1) belongs to $H^2(\Omega) \cap H^1_0(\Omega)$ if $u \in L^2(\Omega)$, and there exists a constant C such that

$$\|y_u\|_{H^2(\Omega)} \leq C\|u\|_{L^2(\Omega)} \quad \forall u \in L^2(\Omega); \tag{5.1}$$

see in Chapter 3 of [20].

Let $\{\mathcal{T}_h\}_h$ be a quasi-uniform family of triangulations of $\bar{\Omega}$ and let Ω_h be the interior of $\cup\{T : T \in \mathcal{T}_h\}$. As usual, we assume that $|\Omega \setminus \Omega_h| \leq ch^2$. This holds if Γ is of class $C^{1,1}$ or if Ω is a polygonal or polyhedral domain. For the discretization of the control, the state and the adjoint state we use the space of linear finite elements $Y_{h0} \subset H^1_0(\Omega)$

$$Y_{h0} = \{y \in C(\bar{\Omega}) : y_h \in P^1(T) \ \forall T \in \mathcal{T}_h, \ y_h \equiv 0 \text{ in } \bar{\Omega} \setminus \Omega_h\}.$$

For the discrete Lagrange multiplier we use the space $\mathcal{M}_h \subset \mathcal{M}(\Omega)$ which is spanned by Dirac measures corresponding to the interior nodes $\{x_j\}_{j=1}^{n_h}$ of the finite element mesh. For every $u \in L^2(\Omega)$, $y_h(u)$ is the unique element in Y_{h0} such that

$$a_A(y_h(u), z_h) = \int_{\Omega_h} u z_h \, \mathrm{d}x \quad \forall z_h \in Y_{h0},$$

where $a_A \colon H^1_0(\Omega) \times H^1_0(\Omega) \to \mathbb{R}$ denotes the bilinear form associated to the elliptic operator A. Problem (P_h) reads like

$$(\mathrm{P}_h) \begin{cases} \min J_h(u) = \dfrac{1}{2}\|y_h(u) - y_d\|^2_{L^2(\Omega_h)} + \dfrac{\nu}{2}\|u\|^2_{L^2(\Omega_h)} \\ u \in Y_{h0}, \ y_h(u) \in \mathcal{Y}_{\mathrm{ab},h}, \end{cases}$$

where

$$\mathcal{Y}_{\mathrm{ab},h} = \{y_h \in Y_{h0} : \ a(x_j) \leq y_h(x_j) \leq b(x_j) \text{ for all } j = 1, \ldots, n_h\}.$$

Proposition 5.2. *Under Assumptions 2.1, 2.2 and 5.1, problem* (P_h) *has a unique solution* $\bar{u}_h \in Y_{h0}$, *with related state* $\bar{y}_h = y_h(\bar{u}_h)$. *Moreover, for every* $h \leq h_0$, *for some* $h_0 > 0$, *there exist* $\bar{\varphi}_h \in Y_{h0}$ *and* $\bar{\mu}_h \in \mathcal{M}_h$ *such that the following optimality system is satisfied.*

$$a_A(\bar{y}_h, z_h) = (\bar{u}_h, z_h) \ \forall z_h \in Y_{h0}, \tag{5.2a}$$

$$a_A(z_h, \bar{\varphi}_h) = (\bar{y}_h - y_d, z_h) + \int_{\Omega_h} z_h \, \mathrm{d}\bar{\mu}_h \ \forall z_h \in Y_{h0}, \tag{5.2b}$$

$$\int_{\Omega_h} (y_h - \bar{y}_h) \, \mathrm{d}\bar{\mu}_h \leq 0 \ \forall y_h \in \mathcal{Y}_{\mathrm{ab},h}, \tag{5.2c}$$

$$\varphi_h + \nu \bar{u}_h = 0, \tag{5.2d}$$

where $(\cdot,\cdot)$ *denotes the inner product in* $L^2(\Omega)$.

Before proving this proposition, we establish a technical lemma that we will use several times in the sequel.

Lemma 5.3. *Under Assumptions 2.1 and 5.1, we have the following estimates*

$$\|y_u - y_h(u)\|_{L^2(\Omega)} \le Ch\|u\|_{H^{-1}(\Omega)} \quad \forall u \in H^{-1}(\Omega), \tag{5.3a}$$

$$\|y_u - y_h(u)\|_{L^2(\Omega)} + h\|y_u - y_h(u)\|_{H^1(\Omega)} \le Ch^2\|u\|_{L^2(\Omega)} \quad \forall u \in L^2(\Omega), \tag{5.3b}$$

$$\|y_u - y_h(u)\|_{L^\infty(\Omega)} \le Ch^{2-\frac{n}{2}}\|u\|_{L^2(\Omega)} \quad \forall u \in L^2(\Omega), \tag{5.3c}$$

where the constants C are independent of h. Moreover if $\{u_h\}_{h>0} \subset L^2(\Omega)$ converges weakly to u in $L^2(\Omega)$, then

$$\lim_{h\to 0}\Big(\|y_u - y_h(u_h)\|_{H^1(\Omega)} + \|y_u - y_h(u_h)\|_{L^\infty(\Omega)}\Big) = 0. \tag{5.4}$$

Proof. For the proof of the estimates (5.3a)–(5.3c) the reader is referred *e.g.* to Chapter 3 in [13]. For the proof of (5.4) it is enough to use the triangle inequality and the above estimates as follows

$$\|y_u - y_h(u_h)\|_{H^1(\Omega)} \le \|y_u - y_{u_h}\|_{H^1(\Omega)} + \|y_{u_h} - y_h(u_h)\|_{H^1(\Omega)}$$
$$\le \|y_u - y_{u_h}\|_{H^1(\Omega)} + Ch\|u_h\|_{L^2(\Omega)}.$$

Now, the weak convergence $u_h \rightharpoonup u$ in $L^2(\Omega)$ implies the strong convergence $u_h \to u$ in $H^{-1}(\Omega)$, which gives the convergence to 0 of the first addend above. For the second addend it is enough to observe that $\{u_h\}_{h>0}$ is bounded in $L^2(\Omega)$. For the convergence in $L^\infty(\Omega)$ we proceed in a similar way, using (5.3c) instead of (5.3b). For the convergence of $y_{u_h} \to y_u$ in $L^\infty(\Omega)$, it is enough to observe that $u_h \to u$ in $W^{-1,p}(\Omega)$ for all $p < 6$, and apply (2.1). □

Proof of Proposition 5.2. Problem (P$_h$) is a finite dimensional strictly convex optimization problem and the constraints define a convex set, so existence and uniqueness of solution follow from the existence of an admissible control. First order necessary and sufficient optimality conditions follow in a standard way from the Slater condition. Therefore, we only need to prove that the Slater condition holds, which also implies that the set of admissible controls is nonempty. To this end, we take $u_0 \in L^2(\Omega)$ satisfying the Slater condition of problem (P), whose existence was shown in the proof of Theorem 2.3. Denote by u_{0h} the L^2-projection of u_0 on Y_{h0}. We know that $u_{h0} \to u_0$ strongly in $L^2(\Omega)$. Hence, from (5.4) we have that $y_h(u_{0h}) \to y_{u_0}$ strongly in $H_0^1(\Omega) \cap C_0(\Omega)$. Since $a(x) < y_{u_0}(x) < b(x)$ $\forall x \in \bar\Omega$, we deduce the existence of $h_0 > 0$ such that every $y_h(u_{h0})$ satisfies the same strict inequalities for all $h \le h_0$. □

Analogously to (2.8), we can write $\bar\mu_h = \mu_h^+ - \mu_h^-$, with both $\mu_h^\pm \ge 0$, and we deduce from (5.2c)

$$\begin{cases} \operatorname{supp}\bar\mu_h^+ \subset \{x_j : \bar y_h(x_j) = b(x_j)\}, \\ \operatorname{supp}\bar\mu_h^- \subset \{x_j : \bar y_h(x_j) = a(x_j)\}. \end{cases} \tag{5.5}$$

Therefore, we have

$$\bar\mu_h = \sum_{j=1}^{n_h} \bar\lambda_j \delta_{x_j}, \quad \text{with} \quad \bar\lambda_j \begin{cases} \ge 0 \text{ if } \bar y_h(x_j) = b(x_j), \\ \le 0 \text{ if } \bar y_h(x_j) = a(x_j), \end{cases} \tag{5.6}$$

where δ_{x_j} denotes the Dirac measure centered at x_j.

In the next theorem, we analyze the convergence of the solutions of problems (P$_h$) as well as the convergence of the optimality system (5.2a)–(5.2d).

Theorem 5.4. *Let h_0 be as in Proposition 5.2. Under the Assumptions 2.1, 2.2 and 5.1, for every $h \le h_0$ the system (5.2a)–(5.2d) has a unique solution and the following convergence holds*

$$(\bar u_h, \bar y_h, \bar\varphi_h, \bar\mu_h) \to (\bar u, \bar y, \bar\varphi, \bar\mu)$$

strongly in $L^2(\Omega) \times [H_0^1(\Omega) \cap C(\bar\Omega)] \times W_0^{1,s}(\Omega) \times W^{-1,s}(\Omega)$ as $h \to 0$ for all $s < \frac{n}{n-1}$.

Proof. In a similar way to Theorem 2.3, the uniqueness of $\bar\varphi_h$ and $\bar\mu_h$ follows from (5.2d) and (5.2b), respectively. Let us prove the convergence of $(\bar u_h, \bar y_h, \bar\varphi_h, \bar\mu_h)$. First, we observe that the optimality of $\bar u_h$ implies

$$\frac{\nu}{2}\|\bar u_h\|^2_{L^2(\Omega)} \le J_h(\bar u_h) \le J_h(u_{0h}),$$

where u_{0h} are the admissible discrete controls found in the Proof of Proposition 5.2. The convergences indicated in the afore-mentioned proposition imply that $\{u_{0h}\}$ and $\{y_{0h}\}$ are bounded in $L^2(\Omega)$, and hence $\{\bar u_h\}$ is bounded in $L^2(\Omega)$. Now, taking a subsequence if necessary, we can assume that $\{\bar u_h\}_{h>0}$ converges weakly in $L^2(\Omega)$. With (5.4), this leads to the strong convergence of the associated discrete states $\{\bar y_h\}_{h>0}$ in $H^1_0(\Omega) \cap C_0(\Omega)$. The proof of the convergence of $\{\bar u_h\}_{h>0}$ to $\bar u$, the solution of (P), is standard. Even more, due to the structure of the cost functional, we have that $\bar u_h \to \bar u$ strongly in $L^2(\Omega)$. We also have that $\bar y_h \to \bar y$ strongly in $H^1_0(\Omega)\cap C(\bar\Omega)$; see, for instance, [8] for the details. Moreover, from (5.2d) and (2.3d), we deduce the strong convergence $\bar\varphi_h \to \bar\varphi$ in $L^2(\Omega)$.

Let us prove the boundedness of $\{\bar\mu_h\}_{h>0}$. To this end, we take u_0 and $\{u_{0h}\}_{h\le h_0}$ as in the Proof of Proposition 5.2: $u_{h0} \to u_0$ in $L^2(\Omega)$ and u_{0h} satisfies the Slater condition for problem (P$_h$). Denote by y_{0h} the discrete state associated with u_{0h}. We have that $y_{0h} \to y_{u_0}$ strongly in $H^1_0(\Omega)\cap C_0(\Omega)$. Select a number $\rho > 0$ such that

$$a(x_j) < y_{h0}(x_j) - \rho < y_{h0}(x_j) + \rho < b(x_j), \quad j = 1, \dots n_h,\ \forall h \le h_0,$$

which is possible because u_0 also satisfies the Slater condition for problem (P). Consider the element $y_h \in Y_{h0}$ given by

$$y_h(x_j) = \begin{cases} +\rho \text{ if } \bar\lambda_j \ge 0, \\ -\rho \text{ if } \bar\lambda_j < 0. \end{cases}$$

Then, $y_h + y_{0h} \in \mathcal{Y}_{\text{ab},h}$ and using (5.2c) we get

$$\begin{aligned}\rho\|\bar\mu_h\|_{\mathcal{M}(\Omega)} = \rho\sum_{j=1}^{n_h}|\bar\lambda_j| &= \int_{\Omega_h} y_h \,\mathrm{d}\bar\mu_h \le \int_{\Omega_h}(\bar y_h - y_{0h})\,\mathrm{d}\bar\mu_h \\ &= a_A(\bar y_h - y_{0h}, \bar\varphi_h) - \int_{\Omega_h}(\bar y_h - y_d)(\bar y_h - y_{0h})\,\mathrm{d}x \\ &= \int_{\Omega_h}(\bar u_h - u_{0h})\bar\varphi_h\,\mathrm{d}x - \int_{\Omega_h}(\bar y_h - y_d)(\bar y_h - y_{0h})\,\mathrm{d}x \le C.\end{aligned}$$

Since $\{\bar\mu_h\}_{h>0}$ is bounded in $\mathcal{M}(\Omega)$, we can take a subsequence such that $\bar\mu_h \overset{*}{\rightharpoonup} \bar\mu$ in $\mathcal{M}(\Omega)$. From the compactness of the embedding $\mathcal{M}(\Omega) \subset W^{-1,s}(\Omega)$ for every $1 \le s < \frac{n}{n-1}$, we infer the strong convergence $\bar\mu_h \to \bar\mu$ in $W^{-1,s}(\Omega)$ for every $1 \le s < \frac{n}{n-1}$. From here, using $W^{1,s}$ stability of the Ritz-projection, see ([5], Sect. 7.5), we deduce the strong convergence $\bar\varphi_h \to \bar\varphi$ in $W^{1,s}_0(\Omega)$ for every $1 \le s < \frac{n}{n-1}$. Thus, we have proved that $(\bar u_h, \bar y_h, \bar\varphi_h, \bar\mu_h) \to (\bar u, \bar y, \bar\varphi, \bar\mu)$ strongly in $L^2(\Omega)\times[H^1_0(\Omega)\cap C(\bar\Omega)]\times W^{1,s}_0(\Omega)\times W^{-1,s}(\Omega)$ as $h \to 0$. Observe that no subsequence is necessary because any subsequence has the same limit. □

Corollary 5.5. *There exists $\bar h_0 \le h_0$ and an open set $\Omega_0 \subset \bar\Omega_0 \subset \Omega$ such that* $\operatorname{supp}\bar\mu \subset \Omega_0$ *and* $\operatorname{supp}\bar\mu_h \subset \Omega_0$ *for every $h \le \bar h_0$.*

Proof. From (2.2a)–(2.2c) we deduce the existence of $\varepsilon > 0$ and $\rho > 0$ such that

$$a(x) + \rho < \bar y(x) < b(x) - \rho \quad \text{if } d(x, \Gamma) \le \varepsilon.$$

From the uniform convergence $\bar y_h \to \bar y$ in $C_0(\Omega)$, we deduce the existence of $\bar h_0 \in (0, h_0]$ such that

$$a(x) + \frac{\rho}{2} < \bar y_h(x) < b(x) - \frac{\rho}{2} \quad \text{if } d(x,\Gamma) \le \varepsilon,\ \forall h \le \bar h_0.$$

Now, we set

$$\Omega_0 = \{x \in \Omega : d(x, \Gamma) > \varepsilon\}.$$

Obviously, if $\bar{y}_h(x_j) = a(x_j)$ or $\bar{y}_h(x_j) = b(x_j)$, then $x_j \in \Omega_0$. Hence, the proof is concluded by (5.5). □

To every control $u \in L^2(\Omega)$ we will relate $\varphi_0(u) \in H_0^1(\Omega)$ and $\varphi_{0,h}(u) \in Y_{h0}$ the unique solutions of, respectively

$$\begin{aligned} a_A(z, \varphi_0(u)) &= (y_u - y_d, z) \quad \forall z \in H_0^1(\Omega) \\ a_A(z_h, \varphi_{0,h}(u)) &= (y_h(u) - y_d, z_h) \quad \forall z_h \in Y_{h0} \end{aligned}$$

and for every $\mu \in \mathcal{M}(\Omega)$, we define $\varphi(\mu) \in W^{1,s}(\Omega)$ for all $s < n/(n-1)$ and $\varphi_h(\mu) \in Y_{h0}$ as the unique solutions of (2.6) and

$$a_A(z_h, \varphi_h(\mu)) = \int_\Omega z_h \,\mathrm{d}\mu \quad \forall z_h \in Y_{h0},$$

respectively. In this way, we can also split $\bar{\varphi}_h = \varphi_{0,h}(\bar{u}_h) + \varphi_h(\bar{\mu}_h)$ and $\bar{\varphi} = \varphi_0(\bar{u}) + \varphi(\bar{\mu})$.

Theorem 5.6. *Let Ω_0 be as in Corollary 5.5 and assume that $a, b \in W^{2,\infty}(\Omega_0)$. Let $\bar{u}$ and $\bar{u}_h$ be the solutions of problems* (P) *and* (P$_h$)*, respectively. Then, there exists $C > 0$ independent of $h \leq \bar{h}_0$*

$$\begin{aligned} \|\bar{u} - \bar{u}_h\|^2_{L^2(\Omega)} \leq C\Big(&\|\varphi_0(\bar{u}) - \varphi_{0,h}(\bar{u})\|^2_{L^2(\Omega)} + \|\varphi(\bar{\mu}) - \varphi_h(\bar{\mu})\|^2_{L^2(\Omega)} \\ &+ \|\bar{y} - y_h(\bar{u})\|_{L^\infty(\Omega_0)} + h^2 \left(\|a\|_{W^{2,\infty}(\Omega_0)} + \|b\|_{W^{2,\infty}(\Omega_0)}\right)\Big). \end{aligned}$$

Proof. Since the problem is linear quadratic, we have

$$(\varphi_0(\bar{u}) - \varphi_0(\bar{u}_h), \bar{u} - \bar{u}_h) = a_A(\bar{y} - y_{\bar{u}_h}, \varphi_0(\bar{u}) - \varphi_0(\bar{u}_h)) = (\bar{y} - y_{\bar{u}_h}, \bar{y} - y_{\bar{u}_h}) \geq 0,$$

and hence

$$\begin{aligned} \nu\|\bar{u} - \bar{u}_h\|^2_{L^2(\Omega)} &\leq (\nu(\bar{u} - \bar{u}_h) + \varphi_0(\bar{u}) - \varphi_0(\bar{u}_h), \bar{u} - \bar{u}_h) \\ &= (\nu\bar{u} + \varphi_0(\bar{u}), \bar{u} - \bar{u}_h) + (\varphi_{0,h}(\bar{u}_h) - \varphi_0(\bar{u}_h), \bar{u} - \bar{u}_h) \\ &\quad - (\nu\bar{u}_h + \varphi_{0,h}(\bar{u}_h), \bar{u} - \bar{u}_h) \\ &= -(\varphi(\bar{\mu}), \bar{u} - \bar{u}_h) + (\varphi_{0,h}(\bar{u}_h) - \varphi_0(\bar{u}_h), \bar{u} - \bar{u}_h) \\ &\quad + (\varphi_h(\bar{\mu}_h), \bar{u} - \bar{u}_h) \\ &= -(\varphi(\bar{\mu}) - \varphi_h(\bar{\mu}), \bar{u} - \bar{u}_h) + (\varphi_{0,h}(\bar{u}_h) - \varphi_0(\bar{u}_h), \bar{u} - \bar{u}_h) \\ &\quad + (\varphi_h(\bar{\mu}_h) - \varphi_h(\bar{\mu}), \bar{u} - \bar{u}_h). \end{aligned}$$

Therefore, we have that

$$\|\bar{u} - \bar{u}_h\|^2_{L^2(\Omega)} \leq C\Big(\|\varphi(\bar{\mu}) - \varphi_h(\bar{\mu})\|^2_{L^2(\Omega)} + \|\varphi_0(\bar{u}_h) - \varphi_{0,h}(\bar{u}_h)\|^2_{L^2(\Omega)} + (\varphi_h(\bar{\mu}_h) - \varphi_h(\bar{\mu}), \bar{u} - \bar{u}_h)\Big).$$

We have just to get an estimate for the last term. By means of the discrete state equation, the definition of $\varphi_h(\bar{\mu})$ and $\varphi_h(\bar{\mu}_h)$ and the decomposition of the measures, we obtain

$$\begin{aligned} (\varphi_h(\bar{\mu}_h) - \varphi_h(\bar{\mu}), \bar{u} - \bar{u}_h) &= a_A(y_h(\bar{u}) - \bar{y}_h, \varphi_h(\bar{\mu}_h) - \varphi_h(\bar{\mu})) \\ &= \langle \bar{\mu}_h - \bar{\mu}, y_h(\bar{u}) - \bar{y}_h \rangle \\ &= \langle \bar{\mu}^+, \bar{y}_h - y_h(\bar{u}) \rangle - \langle \bar{\mu}^-, \bar{y}_h - y_h(\bar{u}) \rangle \\ &\quad + \langle \bar{\mu}_h^+, y_h(\bar{u}) - \bar{y}_h \rangle - \langle \bar{\mu}_h^-, y_h(\bar{u}) - \bar{y}_h \rangle. \end{aligned}$$

For the first two addends we introduce the nodal interpolation operator $I_h\colon C_0(\Omega) \to Y_{h0}$ and use that $\bar{y} = b$ on $\operatorname{supp}\mu^+$, $\bar{y} = a$ on $\operatorname{supp}\mu^-$, $I_h a \leq \bar{y}_h \leq I_h b$ and the estimates for the error of interpolation to obtain

$$\begin{aligned}
&\langle \bar{\mu}^+, \bar{y}_h - y_h(\bar{u})\rangle - \langle \bar{\mu}^-, \bar{y}_h - y_h(\bar{u})\rangle \\
&\leq \langle \bar{\mu}^+, b - y_h(\bar{u})\rangle + \langle \bar{\mu}^+, I_h b - b\rangle + \langle \bar{\mu}^-, -a + y_h(\bar{u})\rangle + \langle \bar{\mu}^-, a - I_h a\rangle \\
&= \langle \bar{\mu}^+, b - \bar{y}\rangle - \langle \bar{\mu}^-, a - \bar{y}\rangle + \langle \bar{\mu}^+ - \bar{\mu}^-, \bar{y} - y_h(\bar{u})\rangle + \langle \bar{\mu}^+, I_h b - b\rangle + \langle \bar{\mu}^-, a - I_h a\rangle \\
&= \langle \bar{\mu}, \bar{y} - y_h(\bar{u})\rangle \ + \langle \bar{\mu}^+, I_h b - b\rangle + \langle \bar{\mu}^-, a - I_h a\rangle \\
&\leq \|\bar{\mu}\|_{\mathcal{M}(\Omega)} \Big(\|\bar{y} - y_h(\bar{u})\|_{L^\infty(\Omega_0)} + Ch^2 (\|a\|_{W^{2,\infty}(\Omega_0)} + \|b\|_{W^{2,\infty}(\Omega_0)}) \Big).
\end{aligned}$$

To finish, we use that $\bar{y}_h = b$ on $\operatorname{supp}\bar{\mu}_h^+$, $\bar{y} - b \leq 0$, $\bar{y}_h = a$ on $\operatorname{supp}\bar{\mu}_h^-$ and $\bar{y} - a \geq 0$ to obtain

$$\begin{aligned}
&\langle \bar{\mu}_h^+, y_h(\bar{u}) - \bar{y}_h\rangle - \langle \bar{\mu}_h^-, y_h(\bar{u}) - \bar{y}_h\rangle = \langle \bar{\mu}_h^+, y_h(\bar{u}) - b\rangle - \langle \bar{\mu}_h^-, y_h(\bar{u}) - a\rangle \\
&= \langle \bar{\mu}_h^+, y_h(\bar{u}) - \bar{y}\rangle + \langle \bar{\mu}_h^+, \bar{y} - b\rangle - \langle \bar{\mu}_h^-, y_h(\bar{u}) - \bar{y}\rangle - \langle \bar{\mu}_h^-, \bar{y} - a\rangle \leq \langle \bar{\mu}_h, y_h(\bar{u}) - \bar{y}\rangle \\
&\leq \|\bar{\mu}_h\|_{\mathcal{M}(\Omega)} \|\bar{y} - y_h(\bar{u})\|_{L^\infty(\Omega_0)}
\end{aligned}$$

which completes the proof. □

Corollary 5.7. *Under the assumptions of Theorem 5.6, there exists $C > 0$ independent of $h \leq \bar{h}_0$ such that*

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)} \leq Ch|\log h|. \tag{5.7}$$

Proof. Usual finite element error estimates for regular problems lead to the estimate:

$$\|\varphi_0(\bar{u}) - \varphi_{0,h}(\bar{u})\|_{L^2(\Omega)} \leq Ch^2 \|\bar{u}\|_{L^2(\Omega)}.$$

Thanks to the regularity stated in our main result, Theorem 3.1, we have that $\bar{\mu} \in H^{-1}(\Omega)$ and $\varphi(\bar{\mu}) \in H_0^1(\Omega)$. Again usual finite element error estimates lead to

$$\|\varphi(\bar{\mu}) - \varphi_h(\bar{\mu})\|_{L^2(\Omega)} \leq Ch \|\varphi(\bar{\mu})\|_{H_0^1(\Omega)} \leq Ch \|\bar{\mu}\|_{H^{-1}(\Omega)}.$$

The estimates for the final term follow from ([30], Thm. 5.1), the error estimates for the error of interpolation, (5.3b) and (3.3).

Indeed, let us take and open set Ω' satisfying $\bar{\Omega}_0 \subset \Omega' \subset \bar{\Omega}' \subset \Omega$, then for every $p < \infty$ we have

$$\begin{aligned}
\|\bar{y} - y_h(\bar{u})\|_{L^\infty(\Omega_0)} &\leq C\Big(|\log h| \|\bar{y} - I_h \bar{y}\|_{L^\infty(\Omega')} + \|\bar{y} - y_h(\bar{u})\|_{L^2(\Omega')} \Big) \\
&\leq C\Big(|\log h| h^{2-\frac{n}{p}} \|\bar{y}\|_{W^{2,p}(\Omega')} + h^2 \|\bar{u}\|_{L^2(\Omega)} \Big) \\
&\leq C\Big(p|\log h| h^{2-\frac{n}{p}} \|\bar{u}\|_{L^\infty(\Omega)} + h^2 \|\bar{u}\|_{L^2(\Omega)} \Big).
\end{aligned}$$

Now, taking $p = |\log h|$,

$$\|\bar{y} - y_h(\bar{u})\|_{L^\infty(\Omega_0)} \leq Ch^2 |\log h|^2 \|\bar{u}\|_{L^\infty(\Omega)}$$

and the proof is complete. □

TABLE 1. Coefficients for $\hat{y}_{d1}$, $\hat{y}_{d2}$ and $\hat{b}_0$.

	$\hat{y}_{d1}$	$\hat{y}_{d2}$	$\hat{b}_0$
r^6	-31.434189678717527	0	-31.434189678717527
r^5	$+46.159111917905442$	$+12.13291524916886$	$+ 46.159111917905442$
r^4	-18.560977582095056	-17.996278998962907	-18.560977582095056
r^3	0	$+7.3349388302866316$	0
r^2	-2.6404719330122726	-1.1774263620970444	0
r	$+1.6617280290445962$	0	0
1	$+0.26726826901485923$	$+0.5$	$+0.49$

6. Numerical evidence

In this section we present the numerical results obtained for a control problem in dimension 3, that confirm our theoretical error estimates. The reader is referred to [15] for an example in dimension 2.

Let Ω be the unit ball in $\mathbb{R}^3$ and Γ its boundary. We are concerned with the problem

$$\min \frac{1}{2}\|y_u - y_d\|^2 + \frac{\nu}{2}\|u\|^2,$$

$$-\Delta y = u \text{ in } \Omega,\ \ y = 0 \text{ on } \Gamma,$$

$$a \leq y \leq b \text{ in } \bar{\Omega}.$$

We build an example with spherical symmetry. Let $r = |x|$ be the euclidean distance of x to the origin. All our data are piecewise polynomial in r. We fix some of the parameters that appear in the functions, and the others are computed imposing the optimality conditions and some differentiability properties. For the convenience of the reader, we have written the coefficients of the polynomials (computed in double precision) in Table 1, starting in every case with the coefficient of r^6.

Let us take $\nu = 10^{-4}$ and define y_d, a and b as follows:

$$y_d(x) = \begin{cases} \hat{y}_{d1}(r) & \text{if } r \leq 0.25, \\ \hat{y}_{d2}(r) & \text{if } 0.25 \leq r \leq 0.75, \\ 0.375r^2 - 1.125r + 0.75 & \text{if } r \geq 0.75, \end{cases}$$

$$a(x) = -1 \text{ and } b(x) = b_0(x) + \begin{cases} (r-0.25)^2 & \text{if } r < 0.25, \\ 0 & \text{if } 0.25 \leq r \leq 0.75, \\ (r-0.75)^2 & \text{if } r > 0.75, \end{cases}$$

$$b_0(x) = \begin{cases} \hat{b}_0(r) & \text{if } r \leq 0.485, \\ 0.375r^2 - 1.125r + 0.75 & \text{if } r \geq 0.485. \end{cases}$$

This problem has a unique solution

$$\bar{u}(x) = \begin{cases} \hat{u}(r) & \text{if } r \leq 0.485, \\ -2.25 + \dfrac{2.25}{r} & \text{if } r \geq 0.485, \end{cases}$$

TABLE 2. Coefficients for $\hat{u}$ and $\hat{\mu}_1$ and $\hat{\mu}_2$.

	$\hat{u}$	$\hat{\mu}_1$	$\hat{\mu}_2$
r^6	0	$+31.434189678717527$	0
r^5	0	-34.026196668736581	$+12.13291524916886$
r^4	$+1320.235966506136$	$+0.56469858313214871$	-17.996278998962907
r^3	-1384.7733575371633	$+7.3349388302866316$	7.3349388302866316
r^2	$+371.21955164190115$	$+1.4630455709152281$	-1.5524263620970444
r	0	-1.6617280290445962	$+1.125$
1	0	0.23273173098514077	-0.25

TABLE 3. Mesh size, error and experimental order of convergence

h	$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}$	EOC
0.43	1.56	–
0.26	0.82	1.31
0.15	0.55	0.75
0.084	0.272	1.15
0.043	0.136	1.06
0.022	0.0673	1.05

where $\hat{u}$ is a polynomial whose coefficientes are shown in Table 2. The optimal state is $\bar{y} = b_0$ and the related adjoint state is $\bar{\varphi} = -\nu\bar{u}$. The lower constraint a is not attained. The multiplier $\bar{\mu}$ can be written as the sum of a regular part $\mu_r \in L^\infty(\Omega)$ plus a singular part μ_s. The regular part is a piecewise polynomial function

$$\mu_r(x) = \begin{cases} 0 & \text{if } r < 0.25, \\ \hat{\mu}_1(r) & \text{if } 0.25 < r < 0.485, \\ \hat{\mu}_2(r) & \text{if } 0.485 < r < 0.75, \\ 0 & \text{if } r > 0.75. \end{cases}$$

The coefficients of the corresponding polynomials $\hat{\mu}_i$ are given in Table 2. The singular part can be written as

$$\langle \mu_s, z \rangle = \iint_{r=0.485} \mu_0 z(x) \mathrm{d}\sigma(x) \ \forall z \in H_0^1(\Omega)$$

with $\mu_0 = -0.000448550576 16469$.

Notice that $b \in W^{2,\infty}(\Omega_0)$ for any open set Ω_0 satisfying $\operatorname{supp} \bar{\mu} \subset \Omega_0 \subset \bar{\Omega}_0 \subset \Omega$, with $0 \notin \bar{\Omega}_0$. Hence the estimate (5.7) holds.

We have solved the finite element approximation of the control problem using an active set strategy as described in [2]. We have obtained the results summarized in Table 3, where *EOC* denotes the experimental order of convergence. These numbers show the sharpness of our theoretical results. Our finest mesh has more then 5×10^6 tetrahedra and almost 9×10^5 nodes.

REFERENCES

[1] N. Arada, E. Casas and F. Tröltzsch, Error estimates for the numerical approximation of a semilinear elliptic control problem. *Comput. Optim. Appl.* **23** (2002) 201–229.

[2] M. Bergounioux and K. Kunisch, Primal-dual strategy for state-constrained optimal control problems. *Comput. Optim. Appl.* **22** (2002) 193–224.

[3] M. Bergounioux and K. Kunisch, On the structure of Lagrange multipliers for state-constrained optimal control problems. *Systems Control Lett.* **48** (2003) 169–176. Optimization and control of distributed systems.

[4] H. Blum and R. Rannacher, On the boundary value problem of the biharmonic operator on domains with angular corners. *Math. Methods Appl. Sci.* **2** (1980) 556–581.

[5] S.C. Brenner and L.R. Scott, *The Mathematical Theory of Finite Element Methods.* Springer-Verlag, New York, Berlin, Heidelberg (1994).

[6] E. Casas, Control of an elliptic problem with pointwise state constraints. *SIAM J. Control Optim.* **24** (1986) 1309–1318.

[7] E. Casas, J.C. de los Reyes and F. Tröltzsch, Sufficient second order optimality conditions for semilinear control problems with pointwise state constraints. *SIAM J. Optim.* **19** (2008) 616–643.

[8] E. Casas and M. Mateos, Uniform convergence of the FEM. Applications to state constrained control problems. *Comput. Appl. Math.* **21** (2002) 67–100.

[9] E. Casas, Error estimates for the numerical approximation of semilinear elliptic control problems with finitely many state contraints. *ESAIM: COCV* **8** (2002) 345–374.

[10] E. Casas and M. Mateos, Numerical approximation of elliptic control problems with finitely many pointwise constraints. *Comput. Optim. Appl.* **51** (2012) 1319–1343.

[11] E. Casas and F. Tröltzsch, Recent advances in the analysis of pointwise state-constrained elliptic optimal control problems. *ESAIM: COCV* **16** (2010) 581–600.

[12] S. Cherednichenko, K. Krumbiegel and A. Rösch, Error estimates for the Lavrentiev regularization of elliptic optimal control problems. *Inverse Problems* **24** (2008) 21.

[13] P.G. Ciarlet, Basic error estimates for elliptic problems, in *Handbook of numerical analysis, Vol. II*, *Handb. Numer. Anal.*, II. North-Holland, Amsterdam (1991) 17–351

[14] G. Dal Maso, F. Murat, L. Orsina and A. Prignet, Renormalized solutions of elliptic equations with general measure data. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **28** (1999) 741–808.

[15] K. Deckelnick and M. Hinze, Convergence of a finite element approximation to a state constrained elliptic control problem. *SIAM J. Numer. Anal.* **35** (2007) 1937–1953.

[16] K. Deckelnick and M. Hinze, Numerical analysis of a control and state constrained elliptic control problem with piecewise constant control approximations, in Proc. of ENUMATH, 2007. *Numer. Math. Advanced Appl.*, edited by K. Kunisch, G. Of and O. Steinbach. Springer, Berlin (2008) 597–604.

[17] M. Degiovanni and M. Scaglia, A variational approach to semilinear elliptic equations with measure data. *Discrete Contin. Dyn. Syst.* **31** (2011) 1233–1248.

[18] D. Gilbarg and N.S. Trudinger, Elliptic partial differential equations of second order. *Classics in Math.* Reprint of the 1998 edition. Springer-Verlag, Berlin (2001).

[19] W. Gong and N. Yan, A mixed finite element scheme for optimal control problems with pointwise state constraints. *J. Sci. Comput.* **46** (2011) 182–203.

[20] P. Grisvard, *Elliptic Problems in Nonsmooth Domains.* Pitman, Boston-London-Melbourne, 1985.

[21] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, *Optimization with PDE constraints*, vol. 23. *Math. Model.: Theory Appl.* Springer, New York (2009).

[22] D. Leykekhman, D. Meidner and B. Vexler, Optimal error estimates for finite element discretization of elliptic optimal control problems with finitely many pointwise state constraints. *Comput. Optim. Appl.* **55** (2013) 769–802.

[23] W. Liu, W. Gong and N. Yan, A new finite element approximation of a state-constrained optimal control problem. *J. Comput. Math.* **27** (2009) 97–114.

[24] C. Meyer, Error estimates for the finite-element approximation of an elliptic control problem with pointwise state and control constraints. *Control Cybernet* **37** (2008) 51–83.

[25] C. Meyer, U. Prüfert and Tröltzsch, On two numerical methods for state-constrained elliptic control problems. *Optim. Methods Softw.* **22** (2007) 871–899.

[26] C. Meyer, Error estimates for the finite-element approximation of an elliptic control problem with pointwise state and control constraints. *Control and Cybernetics* **37** (2008) 51–85.

[27] K. Pieper and B. Vexler, *A priori* error analysis for discretization of sparse elliptic optimal control problems in measure space. *SIAM J. Control Optim.* **51** (2013) 2788–2808.

[28] A. Rösch and S. Steinig, *A priori* error estimates for a state-constrained elliptic optimal control problem. *ESAIM: M2AN* **46** (2012) 1107–1120.

[29] W. Rudin, *Real and Complex Analysis.* McGraw-Hill, London (1970).

[30] A.H. Schatz and L.B. Wahlbin, Interior maximum norm estimates for finite element methods. *Math. Comput.* **31** (1977) 414–442.

[31] G. Stampacchia, Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. *Ann. Inst. Fourier, Grenoble* **15** (1965) 189–258.