

A general theorem on error estimates with application to a quasilinear elliptic optimal control problem *

Eduardo Casas[†] Fredi Tröltzsch[‡]

October 13, 2011

Abstract

A theorem on error estimates for smooth nonlinear programming problems in Banach spaces is proved that can be used to derive optimal error estimates for optimal control problems. This theorem is applied to a class of optimal control problems for quasilinear elliptic equations. The state equation is approximated by a finite element scheme, while different discretization methods are used for the control functions. The distance of locally optimal controls to their discrete approximations is estimated.

AMS subject classification. 35J60, 35J62, 49K20, 49M25, 65N30

Key words. Quasilinear elliptic equations, optimal control problems, optimality conditions, finite element approximation, error estimates

1 Introduction

This paper has two main aims. First, we develop a general theory of error analysis for smooth nonlinear programming problems in Banach spaces, which is applicable in particular to optimal control problems. Second, as main application, we prove new error estimates for optimal control problems governed by a class of quasilinear elliptic equations.

*This work was supported by Spanish Ministerio de Ciencia e Innovación under the project MTM2008-04206.

[†]Departamento 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).

[‡]Institut für Mathematik, Technische Universität Berlin, D-10623 Berlin, Germany (troeltzsch@math.tu-berlin.de).

Roughly speaking, the following question stands behind our general theory: Let a control problem be discretized by a finite element scheme, where the associated grid has mesh size h . How large is the difference between a locally optimal control and its numerical approximation, if the state equation is handled by a finite element method and the controls are discretized appropriately?

In the last decade, the associated error analysis of optimal control problems for partial differential equations (PDEs) has made considerable progress. This refers in particular to the case of elliptic PDEs. Here, various error estimates are known that can be used to quantify the discretization error for the optimal solution(s) of the control problem. For control-constrained elliptic problems, we mention only the early papers for linear-quadratic elliptic control problems by Falk [17] and Geveci [18], the discussion of semilinear elliptic control problems in [2], [8], error estimates for Dirichlet boundary control problems in [9], [14], the concept of variational discretization in [20], and the investigation of superconvergence in [22]. In the recent past, state-constrained control problems received more attention. For instance, we refer to recent contributions by [6], [15], or [16]. This list of papers could be extended; we refer to the references in the survey paper [21].

Nevertheless, there are still important classes of optimal control problems, where reliable a priori error estimates are desired but now yet proved. For instance, because of intrinsic technical difficulties, the class of optimal control problems for quasilinear elliptic equations was not yet considered. The optimal control of coupled systems of quasilinear PDEs leads to additional difficulties.

We observed that very similar ideas are frequently repeated in any new contribution to this field. A general theorem on error estimates might save tremendous work in future investigations on other types of control problems. Therefore, in the first part of the paper, we provide a general tool for deriving error estimates for the optimal control under control constraints. In the second part, our general analysis is applied to an optimal control problem governed by a class of quasilinear elliptic PDEs. We show the error estimates stated in [11] without proof. Our general result should also be applicable to other classes of control problems, in particular if the state equation is of parabolic type. It is needed that the partial differential equation under consideration and its numerical approximation obey certain regularity properties of their solutions.

We consider the following quasilinear optimal control problem (P),

$$(P) \quad \begin{cases} \min J(u) := \int_{\Omega} L(x, y_u(x), u(x)) dx, \\ \alpha \leq u(x) \leq \beta \text{ for a.e. } x \in \Omega, \end{cases}$$

where y_u is the solution of the state equation

$$\begin{cases} -\operatorname{div}[a(x, y(x)) \nabla y(x)] + f(x, y(x)) = u(x) & \text{in } \Omega \\ y(x) = 0 & \text{on } \Gamma. \end{cases} \quad (1.1)$$

We began with the numerical analysis of (P) in [11], where we studied the finite

element approximation of the equation (1.1), its linearization and its adjoint equation. Moreover, we were able to prove that to any strict locally optimal control \bar{u} there exists a sequence of locally optimal controls \bar{u}_h of the associated discretized optimal control problem that converges to \bar{u} as the mesh size h tends to zero. Here, we will prove the error estimates announced in [11].

The objective functional J of (P) will lead to the well-known two-norm discrepancy: It is twice continuously Fréchet-differentiable in $L^\infty(\Omega)$ but not in general in $L^2(\Omega)$. On the other hand, we need a second-order sufficient optimality condition, which can only be satisfied in the norm of $L^2(\Omega)$. This difficulty complicates the estimation of the error and requires the use of two norms.

With the following simpler functional, the two-norm discrepancy does not occur:

$$J(u) := \int_{\Omega} \left\{ \ell(x, y_u(x)) + \frac{\Lambda}{2} u(x)^2 \right\} dx. \quad (1.2)$$

It is twice continuously differentiable in $L^2(\Omega)$.

2 A Unified Theory of Error Estimates

2.1 An Abstract Optimization Problem and Optimality Conditions

Let U_∞ and U_2 be Banach and Hilbert spaces, respectively, endowed with the norms $\|\cdot\|_\infty$ and $\|\cdot\|_2$. We assume that $U_\infty \subset U_2$ with continuous embedding. In particular, the choice $U_\infty = U_2$ is possible. The latter case is of interest for problems, where the two-norm discrepancy does not occur as for the functional (1.2).

We denote by \mathcal{K} a nonempty convex subset of U_∞ that is closed in U_2 . Moreover, an objective function $J : U_\infty \rightarrow \mathbb{R}$ is given. With these quantities, we define the abstract optimization problem

$$(\mathcal{P}) \quad \min_{u \in \mathcal{K}} J(u).$$

For the well-posedness of the problem, we require the next assumption.

(A1) *The function J satisfies the following properties:*

$$\text{If } \{u_k\}_{k=1}^\infty \subset \mathcal{K} \text{ and } u_k \rightharpoonup u \text{ in } U_2, \text{ then } J(u) \leq \liminf_{k \rightarrow \infty} J(u_k). \quad (2.1)$$

$$\left\{ \begin{array}{l} \text{If } \mathcal{K} \text{ is unbounded in } U_2 \text{ and } \{u_k\}_{k=1}^\infty \subset \mathcal{K} \text{ is a sequence with} \\ \|\|u_k\|_2 \rightarrow +\infty, \text{ then } \lim_{k \rightarrow \infty} J(u_k) = +\infty. \end{array} \right. \quad (2.2)$$

Theorem 2.1 *Under the previous assumptions, (P) has at least one solution.*

The proof is standard. In the case of nonconvex optimization, local solutions play an important role. Under a local solution of (\mathcal{P}) we understand an element $\bar{u} \in \mathcal{K}$ such that, with some $\varepsilon > 0$, $J(\bar{u}) \leq J(u)$ holds for all $u \in \mathcal{K} \cap \{u \in U_\infty : \|u - \bar{u}\|_\infty < \varepsilon\}$. In this way, local optimality is understood in the sense of the topology of U_∞ . If the strong inequality $J(\bar{u}) < J(u)$ is satisfied in this set for $u \neq \bar{u}$, then the solution is called a *strict local solution*.

The next well known result provides first order optimality conditions in form of a variational inequality.

Theorem 2.2 *If \bar{u} is a local solution of (\mathcal{P}) and J is directionally differentiable at \bar{u} , both in the sense of U_∞ , then*

$$J'(\bar{u})(u - \bar{u}) \geq 0 \quad \forall u \in \mathcal{K}. \quad (2.3)$$

Notice that any local solution of (\mathcal{P}) in the U_2 sense is also a local solution in the U_∞ sense. Therefore (2.3) holds also for local solutions of (\mathcal{P}) in the U_2 sense.

In the sequel, we shall need stronger differentiability properties of J . Let us mention here once and for all that first- and second-order differentiability of J and notions as class C^1 or C^2 are always to be understood in the sense of the space U_∞ . The same refers to the approximation J_h that will be defined later.

Next, we establish second order optimality conditions. To this end, we need some further assumptions. Let us fix a point $\bar{u} \in \mathcal{K}$ as a reference element. In what follows, $B_2(\bar{u}, r)$ and $B_\infty(\bar{u}, r)$ denote the open ball of radius $r > 0$ centered at \bar{u} in U_2 and U_∞ , respectively.

(A2) *There exists an open subset $\mathcal{A} \subset U_\infty$ covering \mathcal{K} , such that $J : \mathcal{A} \rightarrow \mathbb{R}$ is of class C^2 . Furthermore, there exist constants $r > 0$, M_i , $i = 1, 2$, such that it holds for all $v, v_1, v_2 \in U_\infty$ and $u \in B_2(\bar{u}, r) \cap \mathcal{K}$*

$$|J'(u)v| \leq M_1 \|v\|_2 \quad \text{and} \quad |J''(u)(v_1, v_2)| \leq M_2 \|v_1\|_2 \|v_2\|_2. \quad (2.4)$$

For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $u_1, u_2 \in B_\infty(\bar{u}, r)$ and $v \in U_\infty$

$$\|u_1 - u_2\|_\infty < \delta \quad \Rightarrow \quad \begin{cases} |[J'(u_1) - J'(u_2)]v| \leq \varepsilon \|v\|_2, \\ |[J''(u_1) - J''(u_2)]v^2| \leq \varepsilon \|v\|_2^2. \end{cases} \quad (2.5)$$

Finally, we assume that the quadratic form $Q : v \mapsto J''(\bar{u})v^2$, $Q : U_2 \rightarrow \mathbb{R}$ is a Legendre form according to the definition below.

Remark 2.3 (i) *For the objective functional J of (P) , (A2) is satisfied in $U_\infty = L^\infty(\Omega)$ and $U_2 = L^2(\Omega)$ under appropriate differentiability and Lipschitz*

conditions on L , cf. Section 4.2. This holds, because $\mathcal{K} = \{u \in L^\infty(\Omega) \mid \alpha \leq u(x) \leq \beta \text{ a.e. in } \Omega\}$ is bounded in $L^\infty(\Omega)$.

If \mathcal{K} is unbounded, e.g. $\mathcal{K} = L^\infty(\Omega)$, then the objective functional of (P) is too general. In this case, (A2) can be verified if J has the particular form (1.2).

(ii) By the mean value theorem, the estimate (2.4) implies the upper inequality of (2.5), if u_1, u_2 belong to $B_2(\bar{u}, r) \cap \mathcal{K}$. This follows from $[J'(u_1) - J'(u_2)]v = J''(\hat{u})(u_1 - u_2, v)$ with some $\hat{u} \in [u_1, u_2]$.

By (2.4) and (2.5), the linear or bilinear forms $J'(u)$ and $J''(u)$ can be continuously extended to U_2 or $U_2 \times U_2$ so that the expressions and estimates above make also sense for $v, v_1, v_2 \in U_2$. Condition (2.5) expresses the continuity of the mappings $u \mapsto J'(u)$ and $u \mapsto J''(u)$ from U_∞ to the associated spaces of linear and bilinear forms.

Following Bonnans and Shapiro [3] or Bonnans and Zidani [4] we say that $J''(u)$ is a *Legendre form*, if the following implications hold:

- (i) If $v_k \rightharpoonup v$ in U_2 as $k \rightarrow \infty$, then $J''(u)v^2 \leq \liminf_{k \rightarrow \infty} J''(u)v_k^2$.
- (ii) If additionally $\lim_{k \rightarrow \infty} J''(u)v_k^2 = J''(u)v^2$ holds, then $\|v - v_k\|_2 \rightarrow 0$.

We define the cones $S_{\bar{u}}$ of feasible directions and $C_{\bar{u}}$ of critical directions by

$$\begin{aligned} S_{\bar{u}} &= \{v \in U_\infty : v = \lambda(u - \bar{u}) \text{ for some } \lambda > 0 \text{ and } u \in \mathcal{K}\}, \\ C_{\bar{u}} &= \text{cl}_2(S_{\bar{u}}) \cap \{v \in U_2 : J'(\bar{u})v = 0\}, \end{aligned}$$

where $\text{cl}_2(S_{\bar{u}})$ denotes the closure of $S_{\bar{u}}$ in U_2 . Now we can prove necessary second order optimality conditions under a regularity assumption given in the next theorem.

Theorem 2.4 *Let \bar{u} be a local solution of (P) in U_∞ . Assume that (A2) is satisfied and that $C_{\bar{u}} = \text{cl}_2(C_{\bar{u}})$, where*

$$C_{\bar{u}} = \{v \in S_{\bar{u}} : J'(\bar{u})v = 0\}.$$

Then it holds $J''(\bar{u})v^2 \geq 0$ for all $v \in C_{\bar{u}}$.

Proof. Given $v \in C_{\bar{u}}$, we take a sequence $\{v_k\}_{k=1}^\infty \subset C_{\bar{u}}$ converging to v in U_2 . By definition of $C_{\bar{u}}$, we have

$$v_k = \lambda_k(u_k - \bar{u}), \quad u_k \in \mathcal{K}, \quad \lambda_k > 0, \quad J'(\bar{u})v_k = 0.$$

Since \bar{u} is a local minimum of (P), there exists $\bar{\varepsilon} > 0$ such that $J(\bar{u}) \leq J(u)$ for any $u \in B_\infty(\bar{u}, \bar{\varepsilon}) \subset U_\infty$. We can assume that $\bar{\varepsilon} \leq r$. Then, for $0 < \rho < \min\{1/\lambda_k, \bar{\varepsilon}/\|v_k\|_\infty\}$, the elements $\bar{u} + \rho v_k$ belong to $\mathcal{K} \cap B_\infty(\bar{u}, \bar{\varepsilon})$. The second order Taylor expansion

$$0 \leq J(\bar{u} + \rho v_k) - J(\bar{u}) = \rho J'(\bar{u})v_k + \frac{\rho^2}{2} J''(\bar{u} + \theta_\rho \rho v_k)v_k^2 = \frac{\rho^2}{2} J''(\bar{u} + \theta_\rho \rho v_k)v_k^2$$

leads to

$$J''(\bar{u} + \theta_\rho \rho v_k) v_k^2 \geq 0 \Rightarrow J''(\bar{u}) v_k^2 = \lim_{\rho \searrow 0} J''(\bar{u} + \theta_\rho \rho v_k) v_k^2 \geq 0.$$

Finally, by (2.4), we arrive at $J''(\bar{u}) v^2 = \lim_{k \rightarrow \infty} J''(\bar{u}) v_k^2 \geq 0$. \square

Remark 2.5 *In control problems with constraints of type $\alpha \leq u(x) \leq \beta$ for $x \in X$, we have*

$$C_{\bar{u}} = \{v \in L^2(X) : v(x) = \begin{cases} \geq 0 & \text{if } \bar{u}(x) = \alpha \\ \leq 0 & \text{if } \bar{u}(x) = \beta \\ 0 & \text{if } \bar{d}(x) \neq 0 \end{cases} \},$$

where $\bar{d} \in U_2$ is the Riesz representative of the derivative of J at \bar{u} , more precisely

$$J'(\bar{u})v = \int_X \bar{d}(x)v(x) dx.$$

In this case, the regularity assumption of Theorem 2.4 holds with $U_\infty = L^\infty(X)$ and $U_2 = L^2(X)$.

Indeed, for given $v \in C_{\bar{u}}$ set

$$v_k(x) = \begin{cases} 0 & \text{if } \alpha < \bar{u}(x) < \alpha + \frac{1}{k} \text{ or } \beta - \frac{1}{k} < \bar{u}(x) < \beta, \\ \mathbb{P}_{[-k, +k]}(v(x)) & \text{otherwise.} \end{cases}$$

Let us check first that $\alpha \leq u_k(x) = \bar{u}(x) + \rho_k v_k(x) \leq \beta$, if $0 < \rho_k \leq \min\{1, \beta - \alpha\}/k^2$. If e.g. $\alpha < \bar{u}(x) < \alpha + 1/k$, then $v_k(x) = 0$ so that the inequality above is trivial. If $\bar{u}(x) = \alpha$, then $0 \leq \rho_k v_k(x) \leq \rho_k k \leq \beta - \alpha$, hence $\alpha \leq u_k(x) \leq \beta$.

If $\alpha + 1/k \leq \bar{u}(x) \leq \beta - 1/k$, then

$$\bar{u}(x) + \rho_k v_k(x) \geq \bar{u}(x) - \rho_k k \geq \alpha + \frac{1}{k} - \frac{1}{k^2} k = \alpha,$$

and

$$\bar{u}(x) + \rho_k v_k(x) \leq \bar{u}(x) + \rho_k k \leq \beta - \frac{1}{k} + \frac{1}{k^2} k = \beta.$$

The upper bound β is handled analogously. Consequently, $v_k = \frac{1}{\rho_k}(u_k - \bar{u}) \in S_{\bar{u}}$.

On the other hand, it is obvious that $|v_k(x)| \leq |v(x)|$. Hence, if $\bar{d}(x) \neq 0$, then $v(x) = 0$ and also $v_k(x) = 0$. Consequently, $v_k \in C_{\bar{u}}$ and $v_k \rightarrow v$ in $L^2(X)$, which proves that $v \in cl_2(C_{\bar{u}})$. Since v was taken arbitrarily in $C_{\bar{u}}$, we deduce that $C_{\bar{u}} \subset cl_2(C_{\bar{u}})$. The opposite inclusion is obvious.

Theorem 2.6 *Suppose that assumption (A2) holds. Let $\bar{u} \in \mathcal{K}$ satisfy (2.3) and*

$$J''(\bar{u})v^2 > 0 \quad \forall v \in C_{\bar{u}} \setminus \{0\}. \quad (2.6)$$

Then, there exist $\varepsilon > 0$ and $\delta > 0$ such that

$$J(\bar{u}) + \frac{\delta}{2} \|u - \bar{u}\|_2^2 \leq J(u) \quad \forall u \in \mathcal{K} \cap B_\infty(\bar{u}, \varepsilon). \quad (2.7)$$

Proof. We argue by contradiction and assume that for any positive integer k there exists $u_k \in \mathcal{K}$ such that

$$\|u_k - \bar{u}\|_\infty < \frac{1}{k} \quad \text{and} \quad J(\bar{u}) + \frac{1}{2k} \|u_k - \bar{u}\|_2^2 > J(u_k). \quad (2.8)$$

Setting $\rho_k = \|u_k - \bar{u}\|_2$ and $v_k = (u_k - \bar{u})/\rho_k$, we can assume that $v_k \rightharpoonup v$ in U_2 ; if necessary, we select a subsequence. Let us prove that $v \in C_{\bar{u}}$. From assumption (A2) and (2.3) we deduce

$$J'(\bar{u})v = \lim_{k \rightarrow \infty} J'(\bar{u})v_k = \lim_{k \rightarrow \infty} \rho_k J'(\bar{u})(u_k - \bar{u}) \geq 0.$$

To prove the opposite inequality, we use (2.8) and find

$$\begin{aligned} J'(\bar{u})v &= \lim_{k \rightarrow \infty} \frac{J(\bar{u} + \rho_k v_k) - J(\bar{u})}{\rho_k} = \lim_{k \rightarrow \infty} \frac{J(u_k) - J(\bar{u})}{\rho_k} \leq \\ &\leq \lim_{k \rightarrow \infty} \frac{1}{2k} \|u_k - \bar{u}\|_2 \leq c \lim_{k \rightarrow \infty} \frac{1}{2k} \|u_k - \bar{u}\|_\infty \leq \lim_{k \rightarrow \infty} \frac{c}{2k^2} = 0. \end{aligned}$$

Thus we have that $J'(\bar{u})v = 0$. The first equality above follows from

$$\begin{aligned} \frac{J(\bar{u} + \rho_k v_k) - J(\bar{u})}{\rho_k} &= \frac{J(u_k) - J(\bar{u})}{\rho_k} = J'(\bar{u} + \theta_k(u_k - \bar{u}))v_k \\ &= J'(\bar{u})v_k + [J'(\bar{u} + \theta_k(u_k - \bar{u})) - J'(\bar{u})]v_k. \end{aligned}$$

For arbitrary $\varepsilon > 0$, we deduce from (2.5) and (2.8) the existence of k_ε such that

$$|[J'(\bar{u} + \theta_k(u_k - \bar{u})) - J'(\bar{u})]v_k| \leq \varepsilon \|v_k\|_2 = \varepsilon \quad \forall k \geq k_\varepsilon.$$

Therefore, by (2.4),

$$\lim_{k \rightarrow \infty} \frac{J(\bar{u} + \rho_k v_k) - J(\bar{u})}{\rho_k} = \lim_{k \rightarrow \infty} J'(\bar{u})v_k + \lim_{k \rightarrow \infty} [J'(\bar{u} + \theta_k(u_k - \bar{u})) - J'(\bar{u})]v_k = J'(\bar{u})v.$$

Next, we prove that $v \in cl_2(S_{\bar{u}})$. From $v_k = (u_k - \bar{u})/\rho_k$ and $u_k \in \mathcal{K}$, we conclude $v_k \in S_{\bar{u}} \subset cl_2(S_{\bar{u}})$. The set $cl_2(S_{\bar{u}})$ is closed and convex in U_2 , hence $v \in cl_2(S_{\bar{u}})$. Thus, we obtain $v \in C_{\bar{u}}$.

Invoking again (2.8) and (2.3) we get by a Taylor expansion

$$\begin{aligned} \frac{\rho_k^2}{2k} &= \frac{1}{2k} \|u_k - \bar{u}\|_2^2 > J(u_k) - J(\bar{u}) \\ &= J(\bar{u} + \rho_k v_k) - J(\bar{u}) = \rho_k J'(\bar{u})v_k + \frac{\rho_k^2}{2} [J''(\bar{u})v_k^2 + r_k] \geq \frac{\rho_k^2}{2} [J''(\bar{u})v_k^2 + r_k], \end{aligned}$$

where

$$r_k = [J''(\bar{u} + \theta_k \rho_k v_k) - J''(\bar{u})]v_k^2 = [J''(\bar{u} + \theta_k(u_k - \bar{u})) - J''(\bar{u})]v_k^2.$$

Therefore, it holds

$$J''(\bar{u})v_k^2 + r_k < \frac{1}{k}.$$

Once again, (2.5) and (2.8) imply $|r_k| \rightarrow 0$ when $k \rightarrow \infty$. Then, the above inequality, along with (2.4) and (2.6) leads to

$$0 \leq J''(\bar{u})v^2 \leq \liminf_{k \rightarrow \infty} J''(\bar{u})v_k^2 \leq \limsup_{k \rightarrow \infty} J''(\bar{u})v_k^2 \leq \limsup_{k \rightarrow \infty} \frac{1}{k} = 0,$$

so that $J''(\bar{u})v_k^2 \rightarrow J''(\bar{u})v^2 = 0$. From (2.6), it follows $v = 0$. Finally, using that $J''(\bar{u})$ is a Legendre form, we get that $v_k \rightarrow v = 0$ in U_2 . This contradicts the fact that $\|v_k\|_2 = 1$ for every k . \square

Remark 2.7 *Under the assumption (A2), the optimality condition (2.6) is equivalent to the following one:*

$$\exists \alpha > 0 \text{ such that } J''(\bar{u})v^2 \geq \alpha \|v\|^2 \quad \forall v \in C_{\bar{u}}. \quad (2.9)$$

Indeed, it is obvious that (2.9) implies (2.6). We verify the converse implication by contradiction. Suppose that (2.6) holds, but not (2.9). Then for any positive integer k there exists an element $v_k \in C_{\bar{u}}$ such that

$$J''(\bar{u})v_k^2 < \frac{1}{k} \|v_k\|_2^2 \quad \forall k \geq 1.$$

Re-defining $v_k := v_k / \|v_k\|_2$ yields $J''(\bar{u})v_k^2 < \frac{1}{k}$. By taking a subsequence, denoted in the same way, we can assume that $v_k \rightharpoonup v$ weakly in U_2 . Since $C_{\bar{u}}$ is closed and convex in U_2 , v belongs to $C_{\bar{u}}$. Moreover, $J''(\bar{u})$ is a Legendre form. Therefore (2.6) implies

$$0 \leq J''(\bar{u})v^2 \leq \liminf_{h \rightarrow 0} J''(\bar{u})v_k^2 \leq \limsup_{h \rightarrow 0} J''(\bar{u})v_k^2 \leq \limsup_{h \rightarrow 0} \frac{1}{h} = 0.$$

Consequently, $v = 0$ and $J''(\bar{u})v_k^2 \rightarrow 0$ must hold and hence $\|v_k\|_2 \rightarrow 0$ contradicting the fact that $\|v_k\|_2 = 1$.

Remark 2.8 If $\bar{u} \in \mathcal{K}$ satisfies (2.7) and the regularity assumption $C_{\bar{u}} = cl_2(C_{\bar{u}})$ holds true, then (2.9) is fulfilled by $\alpha = \delta$. Indeed, by (2.7), \bar{u} is a local solution to

$$\min_{u \in \mathcal{K}} J(u) - \frac{\delta}{2} \|u - \bar{u}\|_2^2.$$

In view of Theorem 2.4, the necessary second-order condition

$$J''(\bar{u})v^2 - \delta \|v\|_2^2 \geq 0 \quad \forall v \in C_{\bar{u}}$$

must hold. Consequently, under the regularity condition above, inequality (2.6) holds if and only if (2.7) is satisfied.

2.2 Approximation of (\mathcal{P})

Let $h > 0$ denote a small parameter. In applications to control problems, h is the meshsize of a grid underlying a numerical approximation of (\mathcal{P}) . We consider a family of problems (\mathcal{P}_h) approximating (\mathcal{P}) when $h \rightarrow 0$, which is characterized in some sense by the parameter h .

For any $h > 0$, we consider a family of sets \mathcal{K}_h with the following properties:

(A3) $\mathcal{K}_h \subset \mathcal{K}$ is convex and closed in U_2 . Moreover, for any $u \in \mathcal{K}$ there exist elements $u_h \in \mathcal{K}_h$ such that $\|u - u_h\|_2 \rightarrow 0$ as $h \rightarrow 0$.

We introduce a family of functionals $J_h : U_\infty \rightarrow \mathbb{R}$ satisfying the following assumptions.

(A4) All functions J_h have the following properties:

$$\text{If } \{u_k\}_{k=1}^\infty \subset \mathcal{K}_h \text{ and } u_k \rightarrow u \text{ in } U_2, \text{ then } J_h(u) \leq \liminf_{k \rightarrow \infty} J_h(u_k). \quad (2.10)$$

If \mathcal{K}_h is unbounded in U_2 , then the following properties hold:

$$\text{If } \{u_k\}_{k=1}^\infty \subset \mathcal{K}_h \text{ and } \|u_k\|_2 \rightarrow +\infty, k \rightarrow \infty, \text{ then } \lim_{k \rightarrow \infty} J_h(u_k) = +\infty. \quad (2.11)$$

$$\text{If } u_h \in \mathcal{K}_h \forall h > 0 \text{ and } \|u_h\|_2 \rightarrow \infty, \text{ then } \lim_{h \rightarrow 0} J_h(u_h) = \infty. \quad (2.12)$$

We investigate the family of approximating control problems

$$(\mathcal{P}_h) \quad \min_{u_h \in \mathcal{K}_h} J_h(u_h).$$

To guarantee that the family $\{(\mathcal{P}_h)\}_h$ really approximates problem (\mathcal{P}) , we need a further assumption.

(A5) Let $\{u_h\}_{h>0} \subset \mathcal{K}$ and $u \in \mathcal{K}$ be given.

$$\text{If } u_h \rightarrow u \text{ in } U_2, \text{ then } \lim_{h \rightarrow 0} J(u_h) = J(u); \quad (2.13)$$

$$\text{if } u_h \rightarrow u \text{ in } U_2, \text{ then } J(u) \leq \liminf_{h \rightarrow 0} J_h(u_h) \text{ and } \lim_{h \rightarrow 0} |J_h(u_h) - J(u_h)| = 0. \quad (2.14)$$

Remark 2.9 *If assumption (A1) is satisfied, then the property of weak lower semicontinuity in (2.14) follows just from $\lim_{h \rightarrow 0} |J_h(u_h) - J(u_h)| = 0$. Indeed, we obtain*

$$\liminf_{h \rightarrow 0} J_h(u_h) = \liminf_{h \rightarrow 0} J(u_h) + \lim_{h \rightarrow 0} (J_h(u_h) - J(u_h)).$$

As the last expression tends to zero, we obtain from (A1)

$$\liminf_{h \rightarrow 0} J_h(u_h) = \liminf_{h \rightarrow 0} J(u_h) \geq J(u).$$

As a consequence of assumption (A5) the next result is deduced.

Lemma 2.10 *Under assumptions (A1) and (A5) the following holds:*

$$\text{If } \{u_h\}_{h>0} \subset \mathcal{K} \text{ and } u_h \rightarrow u \text{ in } U_2, \text{ then } \lim_{h \rightarrow 0} J_h(u_h) = J(u). \quad (2.15)$$

Proof. We apply first (2.13) and then (2.14),

$$J(u) = \lim_{h \rightarrow 0} J(u_h) = \lim_{h \rightarrow 0} J_h(u_h) + \lim_{h \rightarrow 0} [J(u_h) - J_h(u_h)] = \lim_{h \rightarrow 0} J_h(u_h).$$

□ Next, we show that the problems (\mathcal{P}_h) approximate problem (\mathcal{P}) . To this aim, we introduce another assumption.

(A6) *If $\{u_k\}_{k=1}^\infty \subset \mathcal{K}$, $u_k \rightharpoonup u$ in U_2 , and $J(u_k) \rightarrow J(u)$, then $\|u_k - u\|_2 \rightarrow 0$.*

Theorem 2.11 *Assume that (A1) and (A3)–(A5) hold. Then, for all $h > 0$, the problem (\mathcal{P}_h) has at least one (global) solution \bar{u}_h . Any sequence of solutions $\{\bar{u}_h\}_{h>0}$ contains a subsequence, denoted for convenience in the same way, that converges weakly in U_2 to a point \bar{u} . Each of these weak limit points is a solution to (\mathcal{P}) . Moreover, $\lim_{h \rightarrow 0} J_h(\bar{u}_h) = \lim_{h \rightarrow 0} J(\bar{u}_h) = J(\bar{u})$. Under assumption (A6), the convergence $\bar{u}_h \rightarrow \bar{u}$ is even strong in U_2 .*

Proof. First we prove that every problem (\mathcal{P}_h) has at least one solution. Thanks to (A3), \mathcal{K}_h is not empty. Let $\{u_k\}_{k=1}^\infty \subset \mathcal{K}_h$ be a minimizing sequence of (\mathcal{P}_h) . Assumption (2.11) implies that $\{u_k\}_{k=1}^\infty$ is bounded in U_2 . Then, selecting a subsequence if necessary, we can assume that $u_k \rightharpoonup \bar{u}_h$ weakly in U_2 . Since \mathcal{K}_h is convex and closed in U_2 by assumption (A3), it holds $\bar{u}_h \in \mathcal{K}_h$. Finally, we conclude from (2.10) that \bar{u}_h is a solution of (\mathcal{P}_h) .

Next, we verify the boundedness $\{\bar{u}_h\}_{h>0}$ in U_2 . This is obvious if \mathcal{K} is bounded in U_2 . If \mathcal{K} is unbounded, we fix an element $u \in \mathcal{K}$. Thanks to assumption (A3), there exists a sequence $\{u_h\}_{h>0}$ with $u_h \in \mathcal{K}_h$, such that $u_h \rightarrow u$ strongly in U_2 . From (2.15) and the fact that \bar{u}_h is a solution of (\mathcal{P}_h) , we get

$$J_h(\bar{u}_h) \leq J_h(u_h) \rightarrow J(u) \quad \Rightarrow \quad \exists M > 0 \text{ such that } J_h(\bar{u}_h) \leq M \quad \forall h > 0.$$

In view of (2.12), this implies that $\{\bar{u}_h\}_{h>0}$ is bounded in U_2 . Therefore, there exist subsequences weakly convergent in U_2 to points \bar{u} . Let us prove that any of these limit points is a solution of (\mathcal{P}) . First of all, $\bar{u} \in \mathcal{K}$ because $\{\bar{u}_h\}_{h>0} \subset \mathcal{K}$ and \mathcal{K} is convex and closed in U_2 . Let \tilde{u} be a solution of (\mathcal{P}) and consider a sequence $\{\tilde{u}_h\}_{h>0}$, with $\tilde{u}_h \in \mathcal{K}_h$, such that $\tilde{u}_h \rightarrow \tilde{u}$ in U_2 . Such a sequence exists thanks to (A3). Then, (2.14), (2.15) and the fact that \bar{u}_h is a solution of (\mathcal{P}_h) lead to

$$J(\bar{u}) \leq \liminf_{h \rightarrow 0} J_h(\bar{u}_h) \leq \limsup_{h \rightarrow 0} J_h(\bar{u}_h) \leq \limsup_{h \rightarrow 0} J_h(\tilde{u}_h) = J(\tilde{u}) = \inf(\mathcal{P}) \leq J(\bar{u}),$$

which proves that \bar{u} is a solution of (\mathcal{P}) and $J_h(\bar{u}_h) \rightarrow J(\bar{u})$. From (2.14) we get

$$\lim_{h \rightarrow 0} J(\bar{u}_h) = \lim_{h \rightarrow 0} J_h(\bar{u}_h) + \lim_{h \rightarrow 0} [J(\bar{u}_h) - J_h(\bar{u}_h)] = J(\bar{u}).$$

Finally, (A6) and the above equality yield the strong convergence $\|\bar{u} - \bar{u}_h\|_2 \rightarrow 0$. \square

The reader might wonder if there is a reciprocal theorem to the previous one. More precisely, we have proved that the solutions of (\mathcal{P}_h) converge to solutions of (\mathcal{P}) . Now, the question is: given a solution \bar{u} of (\mathcal{P}) , can it be approximated by solutions of (\mathcal{P}_h) ? The answer is yes assuming that \bar{u} is the unique solution of (\mathcal{P}) in a neighborhood. Even more, the next theorem states that strict local solution of (\mathcal{P}) can be approximated by local solutions of (\mathcal{P}_h) .

Theorem 2.12 *Under the assumptions (A1) and (A3)–(A6), if \bar{u} is a strict local solution of (\mathcal{P}) in the sense of U_2 , then there exists a sequence $\{\bar{u}_h\}_{h>0}$ such that \bar{u}_h is a local solution of (\mathcal{P}_h) and $\|\bar{u} - \bar{u}_h\|_2 \rightarrow 0$.*

Proof. Since \bar{u} is a strict local solution of (\mathcal{P}) , there exists $\varepsilon > 0$ such that

$$J(\bar{u}) < J(u) \quad \forall u \in \mathcal{K}^\varepsilon \setminus \{\bar{u}\}, \quad \text{where } \mathcal{K}^\varepsilon = \{u \in \mathcal{K} : \|u - \bar{u}\|_2 \leq \varepsilon\}. \quad (2.16)$$

Consider the problems

$$(P^\varepsilon) \quad \min_{u \in \mathcal{K}^\varepsilon} J(u) \quad \text{and} \quad (P_h^\varepsilon) \quad \min_{u_h \in \mathcal{K}_h^\varepsilon} J_h(u_h),$$

where $\mathcal{K}_h^\varepsilon = \mathcal{K}^\varepsilon \cap \mathcal{K}_h$. From (2.16) we know that \bar{u} is the unique solution of (P^ε) . From assumption (A3) we obtain the existence of a sequence $\{u_h\}_{h>0}$, with $u_h \in \mathcal{K}_h$, such that $\|\bar{u} - u_h\|_2 \rightarrow 0$. Therefore, $\|\bar{u} - u_h\|_2 \leq \varepsilon$ for all $h \leq h_\varepsilon$, hence $u_h \in \mathcal{K}_h^\varepsilon$ for every $h \leq h_\varepsilon$. Since $\mathcal{K}_h^\varepsilon$ is nonempty, bounded, convex and closed in U_2 , we can deduce easily from (2.10) and (2.11) that (P_h^ε) has at least one solution \bar{u}_h for $h \leq h_\varepsilon$. Then, applying Theorem 2.11 we deduce that $\|\bar{u} - \bar{u}_h\|_2 \rightarrow 0$. Therefore, there exists $h_0 > 0$ such that $\|\bar{u} - \bar{u}_h\|_2 < \varepsilon$ for $h \leq h_0$. This implies that \bar{u}_h is a local minimum of (\mathcal{P}_h) for $h \leq h_0$, which concludes the proof. \square

Remark 2.13 Often, Theorem 2.11 can be improved by deriving the convergence $\|\bar{u} - \bar{u}_h\|_\infty \rightarrow 0$; see e.g. [2], [8], [11]. In such cases, Theorem 2.12 can be extended: Any strict local solutions of (\mathcal{P}) in the sense of U_∞ can be approximated by local solutions of (\mathcal{P}_h) in the sense of U_∞ .

In the remainder of this section, $\{\bar{u}_h\}_{h>0}$ will denote a sequence of local minima of problems (\mathcal{P}_h) converging strongly in U_2 to a local minimum \bar{u} of (\mathcal{P}) ; see Theorems 2.11 and 2.12. Our next goal is to prove some error estimates for $\|\bar{u} - \bar{u}_h\|_2$. To this end, we need extra hypotheses.

(A7) The functions $J_h : \mathcal{A} \rightarrow \mathbb{R}$ are of class C^1 , where \mathcal{A} is the open set of U_∞ introduced in assumption (A2). Furthermore, with a sequence $\varepsilon_h \rightarrow 0$, it holds

$$|[J'_h(u) - J'(u)](u_h - \bar{u})| \leq \varepsilon_h \|u_h - \bar{u}\|_2, \quad \forall (u, u_h) \in \mathcal{K} \times \mathcal{K}_h \quad (2.17)$$

and

$$\left\{ \begin{array}{l} \bullet \text{ either } \|\bar{u} - \bar{u}_h\|_\infty \rightarrow 0, \\ \bullet \text{ or for any sequence } \{(u_k, v_k)\}_{k=1}^\infty \subset \mathcal{K} \times U_2 \text{ with } \|u_k - \bar{u}\|_2 \rightarrow 0, \\ \quad \text{if } v_k \rightharpoonup v \text{ weakly in } U_2 \Rightarrow J''(u)v^2 \leq \liminf_{k \rightarrow \infty} J''(u_k)v_k^2, \\ \quad \text{if } v_k \rightharpoonup 0 \text{ weakly in } U_2 \Rightarrow \liminf_{k \rightarrow \infty} J''(u_k)v_k^2 \geq \Lambda \liminf_{k \rightarrow \infty} \|v_k\|_2^2 \end{array} \right. \quad (2.18)$$

with some $\Lambda > 0$.

Theorem 2.14 Let the assumptions (A2), (A3) and (A7) be satisfied and let $\{\bar{u}_h\}_{h>0}$ be a sequence of local solutions to (\mathcal{P}_h) converging strongly to \bar{u} in U_2 . If the second-order sufficiency condition (2.6) holds, then there exist $C > 0$ and $h_0 > 0$ such that

$$\|\bar{u} - \bar{u}_h\|_2 \leq C [\varepsilon_h^2 + \|\bar{u} - u_h\|_2^2 + J'(\bar{u})(u_h - \bar{u})]^{1/2} \quad \forall u_h \in \mathcal{K}_h, \quad \forall h < h_0. \quad (2.19)$$

Proof. Since \bar{u}_h is a local minimum of (\mathcal{P}_h) and J_h is C^1 around \bar{u}_h , we have

$$J'_h(\bar{u}_h)(u_h - \bar{u}_h) \geq 0 \quad \forall u_h \in \mathcal{K}_h. \quad (2.20)$$

From this inequality we get

$$J'(\bar{u}_h)(\bar{u} - \bar{u}_h) + [J'_h(\bar{u}_h) - J'(\bar{u}_h)](\bar{u} - \bar{u}_h) + [J'_h(\bar{u}_h) - J'(\bar{u})](u_h - \bar{u}) + J'(\bar{u})(u_h - \bar{u}) \geq 0.$$

Taking $u = \bar{u}_h$ in (2.3) and adding the resulting inequality to the previous one we get

$$\begin{aligned} [J'(\bar{u}_h) - J'(\bar{u})](\bar{u}_h - \bar{u}) &\leq [J'_h(\bar{u}_h) - J'(\bar{u}_h)](\bar{u} - \bar{u}_h) \\ &\quad + [J'_h(\bar{u}_h) - J'(\bar{u})](u_h - \bar{u}) + J'(\bar{u})(u_h - \bar{u}). \end{aligned}$$

We introduce an additional term

$$[J'(\bar{u}_h) - J'(\bar{u})](\bar{u}_h - \bar{u}) \leq [J'_h(\bar{u}_h) - J'(\bar{u}_h)](\bar{u} - \bar{u}_h)$$

$$+[J'_h(\bar{u}_h) - J'(\bar{u}_h)](u_h - \bar{u}) + [J'(\bar{u}_h) - J'(\bar{u})](u_h - \bar{u}) + J'(\bar{u})(u_h - \bar{u}).$$

We use (2.17) for the first and second term of the right-hand side of this inequality, apply the mean value theorem to the third one and estimate the resulting term by (2.4). In this way, we deduce from the above inequality

$$\begin{aligned} & [J'(\bar{u}_h) - J'(\bar{u})](\bar{u}_h - \bar{u}) \\ & \leq \varepsilon_h \|\bar{u} - \bar{u}_h\|_2 + \varepsilon_h \|\bar{u} - u_h\|_2 + M_2 \|\bar{u}_h - \bar{u}\|_2 \|\bar{u} - u_h\|_2 + J'(\bar{u})(u_h - \bar{u}). \end{aligned}$$

Since J is of class C^2 in \mathcal{A} , we have

$$(J'(\bar{u}_h) - J'(\bar{u}))(\bar{u}_h - \bar{u}) = J''(\hat{u}_h)(\bar{u}_h - \bar{u})^2$$

with some \hat{u}_h in the segment $[\bar{u}_h, \bar{u}]$. Hence, from the above two relations we get

$$\begin{aligned} & J''(\hat{u}_h)(\bar{u}_h - \bar{u})^2 \leq \varepsilon_h (\|\bar{u} - \bar{u}_h\|_2 + \|\bar{u} - u_h\|_2) \\ & + M_2 \|\bar{u}_h - \bar{u}\|_2 \|\bar{u} - u_h\|_2 + J'(\bar{u})(u_h - \bar{u}). \end{aligned} \quad (2.21)$$

From our assumptions $\|\bar{u}_h - \bar{u}\|_2 \rightarrow 0$ as $h \rightarrow 0$, hence also $\|\hat{u}_h - \bar{u}\|_2 \rightarrow 0$.

Now, let us argue by contradiction. We suppose that (2.19) is false, then there exists a subsequence $\{h_k\}_{k=1}^\infty$ converging to 0, with $u_{h_k} \in \mathcal{K}_{h_k}$, such that

$$\|\bar{u} - \bar{u}_{h_k}\|_2^2 > k[\varepsilon_{h_k}^2 + \|\bar{u} - u_{h_k}\|_2^2 + J'(\bar{u})(u_{h_k} - \bar{u})],$$

or equivalently

$$\frac{1}{k} > \frac{\varepsilon_{h_k}^2}{\|\bar{u} - \bar{u}_{h_k}\|_2^2} + \frac{\|u_{h_k} - \bar{u}\|_2^2}{\|\bar{u} - \bar{u}_{h_k}\|_2^2} + \frac{J'(\bar{u})(u_{h_k} - \bar{u})}{\|\bar{u} - \bar{u}_{h_k}\|_2^2}. \quad (2.22)$$

Since the three summands of the right hand side are nonnegative, (2.22) implies that all of them converge to zero. Define

$$v_{h_k} := \frac{\bar{u}_{h_k} - \bar{u}}{\|\bar{u}_{h_k} - \bar{u}\|_2}.$$

All v_{h_k} belong to the unit sphere, hence we can assume $v_{h_k} \rightharpoonup v$ in U_2 with some $v \in U_2$. Let us prove that $v \in C_{\bar{u}}$. It is clear that $v_{h_k} \in S_{\bar{u}}$, hence $v \in cl_2(S_{\bar{u}})$ because of the convexity of this set. It remains to prove that $J'(\bar{u})v = 0$. The optimality conditions (2.3) imply that $J'(\bar{u})v = \lim_{k \rightarrow \infty} J'(\bar{u})v_{h_k} \geq 0$. To prove the converse inequality we use (2.4), (2.17) and (2.14) as follows

$$\begin{aligned} & J'(\bar{u})v = \lim_{k \rightarrow \infty} J'(\bar{u})v_{h_k} = \lim_{k \rightarrow \infty} J'_{h_k}(\bar{u}_{h_k})v_{h_k} \\ & + \lim_{k \rightarrow \infty} [J'(\bar{u}_{h_k}) - J'_{h_k}(\bar{u}_{h_k})]v_{h_k} + \lim_{k \rightarrow \infty} [J'(\bar{u}) - J'(\bar{u}_{h_k})]v_{h_k} \end{aligned} \quad (2.23)$$

The first limit above exists due to the weak convergence of v_{h_k} . We verify below that the third and fourth limit exist, hence the second limit must exist as well. Indeed, by assumption (2.17) we have for the third limit

$$\lim_{k \rightarrow \infty} |[J'(\bar{u}_{h_k}) - J'_{h_k}(\bar{u}_{h_k})]v_{h_k}| \leq \lim_{k \rightarrow \infty} \varepsilon_{h_k} \|v_{h_k}\|_2 = \lim_{k \rightarrow \infty} \varepsilon_{h_k} = 0.$$

Thanks to the mean value theorem and (2.4), we obtain for the fourth limit

$$\lim_{k \rightarrow \infty} |[J'(\bar{u}) - J'(\bar{u}_{h_k})]v_{h_k}| \leq \lim_{k \rightarrow \infty} M_2 \|\bar{u} - \bar{u}_{h_k}\|_2 \|v_{h_k}\|_2 = 0.$$

In view of (2.23), this yields the inequality $J'(\bar{u})v \leq \lim_{k \rightarrow \infty} J'_{h_k}(\bar{u}_{h_k})v_{h_k}$. Now, from this inequality and using first (2.20), next (2.4) along with (2.17) and finally (2.22), we obtain

$$\begin{aligned} J'(\bar{u})v &\leq \lim_{k \rightarrow \infty} J'_{h_k}(\bar{u}_{h_k})v_{h_k} \\ &= \lim_{k \rightarrow \infty} \frac{1}{\|\bar{u}_{h_k} - \bar{u}\|_2} [J'_{h_k}(\bar{u}_{h_k})(u_{h_k} - \bar{u}) + J'_{h_k}(\bar{u}_{h_k})(\bar{u}_{h_k} - u_{h_k})] \\ &\leq \lim_{k \rightarrow \infty} \frac{1}{\|\bar{u}_{h_k} - \bar{u}\|_2} J'_{h_k}(\bar{u}_{h_k})(u_{h_k} - \bar{u}) \leq \lim_{k \rightarrow \infty} (M_1 + \varepsilon_{h_k}) \frac{\|u_{h_k} - \bar{u}\|_2}{\|\bar{u}_{h_k} - \bar{u}\|_2} = 0. \end{aligned}$$

Let us show how (2.4) and (2.17) imply the last inequality. For any $u \in B_\infty(\bar{u}, r) \cap \mathcal{K}$, see assumption (A7), and every $v \in U_2$,

$$|J'_{h_k}(u)v| \leq |J'(u)v| + |[J'_{h_k}(u) - J'(u)]v| \leq M_1 \|v\|_2 + \varepsilon_{h_k} \|v\|_2 = [M_1 + \varepsilon_{h_k}] \|v\|_2.$$

We observe that $J''(\bar{u})v^2 \leq \liminf_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2$. Indeed, this is a consequence of (2.18). The inequality is obvious if the second condition of (2.18) is fulfilled. Under the first condition, we also have that $\|\bar{u} - \hat{u}_{h_k}\|_\infty \rightarrow 0$. Because $J''(\bar{u}) : U_2 \times U_2 \rightarrow \mathbb{R}$ is a Legendre form, we get from (2.5)

$$\begin{aligned} \liminf_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2 &= \liminf_{k \rightarrow \infty} [J''(\bar{u})v_{h_k}^2 + (J''(\hat{u}_{h_k}) - J''(\bar{u}))v_{h_k}^2] \\ &= \liminf_{k \rightarrow \infty} J''(\bar{u})v_{h_k}^2 \geq J''(\bar{u})v^2. \end{aligned}$$

Combining this with (2.21) and (2.22), we obtain

$$\begin{aligned} J''(\bar{u})v^2 &\leq \liminf_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2 \leq \limsup_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2 \\ &\leq \lim_{k \rightarrow \infty} \left\{ \frac{\varepsilon_{h_k}}{\|\bar{u}_{h_k} - \bar{u}\|_2} \left(1 + \frac{\|\bar{u} - u_{h_k}\|_2}{\|\bar{u}_{h_k} - \bar{u}\|_2} \right) \right. \\ &\quad \left. + M_2 \frac{\|\bar{u} - u_{h_k}\|_2}{\|\bar{u}_{h_k} - \bar{u}\|_2} + \frac{J'(\bar{u})(u_{h_k} - \bar{u})}{\|\bar{u}_{h_k} - \bar{u}\|_2^2} \right\} = 0. \end{aligned}$$

This inequality, the fact that $v \in C_{\bar{u}}$ and (2.6) imply that $v = 0$. Finally, (2.18) leads to the contradiction. Indeed, under the first assumption of (2.18) and using again (2.5) along with the fact that $\|\bar{u} - \hat{u}_{h_k}\|_\infty \rightarrow 0$, we have

$$\lim_{k \rightarrow \infty} J''(\bar{u})v_{h_k}^2 = \lim_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2 + \lim_{k \rightarrow \infty} [J''(\bar{u}) - J''(\hat{u}_{h_k})]v_{h_k}^2 = \lim_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2 = 0.$$

Thus, we have proved that $J''(\bar{u})v_{h_k}^2 \rightarrow 0$ and $v_{h_k} \rightarrow 0$ weakly in U_2 . Since $J''(\bar{u})$ is a Legendre quadratic form, we get $\|v_{h_k}\|_2 \rightarrow 0$, which contradicts the fact that $\|v_{h_k}\|_2 = 1$ for every k .

Under the second assumption of (2.18), we achieve the contradiction by

$$0 \geq \liminf_{k \rightarrow \infty} J''(\hat{u}_{h_k})v_{h_k}^2 \geq \Lambda \liminf_{k \rightarrow \infty} \|v_{h_k}\|_2^2 = \Lambda > 0. \quad \square$$

3 Application to the optimal control problem (P)

3.1 Main assumptions and known results on the control-to-state mapping

We apply Theorem 2.14 on error estimates to the PDE constrained optimal control problem (P). In this way, we derive the error estimates we have already announced in [11] and illustrate the use of our abstract theory. We rely on the assumptions (H1)-(H4) below.

(H1) Ω is an open, convex and bounded subset of \mathbb{R}^n , $n = 2$ or 3 , with boundary Γ of class $C^{1,1}$. For $n = 2$, Ω is allowed to be polygonal instead of class $C^{1,1}$.

In the next assumptions, $C_M > 0$ denotes a constant depending on a real bound $M > 0$. We assume in (H2)-(H4) that such a constant C_M exists. Although we might use different constants for a , L , and f , we view C_M as the maximum of all of them.

(H2) The function $a : \bar{\Omega} \times \mathbb{R} \rightarrow \mathbb{R}$ is of class C^2 with respect to the second component. For any $M > 0$, it holds for all $x_i \in \bar{\Omega}$ and $|y|, |y_i| \leq M$, $i = 1, 2$,

$$\sum_{j=1}^2 \left| \frac{\partial^j a}{\partial y^j}(x, 0) \right| \leq C_M \quad \forall x \in \Omega, \quad (3.1)$$

$$\left| \frac{\partial^j a}{\partial y^j}(x_2, y_2) - \frac{\partial^j a}{\partial y^j}(x_1, y_1) \right| \leq C_M(|x_2 - x_1| + |y_2 - y_1|), \quad j = 0, 1, 2 \quad (3.2)$$

Moreover, there exists $a_0 > 0$ such that

$$a(x, y) \geq a_0 \quad \forall x \in \Omega, \quad \forall y \in \mathbb{R}. \quad (3.3)$$

(H3) The function $f : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ is measurable with respect to the first variable and twice differentiable with respect to the second. It obeys the following properties:

$$\exists \bar{p} > n \text{ such that } f(\cdot, 0) \in L^{\bar{p}}(\Omega), \quad (3.4)$$

$$\frac{\partial f}{\partial y}(x, y) \geq 0 \text{ for a.e. } x \in \Omega, \quad \forall y \in \mathbb{R}, \quad (3.5)$$

$$\sum_{j=1}^2 \left| \frac{\partial^j f}{\partial y^j}(x, 0) \right| \leq C, \quad \text{for a.a. } x \in \Omega \quad (3.6)$$

with some $C > 0$, and $\forall M > 0 \exists C_M > 0$ such that

$$\left| \frac{\partial^2 f}{\partial y^2}(x, y_1) - \frac{\partial^2 f}{\partial y^2}(x, y_2) \right| \leq C_M |y_1 - y_2| \quad (3.7)$$

for almost all $x \in \Omega$ and all $|y|, |y_1|, |y_2| \leq M$.

(H4) The function $L : \Omega \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is measurable with respect to the first variable and twice differentiable with respect to the others. It fulfills the *convexity condition*

$$\frac{\partial^2 L}{\partial u^2}(x, y, u) \geq \Lambda > 0 \quad \text{for a.a. } x \in \Omega, \forall y, u \in \mathbb{R}. \quad (3.8)$$

Moreover

$$L(\cdot, 0, 0) \in L^1(\Omega), \quad \frac{\partial L}{\partial y}(\cdot, 0, 0) \in L^{\bar{p}}(\Omega), \quad (3.9)$$

$\forall M > 0 \exists C_M > 0$ such that

$$\left| \frac{\partial L}{\partial u}(x_2, y, u) - \frac{\partial L}{\partial u}(x_1, y, u) \right| \leq C_M |x_2 - x_1|, \quad (3.10)$$

$$\|D_{(y,u)}^2 L(x, y_2, u_2) - D_{(y,u)}^2 L(x, y_1, u_1)\| \leq C_M (|y_2 - y_1| + |u_2 - u_1|) \quad (3.11)$$

for all $x_1, x_2 \in \bar{\Omega}$, for a.e. $x \in \Omega$, and all $|y|, |y_i|, |u|, |u_i| \leq M$, $i = 1, 2$, where $\bar{p} > n$ is as in assumption (H3) and $D_{(y,u)}^2 L$ denotes the second derivative of L with respect to (y, u) , i.e. the associated Hessian matrix.

Remark 3.1 *To simplify the presentation, the conditions (H3) and (H4) are required slightly stronger than needed. Continuity of the second-order derivatives of f and L with respect to y and u , uniform w.r. to $x \in \Omega$, would suffice.*

We used similar assumptions as (H1)-(H4) in our papers [10] and [11]. Preparing our error analysis for locally optimal controls, we discussed in [10] second-order sufficient optimality conditions. Consequently, we had to require associated second-order derivatives. In [11], we did not need second-order derivatives of a and f , since our focus was only on convergence of discretizations of the state equation and the adjoint equation.

In (H1), we have now added the possibility of a convex and polygonal domain if $n = 2$. Our assumptions are imposed to assure that the states and adjoint states belong to $W^{2,p}(\Omega)$ for $p > n$.

It follows from Theorem 3.2 below that for every $u \in L^2(\Omega)$, the state equation (1.1) has a unique solution $y_u \in H^2(\Omega) \cap H_0^1(\Omega)$. Moreover, for all bounded sets $\mathcal{B} \subset L^2(\Omega)$, there exists a constant $C_{\mathcal{B}} > 0$ such that

$$\|y_u\|_{H^2(\Omega)} \leq C_{\mathcal{B}} \quad \forall u \in \mathcal{B}. \quad (3.12)$$

The control-to-state mapping $G : L^2(\Omega) \rightarrow H_0^1(\Omega) \cap H^2(\Omega)$, $G : u \mapsto y_u$ that assigns to u the unique solution y of (1.1), is now well defined. We fix the control spaces $U_2 := L^2(\Omega)$, $U_\infty := L^\infty(\Omega)$, and $Y := H_0^1(\Omega) \cap H^2(\Omega)$ as the state space. Notice that Y is continuously embedded in $C(\bar{\Omega})$.

Let us define the set \mathcal{K} of admissible controls by

$$\mathcal{K} = \{u \in L^\infty(\Omega) : \alpha \leq u(x) \leq \beta \quad \text{for a.a. } x \in \Omega\}.$$

We write the integral functional of (P) as

$$F(y, u) := \int_{\Omega} L(x, y(x), u(x)) dx.$$

Then the reduced objective functional J admits the form

$$J(u) := F(G(u), u) = \int_{\Omega} L(x, y_u(x), u(x)) dx. \quad (3.13)$$

Thanks to our assumptions, F is defined and twice continuously Fréchet differentiable on $C(\bar{\Omega}) \times L^{\infty}(\Omega)$, but it is not in general differentiable on $C(\bar{\Omega}) \times L^2(\Omega)$. In contrast to this, the objective functional (1.2) has better properties, if $L(x, y, u) := \ell(x, y) + \Lambda/2 u^2$ satisfies the assumption (H4). This functional is of class C^2 even in $C(\bar{\Omega}) \times L^2(\Omega)$. Therefore, for (1.2) we can take $U_2 = U_{\infty} := L^2(\Omega)$.

Now we are able to relate the control problem (P) to the abstract optimization problem (\mathcal{P}). (P) can be written in the form

$$(P) \quad \min_{u \in \mathcal{K}} J(u),$$

where \mathcal{K} and J are defined as above.

Theorem 3.2 ([10]) *For every $q > n/2$, the mapping $G : L^q(\Omega) \rightarrow W^{2,q}(\Omega)$, defined by $G(u) = y_u$, is of class C^2 . For any $v \in L^q(\Omega)$, the function $z_v = G'(u)v$ is the unique solution in $W^{2,q}(\Omega) \cap W_0^{1,q}(\Omega)$ of the equation*

$$\begin{cases} -\operatorname{div} \left[a(x, y_u) \nabla z + \frac{\partial a}{\partial y}(x, y_u) z \nabla y_u \right] + \frac{\partial f}{\partial y}(x, y_u) z = v & \text{in } \Omega \\ z = 0 & \text{on } \Gamma. \end{cases} \quad (3.14)$$

Moreover, for any $v_1, v_2 \in L^q(\Omega)$ the function $z = G''(u)[v_1, v_2]$ is the unique solution in $W^{2,q}(\Omega) \cap W_0^{1,q}(\Omega)$ of the following equation, where $z_{v_i} = G'(u)v_i$, $i = 1, 2$:

$$\begin{cases} -\operatorname{div} \left[a(x, y_u) \nabla z + \frac{\partial a}{\partial y}(x, y_u) z \nabla y_u \right] + \frac{\partial f}{\partial y}(x, y_u) z = -\frac{\partial^2 f}{\partial y^2}(x, y_u) z_{v_1} z_{v_2} \\ + \operatorname{div} \left[\frac{\partial a}{\partial y}(x, y_u) (z_{v_1} \nabla z_{v_2} + \nabla z_{v_1} z_{v_2}) + \frac{\partial^2 a}{\partial y^2}(x, y_u) z_{v_1} z_{v_2} \nabla y_u \right] & \text{in } \Omega \\ z = 0 & \text{on } \Gamma. \end{cases} \quad (3.15)$$

We have $n \leq 3$, hence our mapping G is of class C^2 from U_2 to Y . For the next theorem we recall that \bar{p} was defined in (H3).

Theorem 3.3 ([10]) *The functional $J : L^\infty(\Omega) \rightarrow \mathbb{R}$ is of class C^2 . For every $u, v \in L^\infty(\Omega)$ we have*

$$J'(u)v = \int_{\Omega} \left(\frac{\partial L}{\partial u}(x, y_u, u) + \varphi_u \right) v \, dx, \quad (3.16)$$

where $\varphi_u \in W_0^{1,\bar{p}}(\Omega) \cap W^{2,\bar{p}}(\Omega)$ is the unique solution of the adjoint equation

$$\begin{cases} -\operatorname{div} [a(x, y_u) \nabla \varphi] + \frac{\partial a}{\partial y}(x, y_u) \nabla y_u \cdot \nabla \varphi + \frac{\partial f}{\partial y}(x, y_u) \varphi = \frac{\partial L}{\partial y}(x, y_u, u) & \text{in } \Omega \\ \varphi = 0 & \text{on } \Gamma. \end{cases} \quad (3.17)$$

With $z_{v_i} = G'(u)v_i$, $i = 1, 2$, the second derivative J'' is given by

$$\begin{aligned} J''(u)v_1v_2 = & \int_{\Omega} \left\{ \frac{\partial^2 L}{\partial y^2}(x, y_u, u)z_{v_1}z_{v_2} + \frac{\partial^2 L}{\partial y \partial u}(x, y_u, u)(z_{v_1}v_2 + z_{v_2}v_1) \right. \\ & + \frac{\partial^2 L}{\partial u^2}(x, y_u, u)v_1v_2 - \varphi_u \frac{\partial^2 f}{\partial y^2}(x, y_u)z_{v_1}z_{v_2} \\ & \left. - \nabla \varphi_u \left[\frac{\partial a}{\partial y}(x, y_u)(z_{v_1} \nabla z_{v_2} + \nabla z_{v_1} z_{v_2}) + \frac{\partial^2 a}{\partial y^2}(x, y_u)z_{v_1}z_{v_2} \nabla y_u \right] \right\} dx. \end{aligned} \quad (3.18)$$

The property $\varphi_u \in W^{2,\bar{p}}(\Omega)$ is not explicitly mentioned in the associated proof. However, it follows as in the proof of [10, Thm. 2.11]. In the case of a locally optimal control \bar{u} , (3.16) and (2.3) amount to

$$\int_{\Omega} \left(\frac{\partial L}{\partial u}(x, y_{\bar{u}}(x), \bar{u}(x)) + \varphi_{\bar{u}}(x) \right) (u(x) - \bar{u}(x)) \, dx \geq 0 \quad \forall u \in \mathcal{K}. \quad (3.19)$$

A standard discussion of (3.19) yields the following projection result:

Theorem 3.4 ([10]) *If \bar{u} is a local minimum of (P), then the equation*

$$\frac{\partial L}{\partial u}(x, \bar{y}(x), t) + \bar{\varphi}(x) = 0 \quad (3.20)$$

has a unique solution $\bar{t} = \bar{s}(x)$ for every $x \in \bar{\Omega}$. The function $\bar{s} : \bar{\Omega} \rightarrow \mathbb{R}$ is Lipschitz and \bar{u} is related to \bar{s} by the projection formula

$$\bar{u}(x) = \mathbb{P}_{[\alpha, \beta]}(\bar{s}(x)) = \max\{\min\{\beta, \bar{s}(x)\}, \alpha\}. \quad (3.21)$$

Consequently, also \bar{u} is Lipschitz in $\bar{\Omega}$.

In the particular case of (1.2), the projection formula (3.21) simplifies to

$$\bar{u}(x) = \mathbb{P}_{[\alpha, \beta]} \left\{ -\frac{1}{\Lambda} \bar{\varphi}(x) \right\}. \quad (3.22)$$

3.2 Finite element discretization of (P)

For a numerical approximation of (P), we have to discretize the state equation and in general also the set of control functions. First, we introduce the finite element approximation of the quasilinear state equation of (P). To this aim, we consider a family of regular triangulations $\{\mathcal{T}_h\}_{h>0}$ of $\bar{\Omega}$, defined in the standard way, e.g. in [5]. In particular, this definition excludes the so-called hanging nodes. Moreover, this triangulation is supposed to be regular and to satisfy an inverse assumption; see (i) below. Also for $\Omega \subset \mathbb{R}^3$, we shall speak of a triangulation, although we will have a set of tetrahedra T .

With each element $T \in \mathcal{T}_h$, we associate two parameters $\rho(T)$ and $\sigma(T)$, where $\rho(T)$ denotes the diameter of the set T and $\sigma(T)$ is the diameter of the largest ball contained in T . Define the mesh size by

$$h := \max_{T \in \mathcal{T}_h} \rho(T).$$

We suppose that the following standard regularity assumptions are satisfied.

(i) There exist two positive constants ρ and σ such that

$$\frac{\rho(T)}{\sigma(T)} \leq \sigma, \quad \frac{h}{\rho(T)} \leq \rho$$

hold for all $T \in \mathcal{T}_h$ and all $h > 0$.

(ii) Define $\bar{\Omega}_h = \cup_{T \in \mathcal{T}_h} T$, and let Ω_h and Γ_h denote its interior and its boundary, respectively. We assume that $\bar{\Omega}_h$ is convex and that the vertices of \mathcal{T}_h placed on the boundary Γ_h are points of Γ . From [23, estimate (5.2.19)] we know that

$$|\Omega \setminus \Omega_h| \leq Ch^2. \quad (3.23)$$

We will use piecewise linear approximations for the states, thus we set

$$Y_h = \{y_h \in C(\bar{\Omega}) \mid y_h|_T \in \mathcal{P}_1, \text{ for all } T \in \mathcal{T}_h, \text{ and } y_h = 0 \text{ on } \bar{\Omega} \setminus \Omega_h\},$$

where \mathcal{P}_1 is the space of polynomials of degree less or equal than 1.

The discrete version of the state equation is

$$\begin{cases} \text{Find } y_h \in Y_h \text{ such that, for all } z_h \in Y_h, \\ \int_{\Omega_h} [a(x, y_h(x)) \nabla y_h \cdot \nabla z_h + f(x, y_h(x)) z_h] dx = \int_{\Omega_h} u z_h dx. \end{cases} \quad (3.24)$$

By an application of the Brouwer fixed point theorem, we showed the existence of at least one solution to this equation. To our surprise, we were not able to show uniqueness of the solution. To our best knowledge, this is an open question until now. Under the additional requirement of boundedness of a , we were able to show uniqueness for sufficiently small h . The situation is easier for semilinear

elliptic state equations ($a(x, y) = a(x)$) where the solution of (3.24) is unique, [2].

To set up the discretized problem, we define its objective functional by

$$F_h(y_h, u) = \int_{\Omega_h} L(x, y_h(x), u(x)) dx.$$

Remark 3.5 *In what follows, we will derive various estimates of $\|u_h - u\|_{L^2(\Omega_h)}$ for functions $u_h, u \in \mathcal{K}$. We should notice that then $\|u_h - u\|_{L^2(\Omega_h)} \rightarrow 0$ is equivalent to $\|u_h - u\|_{L^2(\Omega)} \rightarrow 0$ independently on how the functions u, u_h are defined on $\Omega \setminus \Omega_h$. This follows immediately from $\|u - u_h\|_{L^2(\Omega \setminus \Omega_h)} \leq Ch^2$ for all $u_h, u \in \mathcal{K}$.*

We also introduce a set $\mathcal{K}_h \subset \mathcal{K}$ of discrete control functions u_h . These functions are only needed on Ω_h , because the integral of F_h is only evaluated there. However, we need $\mathcal{K}_h \subset \mathcal{K}$ to apply the abstract theory. For this reason, discretized functions u_h defined only on Ω_h can be extended arbitrarily to $\Omega \setminus \Omega_h$ such that the resulting function belongs to \mathcal{K} . Later, in the context of error estimates, we will always set $u_h(x) = \bar{u}(x)$ a.e. in $\Omega \setminus \Omega_h$, where \bar{u} is a fixed locally optimal reference control.

With these definitions, we consider the following family of discretized problems depending on the mesh size $h > 0$:

$$(Q_h) \quad \min_{(y_h, u_h) \in Y_h \times \mathcal{K}_h} F_h(y_h, u_h).$$

(Q_h) is a discrete approximation of the optimal control problem (P). As y_h is possibly not uniquely determined by u_h , this problem does not yet fit in the abstract theory. However, the discretized states are locally unique:

Theorem 3.6 ([11, Thm. 3.2]) *Suppose that $n < p \leq \bar{p}$. Then there exist $h_0 > 0$, $\rho_{\bar{u}} > 0$ and $\rho_{\bar{y}} > 0$ such that, for any $h < h_0$ and any u in the closed ball $\bar{B}(\bar{u}, \rho_{\bar{u}})$ of $L^p(\Omega)$, the equation (3.24) has a unique solution $y_h(u)$ in the closed ball $\bar{B}(\bar{y}, \rho_{\bar{y}})$ of $W_0^{1,p}(\Omega)$. Moreover, there holds the estimate*

$$\|y_u - y_h(u)\|_{L^p(\Omega_h)} + h\|y_u - y_h(u)\|_{W^{1,p}(\Omega_h)} \leq C(\bar{u})h^2. \quad (3.25)$$

To simplify the presentation, we shall use the same notation $\bar{B}(\bar{u}, \rho_{\bar{u}})$ and $\bar{B}(\bar{y}, \rho_{\bar{y}})$ for the balls in $L^p(\Omega)$ and $W^{1,p}(\Omega_h)$. This should not lead to confusion. Moreover, as in this theorem, we shall denote the particular solution y_h of (3.24) contained in $\bar{B}(\bar{y}, \rho_{\bar{y}})$ by $y_h(u)$. Introducing a mapping $G_h : \bar{B}(\bar{u}, \rho_{\bar{u}}) \rightarrow \bar{B}(\bar{y}, \rho_{\bar{y}})$ by $G_h : u \mapsto y_h(u)$, we define the family of approximated functionals

$$J_h(u) := F_h(G_h(u), u)$$

and introduce the family of discretized problems

$$(P_h) \quad \min_{u_h \in \mathcal{K}_h \cap \bar{B}(\bar{u}, \rho_{\bar{u}})} J_h(u_h).$$

Theorem 3.7 *Let \bar{u} be a strict local solution of (P). Then there exists a sequence $\{(\bar{y}_h, \bar{u}_h)\}_{h>0}$ of local solutions to (Q_h) such that*

$$\lim_{h \rightarrow 0} \{\|\bar{u}_h - \bar{u}\|_{L^p(\Omega)} + \|\bar{y}_h - y_{\bar{u}}\|_{W^{1,p}(\Omega)}\} = 0 \quad \forall p \leq \bar{p}$$

and $\lim_{h \rightarrow 0} F_h(\bar{y}_h, \bar{u}_h) = F(y_{\bar{u}}, \bar{u}) = J(\bar{u})$, where $\bar{y} = y_{\bar{u}}$.

Proof. Consider the admissible set $\hat{\mathcal{K}} := \mathcal{K} \cap \bar{B}(\bar{u}, \rho_{\bar{u}})$. It is clear that \bar{u} is also a strict local solution to

$$(\hat{P}) \quad \min_{u \in \hat{\mathcal{K}}} J(u).$$

This problem fits to our abstract theory by $(\mathcal{P}) := (\hat{P})$. We consider now the approximated problem

$$(\hat{P}_h) \quad \min_{u_h \in \hat{\mathcal{K}}_h} J_h(u),$$

where $\hat{\mathcal{K}}_h := \mathcal{K}_h \cap \bar{B}(\bar{u}, \rho_{\bar{u}})$. Later, we shall verify that all the assumptions (A1)-(A7) are satisfied for (P), in particular for (\hat{P}) , hence Theorem 2.11 is applicable: There exists a sequence \bar{u}_h of global solutions to (\hat{P}_h) that converges strongly in U_2 to \bar{u} . Consequently, $\bar{u}_h \in B(\bar{u}, \rho_{\bar{u}}/2)$ holds for all sufficiently small $h > 0$. Then \bar{u}_h does not touch the boundary of $B(\bar{u}, \rho_{\bar{u}})$. It is easy to see that therefore (\bar{y}_h, \bar{u}_h) is a local solution to (Q_h) , since the restriction $u \in \bar{B}(\bar{u}, \rho_{\bar{u}})$ is not active.

The convergence in $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} \rightarrow 0$ implies the convergence $\bar{u}_h \rightarrow \bar{u}$ in all spaces $L^p(\Omega)$, $1 \leq p < \infty$, because $\hat{\mathcal{K}}_h \subset \mathcal{K}$ and \mathcal{K} is bounded in $L^\infty(\Omega)$, hence the other statements follow immediately from (3.25). \square

3.3 Finite element discretization of the adjoint equation

The discrete adjoint equation for (P) is the finite element version of the adjoint equation (3.17):

$$\left\{ \begin{array}{l} \text{Find } \varphi_h \in Y_h \text{ such that, for all } \phi_h \in Y_h, \\ \int_{\Omega_h} \left\{ a(x, y_h) \nabla \varphi_h \cdot \nabla \phi_h + \left[\frac{\partial a}{\partial y}(x, y_h) \nabla y_h \cdot \nabla \varphi_h + \frac{\partial f}{\partial y}(x, y_h) \varphi_h \right] \phi_h \right\} dx \\ \qquad \qquad \qquad = \int_{\Omega_h} \frac{\partial L}{\partial y}(x, y_h, u_h) \phi_h dx. \end{array} \right. \quad (3.26)$$

By [11, Thms. 4.1, 5.1], this equation has a unique solution for all sufficiently small $h > 0$, see also Theorem 3.10. In view of the possible non-uniqueness of the solution to the discrete state equation, we associated with $u \in B(\bar{u}, \rho_{\bar{u}})$ the locally unique state $y_h(u)$. We insert $y_h(u)$ for y_h in the discrete adjoint equation above; the associated unique solution φ_h is denoted by $\varphi_h(u)$. Then the following result holds:

Lemma 3.8 ([11]) *For every pair $u, v \in \mathcal{K}$ with $v \in B(\bar{u}, \rho_{\bar{u}})$, there holds*

$$\|y_u - y_h(v)\|_{L^2(\Omega_h)} + \|\varphi_u - \varphi_h(v)\|_{L^2(\Omega_h)} \leq C(h^2 + \|u - v\|_{L^2(\Omega)}) \quad (3.27)$$

$$\|y_u - y_h(v)\|_{H^1(\Omega_h)} + \|\varphi_u - \varphi_h(v)\|_{H^1(\Omega_h)} \leq C(h + \|u - v\|_{L^2(\Omega)}) \quad (3.28)$$

$$\|y_u - y_h(v)\|_{L^\infty(\Omega_h)} + \|\varphi_u - \varphi_h(v)\|_{L^\infty(\Omega_h)} \leq C(h + \|u - v\|_{L^2(\Omega)}). \quad (3.29)$$

3.4 Necessary optimality conditions for (P_h)

The theory of first-order necessary conditions for (P_h) is similar to that of (P) .

Theorem 3.9 ([11]) *For every $h \leq h_0$, the functional $J_h : B(\bar{u}, \rho_{\bar{u}}) \cap L^\infty(\Omega) \rightarrow \mathbb{R}$ is of class C^1 , and its derivative is given by*

$$J'_h(u)v = \int_{\Omega_h} \left(\frac{\partial L}{\partial u}(x, y_h(u), u) + \varphi_h(u) \right) v \, dx, \quad (3.30)$$

where $\varphi_h(u) \in Y_h$ is the unique solution of the adjoint state equation

$$\begin{aligned} & \int_{\Omega_h} [a(x, y_h(u)) \nabla \varphi_h(u) \cdot \nabla \phi_h + \frac{\partial a}{\partial y}(x, y_h(u)) \nabla y_h(u) \cdot \nabla \varphi_h(u) \phi_h] \, dx \\ & + \int_{\Omega_h} \frac{\partial f}{\partial y}(x, y_h(u)) \varphi_h(u) \phi_h \, dx = \int_{\Omega_h} \frac{\partial L}{\partial y}(x, y_h(u), u) \phi_h \, dx \quad \forall \phi_h \in Y_h. \end{aligned} \quad (3.31)$$

From this expression for $J'(u)$, we obtain first-order necessary optimality conditions for the discretized problem. Let $(\bar{y}_h, \bar{u}_h) \in B(\bar{y}, \rho_{\bar{y}}) \times B(\bar{u}, \rho_{\bar{u}})$ be a local solution to (Q_h) in the sense of U_∞ such that \bar{u}_h belongs to $B(\bar{u}, \rho_{\bar{u}}/2)$. The existence of such solutions follows from Theorem 3.7 for sufficiently small $h > 0$. Then \bar{u}_h is also a local solution of (P_h) in the sense of U_∞ that does not touch the boundary of $B(\bar{u}, \rho_{\bar{u}})$ where $J_h(\bar{u}_h)$ is well defined. Therefore, we are justified to apply the following result on necessary optimality conditions that was proved in [11] for the problem (P_h) .

Theorem 3.10 ([11]) *Let $(\bar{y}_h, \bar{u}_h) \in \bar{B}(\bar{y}, \rho_{\bar{y}}) \times \bar{B}(\bar{u}, \rho_{\bar{u}})$ be a local solution of (Q_h) . Then there exists a unique solution $\bar{\varphi}_h$ in Y_h of*

$$\begin{aligned} & \int_{\Omega_h} [a(x, \bar{y}_h) \nabla \bar{\varphi}_h \cdot \nabla \phi_h + \frac{\partial a}{\partial y}(x, \bar{y}_h) \nabla \bar{y}_h \cdot \nabla \bar{\varphi}_h \phi_h] \, dx + \int_{\Omega_h} \frac{\partial f}{\partial y}(x, \bar{y}_h) \bar{\varphi}_h \phi_h \, dx \\ & = \int_{\Omega_h} \frac{\partial L}{\partial y}(x, \bar{y}_h, \bar{u}_h) \phi_h \, dx \quad \forall \phi_h \in Y_h \end{aligned} \quad (3.32)$$

such that

$$\int_{\Omega_h} \left(\frac{\partial L}{\partial u}(x, \bar{y}_h, \bar{u}_h) + \bar{\varphi}_h \right) (u_h - \bar{u}_h) \, dx \geq 0 \quad \forall u_h \in \mathcal{K}_h. \quad (3.33)$$

From (3.33), we derive again projection formulas for \bar{u}_h , if \mathcal{K}_h is the set (4.41) of functions being piecewise constant in Ω_h or the variational discretization $\mathcal{K}_h = \mathcal{K}$ is applied. In the case of the variational discretization $\mathcal{K}_h = \mathcal{K}$, proceeding as in Theorem 3.4, we deduce from (3.33) that

$$\bar{u}_h(x) = \mathbb{P}_{[\alpha, \beta]}(\bar{s}_h(x)) \quad \forall x \in \Omega_h. \quad (3.34)$$

Here, $\bar{s}_h(x)$ is the unique solution t of the equation

$$\frac{\partial L}{\partial u}(x, \bar{y}_h(x), t) + \bar{\varphi}_h(x) = 0 \quad \forall x \in \Omega_h. \quad (3.35)$$

In the case of the integral functional (1.2), this formula admits the form

$$\bar{u}_h(x) = \mathbb{P}_{[\alpha, \beta]}(-\frac{1}{\Lambda} \bar{\varphi}_h(x)), \quad (3.36)$$

hence \bar{u}_h is a piecewise linear function, although it was formally not discretized.

For piecewise constant control functions, formula (3.33) yields

$$\bar{u}_{h|_T} = \mathbb{P}_{[\alpha, \beta]}(\bar{s}_{h|_T}) \quad \forall T \in \mathcal{T}_h, \quad (3.37)$$

where $s_{h|_T}$ is the unique real number satisfying the equation

$$\int_T \left(\frac{\partial L}{\partial u}(x, \bar{y}_h(x), \bar{s}_{h|_T}) + \bar{\varphi}_h(x) \right) dx = 0. \quad (3.38)$$

For piecewise linear controls, we do not have an analogous projection formula.

4 Error estimates for the problem (P)

4.1 Control discretization and error estimate for (P)

We shall verify the general assumptions (A1)–(A7) for our quasilinear control problem (P) so that Theorem 2.14 provides a general tool for deriving error estimates. To apply it, the three different terms in the right-hand side of (2.19) must be chosen appropriately.

Once and for all, we select a fixed locally optimal reference control $\bar{u} \in \mathcal{K}$. The treatment of the first term in (2.19) is fairly clear: To estimate ε_h^2 , an error analysis of the state equation is needed. The two other terms, $\|u_h - \bar{u}\|^2$ and $J'(\bar{u})(u_h - \bar{u})$ need special strategies to arrive at optimal error estimates. In particular, they depend on how the control functions u are discretized. We explain associated methods and their consequences below. Let us notice already here that, within the framework of the abstract theory, we have to work in $\hat{\mathcal{K}}_h$ and $\hat{\mathcal{K}}$ instead of \mathcal{K}_h and \mathcal{K} , respectively.

4.1.1 Variational discretization of controls

If a control $u \in \mathcal{K}$ is not discretized, we set $\mathcal{K}_h = \mathcal{K}$. Then $u|_{\Omega_h}$ is to be inserted in J_h . We extend u_h to $\Omega \setminus \Omega_h$ by $u_h(x) := \bar{u}(x)$ on $\Omega \setminus \Omega_h$. In this case, a discretization of the control enters only implicitly by the necessary optimality conditions. The discretized adjoint state $\bar{\varphi}_h$ is a piecewise linear function. If $\mathcal{K}_h = \mathcal{K}$ and the simplified functional (1.2) is considered, the discretized optimal control \bar{u}_h of (P_h) satisfies the projection formula (3.36). Then, with $\bar{\varphi}_h$, also the associated control \bar{u}_h is piecewise linear on Ω_h . This strategy is called *variational discretization*, cf. Hinze [20].

For the more general objective functional of (P), the projection formula (3.34) has to be used instead. In this case, the associated discretized control \bar{u}_h is not necessarily piecewise linear and has a more complicated structure. Then the variational discretization is only of limited value. Nevertheless, the following result holds:

Theorem 4.1 (Variational discretization, L^2 -estimate) *Assume that (H1)-(H4) are satisfied. Let \bar{u} be a strict locally optimal control of (P) that fulfills the second-order sufficient optimality condition (2.6) and suppose that $\mathcal{K}_h = \mathcal{K}$. Let further $\{(\bar{y}_h, \bar{u}_h)\}$ be a sequence of locally optimal solutions for (Q_h) that converges strongly in $W^{1,p}(\Omega) \times L^p(\Omega)$ to (\bar{y}, \bar{u}) and exists according to Theorem 2.12. Then there is some constant $C > 0$ not depending on h such that*

$$\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} \leq C h^2 \quad \forall h > 0. \quad (4.1)$$

Proof. We can apply Theorem 2.14, since all needed assumptions are satisfied. In (2.19), we set $u_h := \bar{u}$. Then only the term ε_h^2 does not vanish and we obtain instantly the error estimate

$$\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} \leq C \varepsilon_h. \quad (4.2)$$

In (4.59) we will show $\varepsilon_h = C h^2$, hence (2.19) yields the result. \square

4.1.2 Piecewise constant control approximation

We assumed in (H1) that Ω is convex with smooth boundary to obtain the needed regularity of \bar{u} . For formal reasons, we must specify the discretized controls u_h also in $\Omega \setminus \Omega_h$. We define

$$\mathcal{K}_h = \{u_h \in \mathcal{K} : u_h(x) \equiv u_T \in \mathbb{R} \text{ in each triangle } T, u_h(x) = \bar{u}(x) \forall x \in \Omega \setminus \Omega_h\}. \quad (4.3)$$

Notice that we do not need the values of u_h in $\Omega \setminus \Omega_h$; F_h and the discretized equations are only considered in Ω_h . It does not matter that we do not know \bar{u} in advance. For (P), the derivative $J'(\bar{u})$ is obtained by (3.30). We write it in the form

$$J'(\bar{u})u = \int_{\Omega} \bar{d}(x)u(x)dx,$$

where the reduced gradient $\bar{d} \in L^2(\Omega)$ is given by

$$\bar{d}(x) := \frac{\partial L}{\partial u}(x, \bar{y}(x), \bar{u}(x)) + \bar{\varphi}(x). \quad (4.4)$$

For error estimates, we apply Theorem 2.14. Again, we have to define the function u_h in (2.19) in the right way. With piecewise constant controls, we proceed as follows: On any triangle $T \in \mathcal{T}_h$, we define u_h as the piecewise constant interpolate

$$u_h(x) := \frac{1}{|T|} \int_T \bar{u}(x) dx \quad \forall x \in T, \forall T \in \mathcal{T}_h. \quad (4.5)$$

Notice that we have set $u_h(x) = \bar{u}(x)$ for all $x \in \Omega \setminus \Omega_h$.

In view of (3.27), we can select $\varepsilon_h = Ch^2$ in Theorem 2.14. For piecewise constant control approximation, we cannot fully benefit from this. Even $\varepsilon_h = Ch$ would guarantee the error estimate below.

In the next proofs, we denote by $\mathcal{T}_0 \subset \mathcal{T}$ the set of all triangles T , which contain at least one zero x_T of \bar{d} . Notice that we need continuity of \bar{d} to make this definition correct. In all other triangles, \bar{d} is by continuity either everywhere positive or everywhere negative. The set \mathcal{T}_0 will be essential for deriving optimal error estimates.

Theorem 4.2 (Piecewise constant controls, L^2 -estimate) *Assume (H1)-(H4), let a locally optimal control \bar{u} of (P) satisfy the second-order sufficient conditions (2.6). Define the admissible set \mathcal{K}_h by (4.41). Let further \bar{u}_h be a sequence of locally optimal (piecewise constant) solutions to (Q_h) that converges strongly in $L^2(\Omega)$ to \bar{u} . Then there is some constant $C > 0$ not depending on h such that*

$$\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} \leq Ch \quad \forall h > 0. \quad (4.6)$$

Proof. By Theorem 3.3, the adjoint state $\bar{\varphi}$ belongs to $W^{2,\bar{p}}(\Omega)$, since the right-hand side of the adjoint equation (3.17) belongs to $L^{\bar{p}}(\Omega)$. Therefore, $\bar{\varphi}$ belongs even to $C^1(\bar{\Omega})$. Theorem 3.4 yields that \bar{u} is Lipschitz, hence \bar{d} is also Lipschitz. Notice that assumption (3.10) implies that L is Lipschitz with respect to x . The projection formula ensures in particular $\bar{u} \in H^1(\Omega)$, hence it holds

$$\|u_h - \bar{u}\|_{L^2(\Omega)} \leq Ch, \quad (4.7)$$

cf. Ciarlet [13]. Consider first the triangles $T \notin \mathcal{T}_0$. Here, \bar{d} is either positive or negative in T . If $\bar{d}(x) > 0$ in T then a.e. we have $\bar{u}(x) = \alpha$. This follows from a standard pointwise discussion of (3.19), see e.g. [24, Lemma 2.2.6].

By the construction (4.43), it then also holds $u_h(x) = \alpha \forall x \in T$. If $\bar{d}(x) < 0$ a.e. in T , we obtain accordingly $u_h(x) = \beta \forall x \in T$. In either case, we have

$$\int_T \bar{d}(x)(\bar{u}_h(x) - \bar{u}(x)) dx = 0 \quad \forall T \notin \mathcal{T}_0.$$

In each $T \in \mathcal{T}_0$, there exists x_T such that $\bar{d}(x_T) = 0$. We estimate

$$\begin{aligned}
J'(\bar{u})(u_h - \bar{u}) &= \sum_{T \in \mathcal{T}_0} \int_T \bar{d}(x)(u_h(x) - \bar{u}(x)) dx \\
&= \sum_{T \in \mathcal{T}_0} \int_T (\bar{d}(x) - \bar{d}(x_T))(u_h(x) - \bar{u}(x)) dx \\
&\leq L_{\bar{d}} h \sum_{T \in \mathcal{T}_0} \int_T |u_h(x) - \bar{u}(x)| dx \leq C h \|u_h - \bar{u}\|_{L^2(\Omega)} \leq C h^2
\end{aligned} \tag{4.8}$$

where $L_{\bar{d}}$ is the Lipschitz constant of \bar{d} ; we used the inequality (4.45).

Now we apply Theorem 2.14 and use (3.27), (4.45) and (4.46). These estimates show that all three terms appearing in the right-hand side of (2.19) have at least the order h^2 , hence

$$\varepsilon_h^2 + \|u_h - \bar{u}\|_{L^2(\Omega)}^2 + J'(\bar{u})(u_h - \bar{u}) \leq C h^2.$$

Now (2.19) yields $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} \leq C h$ as it was claimed in the theorem. \square

This estimate is optimal, since the order of the L^2 -approximation of a function by step functions is not in general better than h . However, as it is observed numerically, the error $\|\bar{y}_h - \bar{y}\|_{L^\infty(\Omega_h)}$ for the state function has often a higher order. This phenomenon was first observed in [8] and explained analytically by [22].

Our last result leads directly to an associated maximum-norm estimate.

Theorem 4.3 (Piecewise constant controls, L^∞ -estimate) *Under the assumptions of Theorem 4.2, there is a constant $C > 0$ not depending on h such that*

$$\|\bar{u}_h - \bar{u}\|_{L^\infty(\Omega_h)} \leq C h \quad \forall h > 0. \tag{4.9}$$

Proof. From Lemma 3.8, estimate (3.29) and Theorem 4.2, we get

$$\|\varphi_{\bar{u}} - \varphi_h(\bar{u}_h)\|_{C(\bar{\Omega}_h)} \leq C h.$$

Subtracting (3.21) and (3.37), we obtain a.e. on each triangle T

$$\begin{aligned}
|\bar{u}(x) - \bar{u}_h(x)| &= |\mathbb{P}_{[\alpha, \beta]}(\bar{s}(x)) - \mathbb{P}_{[\alpha, \beta]}(\bar{s}_{h|_T})| \\
&\leq |\bar{s}(x) - \bar{s}_{h|_T}| \leq \frac{1}{\Lambda} \{ \|\varphi_{\bar{u}} - \varphi_h(\bar{u}_h)\|_{L^\infty(\Omega_h)} + c \|\bar{y} - y_h(\bar{u}_h)\|_{L^\infty(\Omega_h)} \} \\
&\leq c \{ \|\varphi_{\bar{u}} - \varphi_{\bar{u}_h}\|_{L^\infty(\Omega_h)} + \|\varphi_{\bar{u}_h} - \varphi_h(\bar{u}_h)\|_{L^\infty(\Omega_h)} \\
&\quad + \|\bar{y}_{\bar{u}} - y_{\bar{u}_h}\|_{L^\infty(\Omega_h)} + \|y_{\bar{u}_h} - y_h(\bar{u}_h)\|_{L^\infty(\Omega_h)} \} \tag{4.10}
\end{aligned}$$

$$\leq c [h + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega_h)}] \leq \tilde{C} h. \tag{4.11}$$

The second inequality follows with some effort by an application of the mean value theorem in (3.38) and a Taylor expansion of $L(x, \bar{y}_h(x), s_{h|_T})$ with respect

to the second and third variable. Moreover, Lemma 3.8, the Lipschitz continuity of $y_{\bar{u}}$, $\varphi_{\bar{u}}$, $L(x, y, u)$, and of \bar{s} with respect to x , and the convexity condition of (H4) are needed. In the last line, we applied Theorem 4.2. For details we refer to [11, p.29]. \square

4.1.3 Piecewise linear control approximation

This type of approximation is frequently used in numerical computations, since its implementation is fairly easy. We adopt the notation \bar{d} and \mathcal{T}_0 from above. We define

$$\mathcal{K}_h = \{u \in \mathcal{K} : u|_T \in \mathcal{P}_1 \text{ for all } T \in \mathcal{T}_h \text{ and } u|_{\Omega \setminus \Omega_h} = \bar{u}\}. \quad (4.12)$$

To apply Theorem 2.14, we use the function

$$u_h := \Pi_h \bar{u} := \begin{cases} (\pi_h \bar{u})(x) & \forall x \in \Omega_h \\ \bar{u}(x) & \forall x \in \Omega \setminus \Omega_h, \end{cases} \quad (4.13)$$

where $\pi_h : C(\bar{\Omega}) \rightarrow C(\bar{\Omega}_h)$ associates to $u \in C(\bar{\Omega})$ its standard piecewise linear interpolate on Ω_h . It holds that $\|u - \Pi_h u\|_{C(\bar{\Omega}_h)} \rightarrow 0$, if u is continuous on $\bar{\Omega}$.

In general, we expect a better approximation of \bar{u} by piecewise linear functions than by piecewise constant controls. However, without further assumptions, the improvement is only marginal.

Theorem 4.4 (Piecewise linear controls, low order L^2 -estimate) *Assume that \bar{u} is a local solution to (P) that satisfies the second-order sufficient condition (2.6). Let the assumptions (H1)-(H4) be satisfied and \mathcal{K}_h be the set defined in (4.50). If $\{\bar{u}_h\}$ is a sequence of locally optimal piecewise linear controls to (Q_h) with $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} \rightarrow 0$, then $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} = o(h)$, i.e.*

$$\lim_{h \rightarrow 0} \frac{1}{h} \|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} = 0. \quad (4.14)$$

Proof. We proceed similarly as in the proof of the last theorem, but we define now $u_h = \Pi_h \bar{u}$ as in (4.51). If $T \notin \mathcal{T}_0$, then again either $\bar{u}(x) \equiv \alpha$ or $\bar{u}(x) \equiv \beta$ holds in T so that $u_h(x) = \Pi_h \bar{u}_h(x) = \bar{u}(x)$ holds in T . Now we estimate

$$\begin{aligned} J'(\bar{u})(u_h - \bar{u}) &= \sum_{T \in \mathcal{T}_0} \int_T \bar{d}(x)(u_h(x) - \bar{u}(x)) dx \\ &= \sum_{T \in \mathcal{T}_0} \int_T (\bar{d}(x) - \bar{d}(\xi_T))(u_h(x) - \bar{u}(x)) dx \\ &\leq L h \sum_{T \in \mathcal{T}_0} \int_T |\Pi_h \bar{u}(x) - \bar{u}(x)| dx \leq C h \|\Pi_h \bar{u} - \bar{u}\|_{L^2(\Omega_h)} = o(h^2) \end{aligned}$$

in view of the Lipschitz continuity of \bar{d} and the estimate

$$\|\Pi_h u - u\|_{L^2(\Omega_h)} = o(h) \quad (4.15)$$

for $u \in W^{1,p}(\Omega)$, $p > n$, cf. Brenner and Scott [5]. In Theorem 2.14 we have to deal with the terms ε_h^2 , $\|u_h - \bar{u}\|_{L^2(\Omega_h)}^2$, and $J'(\bar{u})(u_h - \bar{u})$. For ε_h , the finite element analysis yields again $\varepsilon_h \leq C h^2$. By (4.53), the second term is of the order $o(h^2)$, while the last one was shown above to have the order $o(h^2)$.

All three terms appearing in the right-hand side of (2.19) were confirmed to have the order $o(h^2)$, hence $\varepsilon_h^2 + \|u_h - \bar{u}\|_{L^2(\Omega_h)}^2 + J'(\bar{u})(u_h - \bar{u}) = o(h^2)$ so that (2.19) yields $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} = o(h)$ as it was claimed in the theorem. \square

In numerical tests with piecewise linear control approximation, often the higher order $h^{3/2}$ of approximation is observed. The reason is that the boundary of the active set of \bar{u} often has zero measure.

Theorem 4.5 (Piecewise linear controls, optimal L^2 -estimate) *Suppose in addition to the assumptions of Theorem 4.4 that $L(x, y, u) = l(x, y) + (\Lambda/2)u^2$ and the measure of the set \mathcal{T}_* formed by the union of all triangles $T \in \mathcal{T}$, where \bar{d} is not identically zero but vanishes in a subset of T of positive measure, can be estimated by*

$$|\mathcal{T}_*| \leq C h.$$

Then there exists some constant $C > 0$ not depending on h such that

$$\|\bar{u}_h - \bar{u}\|_{L^2(\Omega_h)} \leq C h^{\frac{3}{2}} \quad \forall h > 0. \quad (4.16)$$

Proof. We decompose \mathcal{T}_h in three families of triangles

$$\begin{aligned} \mathcal{T}_1 &= \{T \in \mathcal{T}_h : |\bar{d}(x)| > 0 \text{ for a.a. } x \in T\}, \\ \mathcal{T}_2 &= \{T \in \mathcal{T}_h : \bar{d}(x) = 0 \text{ for all } x \in T\}, \\ \mathcal{T}_3 &= \mathcal{T}_h \setminus (\mathcal{T}_1 \cup \mathcal{T}_2). \end{aligned}$$

If $T \in \mathcal{T}_1$, then either $\bar{u}(x) = \alpha \forall x \in T$ or $\bar{u}(x) = \beta \forall x \in T$, therefore $\bar{u}(x) - u_h(x) \equiv 0$ in T . If $T \in \mathcal{T}_2$, then $\bar{u}(x) = -(1/\Lambda)\bar{\varphi}(x)$ for every $x \in T$, hence $\bar{u}|_T \in H^2(T)$ and we have the interpolation error

$$\|\bar{u} - u_h\|_{L^2(T)} \leq C h^2 \|\bar{u}\|_{H^2(T)} = \frac{C}{\Lambda} h^2 \|\bar{\varphi}\|_{H^2(T)}.$$

Finally, in \mathcal{T}_3 it holds

$$\sum_{T \in \mathcal{T}_3} \int_T |\bar{u} - u_h|^2 dx \leq C h^2 \|\bar{u}\|_{C^{0,1}(\bar{\Omega})}^2 \sum_{T \in \mathcal{T}_3} |T| = C h^2 \|\bar{u}\|_{C^{0,1}(\bar{\Omega})}^2 |\mathcal{T}_*| \leq C h^3.$$

From these estimates we infer

$$\|\bar{u} - u_h\|_{L^2(\Omega_h)} \leq C h^{3/2}.$$

It remains to estimate $J'(\bar{u})(u_h - \bar{u})$. This follows from the above remarks and the fact that \bar{u} and u_h coincide in $\Omega \setminus \Omega_h$. In each $T \in \mathcal{T}_3$ there exists x_T such that $\bar{d}(x_T) = 0$, hence

$$\begin{aligned} J'(\bar{u})(u_h - \bar{u}) &= \sum_{T \in \mathcal{T}} \int_T \bar{d}(\bar{u} - u_h) dx = \sum_{T \in \mathcal{T}_3} \int_T (\bar{d}(x) - \bar{d}(x_T))(\bar{u}(x) - u_h(x)) dx \\ &\leq L_{\bar{d}} h \sum_{T \in \mathcal{T}_3} \int_T |\bar{u} - u_h| dx \leq L_{\bar{d}} h \sum_{T \in \mathcal{T}_3} \sqrt{|T|} \left(\int_T |\bar{u} - u_h|^2 dx \right)^{1/2} \\ &\leq L_{\bar{d}} h \left(\sum_{T \in \mathcal{T}_3} |T| \right)^{1/2} \left(\sum_{T \in \mathcal{T}_3} \int_T |\bar{u} - u_h|^2 dx \right)^{1/2} \leq c h^{3/2} \|\bar{u} - u_h\|_{L^2(\Omega_h)} \leq C h^3. \end{aligned}$$

□

4.2 Verification of the assumptions (A1)–(A7)

Finally, we verify the assumptions (A1)–(A7) for the quasilinear optimal control problem (P). This is a fairly technical task, characteristic for problems with quasilinear equations. The verification would be much easier for the semilinear elliptic problems. In either case, the discussion of error estimates would be much longer without our general Theorem 2.14.

We recall our choice $U_\infty = L^\infty(\Omega)$, $U_2 = L^2(\Omega)$. Notice that we must verify the assumptions for the sets $\hat{\mathcal{K}}$ and $\hat{\mathcal{K}}_h$ instead of \mathcal{K} and \mathcal{K}_h , respectively.

(A1): By Theorem 3.2 and the uniform boundedness of \mathcal{K} , all possible states y_u obey with some $C > 0$

$$|y_u(x)| \leq C \quad \forall x \in \Omega, \quad \forall u \in \mathcal{K}.$$

The mapping $u \mapsto y_u$ is in particular continuous from $L^2(\Omega)$ to $H_0^1(\Omega)$, see Theorem 3.2. Let now $u_k \rightharpoonup u$ in $L^2(\Omega)$. The embedding of $L^2(\Omega)$ in $W^{-1,p}(\Omega)$ is compact for all $p > 1$, if $n = 2$, and for all $1 < p < 6$ for $n = 3$. Therefore, we have $u_k \rightharpoonup u$ in $W^{-1,p}(\Omega)$ and Theorem 2.3 of [10] ensures $y_{u_k} \rightarrow y_u$ in $W^{1,p}(\Omega) \subset C(\bar{\Omega})$. We split

$$\int_{\Omega} L(x, y_{u_k}, u_k) dx = \int_{\Omega} L(x, y_u, u_k) dx + \int_{\Omega} (L(x, y_{u_k}, u_k) - L(x, y_u, u_k)) dx.$$

By the uniform boundedness of y_{u_k} , $y_{u_k} \rightarrow y_u$ in $C(\bar{\Omega})$ and the Lipschitz property (3.11), the second integral tends to zero. Thanks to condition (3.8), the mapping $u \mapsto L(x, y, u)$ is convex.

$$\liminf_{k \rightarrow \infty} \int_{\Omega} L(x, y_{u_k}, u_k) dx = \liminf_{k \rightarrow \infty} \int_{\Omega} L(x, y_u, u_k) dx \geq \int_{\Omega} L(x, y_u, u) dx.$$

The functional above is defined on \mathcal{K} , but not in general on $L^2(\Omega)$. In view of this, general results on lower semicontinuity of convex and continuous functionals defined on the whole space are not directly applicable. Therefore, to obtain

the property of lower semicontinuity above, we apply the Mazur theorem and select a sequence of convex combinations of the u_k that converges strongly to u . Notice that also the convex combinations belong to \mathcal{K} . Then the property follows by the convexity of $u \mapsto L(x, y, u)$.

(A2): Here, we take $\mathcal{A} = L^\infty(\Omega)$ and select any fixed real number $r > 0$. It follows from Theorem 3.2 and assumptions (H4) on L that J is of class C^2 in $L^\infty(\Omega)$. The first-order derivative $J'(u)$ is given by (3.16). We estimate

$$|J'(u)v| \leq \int_{\Omega} \left| \frac{\partial L}{\partial u}(x, y_u, u) + \varphi_u \right| |v| dx.$$

The inequality $|J'(u)v| \leq c\|v\|_{L^2(\Omega)}$ follows immediately, since y_u , u , and φ_u are all bounded and measurable. The boundedness of φ_u is obtained from Theorem 3.3.

To discuss $J''(u)$, we employ its representation (3.18). The functions $z_{v_i} = G'(u)v_i$ were introduced in Theorem 3.2. Thanks to $\mathcal{K} \subset L^\infty(\Omega)$, the functions y_u and φ_u belong to $W^{2,\bar{p}}(\Omega) \subset C^1(\bar{\Omega})$. Therefore, they and their gradients are bounded and measurable. The linear mappings $v_i \mapsto z_{v_i}$ and $v_i \mapsto \nabla z_{v_i}$ are continuous from $L^2(\Omega)$ to $L^2(\Omega)^n$, as the reader may verify with Theorem 3.2. Therefore, the second part of (2.4) is easy to confirm. Let us consider the most difficult term in (3.18),

$$\int_{\Omega} \left| \nabla \varphi_u \frac{\partial a}{\partial y}(x, y_u) z_{v_1} \nabla z_{v_2} \right| dx \leq \|\varphi_u\|_{W^{2,\bar{p}}(\Omega)} \left\| \frac{\partial a}{\partial y}(x, y_u) \right\|_{\infty} \|z_{v_1}\|_{L^2(\Omega)} \|z_{v_2}\|_{H_0^1(\Omega)}.$$

It is clear that it can be estimated against $\|v_1\|_{L^2(\Omega)}\|v_2\|_{L^2(\Omega)}$.

Next, we verify (2.5).

Lemma 4.6 *Under the assumptions (H1)-(H4), the objective functional J of problem (P) obeys the estimates*

$$|(J'(u_1) - J'(u_2))v| \leq C(r) \|u_1 - u_2\|_{L^\infty(\Omega)} \|v\|_{L^2(\Omega)} \quad (4.17)$$

$$|(J''(u_1) - J''(u_2))[v_1, v_2]| \leq C(r) \|u_1 - u_2\|_{L^\infty(\Omega)} \|v_1\|_{L^2(\Omega)} \|v_2\|_{L^2(\Omega)} \quad (4.18)$$

for all $u_i \in B_\infty(\bar{u}, r)$, $v, v_i \in L^2(\Omega)$, if $r > 0$ is sufficiently small. The constant $C(r) > 0$ does not depend on u_i , v , and v_i .

Proof. Thanks to the representations of J' and J'' , it is sufficient to verify

$$\|\varphi_{u_1} - \varphi_{u_2}\|_{W^{2,\bar{p}}(\Omega)} \leq C(r) \|u_1 - u_2\|_{L^\infty(\Omega)} \quad \forall u_i \in \mathcal{K} \cap B_\infty(\bar{u}, r). \quad (4.19)$$

Write for short $\varphi_i := \varphi_{u_i}$ and $y_i := y_{u_i}$, $i = 1, 2$; then

$$\begin{aligned} & -\operatorname{div} [a(x, y_1)(\nabla(\varphi_1 - \varphi_2)) + \frac{\partial a}{\partial y}(x, y_1)\nabla y_1 \cdot \nabla(\varphi_1 - \varphi_2) + \frac{\partial f}{\partial y}(x, y_1)(\varphi_1 - \varphi_2)] \\ & = -\operatorname{div} [(a(x, y_2) - a(x, y_1))\nabla\varphi_2] + \left[\frac{\partial a}{\partial y}(x, y_2)\nabla y_2 - \frac{\partial a}{\partial y}(x, y_1)\nabla y_1 \right] \cdot \nabla\varphi_2 \\ & \quad + \left[\frac{\partial f}{\partial y}(x, y_2) - \frac{\partial f}{\partial y}(x, y_1) \right] \varphi_2. \end{aligned}$$

The right-hand side can be estimated in $L^p(\Omega)$ by $\|u_2 - u_1\|_{L^p(\Omega)}$, cf. [11, (4.22)-(4.24)]. Now the $W^{2,p}(\Omega)$ -estimate (4.57) follows from regularity results by [19] for smooth domains. \square

It remains to show that $J''(u)$ is a Legendre form. To this aim, we select a sequence $v_k \rightarrow v$ in $L^2(\Omega)$, $k \rightarrow \infty$, and consider (3.18) with $v_1 = v_2 := v_k$

$$\begin{aligned} J''(u)v_k^2 &= \int_{\Omega} \left\{ \frac{\partial^2 L}{\partial y^2}(x, y_u, u)z_{v_k}^2 + 2 \frac{\partial^2 L}{\partial y \partial u}(x, y_u, u)z_{v_k}v_k \right. \\ &\quad \left. - \nabla \varphi_u \cdot \left[2 \frac{\partial a}{\partial y}(x, y_u)(z_{v_k} \nabla z_{v_k}) + \frac{\partial^2 a}{\partial y^2}(x, y_u)z_{v_k}^2 \nabla y_u \right] \right. \\ &\quad \left. - \varphi_u \frac{\partial^2 f}{\partial y^2}(x, y_u)z_{v_k}^2 + \frac{\partial^2 L}{\partial u^2}(x, y_u, u)v_k^2 \right\} dx. \end{aligned} \quad (4.20)$$

The mapping $v \mapsto z_v$ is linear and continuous, hence also weakly continuous, hence $z_{v_k} \rightharpoonup z_v$ in $H^2(\Omega)$. The embedding of $H^2(\Omega)$ into $C(\bar{\Omega}) \cap H_0^1(\Omega)$ is compact, hence $z_{v_k} \rightarrow z_v$ in $C(\bar{\Omega})$ and $\nabla z_{v_k} \rightarrow \nabla z_v$ in $L^2(\Omega)$. Therefore, all expressions above except the last one tend to the terms associated with the limit v . By the convexity condition of (H4), it holds $\frac{\partial^2 L}{\partial u^2}(x, y_u, u) \geq \Lambda$ a.e. in Ω ; hence the integral of the last term is lower semicontinuous. Altogether, we obtain

$$\liminf_{k \rightarrow \infty} J''(u)v_k^2 \geq J''(u)v.$$

If, in addition, $J''(u)v_k^2 \rightarrow J''(u)v$, then it follows

$$\begin{aligned} \Lambda \|v_k - v\|_{L^2(\Omega)}^2 &\leq \int_{\Omega} \frac{\partial^2 L}{\partial u^2}(x, y_u, u)(v_k - v)^2 dx \\ &= \int_{\Omega} \frac{\partial^2 L}{\partial u^2}(x, y_u, u)v_k^2 dx + \int_{\Omega} \frac{\partial^2 L}{\partial u^2}(x, y_u, u)v^2 dx - 2 \int_{\Omega} \frac{\partial^2 L}{\partial u^2}(x, y_u, u)v_k v dx. \end{aligned}$$

By our additional assumption, the first integral tends to the value of the second one. The same holds true for the third one by weak convergence. Therefore, the right-hand side converges to zero. This implies $\|v_k - v\|_{L^2(\Omega)} \rightarrow 0$, hence $J''(u)$ is a Legendre form and all requirements of (A2) are confirmed.

(A3): For our 3 cases of control discretization, this approximation condition is satisfied. This is trivial for the variational discretization. Moreover, any function $u \in \mathcal{K}$ can be approximated by step functions u_h via the projection (4.43) (insert $\bar{u} := u$ there); it holds $u_h \rightarrow u$ in $L^2(\Omega)$ and $u_h \in \mathcal{K}_h$ for sufficiently small $h > 0$.

For piecewise linear control discretization, we mention that $W^{1,p}(\Omega)$ is dense in $L^2(\Omega)$. To any fixed $u \in \mathcal{K}$ and $\varepsilon > 0$ there exists $u_\varepsilon \in W^{1,p}(\Omega)$ such that $\|u - u_\varepsilon\|_{L^2(\Omega)} < \varepsilon/2$. The function $\hat{u}_\varepsilon(x) := \mathbb{P}_{[\alpha, \beta]} u_\varepsilon(x)$ belongs to $W^{1,p}(\Omega) \cap \mathcal{K}$ and satisfies

$$|u(x) - \hat{u}_\varepsilon(x)| = |\mathbb{P}_{[\alpha, \beta]} u(x) - \mathbb{P}_{[\alpha, \beta]} u_\varepsilon(x)| \leq |u(x) - u_\varepsilon(x)| \quad \text{in } \Omega$$

so that also $\|u - \hat{u}_\varepsilon\|_{L^2(\Omega)} < \varepsilon/2$. The function $u_{\varepsilon, h} := \Pi_h \hat{u}_\varepsilon$ belongs to \mathcal{K}_h and it holds $\|\hat{u}_\varepsilon - u_{\varepsilon, h}\|_{L^2(\Omega)} < \varepsilon/2$ for all sufficiently small $h > 0$. By the triangle

inequality, we have $\|u - u_{\varepsilon, h}\|_{L^2(\Omega)} < \varepsilon$ for all $h < h_0(\varepsilon)$ and this finally confirms (A3).

(A4): Take, as in (A4), $u_k \in \mathcal{K}_h \cap B(\bar{u}, \rho_{\bar{u}})$ with $u_k \rightharpoonup u$ in $L^2(\Omega)$; we write for short $y_k = y_h(u_k)$. We have to show the lower semicontinuity property

$$J_h(u) \leq \liminf_{k \rightarrow \infty} J_h(u_k)$$

for each fixed $h > 0$. The sequence $\{u_k\}$ is bounded in $L^\infty(\Omega)$, hence also $\{y_k\}$ in the finite-dimensional space Y_h . Consequently, there exists a subsequence of $\{y_k\}$ denoted in the same way that converges strongly in $H_0^1(\Omega)$ to some $y_h \in Y_h$. This is a solution to the discretized quasilinear equation associated with $u \in \bar{B}(\bar{u}, \rho_{\bar{u}})$. By local uniqueness, we have $y_h = y_h(u)$, hence all such subsequences converge to the same limit $y_h(u)$ so that $\|y_h(u_k) - y_h(u)\|_{L^2(\Omega_h)} \rightarrow 0$. The sequence $\{y_h(u_k)\}$ is bounded in $L^\infty(\Omega)$. Now the lower semicontinuity of J_h can be verified as the one of J .

Since our set \mathcal{K} is bounded, the second requirement of (A4) need not be verified.

(A5): The condition (2.13) follows immediately from $J(u) = F(G(u), u)$ and the continuity of the mapping $G : L^2(\Omega) \rightarrow H^2(\Omega)$.

Let us show that $|J_h(u_h) - J(u_h)| \rightarrow 0$ as $h \rightarrow 0$. Indeed, we obtain

$$\begin{aligned} |J_h(u_h) - J(u_h)| &\leq \\ &\int_{\Omega_h} |L(x, y_h(u_h), u_h) - L(x, y_{u_h}, u_h)| dx + \int_{\Omega \setminus \Omega_h} |L(x, y_{u_h}, u_h)| dx \\ &\leq C \|y_h(u_h) - y_{u_h}\|_{L^\infty(\Omega)} + \int_{\Omega \setminus \Omega_h} |L(x, 0, 0)| dx + c h^2 \rightarrow 0 \text{ as } h \rightarrow 0. \end{aligned}$$

This follows from the finite element estimate (3.24), since all u_h are bounded, by Lemma 3.8 also $\|y_{u_h}\|_{L^\infty(\Omega)}$ and, in view of the estimate (3.26), also $\|y_h(u_h)\|_{L^\infty(\Omega)}$. Notice that $\|L(\cdot, 0, 0)\|_{L^1(\Omega \setminus \Omega_h)} \rightarrow 0$ as $h \rightarrow 0$. Now the semicontinuity requirement (2.14) is obtained by Remark 2.9, because (A1) has already been confirmed.

(A6): Assume that $u_k \rightharpoonup u$, $u_k \in \mathcal{K}$. We write $y_k := y_{u_k}$ and split

$$\begin{aligned} J(u_k) - J(u) &= \int_{\Omega} (L(x, y_k, u_k) - L(x, y, u)) dx \\ &= \int_{\Omega} (L(x, y_k, u_k) - L(x, y, u_k)) dx + \int_{\Omega} (L(x, y, u_k) - L(x, y, u)) dx \\ &= \underbrace{\int_{\Omega} (L(x, y_k, u_k) - L(x, y, u_k)) dx}_I + \underbrace{\int_{\Omega} \frac{\partial L}{\partial u}(x, y, u)(u_k - u) dx}_{II} \\ &\quad + \underbrace{\frac{1}{2} \int_{\Omega} \frac{\partial^2 L}{\partial u^2}(x, y, u_\theta)(u_k - u)^2 dx}_{III} \end{aligned}$$

with some $u_\theta \in [u_k, u]$. Assume $J(u_k) \rightarrow J(u)$ as in (A6). Then the left-hand side and the integrals I, II converge to zero (cf. the verification of (A5) above), hence also the integral III. By the convexity condition in (H4), we deduce $\|u_k - u\|_{L^2(\Omega)} \rightarrow 0$.

(A7): In view of the possible non-uniqueness of the discrete state y_h , we restricted the discussion of (A1)-(A7) to the set $\tilde{\mathcal{K}} = \mathcal{K} \cap \bar{B}(\bar{u}, \rho_{\bar{u}})$. This is compatible with our needs, since we perform a local analysis around \bar{u} . The set $\tilde{\mathcal{K}}$ is also a convex and closed set of $L^\infty(\Omega)$. Therefore, our theory is applicable to this set instead of \mathcal{K} . Let us confirm first (2.17). We obtain

$$\begin{aligned} |(J'_h(u) - J'(u))(u_h - \bar{u})| &\leq \int_{\Omega_h} \left| \varphi_h(u) - \varphi_u + \frac{\partial L}{\partial u}(x, y_h(u), u) \right. \\ &\quad \left. - \frac{\partial L}{\partial u}(x, y_u, u) \right| |u_h - \bar{u}| dx + \int_{\Omega \setminus \Omega_h} \left| \varphi_u + \frac{\partial L}{\partial u}(x, y_u, u) \right| |u_h - \bar{u}| dx \end{aligned}$$

and, by Lemma 3.8,

$$\|\varphi_h(u) - \varphi_u\|_{L^2(\Omega_h)} + \|y_h(u) - y_u\|_{L^2(\Omega_h)} \leq C h^2.$$

Therefore, the integral on Ω_h can be estimated by $ch^2\|u_h - \bar{u}\|_{L^2(\Omega)}$. The integral on $\Omega \setminus \Omega_h$ cannot be handled this way, but it vanishes. Notice that $u_h(x) = \bar{u}(x)$ was assumed a.e. on $\Omega \setminus \Omega_h$ for the types of control discretization we consider. Therefore, (2.17) is satisfied with

$$\varepsilon(h) = C h^2. \quad (4.21)$$

The confirmation of (2.18) is more delicate. It was shown in [11] that, under our assumptions on (P), the convergence $\bar{u}_h \rightarrow \bar{u}$ in $L^2(\Omega)$ implies $\bar{u}_h \rightarrow \bar{u}$ in $L^\infty(\Omega)$, provided that the controls are not discretized (variational discretization) or taken as piecewise constant functions. This follows from the available projection formulas for \bar{u}_h . The situation is more difficult for piecewise linear control approximation. Then (2.18) needs the discussion of $J''(u_k)v_k^2$ for weakly converging sequences $\{v_k\}$. In (3.18) there is one delicate term, namely $\int_{\Omega} \frac{\partial^2 L}{\partial u^2}(x, y_{u_k}, u_k)v_k^2 dx$. Here, we cannot really control the convergence of v_k^2 . It was shown in [7, pp. 149-150] by Egorov's theorem that condition (2.18) is satisfied for piecewise linear control approximation. In [7], the equation was semilinear but this method does not depend on the particular type of the equation.

Acknowledgement. The authors are grateful to one of the referees for his detailed comments, in particular for pointing out a technical problem in the first version of the paper.

References

- [1] W. Alt. On the approximation of infinite optimization problems with an application to optimal control problems. *Appl. Math. Optim.*, 12:15–27, 1984.

- [2] N. Arada, E. Casas, and F. Tröltzsch. Error estimates for the numerical approximation of a semilinear elliptic control problem. *Comput. Optim. Appl.*, 23:201–229, 2002.
- [3] F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York, 2000.
- [4] J.F. Bonnans and H. Zidani. Optimal control problems with partially polyhedral constraints. *SIAM J. Control Optim.*, 37(6):1726–1741, 1999.
- [5] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*. Springer-Verlag, New York, 1994.
- [6] E. Casas. Error Estimates for the Numerical Approximation of Semilinear Elliptic Control Problems with Finitely Many State Constraints. *ESAIM Control Optim. Calc. Var.*, 8:345–374, 2002
- [7] E. Casas. Using piecewise linear functions in the approximation of semilinear elliptic control problems. *Adv. Comput. Math.*, 26:137–153, 2007.
- [8] E. Casas, M. Mateos, and F. Tröltzsch. Error estimates for the numerical approximation of boundary semilinear elliptic control problems. *Comput. Optim. Appl.*, 31:193–220, 2005.
- [9] E. Casas and J.P. Raymond. Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations. *SIAM J. Control Optim.*, 45:1586–1611, 2006.
- [10] E. Casas and F. Tröltzsch. First- and second-order optimality conditions for a class of optimal control problems with quasilinear elliptic equations. *SIAM J. Control Optim.*, 48:688–718, 2009.
- [11] E. Casas and F. Tröltzsch. Numerical analysis of some optimal control problems governed by a class of quasilinear elliptic equations. *ESAIM Control Optim. Calc. Var.*, DOI: 10.1051/cocv/2009010.
- [12] E. Casas and F. Tröltzsch. Error estimates for the finite-element approximation of a semilinear elliptic control problem. *Control Cybernet.*, 31:695–712, 2002.
- [13] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.
- [14] K. Deckelnick, A. Günther and M. Hinze. Finite element approximation of Dirichlet boundary control for elliptic PDEs on two- and three-dimensional curved domains. *SIAM J. Control Optim.*, 48(4):2798–2819, 2009.
- [15] K. Deckelnick and M. Hinze. Convergence of a finite element approximation to a state constrained elliptic control problem. *SIAM J. Numer. Anal.*, 45:1937–1953, 2007.

- [16] K. Deckelnick and M. Hinze. Numerical analysis of a control and state constrained elliptic control problem with piecewise constant control approximations. *Proceedings of the ENUMATH*, 2007.
- [17] F. Falk. Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.*, 44:28–47, 1973.
- [18] T. Geveci. On the approximation of the solution of an optimal control problem governed by an elliptic equation. *RAIRO Anal. Numér.*, 13:313–328, 1979.
- [19] P. GRISVARD, *Elliptic Problems in Nonsmooth Domains*, Pitman, Boston, 1985.
- [20] M. Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *Comput. Optim. Appl.*, 30:45–63, 2005.
- [21] M. Hinze and F. Tröltzsch. Discrete concepts versus error analysis in pde constrained optimization. *GAMM-Mitteilungen*, March 2011.
- [22] C. Meyer and A. Rösch. Superconvergence properties of optimal control problems. *SIAM J. Control Optim.*, 43:970–985, 2004.
- [23] P.A. Raviart and J.M. Thomas. *Introduction à L'analyse Numérique des Equations aux Dérivées Partielles*. Masson, Paris, 1983.
- [24] F. Tröltzsch. *Optimal Control of Partial Differential Equations. Theory, Methods and Applications*, volume 112. American Math. Society, Providence, 2010.