ESAIM: M2AN 59 (2025) 3159–3204
https://doi.org/10.1051/m2an/2025086

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

NUMERICAL ANALYSIS OF A NONSMOOTH QUASILINEAR ELLIPTIC CONTROL PROBLEM

II. FINITE ELEMENT DISCRETIZATION AND ERROR ESTIMATES

Christian Clason[1,2], Huu Nhu Vu[3,*] and Arnd Rösch[1]

Abstract. In this paper, we carry out the numerical analysis of a nonsmooth quasilinear elliptic optimal control problem, where the coefficient in the divergence term of the corresponding state equation is not differentiable with respect to the state variable. Despite the lack of differentiability of the nonlinearity in the quasilinear elliptic equation, the corresponding control-to-state operator is of class C^1 but not of class C^2. Analogously, the discrete control-to-state operators associated with the approximated control problems are proven to be of class C^1 only. By using an explicit second-order sufficient optimality condition, we prove *a priori* error estimates for a variational approximation, a piecewise constant approximation, and a continuous piecewise linear approximation of the continuous optimal control problem. The numerical tests confirm these error estimates.

Mathematics Subject Classification. 35J62, 49J20, 49J52, 65N15, 65N30.

Received March 4, 2023. Accepted October 1, 2025.

1. Introduction

We investigate the nonsmooth quasilinear elliptic optimal control problem

$$\begin{cases} \min\limits_{u\in L^\infty(\Omega)} j(u) := \displaystyle\int_\Omega L(x, y_u(x))\,\mathrm{d}x + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)} & \\ \quad \text{s.t. } -\operatorname{div}[(b + a(y_u))\nabla y_u] = u & \text{in } \Omega, \quad y_u = 0 \text{ on } \partial\Omega, \\ \quad \alpha \le u(x) \le \beta & \text{a.e. } x \in \Omega. \end{cases} \tag{P}$$

Here Ω is a two-dimensional bounded, convex and polygonal domain; $L : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function that is of class C^2 with respect to (w.r.t.) the second variable; $b : \overline{\Omega} \to \mathbb{R}$ is a Lipschitz continuous function; $a : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous but not differentiable; and constants $\alpha, \beta, \nu \in \mathbb{R}$ satisfying $\beta > \alpha$ and $\nu > 0$. We refer to Section 2 for the precise assumptions on the data of (P).

Keywords and phrases. Optimal control, nonsmooth optimization, quasilinear elliptic equation, piecewise differentiable function, sufficient optimality condition, error estimate, finite element approximation.

[1] Faculty of Mathematics, University of Duisburg-Essen, Thea-Leymann-Strasse 9, 45127 Essen, Germany.

[2] Department of Mathematics and Scientific Computing, NAWI Graz, University of Graz, 8010 Graz, Austria.

[3] Faculty of Fundamental Sciences, PHENIKAA University, Duong Noi, Hanoi 12116, Vietnam.

*Corresponding author: nhu.vuhuu@phenikaa-uni.edu.vn

© The authors. Published by EDP Sciences, SMAI 2025

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The state equation in the optimal control problem (P) occurs, for instance, in models of heat conduction in which the coefficient in the divergence term acts as the heat conductivity and is a function of two variables: the temperature y variable and the spatial coordinate x variable; see, *e.g.*, [5, 46]. When the data belong to class C^2, the numerical analysis of the discrete approximation of optimal control problems governed by such state equations were studied by Casas *et al.* in [13, 14] for distributed control and in [10] for Neumann control.

Let us briefly comment on other works concerning the error analysis of optimal control problems governed by partial differential equations (PDEs), in particular by elliptic PDEs. For control-constrained elliptic problems, we refer to the early papers [31, 32] for linear elliptic control problems; to [2, 15] for semilinear elliptic problems. For state-constrained control problems, we mention only the recent contributions [28, 43] and refer to the survey paper [36] for further references. Although the error analysis for smooth PDE-constrained problems has been intensively investigated, there are very few contributions on this topic for nonsmooth PDE-constrained optimal control. Here we want to mention the work [37] concerning error analysis for optimal control of a coupled PDE–ODE system, where the nonsmooth nonlinearity acts on a semilinear ODE. For works related to optimal control of obstacle problems, we refer to [19, 42] and the references therein. Based on a quadratic growth condition, *a priori* error estimates were established in [37, 42]. To the authors' best knowledge, this is the first work that exploits a second-order sufficient optimality condition to show *a priori* error estimates for the discretization of optimal control problems governed by nonsmooth PDEs.

In this paper, our main aim is to derive the convergence analysis and error estimates of the discretization of (P) under the explicit second-order sufficient optimality conditions established in [24] (first derived in general form in [22]). As the coefficient a in the state equation is Lipschitz continuous but not differentiable, there are two major difficulties in deriving the error analysis. The first issue arises in studying error estimates of the discretization of the adjoint state equation; the other is the lack of the second-order differentiability of the cost functional. We therefore cannot apply an abstract theorem on error estimates shown in [14] and in [17]. To deal with the first issue, we introduce the function $Z_{y,\hat{y}}$ defined in (27) that measures pointwise the difference between the gradients of the superposition mappings of a associated with two distinct states y and $\hat{y}$. This allows us to derive thereafter both L^2- and an H_0^1-error estimates for the approximation of the adjoint state equation. Let us emphasize that the solutions to the linearized state equation can only be shown to be in a $W^{1,p}$-space (see Rem. 2.4) and thus we cannot directly employ the standard duality argument based on the Aubin–Nitsche trick; see, *e.g.*, [7, 11, 26]. In order to derive L^2-error estimates, we will instead introduce an *adjusted* linearized state equation in (55) below. Under the assumption that the jump functional Σ defined in (14) to be finite at the optimal state $\bar{y}$, we show that the solutions to this adjusted equation belong to a fractional Sobolev space of order greater than 1. We can then apply standard interpolation error estimates to obtain the desired result. For handling the second issue, we will exploit an structural assumption on the optimal state and employ an explicit formula of a second-order generalized derivative of the objective functional, formulated in [24]. Based on the second-order sufficient optimality conditions for (P) from [22, 24], we then prove general error estimates for variational, piecewise constant, and continuous piecewise linear approximations of the optimal control, which generalize those of Theorem 2.14 in [14] and of of Lemma 5.2 in [17], and can be applied for the case where the cost functional j is of class C^1 but not necessarily C^2; see Theorem 5.10.

The plan of the paper is as follows. This section ends with our notation. In the next section, we make the assumptions for (P) and provide some preliminary results from [22, 24]. Section 3 is devoted to the numerical approximation of the state equation by finite elements and the local well-posedness and differentiability of the discrete counterpart of the control-to-state operator. In Section 4, the error analysis of the adjoint state equation is investigated. Finally, the main results of the paper are presented in Section 5. There, the convergence and error estimates of local minima of discrete optimal control problems are, respectively, stated in Section 5.1 and Section 5.2. The numerical tests illustrating the obtained results are given in Section 5.3. Finally, the verification of a structural assumption and a computation of the jump functional for a specific situation are carried out in Appendix A, while the regularity of solutions in a fractional Sobolev space to the adjusted linearized state equation is shown in Appendix B.

Notation. We denote by $B_X(u,\rho)$ and $\overline{B}_X(u,\rho)$ the open and closed balls in a Banach space X of radius $\rho>0$ centered at $u\in X$, respectively. For Banach spaces X and Y, the notation $X\hookrightarrow(\Subset)Y$ is understood that X is continuously (compactly) embedded in Y. Let X be a Banach space with its dual X^*, the symbol $\langle\cdot,\cdot\rangle_{X^*,X}$ stands for the dual product of X and X^*. For a given function $g:\overline{\Omega}\to\mathbb{R}$ and a subset $A\subset\mathbb{R}$, $\{g\in A\}$ denotes the set of all points $x\in\overline{\Omega}$ for which $g(x)\in A$. For functions g_1,g_2 and subsets $A_1,A_2\subset\mathbb{R}$, we set $\{g_1\in A_1,g_2\in A_2\}:=\{g_1\in A_1\}\cap\{g_2\in A_2\}$. For any set $\omega\subset\overline{\Omega}$, we denote by $\mathbb{1}_\omega$ the characteristic function of ω, *i.e.*, $\mathbb{1}_\omega(x)=1$ if $x\in\omega$ and $\mathbb{1}_\omega(x)=0$ otherwise. For real numbers α,β with $\alpha\le\beta$, the symbol $\mathrm{Proj}_{[\alpha,\beta]}(t)$ denotes the projection of t onto the interval $[\alpha,\beta]$ if $t\in\mathbb{R}$, and the pointwise a.e. projection of t onto the interval $[\alpha,\beta]$ if t is a measurable function. We write the symbol C for a generic positive constant, which may be different at different places of occurrence and the notation, *e.g.*, C_ξ for a constant depending only on the parameter ξ. For a measurable two-dimensional subset M, by $\mathrm{meas}_{\mathbb{R}^2}(M)$, we denote the two-dimensional Lebesgue measure of M. Finally, the symbol $\mathcal{H}^1$ denotes the one-dimensional Hausdorff measure on $\mathbb{R}^2$ that is scaled as in Definition 2.1 of [30].

2. Main assumptions and preliminary results

In this section we first present assumptions, which will be used in the whole paper, and then state some preliminary results on the state equation, the adjoint state equation, as well as the first-order and explicit second-order sufficient optimality conditions from [22, 24].

We first address the salient point, which is the structure of the nondifferentiable nonlinearity a. In this work, we assume that a is defined by

$$a(t):=\mathbb{1}_{(-\infty,\bar t]}(t)a_0(t)+\mathbb{1}_{(\bar t,\infty)}a_1(t)\quad\text{for all } t\in\mathbb{R},\tag{1}$$

for a given number $\bar t\in\mathbb{R}$ and given functions $a_0\in C^2((-\infty,\bar t])$ and $a_1\in C^2([\bar t,\infty))$ with $a_0(\bar t)=a_1(\bar t)$. Obviously, a is locally Lipschitz continuous on $\mathbb{R}$ and twice continuously differentiable on $(-\infty,\bar t)\cup(\bar t,\infty)$, but not even of class C^1 in general. However, a is directionally differentiable and its directional derivative is determined as follows

$$a'(t;s)=\mathbb{1}_{(-\infty,\bar t)}(t)a_0'(t)s+\mathbb{1}_{(\bar t,\infty)}(t)a_1'(t)s+\mathbb{1}_{\{\bar t\}}(t)[\mathbb{1}_{(0,\infty)}(s)a_1'(\bar t)s+\mathbb{1}_{(-\infty,0)}(s)a_0'(\bar t)s]\quad\text{for } t,s\in\mathbb{R}.\tag{2}$$

Remark 2.1. Let us emphasize that the results and the underlying analysis in this paper can be applied to the situation in which the function a is continuous and is twice continuously differentiable on finitely many intervals (*i.e.*, a finitely PC^2 function; see [22] for a precise definition). However, in order to keep the presentation concise and to be able to focus on the main arguments, we restrict the presentation to the simplest such situation given by (1).

The following assumptions shall hold throughout the following.

(A1) $\Omega\subset\mathbb{R}^2$ is an open bounded convex polygonal.
(A2) The Lipschitz continuous function $b:\overline{\Omega}\to\mathbb{R}$ satisfies $b(x)\ge\underline{b}>0$ for all $x\in\overline{\Omega}$.
(A3) $a:\mathbb{R}\to\mathbb{R}$ is nonnegative and given by (1).
(A4) $L:\Omega\times\mathbb{R}\to[0,+\infty)$ is a Carathéodory function that is of class C^2 w.r.t. the second variable with $L(\cdot,0)\in L^1(\Omega)$. Besides, for any $M>0$, there exist $C_M>0$ and $\psi_M\in L^{\bar p}(\Omega)$ ($\bar p>2$) such that $|\frac{\partial L}{\partial y}(x,y)|\le\psi_M(x)$ and $|\frac{\partial^2 L}{\partial y^2}(x,y)|\le C_M$ for all $y\in\mathbb{R}$ with $|y|\le M$, and a.e. $x\in\Omega$.

In the remainder of this subsection, we state some known results for the state equation, the adjoint state equation, and the optimality conditions for (P); see, *e.g.*, [22, 24]. Let us first study the state equation

$$-\operatorname{div}[(b+a(y))\nabla y]=u\ \text{in }\Omega,\quad y=0\ \text{ on }\partial\Omega.\tag{3}$$

Theorem 2.2 ([24], Thm. 3.3 and cf. [22], Thms. 3.1 and 3.5)**.** *Assume that Assumptions (A1) to (A3) are verified. Then, the control-to-state mapping $S : W^{-1,p}(\Omega) \ni u \mapsto y_u \in W_0^{1,p}(\Omega)$ is of class C^1, where y_u is the unique solution to (3). Moreover, for any $u, v \in W^{-1,p}(\Omega)$ with $p > 2$ and $y_u := S(u)$, $z_v := S'(u)v$ uniquely satisfies*

$$-\operatorname{div}\big[(b + a(y_u))\nabla z_v + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u) z_v \nabla y_u\big] = v \quad \text{in } \Omega, \quad z_v = 0 \quad \text{on } \partial\Omega. \tag{4}$$

Consequently, for any bounded set $V \subset L^2(\Omega)$, there exists a constant $C_V > 0$ such that

$$\|S(u_1) - S(u_2)\|_{H_0^1(\Omega)} + \|S(u_1) - S(u_2)\|_{C(\overline{\Omega})} \leq C_V \|u_1 - u_2\|_{L^2(\Omega)} \quad \text{for all } u_1, u_2 \in V. \tag{5}$$

Furthermore, a number $p_* > 2$ exists and satisfies that for any $p \in [2, p_*)$ and for any bounded set $U \subset L^p(\Omega)$, there hold $S(u) \in W_0^{1,p}(\Omega) \cap W^{2,p}(\Omega)$ and $\|S(u)\|_{W^{2,p}(\Omega)} \leq C_U$ for all $u \in U$.

Proof. The continuous differentiability of S and the derivation of (4) as well as the $W^{2,p}(\Omega)$-regularity of solutions to (3) follow from Theorem 3.3 in [24]. It remains to prove (5). To this end, we first deduce from the compact embedding $L^2(\Omega) \Subset W^{-1,p_0}(\Omega)$ for some $p_0 \in (2, \infty)$ and the continuous differentiability of S that

$$C'_V := \sup\Big\{ \|S'(u)\|_{\mathbb{L}(W^{-1,p_0}(\Omega), W_0^{1,p_0}(\Omega))} \mid u \in V \Big\} < \infty.$$

This, along with the mean value theorem, yields

$$\|S(u_1) - S(u_2)\|_{W_0^{1,p_0}(\Omega)} \leq C'_V \|u_1 - u_2\|_{W^{-1,p_0}(\Omega)} \quad \text{for all } u_1, u_2 \in V.$$

The continuous embeddings $L^2(\Omega) \hookrightarrow W^{-1,p_0}(\Omega)$ and $W_0^{1,p_0}(\Omega) \hookrightarrow H_0^1(\Omega) \cap C(\overline{\Omega})$ then imply (5). □

We now investigate the adjoint state equation

$$-\operatorname{div}[(b + a(y_u))\nabla\varphi] + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u) \nabla y_u \cdot \nabla\varphi = v \quad \text{in } \Omega, \quad \varphi = 0 \quad \text{on } \partial\Omega \tag{6}$$

for $u \in W^{-1,p}(\Omega)$, $p > 2$, $v \in H^{-1}(\Omega)$, and $y_u := S(u)$.

Theorem 2.3 ([24], Thm. 3.4 and cf. [22], Lem. 4.1)**.** *Assume that Assumptions (A1) to (A3) are satisfied. Let $p, q > 2$ be arbitrary. Then, for any $u \in W^{-1,p}(\Omega), v \in H^{-1}(\Omega)$, there exists a unique $\varphi \in H_0^1(\Omega)$ which solves (6). Moreover, if U is a bounded subset in $L^p(\Omega)$, then for any $u \in U$ and any $v \in L^q(\Omega)$, the solution φ to (6) is an element of $H^2(\Omega) \cap W^{1,\infty}(\Omega)$ and there holds $\|\varphi\|_{H^2(\Omega)} + \|\varphi\|_{W^{1,\infty}(\Omega)} \leq C_U \|v\|_{L^q(\Omega)}$. Furthermore, if $u \in L^p(\Omega)$ and $v \in L^r(\Omega)$ with $r \in (2, p_*)$, then $\varphi \in W^{2,r}(\Omega)$, where p_* is given as in Theorem 2.2.*

Remark 2.4 ([24], Rem. 3.5)**.** In spite of the $W^{2,p}$-regularity of the state and adjoint state, the function $z_v := S'(u)v$ determined in (4) belongs to $W^{1,p}(\Omega)$ only. This fact is due to the nondifferentiability of the function a.

The optimal control problem (P) can be transferred in the following form

$$\min_{u \in \mathcal{U}_{ad}} j(u) = \int_\Omega L(x, S(u)(x))\,\mathrm{d}x + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)}, \tag{P}$$

where the admissible set is defined as

$$\mathcal{U}_{ad} := \{u \in L^\infty(\Omega) \mid \alpha \leq u(x) \leq \beta \qquad \text{for a.e. } x \in \Omega\}.$$

Thanks to Assumptions (A1) to (A4), the cost functional $j : L^2(\Omega) \to \mathbb{R}$ is first-order continuously differentiable and satisfies

$$j'(u)v = \int_\Omega (\varphi_u + \nu u) v \,\mathrm{d}x \quad \text{for } u, v \in L^2(\Omega) \tag{7}$$

with $\varphi_u \in H_0^1(\Omega)$ solving (6) corresponding to the right-hand side term v substituted by $\frac{\partial L}{\partial y}(\cdot, S(u))$; see Theorem 4.2 of [22]. We have the following first-order necessary optimality conditions from Theorem 4.3 in [22].

Theorem 2.5 ([22], Thm. 4.3). *Let Assumptions (A1) to (A4) hold. Then (P) admits at least one local minimizer $\bar{u}$. Furthermore, an adjoint state $\bar{\varphi} \in H_0^1(\Omega)$ exists and fulfills the following first-order optimality conditions*

$$-\operatorname{div}[(b + a(\bar{y}))\nabla \bar{y}] = \bar{u} \qquad \text{in } \Omega, \quad \bar{y} = 0 \ \text{ on } \partial\Omega, \tag{8a}$$

$$-\operatorname{div}[(b + a(\bar{y}))\nabla \bar{\varphi}] + \mathbb{1}_{\{\bar{y} \neq \bar{t}\}} a'(\bar{y}) \nabla \bar{y} \cdot \nabla \bar{\varphi} = \frac{\partial L}{\partial y}(x, \bar{y}) \qquad \text{in } \Omega, \quad \bar{\varphi} = 0 \ \text{ on } \partial\Omega, \tag{8b}$$

$$\int_\Omega (\bar{\varphi} + \nu \bar{u})(u - \bar{u}) \, \mathrm{d}x \geq 0 \qquad \text{for all } u \in \mathcal{U}_{ad}, \tag{8c}$$

with $\bar{y} := S(\bar{u})$. Moreover, $\bar{y} \in W^{2,p}(\Omega)$ and $\bar{\varphi} \in W^{2,r}(\Omega)$ for any $p, r \in (2, p_*)$ and $r \leq \bar{p}$ with $\bar{p}$ and p_*, respectively, defined in Assumption (A4) and Theorem 2.2. Consequently, $\bar{y}, \bar{\varphi} \in C^1(\overline{\Omega})$ and $\bar{u} \in C^{0,1}(\overline{\Omega})$.

Assume now that $\bar{\varphi} \in H_0^1(\Omega)$ fulfills (8). The *critical cone* of the problem (P) at $\bar{u}$ is defined by

$$\mathcal{C}(\mathcal{U}_{ad}; \bar{u}) := \big\{ v \in L^2(\Omega) \mid v \geq 0 \ \text{ if } \bar{u} = \alpha, v \leq 0 \ \text{ if } \bar{u} = \beta, v = 0 \ \text{ if } \bar{\varphi} + \nu \bar{u} \neq 0 \text{ a.e. in } \Omega \big\}. \tag{9}$$

In the rest of this subsection, we shall provide second-order sufficient optimality conditions for (P). On that account, we need introduce the curvature functional of j, which can be separated into three contributions. For any $(u, y, \varphi) \in L^2(\Omega) \times H^1(\Omega) \times W^{1,\infty}(\Omega)$, the smooth part and the first-order nonsmooth part of the curvature in direction $(v_1, v_2) \in L^2(\Omega)^2$ are defined by

$$\begin{aligned} Q_s(u, y, \varphi; v_1, v_2) &:= \frac{1}{2} \int_\Omega \frac{\partial^2 L}{\partial y^2}(\cdot, y) z_{v_1} z_{v_2} \, \mathrm{d}x + \frac{\nu}{2} \int_\Omega v_1 v_2 \, \mathrm{d}x - \frac{1}{2} \int_\Omega \mathbb{1}_{\{y \neq \bar{t}\}} a''(y) z_{v_1} z_{v_2} \nabla y \cdot \nabla \varphi \, \mathrm{d}x, \\ Q_1(u, y, \varphi; v_1, v_2) &:= -\frac{1}{2} \int_\Omega [a'(y; z_{v_1}) \nabla z_{v_2} + a'(y; z_{v_2}) \nabla z_{v_1}] \cdot \nabla \varphi \, \mathrm{d}x \end{aligned}$$

with $z_{v_i} := S'(u) v_i$, $i = 1, 2$. The critical part of the curvature is of course the second-order nonsmooth part involving some additional notation. Let $\delta > 0$ be arbitrary but fixed and set

$$\begin{cases} \Omega^2_{y,\hat{y}} := \{\hat{y} \in (\bar{t}, \bar{t} + \delta), y \in (\bar{t} - \delta, \bar{t}]\}, \\ \Omega^3_{y,\hat{y}} := \{\hat{y} \in (\bar{t} - \delta, \bar{t}), y \in [\bar{t}, \bar{t} + \delta)\} \end{cases} \tag{10}$$

for given functions $y, \hat{y} \in C(\overline{\Omega})$. For any $s \in \mathbb{R}$, $u, v \in L^2(\Omega)$, $y \in C(\overline{\Omega}) \cap H^1(\Omega)$, and $\varphi \in W^{1,\infty}(\Omega)$, we define

$$\begin{cases} \zeta_0(u, y; s, v) := -\{a'\}_{\bar{t}+0}^{\bar{t}-0} (\bar{t} - S(u + sv)) \mathbb{1}_{\Omega^3_{S(u+sv),y}}, \quad \zeta_1(u, y; s, v) := \{a'\}_{\bar{t}+0}^{\bar{t}-0} (\bar{t} - S(u + sv)) \mathbb{1}_{\Omega^2_{S(u+sv),y}}, \\ \zeta(u, y; s, v) := \zeta_0(u, y; s, v) + \zeta_1(u, y; s, v) = \{a'\}_{\bar{t}+0}^{\bar{t}-0} (\bar{t} - S(u + sv)) \Big[\mathbb{1}_{\Omega^2_{S(u+sv),y}} - \mathbb{1}_{\Omega^3_{S(u+sv),y}} \Big]. \end{cases} \tag{11}$$

Here $\{a'\}_{\bar{t}+0}^{\bar{t}-0}$ denotes the difference between the one-sided derivatives of a at $\bar{t}$ from left and right, *i.e.*,

$$\{a'\}_{\bar{t}+0}^{\bar{t}-0} := \lim_{t \to \bar{t}^-} a'(t) - \lim_{t \to \bar{t}^+} a'(t) = a_0'(\bar{t}) - a_1'(\bar{t}).$$

We then determine for any $\{s_n\} \in c_0^+ := \{\{s_n\} \subset (0, \infty) \mid s_n \to 0\}$ and $v \in L^2(\Omega)$ the term

$$\begin{aligned} \tilde{Q}(u, y, \varphi; \{s_n\}, v) &:= \liminf_{n \to \infty} \frac{1}{s_n^2} \int_\Omega \sum_{i=0}^{1} \zeta_i(u, y; s_n, v) \nabla y \cdot \nabla \varphi \, \mathrm{d}x \\ &= \{a'\}_{\bar{t}+0}^{\bar{t}-0} \liminf_{n \to \infty} \frac{1}{s_n^2} \int_\Omega (\bar{t} - S(u + s_n v)) \Big[\mathbb{1}_{\Omega^2_{S(u+s_n v),y}} - \mathbb{1}_{\Omega^3_{S(u+s_n v),y}} \Big] \nabla y \cdot \nabla \varphi \, \mathrm{d}x. \end{aligned} \tag{12}$$

The second-order nonsmooth part of the curvature in direction $v \in L^2(\Omega)$ is thus defined as

$$Q_2(u,y,\varphi;v) := \inf\Big\{\tilde{Q}(u,y,\varphi;\{s_n\},v) \mid \{s_n\} \in c_0^+\Big\}.$$

We finally identify the total curvature in direction v as

$$Q(u,y,\varphi;v) := Q_s(u,y,\varphi;v,v) + Q_1(u,y,\varphi;v,v) + Q_2(u,y,\varphi;v). \tag{13}$$

Remark 2.6. The definitions of the sets $\Omega^2_{y,\hat{y}}$ and $\Omega^3_{y,\hat{y}}$ in (10) are derived from the ones for $\Omega^{1,2}_{y,\hat{y}}$ and $\Omega^{0,3}_{y,\hat{y}}$, respectively, in Lemma 3.3 in [22] for the situation in which $K := 1$, $t_0 := -\infty$, $t_1 := \bar{t}$, and $t_2 := \infty$. Analogously, the definitions of Q_s, Q_1, and Q_2 in this section can be obtained from those in Section 5.1 in [22].

Thanks to Proposition 5.6 and Lemma 5.7 in [22], we have the weak lower semicontinuity of Q_2 in the last variable and there holds

$$|Q_2(u,S(u),\varphi;v)| \le \Sigma(S(u))\|\nabla\varphi\|_{L^\infty(\Omega)}\|S'(u)v\|^2_{L^\infty(\Omega)} \quad \text{for all } u,v \in L^2(\Omega) \text{ and } \varphi \in W^{1,\infty}(\Omega).$$

Here $\Sigma : C(\overline{\Omega}) \cap W^{1,1}(\Omega) \to \mathbb{R} \cup \{\infty\}$ is the *jump functional* and defined by

$$\Sigma(y) := \sigma_0 \limsup_{r\to 0^+} \frac{1}{r} \sum_{m=1}^{2} \int_\Omega \left[\mathbb{1}_{\{0<|y-\bar{t}|\le r\}} |\partial_{x_m} y|\right] \mathrm{d}x, \quad y \in W^{1,1}(\Omega) \cap C(\overline{\Omega}) \tag{14}$$

with

$$\sigma_0 := \left|\{a'\}_{\bar{t}+0}^{\bar{t}-0}\right| = |a_0'(\bar{t}) - a_1'(\bar{t})|. \tag{15}$$

Moreover, if follows from Corollary 5.5 in [22] that, for any $u \in L^2(\Omega)$, $\{s_n\} \in c_0^+$ and $v_n \rightharpoonup v$ in $L^2(\Omega)$, there holds

$$\liminf_{n\to\infty} \frac{1}{s_n^2} \int_\Omega \sum_{i=0}^{1} \zeta_i(u,S(u);s_n,v_n)\nabla S(u)\cdot\nabla\varphi \,\mathrm{d}x = \tilde{Q}(u,S(u),\varphi;\{s_n\},v) \ge Q_2(u,S(u),\varphi;v), \tag{16}$$

provided that $\Sigma(S(u)) < \infty$.

Theorem 2.7 (Explicit second-order sufficient optimality conditions, [24], Thm. 3.24)**.** *Assume that Assumptions (A1) to (A4) are fulfilled. Let $\bar{u}$ be an admissible control of (P) such that $\{\bar{y} = \bar{t}\}$ decomposes into finitely many connected components and that on each such connected component $\mathcal{C}$, either*

$$\nabla\bar{y}(x) \ne 0 \quad \text{for all} \quad x \in \mathcal{C} \tag{17}$$

or

$$\begin{cases} \nabla\bar{y} = 0 \ \text{ on } \mathcal{C}, \\ \mathrm{meas}_{\mathbb{R}^2}(\{0<|\bar{y}-\bar{t}|<r\} \cap \mathcal{C}^{\varepsilon_0}) \le c_s r, \quad \text{for all } r \in (0,r_0), \quad \text{for some constants } \varepsilon_0, r_0 > 0 \end{cases} \tag{18}$$

holds with $\bar{y} := S(\bar{u})$ and

$$\mathcal{C}^{\varepsilon_0} := \big\{x \in \overline{\Omega} \mid \mathrm{dist}(x,\mathcal{C}) < \varepsilon_0\big\}.$$

Assume further that there exists a $\bar{\varphi} \in W_0^{1,\bar{p}}(\Omega) \cap W^{1,\infty}(\Omega)$, with $\bar{p}$ defined in Assumption (A4), that together with $\bar{u},\bar{y}$ fulfills (8) and

$$\begin{aligned} Q(\bar{u},\bar{y},\bar{\varphi};v) = {} & \frac{1}{2}\int_\Omega \frac{\partial^2 L}{\partial y^2}(\cdot,\bar{y}) z_v^2 \,\mathrm{d}x + \frac{\nu}{2}\int_\Omega v^2 \,\mathrm{d}x - \frac{1}{2}\int_\Omega \mathbb{1}_{\{\bar{y}\ne\bar{t}\}} a''(\bar{y}) z_v^2 \nabla\bar{y}\cdot\nabla\bar{\varphi}\,\mathrm{d}x - \int_\Omega a'(\bar{y};z_v)\nabla z_v\cdot\nabla\bar{\varphi}\,\mathrm{d}x \\ & + \frac{1}{2}[a_0'(\bar{t}) - a_1'(\bar{t})] \int_{\{\bar{y}=\bar{t}\}} \mathbb{1}_{\{|\nabla\bar{y}|>0\}} z_v^2 \frac{\nabla\bar{y}\cdot\nabla\bar{\varphi}}{|\nabla\bar{y}|}\,\mathrm{d}\mathcal{H}^1(x) > 0 \quad \text{for all } v \in \mathcal{C}(\mathcal{U}_{ad};\bar{u}) \setminus \{0\} \end{aligned} \tag{19}$$

with $z_v := S'(\bar{u})v$. Then constants $c_0, \rho_0 > 0$ exist and fulfill

$$j(\bar{u}) + c_0\|u-\bar{u}\|^2_{L^2(\Omega)} \le j(u) \qquad \text{for all } u \in \mathcal{U}_{ad} \cap \overline{B}_{L^2(\Omega)}(\bar{u},\rho_0).$$

The following will be used later to show the error estimates for the approximation of (P).

Proposition 2.8 ([24], Thm. 3.20). *Let $\bar u \in L^2(\Omega)$ be arbitrary and let $\bar y := S(\bar u)$. Assume that $\{\bar y = \bar t\}$ decomposes into finitely many connected components and that on each such connected component $\mathcal{C}$, either (17) or (18) is fulfilled. Let $v \in L^2(\Omega)$ and $\bar\varphi \in C^1(\overline\Omega) \cap W^{2,1}(\Omega)$. Then, for any $\{s_n\} \in c_0^+$ and $\{v_n\} \subset L^2(\Omega)$ such that $v_n \rightharpoonup v$ in $L^2(\Omega)$, $y_n := S(\bar u + s_n v_n) \to \bar y$ in $C^1(\overline\Omega)$, and $(y_n - \bar y)/s_n \to w$ in $W_0^{1,p}(\Omega)$ for some $p > 2$, there hold*

$$\tilde Q(\bar u, \bar y, \bar\varphi; \{s_n\}, v) = \frac{1}{2}[a_0'(\bar t) - a_1'(\bar t)] \int_{\{\bar y = \bar t\}} \mathbb{1}_{\{|\nabla \bar y| > 0\}} w^2 \frac{\nabla \bar y \cdot \nabla \bar\varphi}{|\nabla \bar y|} \,\mathrm{d}\mathcal{H}^1(x) \tag{20}$$

and

$$\lim_{n\to\infty} \frac{1}{s_n^2} \int_\Omega (2\bar t - \bar y - y_n) \Big[\mathbb{1}_{\Omega^2_{y_n,\bar y}} - \mathbb{1}_{\Omega^3_{y_n,\bar y}}\Big] \nabla \bar y \cdot \nabla \bar\varphi \,\mathrm{d}x = 0. \tag{21}$$

Finally, the wellposedness of Σ, defined in (14), is guaranteed by either (17) or (18) as stated below.

Proposition 2.9 ([24], Prop. 3.21). *Let $\bar y \in C^1(\overline\Omega)$ be such that $\{\bar y = \bar t\}$ decomposes into finitely many connected components and that on each such connected component $\mathcal{C}$, either (17) or (18) is fulfilled. Then $\Sigma(\bar y) < \infty$.*

3. Analysis of the discrete state equation

In this section, we study the discrete version of the state equation (3) and show error estimates of solutions to the discrete state equation (22), local uniqueness of these solutions, and local differentiability of the solution operators of (22). To this end, we introduce a family of regular triangulations $\{\mathcal{T}_h\}_{h>0}$: $\overline\Omega = \bigcup_{T\in\mathcal{T}_h} T$ for all $h > 0$. For each element $T \in \mathcal{T}_h$, we denote by $\varrho(T)$ and $\delta(T)$ the diameter of T and the diameter of the largest ball contained in T, respectively. The mesh size of $\mathcal{T}_h$ will be denoted by $h := \max_{T\in\mathcal{T}_h} \varrho(T)$. This triangulation is assumed to be *regular* in the sense that there exist $\bar\varrho, \bar\delta > 0$ such that $\frac{\varrho(T)}{\delta(T)} \le \bar\delta$ and $\frac{h}{\varrho(T)} \le \bar\varrho$ for all $T \in \mathcal{T}_h$ and $h > 0$; see, *e.g.*, [20].

We will employ the standard continuous piecewise linear finite elements for the state y and set

$$V_h := \big\{v_h \in C(\overline\Omega) \mid v_{h|_T} \in \mathcal{P}_1 \quad \text{for all } T \in \mathcal{T}_h, v_h = 0 \quad \text{on } \partial\Omega\big\},$$

where $\mathcal{P}_1$ stands for the space of polynomials of degree equal at most 1. The discrete approximation of the state equation (3) for $y_h \in V_h$ is then

$$\int_\Omega (b + a(y_h))\nabla y_h \cdot \nabla v_h \,\mathrm{d}x = \int_\Omega u v_h \,\mathrm{d}x \qquad \text{for all } v_h \in V_h. \tag{22}$$

While the existence of solutions to (22) follows from Theorem 3.1 in [9], the uniqueness of solutions is still an open problem. However, if $a : \mathbb{R} \to R$ is assumed to be bounded, then we have uniqueness provided that h is small enough; see Theorem 4.1 in [9]. Below, we provide some error estimates for solutions to (22) that are sufficiently close to the solutions of (3).

In what follows, we fix $\bar u \in L^2(\Omega)$ and set $\bar y := S(\bar u)$. From Theorem 2.2 and the continuous embedding $L^2(\Omega) \hookrightarrow W^{-1,p}(\Omega)$ for any $p > 1$, we then have $\bar y \in W_0^{1,p}(\Omega) \cap H^2(\Omega)$.

Theorem 3.1 ([13], Thm. 3.1). *Let $\rho_0 > 0$ be arbitrary but fixed, $U := \overline{B}_{L^2(\Omega)}(\bar u, \rho_0)$, and let $p \ge 2$. Assume that Assumptions (A1) to (A3) are fulfilled. Then there exists a constant $h_0 > 0$ such that for any $u \in U$ and $h < h_0$, there exists at least one solution $y_h(u)$ to (22) satisfying for $y_u := S(u)$*

$$\|y_u - y_h(u)\|_{L^2(\Omega)} + h\|y_u - y_h(u)\|_{H_0^1(\Omega)} + h\|y_u - y_h(u)\|_{L^\infty(\Omega)} \le C_U h^2, \tag{23}$$

$$\|y_u - y_h(u)\|_{W_0^{1,p}(\Omega)} \le C_{U,p} h^{2/p}. \tag{24}$$

Proof. The estimates for the norms in L^2, H_0^1, and $W^{1,p}$ are shown in Theorem 3.1 in [13], while the estimate for the L^∞ norm can be obtained similar to estimate (3.11) in [13]. □

The following theorem guarantees the local uniqueness of solutions to (22). Its proof is similar to that of Theorem 4.2 in [10] with slight modifications and is thus omitted here.

Theorem 3.2. *Let $p > 2$ be arbitrary and let h_0 be defined in Theorem 3.1. Under Assumptions (A1) to (A3), there exist $h_1 \in (0, h_0)$, $\rho > 0$, and $\kappa_\rho > 0$ such that for any $h < h_1$ and any $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho)$, (22) admits a unique solution in $\overline{B}_{W_0^{1,p}(\Omega)}(\bar{y}, \kappa_\rho) \cap V_h$.*

From now on, let us fix $\tilde{p} \geq 4$ and let h_0, h_1, ρ, and κ_ρ be the constants defined in Theorems 3.1 and 3.2 for $p = \tilde{p}$. In the rest of this section, we shall investigate the differentiability of the discrete solution operator

$$S_h : B_{L^2(\Omega)}(\bar{u}, \rho) \ni u \mapsto y_h(u) \in \overline{B}_{W_0^{1,\tilde{p}}(\Omega)}(\bar{y}, \kappa_\rho) \cap V_h, \tag{25}$$

where $y_h(u)$ is the unique solution to (22) in $\overline{B}_{W_0^{1,\tilde{p}}(\Omega)}(\bar{y}, \kappa_\rho)$ from Theorem 3.2.

For any $y, \hat{y} \in C(\overline{\Omega}) \cap W^{1,1}(\Omega)$, we define functions $T_{y,\hat{y}}$ and $Z_{y,\hat{y}}$ on Ω *via*

$$T_{y,\hat{y}} := \mathbb{1}_{\{\hat{y} \neq \bar{t}\}}[a(y) - a(\hat{y}) - a'(\hat{y})(y - \hat{y})] \tag{26}$$

and

$$Z_{y,\hat{y}} := \mathbb{1}_{\{y \neq \bar{t}\}} a'(y)\nabla y - \mathbb{1}_{\{\hat{y} \neq \bar{t}\}} a'(\hat{y})\nabla \hat{y}. \tag{27}$$

In order to prove the differentiability of S_h, we need the following lemmas.

Lemma 3.3 ([22], Lem. 3.3)**.** *Let Assumption (A3) be fulfilled. Assume that $y_n \to y$ in $W_0^{1,p}(\Omega)$ as $n \to \infty$ with $p > 2$. Then*

$$\frac{1}{\|y_n - y\|_{W_0^{1,p}(\Omega)}} \|T_{y_n,y} \nabla y\|_{L^p(\Omega)} \to 0 \quad \text{as } n \to \infty.$$

Lemma 3.4. *Let Assumption (A3) be fulfilled and let $y, \hat{y} \in C(\overline{\Omega}) \cap W^{1,1}(\Omega)$ and $M > 0$ be arbitrary such that $\|y - \hat{y}\|_{C(\overline{\Omega})} < \delta$ with δ defined in (10) and $\|y\|_{C(\overline{\Omega})}, \|\hat{y}\|_{C(\overline{\Omega})} \leq M$. Then*

$$Z_{y,\hat{y}} = Z^{(1)}_{y,\hat{y}} + Z^{(2)}_{y,\hat{y}} + Z^{(3)}_{y,\hat{y}} + Z^{(4)}_{y,\hat{y}}, \tag{28}$$

for

$$\begin{aligned}
Z^{(1)}_{y,\hat{y}} &:= \mathbb{1}_{\{\hat{y} \in (-\infty, \bar{t}), y \in (-\infty, \bar{t})\}}[a_0'(y)\nabla y - a_0'(\hat{y})\nabla \hat{y}] + \mathbb{1}_{\{\hat{y} \in (\bar{t}, \infty), y \in (\bar{t}, \infty)\}}[a_1'(y)\nabla y - a_1'(\hat{y})\nabla \hat{y}], \\
Z^{(2)}_{y,\hat{y}} &:= \mathbb{1}_{\{\hat{y} = \bar{t}\}} \mathbb{1}_{\{y \neq \bar{t}\}} a'(y)\nabla y, \qquad Z^{(3)}_{y,\hat{y}} := [a_0'(\bar{t}) - a_1'(\bar{t})]\Big[\mathbb{1}_{\Omega^2_{y,\hat{y}}} - \mathbb{1}_{\Omega^3_{y,\hat{y}}}\Big]\nabla \hat{y},
\end{aligned}$$

and

$$\begin{aligned}
Z^{(4)}_{y,\hat{y}} := \mathbb{1}_{\Omega^2_{y,\hat{y}}}&[a_0'(y)\nabla(y - \hat{y}) + (a_0'(y) - a_0'(\bar{t}))\nabla \hat{y} + (a_1'(\bar{t}) - a_1'(\hat{y}))\nabla \hat{y}] \\
&+ \mathbb{1}_{\Omega^3_{y,\hat{y}}}[a_1'(y)\nabla(y - \hat{y}) + (a_1'(y) - a_1'(\bar{t}))\nabla \hat{y} + (a_0'(\bar{t}) - a_0'(\hat{y}))\nabla \hat{y}],
\end{aligned}$$

with the sets $\Omega^2_{y,\hat{y}}$ and $\Omega^3_{y,\hat{y}}$ defined as in (10). Moreover, there exists a constant $C_M > 0$ such that a.e. in Ω,

$$\begin{cases}
\left|Z^{(1)}_{y,\hat{y}}\right| + \left|Z^{(2)}_{y,\hat{y}}\right| \leq C_M[|y - \hat{y}||\nabla \hat{y}| + |\nabla(y - \hat{y})|], \\
\qquad\qquad \left|Z^{(4)}_{y,\hat{y}}\right| \leq C_M[|y - \hat{y}||\nabla \hat{y}| + |\nabla(y - \hat{y})|]\Big(\mathbb{1}_{\Omega^2_{y,\hat{y}}} + \mathbb{1}_{\Omega^3_{y,\hat{y}}}\Big).
\end{cases} \tag{29}$$

Consequently, $Z_{y,\hat{y}} \to 0$ in $L^p(\Omega)$ as $y \to \hat{y}$ in $W^{1,p}(\Omega) \cap C(\overline{\Omega})$ for any $p \geq 1$.

Remark 3.5. As will be shown later in the proof of Lemma 5.9, the decomposition (28) plays an important role in proving a link between $Z_{y_n,\bar{y}}$ and the curvature functional Q defined in (13) for some $y_n, \bar{y}$ with $y_n \to \bar{y}$ strongly in $C^1(\overline{\Omega})$. Specifically, $Z^{(1)}_{y_n,\bar{y}}$ and $Z^{(2)}_{y_n,\bar{y}}$ contribute to the smooth part Q_s and the first-order nonsmooth part Q_1; see (97) and (98). Meanwhile, $Z^{(3)}_{y_n,\bar{y}}$ corresponds to the second-order nonsmooth part Q_2; see (101).

Proof of Lemma 3.4. We first prove the identity (28). To this end, we now decompose the domain Ω into a disjoint union of its subsets as follows

$$\begin{aligned}\Omega &= \{\hat{y} \in (-\infty, \bar{t})\} \cup \{\hat{y} \in (\bar{t}, \infty)\} \cup \{\hat{y} = \bar{t}\}\\ &= \{\hat{y} \in (-\infty, \bar{t}), y \in (-\infty, \bar{t})\} \cup \{\hat{y} \in (-\infty, \bar{t}), y \in [\bar{t}, \infty)\}\\ &\cup \{\hat{y} \in (\bar{t}, \infty), y \in (-\infty, \bar{t}]\} \cup \{\hat{y} \in (\bar{t}, \infty), y \in (\bar{t}, \infty)\} \cup \{\hat{y} = \bar{t}\}.\end{aligned}$$

In view of the assumption that $\|y - \hat{y}\|_{C(\overline{\Omega})} < \delta$ and the definition of $\Omega^2_{y,\hat{y}}$ and $\Omega^3_{y,\hat{y}}$ in (10), there hold

$$\{\hat{y} \in (-\infty, \bar{t}), y \in [\bar{t}, \infty)\} = \Omega^3_{y,\hat{y}} \quad \text{and} \quad \{\hat{y} \in (\bar{t}, \infty), y \in (-\infty, \bar{t}]\} = \Omega^2_{y,\hat{y}}.$$

We thus have

$$\Omega = \{\hat{y} \in (-\infty, \bar{t}), y \in (-\infty, \bar{t})\} \cup \{\hat{y} \in (\bar{t}, \infty), y \in (\bar{t}, \infty)\} \cup \{\hat{y} = \bar{t}\} \cup \Omega^2_{y,\hat{y}} \cup \Omega^3_{y,\hat{y}}.$$

In order to prove (28), it suffices to show (28) on each component in the disjoint decomposition of Ω described above.

- For a.e. $x \in \{\hat{y} \in (-\infty, \bar{t}), y \in (-\infty, \bar{t})\}$, we deduce from the definition of $Z_{y,\hat{y}}$ in (27) and of a in (1) that

  $$Z_{y,\hat{y}}(x) = a_0'(y(x))\nabla y(x) - a_0'(\hat{y}(x))\nabla\hat{y}(x) = Z^{(1)}_{y,\hat{y}}(x)$$

  and

  $$Z^{(2)}_{y,\hat{y}}(x) = Z^{(3)}_{y,\hat{y}}(x) = Z^{(4)}_{y,\hat{y}}(x) = 0.$$

  These identities show (28).
- For a.e. $x \in \{\hat{y} \in (\bar{t}, \infty), y \in (\bar{t}, \infty)\}$, we similarly have (28).
- For a.e. $x \in \{\hat{y} = \bar{t}\}$, there holds $Z^{(1)}_{y,\hat{y}}(x) = Z^{(3)}_{y,\hat{y}}(x) = Z^{(4)}_{y,\hat{y}}(x) = 0$. Moreover, we deduce from the definition of $Z_{y,\hat{y}}$ in (27) that

  $$Z_{y,\hat{y}}(x) = \mathbb{1}_{\{y \neq \bar{t}\}}(x)a'(y(x))\nabla y(x) = Z^{(2)}_{y,\hat{y}}(x),$$

  which then verifies (28).
- For a.e. $x \in \Omega^2_{y,\hat{y}}$, there hold $Z^{(1)}_{y,\hat{y}}(x) = Z^{(2)}_{y,\hat{y}}(x) = 0$ and

  $$\begin{aligned}&Z^{(3)}_{y,\hat{y}}(x) + Z^{(4)}_{y,\hat{y}}(x)\\ &= [a_0'(\bar{t}) - a_1'(\bar{t})]\nabla\hat{y}(x) + [a_0'(y(x))\nabla(y(x) - \hat{y}(x)) + (a_0'(y(x)) - a_0'(\bar{t}))\nabla\hat{y}(x) + (a_1'(\bar{t}) - a_1'(\hat{y}(x)))\nabla\hat{y}(x)]\\ &= a_0'(y(x))\nabla y(x) - a_1'(\hat{y}(x))\nabla\hat{y}(x)\\ &= Z_{y,\hat{y}}(x),\end{aligned}$$

  where we have just used (10) and the definition of $Z_{y,\hat{y}}$ and of a, respectively, in (27) and in (1) to derive the last identity. We therefore have (28).
- For a.e. $x \in \Omega^3_{y,\hat{y}}$, we analogously have (28).

We have shown the identity (28). It remains to show (29) and the limit $Z_{y,\hat{y}} \to 0$ in $L^p(\Omega)$. A simple computation yields

$$Z^{(1)}_{y,\hat{y}} := \mathbb{1}_{\{\hat{y} \in (-\infty, \bar{t}), y \in (-\infty, \bar{t})\}}[a_0'(y)\nabla(y - \hat{y}) - (a_0'(y) - a_0'(\hat{y}))\nabla\hat{y}]$$

$$+ \mathbb{1}_{\{\hat{y}\in(\bar{t},\infty),y\in(\bar{t},\infty)\}}[a_1'(y)\nabla(y-\hat{y}) - (a_1'(y)-a_1'(\hat{y}))\nabla\hat{y}],$$

which, along with the continuity of a_0' and a_1', shows that $Z^{(1)}_{y,\hat{y}}$ satisfies (29). For $Z^{(2)}_{y,\hat{y}}$, we employ the fact that $\nabla\hat{y} = 0$ a.e. in $\{\hat{y} = \bar{t}\}$ (see; *e.g.*, [18], Rem. 2.6) and have

$$Z^{(2)}_{y,\hat{y}} = \mathbb{1}_{\{\hat{y}=\bar{t}\}}\mathbb{1}_{\{y\neq\bar{t}\}}a'(y)[\nabla y - \nabla\hat{y}] = \mathbb{1}_{\{\hat{y}=\bar{t}\}}\big[\mathbb{1}_{\{y<\bar{t}\}}a_0'(y) + \mathbb{1}_{\{y>\bar{t}\}}a_1'(y)\big][\nabla y - \nabla\hat{y}].$$

From this and the continuity of a_0' and a_1', we conclude that $Z^{(2)}_{y,\hat{y}}$ verifies (29). For $Z^{(4)}_{y,\hat{y}}$, by combining the continuity of a_0' and a_1' with the estimates

$$|y(x)-\bar{t}|, |\hat{y}(x)-\bar{t}| \leq |y(x)-\hat{y}(x)| \quad \text{for a.e. } x \in \Omega^2_{y,\hat{y}} \cup \Omega^3_{y,\hat{y}}$$

due to the definition of $\Omega^2_{y,\hat{y}}$ and $\Omega^3_{y,\hat{y}}$, it follows that $Z^{(4)}_{y,\hat{y}}$ satisfies (29). Finally, the claimed convergence of $Z_{y,\hat{y}}$ in $L^p(\Omega)$ follows from (28), (29), the fact that $\mathbb{1}_{\Omega^2_{y,\hat{y}}}, \mathbb{1}_{\Omega^3_{y,\hat{y}}} \to 0$ a.e. in Ω as $y \to \hat{y}$ in $C(\overline{\Omega})$, and Lebesgue's dominated convergence theorem. □

For any $h \in (0, h_1)$ and $y_h \in V_h$, we now define the operator $D_{h,y_h} : V_h \to V_h^*$ *via*

$$\langle D_{h,y_h}w_h, z_h\rangle := \int_\Omega \big[(b + a(y_h))\nabla w_h + \mathbb{1}_{\{y_h\neq\bar{t}\}}a'(y_h)\nabla y_h w_h\big]\cdot\nabla z_h \,\mathrm{d}x,\ w_h, z_h \in V_h. \tag{30}$$

Lemma 3.6. *Let all assumptions of Theorem 3.1 hold. Then for any $h \in (0, h_1)$ and any $\{y_h^k\} \subset V_h$ converging to $y_h \in V_h$ in $H_0^1(\Omega)$ as $k \to \infty$, there holds $\|D_{h,y_h^k} - D_{h,y_h}\|_{\mathbb{L}(V_h,V_h^*)} \to 0$.*

Proof. Let $w_h, v_h \in V_h$ be arbitrary such that $\|w_h\|_{H_0^1(\Omega)}, \|v_h\|_{H_0^1(\Omega)} \leq 1$ and $h \in (0, h_1)$ be arbitrary but fixed. Assume that $\{y_h^k\} \subset V_h$ converges to $y_h \in V_h$ in $H_0^1(\Omega)$ as $k \to \infty$. By virtue of the inverse inequality ([20], Thm. 3.2.6), we deduce that $y_h^k \to y_h$ in $W_0^{1,\tilde{p}}(\Omega)$ and hence in $C(\overline{\Omega})$ as $k \to \infty$. We can therefore assume that $\|y_h^k - y_h\|_{C(\overline{\Omega})} < \delta$ for all $k \in \mathbb{N}$ large enough. On the other hand, we have

$$\Big\langle \Big(D_{h,y_h^k} - D_{h,y_h}\Big)w_h, v_h\Big\rangle = \int_\Omega \Big[\big(a(y_h^k) - a(y_h)\big)\nabla w_h + Z_{y_h^k,y_h}w_h\Big]\cdot\nabla v_h \,\mathrm{d}x.$$

Together with the Hölder inequality, this yields that

$$\begin{aligned}\Big\|\Big(D_{h,y_h^k} - D_{h,y_h}\Big)w_h\Big\|_{V_h^*} &\leq \big\|a(y_n^k) - a(y_h)\big\|_{L^\infty(\Omega)}\|\nabla w_h\|_{L^2(\Omega)} + \Big\|Z_{y_h^k,y_h}\Big\|_{L^{\tilde{p}}(\Omega)}\|w_h\|_{L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)}\\ &\leq \big\|a(y_h^k) - a(y_h)\big\|_{L^\infty(\Omega)} + C\Big\|Z_{y_h^k,y_h}\Big\|_{L^{\tilde{p}}(\Omega)},\end{aligned}$$

where we have employed the continuous embedding $H_0^1(\Omega) \hookrightarrow L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)$ and the fact that $\|w_h\|_{H_0^1(\Omega)} \leq 1$ to obtain the last inequality. We thus have

$$\Big\|D_{h,y_h^k} - D_{h,y_h}\Big\|_{\mathbb{L}(V_h,V_h^*)} \leq \big\|a(y_h^k) - a(y_h)\big\|_{L^\infty(\Omega)} + C\Big\|Z_{y_h^k,y_h}\Big\|_{L^{\tilde{p}}(\Omega)}.$$

The first term on the right-hand side of the last estimate tends to zero as $k \to \infty$ since $y_h^k \to y_h$ in $C(\overline{\Omega})$ as $k \to \infty$. Moreover, the second term tends to zero as a result of Lemma 3.4. Hence $\|D_{h,y_h^k} - D_{h,y_h}\|_{\mathbb{L}(V_h,V_h^*)} \to 0$ as $k \to \infty$. □

Lemma 3.7. *Let all assumptions of Theorem 3.1 hold. Then there exists a constant $h_2 \in (0, h_1)$ such that for any $h \in (0, h_2)$ and any $y_h \in \overline{B}_{W_0^{1,\tilde{p}}(\Omega)}(\bar{y}, \kappa_\rho) \cap V_h$, the operator $D_{h,y_h} : V_h \to V_h^*$ is an isomorphism.*

Proof. Since V_h is finite-dimensional and D_{h,y_h} is linear, it suffices to prove that there exists an $h_2 \in (0, h_1)$ such that for any $h \in (0, h_2)$ and $y_h \in \overline{B}_{W_0^{1,\tilde{p}}(\Omega)}(\bar{y}, \kappa_\rho) \cap V_h$, the equation

$$D_{h,y_h} w_h = 0 \tag{31}$$

admits the unique solution $w_h = 0$. We argue by contradiction. Assume for any $k \geq 1$ that there exist $h_k \in (0, h_1)$, $y_{h_k} \in \overline{B}_{W_0^{1,\tilde{p}}(\Omega)}(\bar{y}, \kappa_\rho) \cap V_{h_k}$, and $w_{h_k} \in V_h \backslash \{0\}$ such that $h_k \to 0^+$ and w_{h_k} solves (31) for $h = h_k$ and $y_h = y_{h_k}$. By setting $\hat{w}_{h_k} := \frac{w_{h_k}}{\|w_{h_k}\|_{L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)}}$, we deduce that

$$\|\hat{w}_{h_k}\|_{L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)} = 1 \quad \text{and} \quad D_{h_k,y_{h_k}} \hat{w}_{h_k} = 0. \tag{32}$$

Furthermore, as a result of the embedding $W_0^{1,\tilde{p}}(\Omega) \Subset C(\overline{\Omega})$, there hold that $\|y_{h_k}\|_{C(\overline{\Omega})} \leq M$ for all $k \geq 1$ and some constant $M > 0$ independent of k and that

$$y_{h_k} \to y \ \text{ in } C(\overline{\Omega}) \quad \text{for some } y \in W_0^{1,\tilde{p}}(\Omega). \tag{33}$$

Testing the second equation in (32) by $\hat{w}_{h_k}$, Hölder's inequality thus gives

$$\underline{b}\|\nabla \hat{w}_{h_k}\|_{L^2(\Omega)} \leq \|\nabla y_{h_k}\|_{L^{\tilde{p}}(\Omega)} \left\| \mathbb{1}_{\{y_{h_k} \neq \bar{t}\}} a'(y_{h_k}) \right\|_{L^\infty(\Omega)} \|\hat{w}_{h_k}\|_{L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)} \leq C_{M,\rho}$$

for some constant $C_{M,\rho} > 0$. From this and the compact embedding $H_0^1(\Omega) \Subset L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)$, a subsequence argument shows that we can assume that

$$\hat{w}_{h_k} \rightharpoonup \hat{w} \ \text{ in } H_0^1(\Omega) \quad \text{and} \quad \hat{w}_{h_k} \to \hat{w} \ \text{ in } L^{2\tilde{p}/(\tilde{p}-2)}(\Omega) \tag{34}$$

for some $\hat{w} \in H_0^1(\Omega)$. Moreover, there exist an element $\mathbb{b} \in L^{\tilde{p}}(\Omega)^2$ and a subsequence of $\{\mathbb{b}_k\}$ with $\mathbb{b}_k := \mathbb{1}_{\{y_{h_k} \neq \bar{t}\}} a'(y_{h_k}) \nabla y_{h_k}$, denoted in the same way, such that $\mathbb{b}_k \rightharpoonup \mathbb{b}$ weakly in $L^{\tilde{p}}(\Omega)^2$. By fixing any $v \in H^2(\Omega) \cap H_0^1(\Omega)$ and testing the last equation in (32) with $v_{h_k} := \Pi_{h_k} v \in V_{h_k}$, where Π_{h_k} is the interpolation operator, we have

$$\int_\Omega [(b + a(y_{h_k}))\nabla \hat{w}_{h_k} + \hat{w}_{h_k} \mathbb{b}_k] \cdot \nabla v_{h_k} \, \mathrm{d}x = 0 \quad \text{for all } k \geq 1.$$

Letting $k \to \infty$ and exploiting the limits (33), (34), $\mathbb{b}_k \rightharpoonup \mathbb{b}$ in $L^{\tilde{p}}(\Omega)^N$, and $v_{h_k} \to v$ in $H_0^1(\Omega)$, we can conclude that $\int_\Omega [(b + a(y))\nabla \hat{w} + \hat{w}\mathbb{b}] \cdot \nabla v \, \mathrm{d}x = 0$. From this and the density of $H^2(\Omega) \cap H_0^1(\Omega)$ in $H_0^1(\Omega)$, we deduce that $\hat{w}$ satisfies $\langle \mathcal{S}\hat{w}, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = 0$ for all $v \in H_0^1(\Omega)$, where

$$\langle \mathcal{S}w, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} := \int_\Omega [a_* \nabla w + w\mathbb{b}] \cdot \nabla v \, \mathrm{d}x = 0 \quad \text{for } w, v \in H_0^1(\Omega)$$

with $a_*(x) := b(x) + a(y(x)) \geq \underline{b} > 0$ for all $x \in \overline{\Omega}$. An argument similar to that in the proof of Theorem 2.6 (1) in [9] shows that $\mathcal{S}$ is injective as a mapping from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$. Consequently, there holds $\hat{w} = 0$. This contradicts the fact that $\|\hat{w}\|_{L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)} = \lim_{k\to\infty} \|\hat{w}_{h_k}\|_{L^{2\tilde{p}/(\tilde{p}-2)}(\Omega)} = 1$. □

As a consequence of Lemmas 3.6 and 3.7 and the implicit function theorem, we obtain the differentiability of S_h.

Theorem 3.8. *Let all assumptions of Theorem 3.1 hold. Then, for any $h \in (0, h_2)$, the operator S_h defined in (25) is of class C^1. Moreover, for any $u \in B_{L^2(\Omega)}(\bar{u}, \rho)$, let $y_h(u) := S_h(u)$. Then for any $v \in L^2(\Omega)$, the Fréchet derivative $S_h'(u)v =: z_h$ is the unique solution to*

$$\int_\Omega [(b + a(y_h(u)))\nabla z_h + \mathbb{1}_{\{y_h(u)\neq \bar{t}\}} a'(y_h(u)) z_h \nabla y_h(u)] \cdot \nabla w_h \,\mathrm{d}x = \int_\Omega v w_h \,\mathrm{d}x \quad \text{for all } w_h \in V_h. \tag{35}$$

Proof. We first consider for any $h \in (0, h_2)$ the mapping $F_h : B_{L^2(\Omega)}(\bar{u}, \rho) \times V_h \to V_h^*$ defined *via*

$$\langle F_h(u, y_h), v_h\rangle = \int_\Omega (b + a(y_h))\nabla y_h \cdot \nabla v_h - u v_h \,\mathrm{d}x, \quad u \in B_{L^2(\Omega)}(\bar{u}, \rho), y_h, v_h \in V_h. \tag{36}$$

Clearly, $F_h(u, y_h(u)) = 0$ and F_h is continuously partially differentiable in u. We now prove that F_h is partially differentiable in y_h with $\frac{\partial F_h}{\partial y_h}(u, y_h) = D_{h,y_h}$, where D_{h,y_h} is defined in (30). We thus derive the differentiability of S_h according to Lemmas 3.6 and 3.7 as well as a simple computation. To this end, by taking any $v_h \in V_h$ and $\{w_h^k\} \subset V_h$ with $\|w_h^k\|_{H_0^1(\Omega)} \to 0$ as $k \to \infty$ and $\|v_h\|_{H_0^1(\Omega)} \le 1$, we deduce from a straightforward computation that

$$\big\langle F_h\big(u, y_h + w_h^k\big) - F_h(u, y_h) - D_{h,y_h} w_h^k, v_h\big\rangle = \int_\Omega \Big[T_{y_h^k, y_h} \nabla y_h + \big(a\big(y_h^k\big) - a(y_h)\big)\nabla w_h^k\Big] \cdot \nabla v_h \,\mathrm{d}x,$$

where $y_h^k := y_h + w_h^k$ and $T_{y,\tilde{y}}$ is defined in (26). This gives

$$\big\|F_h\big(u, y_h + w_h^k\big) - F_h(u, y_h) - D_{h,y_h} w_h^k\big\|_{V_h^*} \le \Big\|T_{y_h^k, y_h} \nabla y_h\Big\|_{L^2(\Omega)} + \big\|a(y_h^k) - a(y_h)\big\|_{L^\infty(\Omega)} \big\|w_h^k\big\|_{H_0^1(\Omega)}.$$

Moreover, in view of inverse estimates ([20], Thm. 3.2.6), we have $\|w_h^k\|_{W_0^{1,\tilde{p}}(\Omega)} \le C h^{-2(\frac{1}{2} - \frac{1}{\tilde{p}})} \|w_h^k\|_{H_0^1(\Omega)}$. Consequently, $y_h^k \to y_h$ in $W_0^{1,\tilde{p}}(\Omega)$ and hence in $C(\overline{\Omega})$ as $k \to \infty$. Moreover, an inverse estimate and Hölder's inequality imply that

$$\begin{aligned}
&\frac{1}{\|w_h^k\|_{H_0^1(\Omega)}} \big\|F_h(u, y_h + w_h^k) - F_h(u, y_h) - D_{h,y_h} w_h^k\big\|_{V_h^*} \le \frac{1}{\|w_h^k\|_{H_0^1(\Omega)}} \Big\|T_{y_h^k, y_h} \nabla y_h\Big\|_{L^2(\Omega)} + \big\|a(y_h^k) - a(y_h)\big\|_{L^\infty(\Omega)} \\
&\quad \le \frac{C h^{-2(\frac{1}{2} - \frac{1}{\tilde{p}})}}{\|w_h^k\|_{W_0^{1,\tilde{p}}(\Omega)}} \Big\|T_{y_h^k, y_h} \nabla y_h\Big\|_{L^{\tilde{p}}(\Omega)} + \big\|a(y_h^k) - a(y_h)\big\|_{L^\infty(\Omega)} \to 0,
\end{aligned}$$

where we have just exploited Lemma 3.3 and the continuity of a to obtain the last limit. This gives that $\frac{\partial F_h}{\partial y_h}(u, y_h) = D_{h,y_h}$. We have shown that $F_h(u, S_h(u)) = 0$ and $\frac{\partial F_h}{\partial y_h}(u, y_h) = D_{h,y_h}$. We then deduce from the Implicit Function Theorem and Lemmas 3.6 and 3.7 that S_h is of class C^1. Finally, (35) follows from (30) and (36). □

4. Numerical analysis of the adjoint state equation

In this section, we will carry out the numerical analysis of the adjoint equation (6). For any $h \in (0, h_2)$, $u \in B_{L^2(\Omega)}(\bar{u}, \rho)$, $v \in L^2(\Omega)$ and $y_h := S_h(u)$, we approximate (6) using the triangulation $\mathcal{T}_h$ by

$$\int_\Omega (b + a(y_h))\nabla \varphi_h \cdot \nabla w_h + \mathbb{1}_{\{y_h \neq \bar{t}\}} a'(y_h) w_h \nabla y_h \cdot \nabla \varphi_h \,\mathrm{d}x = \int_\Omega v w_h \,\mathrm{d}x \quad \text{for all } w_h \in V_h. \tag{37}$$

From the bijectivity of D_{h,y_h} shown in Lemma 3.7, we deduce the existence and uniqueness of solutions to (37).

Theorem 4.1. *Let all assumptions of Theorem 3.8 hold. Then for all $h \in (0, h_2)$, $u \in B_{L^2(\Omega)}(\bar{u}, \rho)$, and $v \in L^2(\Omega)$, there exists a unique solution $\varphi_h \in V_h$ to (37).*

In order to derive error estimates for the full approximation (37) of (6), we first consider the continuous problem (6) with $y_h(u)$ in place of y_u.

Lemma 4.2. *Let all assumptions of Theorem 3.8 hold. Then for any $h \in (0, h_2)$, $u \in B_{L^2(\Omega)}(\bar{u}, \rho)$, $y_h := S_h(u)$, and $v \in L^2(\Omega)$, the equation*

$$-\operatorname{div}[(b + a(y_h))\nabla\tilde{\varphi}] + \mathbb{1}_{\{y_h \neq \bar{t}\}} a'(y_h)\nabla y_h \cdot \nabla\tilde{\varphi} = v \quad \text{in } \Omega, \quad \tilde{\varphi} = 0 \quad \text{on } \partial\Omega, \tag{38}$$

has a unique solution $\tilde{\varphi}$ in $H^2(\Omega) \cap H_0^1(\Omega)$. Moreover,

$$\|\varphi - \tilde{\varphi}\|_{H^2(\Omega)} \le C_\rho h \|v\|_{L^2(\Omega)} \quad \text{and} \quad \|\varphi - \tilde{\varphi}\|_{L^2(\Omega)} \le C_\rho h^2 \|v\|_{L^2(\Omega)} \tag{39}$$

for some constant C_ρ independent of u, v, and h, where φ is the unique solution to (6).

Proof. From Theorem 2.2, the continuous embedding $W_0^{1,\tilde{p}}(\Omega) \hookrightarrow C(\overline{\Omega})$, and (24) for $p := \tilde{p} \ge 4$, there holds

$$\left\|\mathbb{1}_{\{y_h \neq \bar{t}\}} a'(y_h)\nabla y_h\right\|_{L^{\tilde{p}}(\Omega)} + \|b + a(y_h(u))\|_{W^{1,\tilde{p}}(\Omega)} \le C_\rho \quad \text{for all } h \in (0, h_2), u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho).$$

A standard argument then proves the existence of solutions $\tilde{\varphi}$ to (38) in $H^2(\Omega) \cap H_0^1(\Omega)$; see, *e.g.*, Theorem 2.6 in [9] and the proof of Lemma 4.1 in [22]. Moreover, we have

$$\|\tilde{\varphi}\|_{H^2(\Omega)} \le C_\rho \|v\|_{L^2(\Omega)}. \tag{40}$$

Setting $\psi := \varphi - \tilde{\varphi}$ and subtracting the equations corresponding to φ and $\tilde{\varphi}$ yields

$$-\operatorname{div}[(b + a(y_u))\nabla\psi] + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u)\nabla y_u \cdot \nabla\psi = g_{u,h} \quad \text{in } \Omega, \quad \psi = 0 \quad \text{on } \partial\Omega, \tag{41}$$

with

$$g_{u,h} := \operatorname{div}[(a(y_u) - a(y_h))\nabla\tilde{\varphi}] + Z_{y_h, y_u} \cdot \nabla\tilde{\varphi}.$$

By the chain rule ([33], Thm. 7.8) and the fact that $y_u, y_h \in W_0^{1,\tilde{p}}(\Omega)$ and that $\tilde{\varphi} \in H^2(\Omega)$, we can write

$$g_{u,h} = (a(y_u) - a(y_h))\Delta\tilde{\varphi}. \tag{42}$$

Similar to (40), there holds

$$\|\psi\|_{H^2(\Omega)} \le C_\rho \|g_{u,h}\|_{L^2(\Omega)} \le C_\rho \|a(y_u) - a(y_h)\|_{L^\infty(\Omega)} \|\Delta\tilde{\varphi}\|_{L^2(\Omega)}.$$

Combining this with the Lipschitz continuity of a on bounded sets, the L^∞-estimate in (23), and (40) yields the first estimate in (39). To show the second estimate, set $z_{u,\psi} := S'(u)\psi$ and note that $\psi = S'(u)^* g_{u,h}$. We then deduce from (42) that

$$\|\psi\|_{L^2(\Omega)}^2 = \int_\Omega g_{u,h} z_{u,\psi} \, \mathrm{d}x \le C_\rho \|z_{u,\psi}\|_{L^\infty(\Omega)} \|\Delta\tilde{\varphi}\|_{L^2(\Omega)} \|y_u - y_h\|_{L^2(\Omega)}. \tag{43}$$

By Theorem 2.2 and the compact embedding $L^2(\Omega) \Subset W^{-1,\tilde{p}}(\Omega)$, we have

$$\sup\left\{\|S'(u)\|_{\mathbb{L}(W^{-1,\tilde{p}}(\Omega), W_0^{1,\tilde{p}}(\Omega))} \mid u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho)\right\} \le C_\rho. \tag{44}$$

The continuous embeddings $W_0^{1,\tilde{p}}(\Omega) \hookrightarrow L^\infty(\Omega)$ and $L^2(\Omega) \hookrightarrow W^{-1,\tilde{p}}(\Omega)$ therefore yield

$$\|z_{u,\psi}\|_{L^\infty(\Omega)} \le C \|z_{u,\psi}\|_{W_0^{1,\tilde{p}}(\Omega)} \le C_\rho \|\psi\|_{W^{-1,\tilde{p}}(\Omega)} \le C_\rho \|\psi\|_{L^2(\Omega)}.$$

The inequality (43) thus yields

$$\|\psi\|_{L^2(\Omega)} \le C_\rho \|\Delta\tilde{\varphi}\|_{L^2(\Omega)} \|y_u - y_h\|_{L^2(\Omega)}.$$

This, (23) and (40) yield the last estimate in (39). □

Below, we shall estimate the term $Z_{y_h,y}$ defined in (27). We first observe from the L^∞-error estimate in (23) that

$$\|S(u) - S_h(u)\|_{L^\infty(\Omega)} \leq C_\infty h \quad \text{for all } u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}, h \in (0, h_2) \tag{45}$$

for some positive constant C_∞. For any $y \in W^{1,1}(\Omega) \cap C(\overline{\Omega})$ and $r > 0$, let

$$V(y, r) := \sigma_0 \sum_{m=1}^{2} \mathbb{1}_{\{0<|y-\bar{t}|\leq r\}} |\partial_{x_m} y| \quad \text{and} \quad \Sigma_r(y) := \frac{1}{r} V(y, r) \tag{46}$$

with σ_0 determined as in (15).

Proposition 4.3. *Let $r > 0$, $y \in W^{1,1}(\Omega) \cap C(\overline{\Omega})$, and $\hat{y} \in W^{1,\infty}(\Omega)$ be arbitrary and let $\kappa := r + \|y - \hat{y}\|_{C(\overline{\Omega})}$. Then*

(i) $V(y, r) \leq V(\hat{y}, \kappa) + \sigma_0 \sum_{m=1}^{2} |\partial_{x_m} y - \partial_{x_m} \hat{y}|$ *for a.e. in* Ω;
(ii) $\|V(\hat{y}, r)\|^2_{L^2(\Omega)} \leq 2r\sigma_0 \|\nabla \hat{y}\|_{L^\infty(\Omega)} \|\Sigma_r(\hat{y})\|_{L^1(\Omega)}$.

Proof. The proof of the second claim is straightforward. It remains to prove the first assertion. To this end, we now observe that $\{0 < |y - \bar{t}| \leq r\} \subset \{|\hat{y} - \bar{t}| \leq \kappa\}$ and $|\partial_{x_m} y| \leq |\partial_{x_m} \hat{y}| + |\partial_{x_m} y - \partial_{x_m} \hat{y}|$. There thus holds

$$\begin{aligned} V(y, r) &\leq \sum_{m=1}^{2} \sigma_0 \mathbb{1}_{\{|\hat{y}-\bar{t}|\leq\kappa\}} [|\partial_{x_m} \hat{y}| + |\partial_{x_m} y - \partial_{x_m} \hat{y}|] \\ &= \sum_{m=1}^{2} \sigma_0 \mathbb{1}_{\{|\hat{y}-\bar{t}|\leq\kappa\}} |\partial_{x_m} \hat{y}| + \sum_{m=1}^{2} \sigma_0 \mathbb{1}_{\{|\hat{y}-\bar{t}|\leq\kappa\}} |\partial_{x_m} y - \partial_{x_m} \hat{y}| \\ &= V(\hat{y}, \kappa) + \sigma_0 \sum_{m=1}^{2} \mathbb{1}_{\{|\hat{y}-\bar{t}|\leq\kappa\}} |\partial_{x_m} y - \partial_{x_m} \hat{y}|, \end{aligned}$$

where we have employed the fact that $\nabla \hat{y}$ vanishes a.e. in $\{\hat{y} = \bar{t}\}$ in order to obtain the last identity. This yields the first claim. □

Lemma 4.4. *There exist an $h_3 \in (0, h_2]$ and a constant $L_\rho > 0$ such that for all $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}$ and $h \in (0, h_3)$, there hold*

$$\|Z_{y_h,y}\|_{L^2(\Omega)} \leq L_\rho h + \left\| V\left(y, \|y_h - y\|_{L^\infty(\Omega)}\right) \right\|_{L^2(\Omega)} \tag{47}$$

and

$$\|Z_{y,\bar{y}}\|_{L^2(\Omega)} \leq L_\rho \|u - \bar{u}\|_{L^2(\Omega)} + \left\| V\left(\bar{y}, \|y - \bar{y}\|_{L^\infty(\Omega)}\right) \right\|_{L^2(\Omega)} \tag{48}$$

with $y := S(u)$ and $y_h := S_h(u)$.

Proof. By Theorem 2.2 (see also [22], Thm. 3.1), there exists a constant $M_{1,\rho}$ such that

$$\|S(u)\|_{W^{1,\infty}(\Omega)} \leq M_{1,\rho} \quad \text{for all } u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}. \tag{49}$$

Setting $h_3 := \min\{h_2, \delta 2^{-1} C_\infty^{-1}\}$ and exploiting (45) shows that $\|y - y_h\|_{C(\overline{\Omega})} < \delta$ for any $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}$ and $h \in (0, h_3)$. From the definition of $Z_{y_h,y}$ in (27) and Lemma 3.4, we arrive at

$$Z_{y_h,y} = Z^{(1)}_{y_h,y} + Z^{(2)}_{y_h,y} + Z^{(3)}_{y_h,y} + Z^{(4)}_{y_h,y}. \tag{50}$$

By (29), (49), and Theorem 3.1, we have

$$\left\|Z^{(k)}_{y_h,y}\right\|_{L^2(\Omega)} \le L_\rho h \quad \text{for } k=1,2,4. \tag{51}$$

On the other hand, we have

$$\begin{cases} \Omega^2_{y_h,y} = \{y \in (\bar t, \bar t+\delta), y_h \in (\bar t-\delta, \bar t]\} \subset \left\{0 < y-\bar t \le \|y_h-y\|_{L^\infty(\Omega)}\right\}, \\ \Omega^3_{y_h,y} = \{y \in (\bar t-\delta, \bar t), y_h \in [\bar t, \bar t+\delta)\} \subset \left\{0 < \bar t-y \le \|y_h-y\|_{L^\infty(\Omega)}\right\}, \end{cases} \tag{52}$$

which together with the definitions of $Z^{(3)}_{y_h,y}$ in Lemma 3.4 and of V in (46), and (15), show that

$$\left|Z^{(3)}_{y_h,y}\right| \le V\Big(y, \|y_h-y\|_{L^\infty(\Omega)}\Big) \tag{53}$$

a.e. in Ω. Combining this with (50) and (51), we obtain (47).

For (48), we first see that

$$\left\|Z^{(k)}_{y,\bar y}\right\|_{L^2(\Omega)} \le C\|y-\bar y\|_{H^1_0(\Omega)} \le C\|u-\bar u\|_{L^2(\Omega)} \quad \text{for } k=1,2,4,$$

which is similar to (51) and is derived by using (29), (49), Theorem 2.2, and the continuity of the mapping $S : L^2(\Omega) \hookrightarrow W^{-1,\tilde p}(\Omega) \to W^{1,\tilde p}_0(\Omega) \hookrightarrow H^1_0(\Omega)$. Finally, similar to (53), one has

$$\left|Z^{(3)}_{y,\bar y}\right| \le V\Big(y, \|y-\bar y\|_{L^\infty(\Omega)}\Big)$$

a.e. in Ω. We thus obtain (48). □

From now on, let us fix constants γ, p_0, and p_1 satisfying

$$1-\frac{2}{p_0} \le \gamma < \min\left\{\frac{1}{p_0}, 1-\frac{2}{p_1}\right\} < \frac12, \quad p_0 > 2 = N, \quad \text{and} \quad p_1 \in (2, p_*), \tag{54}$$

where p_* is given in Theorem 2.2.

In order to derive an L^2-error estimate of $\tilde\varphi - \varphi_h$, we cannot directly employ a duality argument based on the Aubin–Nitsche trick since the linearized state equation (4) admits solutions belonging to $W^{1,p}(\Omega)$ only due to the nondifferentiability of the function a; see Remark 2.4. To overcome this difficulty, we now consider the following *adjusted* linearized state equation:

$$-\operatorname{div}\big[(b+a(y_u))\nabla \tilde z_{u,v} + \mathbb{1}_{\{\bar y \ne \bar t\}} a'(\bar y)\tilde z_{u,v}\nabla \bar y\big] = v \ \text{ in } \Omega, \quad \tilde z_{u,v} = 0 \ \text{ on } \partial\Omega \tag{55}$$

for all $u \in \mathcal{U}_{ad}$ and v belonging to $W^{-1+\gamma,2}(\Omega)$, the dual space of a Sobolev space of fractional order. Compared to (4), we have here partly linearized the state equation (3) at the optimal state $\bar y$ by replacing the vector-valued function $\mathbb{1}_{\{y_u \ne \bar t\}} a'(y_u)\nabla y_u$ in (4) by $\mathbb{1}_{\{\bar y \ne \bar t\}} a'(\bar y)\nabla \bar y$. We can therefore exploit assumptions imposed on $\bar y$ to derive the necessary regularity of solutions $\tilde z_{u,v}$ to (55). Indeed, as we will see later in Proposition B.1 in Appendix B, the finiteness of $\Sigma(\bar y)$ implies that

$$\mathbb{1}_{\{\bar y \ne \bar t\}} a'(\bar y)\nabla \bar y \in (W^{\gamma,p_0}(\Omega))^2 = (W^{\gamma,p_0}_0(\Omega))^2,$$

where the constants γ and p_0 are fixed and satisfy (54). From this and the $W^{1+\gamma,2}$-regularity of solutions to (55), we can then show that $\tilde z_{u,v} \in W^{1+\gamma,2}(\Omega)$ whenever $v \in W^{-1+\gamma,2}(\Omega)$. This regularity of solutions to (55) will play an important role in establishing *a priori* error estimates for the discretization of the adjoint state equation (6).

Theorem 4.5. *If $\Sigma(\bar{y}) < \infty$, then there exist constants $\bar{h} := \bar{h}(\bar{u}) \in (0, h_3)$, $\bar{\rho} := \bar{\rho}(\bar{u}) \leq \min\{\rho, \hat{\rho}, \rho_{\tilde{p}}\}$ (with $\hat{\rho}$ and $\rho_{\tilde{p}}$ being the constants in Proposition B.3 associated with $p := \tilde{p}$), and $C_{\bar{u}} > 0$ such that*

$$\|\varphi - \varphi_h\|_{L^2(\Omega)} \leq C_{\bar{u}} \varepsilon_h^u \|v\|_{L^2(\Omega)} \quad \text{and} \quad \|\varphi - \varphi_h\|_{H_0^1(\Omega)} \leq C_{\bar{u}} h \|v\|_{L^2(\Omega)}$$

for all $h \in (0, \bar{h})$, $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\rho}) \cap \mathcal{U}_{ad}$, and $v \in L^2(\Omega)$, where

$$\varepsilon_h^u := h^{1+\gamma} + h\left\|Z_{S(u),\bar{y}}\right\|_{L^2(\Omega)} + h\left\|V\Big(S(u), \|S_h(u) - S(u)\|_{L^\infty(\Omega)}\Big)\right\|_{L^2(\Omega)}, \tag{56}$$

and φ and φ_h are the unique solutions to (6) and (37), respectively.

Proof. Let $\tilde{\varphi}$ be the solution of (38). To simplify the notation, set $y_u := S(u)$ and $y_h := S_h(u)$ for any $h \in (0, h_3)$ and $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \min\{\rho, \hat{\rho}, \rho_{\tilde{p}}\}) \cap \mathcal{U}_{ad}$. We divide the proof into three steps.

Step 1: Existence of a constant $C_{1,h_3,\rho}$ such that

$$\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)} \leq C_{1,h_3,\rho}\left(h^\gamma + \|Z_{y_u,\bar{y}}\|_{L^2(\Omega)} + \left\|V\Big(y_u, \|y_h - y_u\|_{L^\infty(\Omega)}\Big)\right\|_{L^2(\Omega)}\right)\|\tilde{\varphi} - \varphi_h\|_{H_0^1(\Omega)} \tag{57}$$

for all $h \in (0, h_3)$, $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \min\{\rho, \hat{\rho}, \rho_{\tilde{p}}\}) \cap \mathcal{U}_{ad}$, and $v \in L^2(\Omega)$.

To prove (57), first let $\hat{z}$ be the unique solution to the adjusted linearized state equation (55) corresponding to $v := \tilde{\varphi} - \varphi_h$. Since $\tilde{\varphi} - \varphi_h \in H_0^1(\Omega)$, one has $\tilde{\varphi} - \varphi_h \in W^{-1+\gamma,2}(\Omega)$ due to the embeddings $H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow W^{-1+\gamma,2}(\Omega)$. From Proposition B.3 and the finiteness of $\Sigma(\bar{y})$, we have that $\hat{z} \in W^{1+\gamma,2}(\Omega)$ and that

$$\|\hat{z}\|_{W^{1+\gamma,2}(\Omega)} \leq C\|\tilde{\varphi} - \varphi_h\|_{W^{-1+\gamma,2}(\Omega)} \leq C\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)}. \tag{58}$$

Moreover, we also have that

$$\hat{z} = S'(u)(\tilde{\varphi} - \varphi_h - \operatorname{div}(Z_{y_u,\bar{y}}\hat{z})); \tag{59}$$

see the identity (B.16) in the proof of Proposition B.3 for $v := \tilde{\varphi} - \varphi_h$. Testing (4) for $v := \tilde{\varphi} - \varphi_h - \operatorname{div}(Z_{y_u,\bar{y}}\hat{z})$ by $\tilde{\varphi} - \varphi_h$ and using (59) thus yields

$$\begin{aligned}\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)}^2 + \int_\Omega Z_{y_u,\bar{y}} \cdot \nabla(\tilde{\varphi} - \varphi_h)\hat{z}\,\mathrm{d}x \\ = \int_\Omega (b + a(y_u))\nabla\hat{z} \cdot \nabla(\tilde{\varphi} - \varphi_h) + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u)\hat{z}\nabla y_u \cdot \nabla(\tilde{\varphi} - \varphi_h)\,\mathrm{d}x.\end{aligned} \tag{60}$$

Furthermore, applying assertion (i) in Proposition B.3 for $p := \tilde{p}$ and using the embedding $L^2(\Omega) \hookrightarrow W^{-1,\tilde{p}}(\Omega)$ (due to $N = 2$) shows that

$$\|\hat{z}\|_{W_0^{1,\tilde{p}}(\Omega)} \leq C_\rho\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)} \quad \text{for all } h \in (0, h_3), u \in B_{L^2(\Omega)}(\bar{u}, \rho_{\tilde{p}}) \cap \mathcal{U}_{ad}, v \in L^2(\Omega). \tag{61}$$

Consider for any $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \min\{\rho, \hat{\rho}, \rho_{\tilde{p}}\}) \cap \mathcal{U}_{ad}$ and $h \in (0, h_3)$ the bilinear operators $B_u : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ and $B_{u,h} : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ defined *via*

$$\begin{aligned}B_u(z, w) &:= \int_\Omega \left[(b + a(y_u))\nabla z + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u) z \nabla y_u\right] \cdot \nabla w\,\mathrm{d}x, \\ B_{u,h}(z, w) &:= \int_\Omega \left[(b + a(y_h))\nabla z + \mathbb{1}_{\{y_h \neq \bar{t}\}} a'(y_h) z \nabla y_h\right] \cdot \nabla w\,\mathrm{d}x.\end{aligned}$$

From this and (60), we obtain for any $w_h \in V_h$ that

$$\|\tilde{\varphi} - \varphi_h\|^2_{L^2(\Omega)} + \int_\Omega Z_{y_u,\bar{y}} \cdot \nabla(\tilde{\varphi} - \varphi_h)\hat{z}\,\mathrm{d}x = B_u(\hat{z}, \tilde{\varphi} - \varphi_h) = B_u(\hat{z} - w_h, \tilde{\varphi} - \varphi_h) + B_u(w_h, \tilde{\varphi} - \varphi_h)$$
$$= B_u(\hat{z} - w_h, \tilde{\varphi} - \varphi_h) + [B_u(w_h, \tilde{\varphi} - \varphi_h) - B_{u,h}(w_h, \tilde{\varphi} - \varphi_h)],$$

where we have used the fact that $B_{u,h}(w_h, \tilde{\varphi} - \varphi_h) = 0$ which follows from combining (37) with (38). We now estimate the first integral in the left-hand side and two summands in the right-hand side of the above identity. From Hölder's inequality, there holds

$$\begin{aligned}
&\|\tilde{\varphi} - \varphi_h\|^2_{L^2(\Omega)} - \|Z_{y_u,\bar{y}}\|_{L^2(\Omega)}\|\hat{z}\|_{L^\infty(\Omega)}\|\tilde{\varphi} - \varphi_h\|_{H^1_0(\Omega)} \\
&\quad \le \|b + a(y_u)\|_{L^\infty(\Omega)}\|\hat{z} - w_h\|_{H^1_0(\Omega)}\|\tilde{\varphi} - \varphi_h\|_{H^1_0(\Omega)} \\
&\qquad + \left\|\mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u)\nabla y_u\right\|_{L^\infty(\Omega)}\|\hat{z} - w_h\|_{L^2(\Omega)}\|\tilde{\varphi} - \varphi_h\|_{H^1_0(\Omega)} \\
&\qquad + \|a(y_u) - a(y_h)\|_{L^4(\Omega)}\|w_h\|_{W^{1,4}_0(\Omega)}\|\tilde{\varphi} - \varphi_h\|_{H^1_0(\Omega)} \\
&\qquad + \|Z_{y_h,y_u}\|_{L^2(\Omega)}\|w_h\|_{L^\infty(\Omega)}\|\tilde{\varphi} - \varphi_h\|_{H^1_0(\Omega)}
\end{aligned} \tag{62}$$

for all $w_h \in V_h$. Moreover, we have from Assumption (A3), the continuous embedding $H^{N/4}(\Omega) \hookrightarrow L^4(\Omega)$ with $N = 2$ (see, *e.g.*, [34], Thm. 1.4.4.1), interpolation theory ([7], Thm. 14.2.7), and (23) that

$$\begin{aligned}
\|a(y_u) - a(y_h)\|_{L^4(\Omega)} &\le C\|y_u - y_h\|_{L^4(\Omega)} \le C\|y_u - y_h\|_{H^{N/4}(\Omega)} \\
&\le C\|y_u - y_h\|^{1-\frac{2}{4}}_{L^2(\Omega)}\|y_u - y_h\|^{\frac{2}{4}}_{H^1_0(\Omega)} \le Ch^{\frac{3}{2}}.
\end{aligned} \tag{63}$$

We then deduce from this, (62), (49), the assumptions on b and a, and (47) that

$$\begin{aligned}
\|\tilde{\varphi} - \varphi_h\|^2_{L^2(\Omega)} \le C_\rho\|\tilde{\varphi} - \varphi_h\|_{H^1_0(\Omega)} \Bigg[&\|\hat{z} - w_h\|_{H^1_0(\Omega)} + h^{\frac{3}{2}}\|w_h\|_{W^{1,4}_0(\Omega)} \\
&+ \left(h + \left\|V\left(y_u, \|y_h - y_u\|_{L^\infty(\Omega)}\right)\right\|_{L^2(\Omega)}\right)\|w_h\|_{L^\infty(\Omega)} + \|Z_{y_u,\bar{y}}\|_{L^2(\Omega)}\|\hat{z}\|_{L^\infty(\Omega)}\Bigg]
\end{aligned} \tag{64}$$

for all $w_h \in V_h$. Moreover, from standard interpolation error estimates (see, *e.g.*, [7], Thm. 4.4.20) and applying the estimates (4.4.21) for $(p, s, m) := (4, 1, 1)$ and (4.4.22) for $(p, s, m, n) := (\tilde{p}, 0, 1, 2)$, we have that

$$\begin{cases}
\|\hat{z} - \Pi_h\hat{z}\|_{W^{1,4}_0(\Omega)} \le C_1\|\hat{z}\|_{W^{1,4}_0(\Omega)} \le C_1\|\hat{z}\|_{W^{1,\tilde{p}}_0(\Omega)}, \\
\|\hat{z} - \Pi_h\hat{z}\|_{L^\infty(\Omega)} \le C_1 h^{1-\frac{2}{\tilde{p}}}\|\hat{z}\|_{W^{1,\tilde{p}}_0(\Omega)},
\end{cases} \tag{65}$$

where we have used the fact that $\tilde{p} \ge 4$ to derive the first line in (65). On the other hand, from Theorem 6.1 in [29] and arguments similar to the ones in Example 3 in [29], we also obtain that

$$\|\hat{z} - \Pi_h\hat{z}\|_{H^1_0(\Omega)} \le C_2 h^\gamma\|\hat{z}\|_{W^{1+\gamma,2}(\Omega)}; \tag{66}$$

see also the error estimates for the interpolation operator $\Pi_h : W^{1+\gamma,2}(\Omega) \to V_h$ in the proof of Lemma 3.1 in [26]. Combining (65) with (66), the triangle inequality, and the embedding $W^{1,\tilde{p}}_0(\Omega) \hookrightarrow L^\infty(\Omega) \cap W^{1,4}_0(\Omega)$ (due to $\tilde{p} \ge 4$) as well as estimates (58) and (61) gives

$$\|\hat{z} - \Pi_h\hat{z}\|_{H^1_0(\Omega)} \le C_3 h^\gamma\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)}, \quad \|\hat{z}\|_{L^\infty(\Omega)} + \|\Pi_h\hat{z}\|_{L^\infty(\Omega)} + \|\Pi_h\hat{z}\|_{W^{1,4}_0(\Omega)} \le C_4\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)}.$$

By choosing $w_h := \Pi_h\hat{z}$ in (64) and using these above estimates and the fact that $\gamma < \frac{1}{2}$, we obtain (57).

Step 2: Existence of a constant $C_{2,h_3,\rho}$ such that

$$\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)} \le C_{2,h_3,\rho}\Big(h\|v\|_{L^2(\Omega)} + \|\tilde{\varphi}-\varphi_h\|_{L^2(\Omega)}\Big) \tag{67}$$

for all $h \in (0,h_3)$, $u \in \overline{B}_{L^2(\Omega)}(\bar{u},\rho)\cap\mathcal{U}_{ad}$, and $v \in L^2(\Omega)$.

To show this, we first consider for any $u \in \overline{B}_{L^2(\Omega)}(\bar{u},\rho)\cap\mathcal{U}_{ad}$ and $h \in (0,h_3)$, the bilinear mapping $S_{u,h} : H_0^1(\Omega)\times H_0^1(\Omega) \to \mathbb{R}$ defined by

$$S_{u,h}(\omega,\psi) := \int_\Omega (b + a(y_h))\nabla\omega\cdot\nabla\psi\,\mathrm{d}x.$$

From Assumptions (A2) and (A3), we obtain

$$\underline{b}\|\tilde{\varphi}-\varphi_h\|^2_{H_0^1(\Omega)} \le S_{u,h}(\tilde{\varphi}-\varphi_h,\tilde{\varphi}-\varphi_h) = S_{u,h}(\tilde{\varphi}-\varphi_h,\tilde{\varphi}-\Pi_h\tilde{\varphi}) + S_{u,h}(\tilde{\varphi}-\varphi_h,\Pi_h\tilde{\varphi}-\varphi_h). \tag{68}$$

Moreover, the Cauchy–Schwarz inequality, the uniform boundedness of $\{y_h\}$ on $C(\overline{\Omega})$, and Assumptions (A2) and (A3) yield that

$$\begin{aligned} S_{u,h}(\tilde{\varphi}-\varphi_h,\tilde{\varphi}-\Pi_h\tilde{\varphi}) &\le C_\rho\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)}\|\tilde{\varphi}-\Pi_h\tilde{\varphi}\|_{H_0^1(\Omega)} \\ &\le C_\rho h\|\tilde{\varphi}\|_{H^2(\Omega)}\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)} \le C_\rho h\|v\|_{L^2(\Omega)}\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)}, \end{aligned} \tag{69}$$

where we have exploited the interpolation error [20] and (40) in order to obtain the last two estimates. Now using (37) and (38), we deduce from Hölder's inequality that

$$\begin{aligned} S_{u,h}(\tilde{\varphi}-\varphi_h,\Pi_h\tilde{\varphi}-\varphi_h) &= -\int_\Omega \mathbb{1}_{\{y_h\neq\bar{t}\}}a'(y_h)\nabla y_h\cdot\nabla(\tilde{\varphi}-\varphi_h)(\Pi_h\tilde{\varphi}-\varphi_h)\,\mathrm{d}x \\ &\le \big\|\mathbb{1}_{\{y_h\neq\bar{t}\}}a'(y_h)\big\|_{L^\infty(\Omega)}\|\nabla y_h\|_{L^4(\Omega)}\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)}\|\Pi_h\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)}. \end{aligned}$$

Combining this with the uniform boundedness in $C(\overline{\Omega})$ of $\{y_h\}$ and the embedding $W_0^{1,\tilde{p}}(\Omega)\hookrightarrow W_0^{1,4}(\Omega)$, we obtain from a triangle inequality that

$$S_{u,h}(\tilde{\varphi}-\varphi_h,\Pi_h\tilde{\varphi}-\varphi_h) \le C_\rho\Big(\|\bar{y}\|_{W_0^{1,\tilde{p}}(\Omega)}+\kappa_\rho\Big)\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)}\|\Pi_h\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)}$$

with κ_ρ being the constant in Theorem 3.2. The combination of a triangle inequality and the embedding $H_0^1(\Omega)\hookrightarrow L^4(\Omega)$ with Theorem 3.1.6 in [20] further implies that

$$\begin{aligned} \|\Pi_h\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)} \le \|\Pi_h\tilde{\varphi}-\tilde{\varphi}\|_{L^4(\Omega)} + \|\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)} &\le C\|\Pi_h\tilde{\varphi}-\tilde{\varphi}\|_{H_0^1(\Omega)} + \|\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)} \\ &\le Ch\|\tilde{\varphi}\|_{H^2(\Omega)} + \|\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)} \\ &\le Ch\|v\|_{L^2(\Omega)} + \|\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)}, \end{aligned}$$

where we have used (40) to obtain the last inequality. Similar to (63), we find that

$$\|\tilde{\varphi}-\varphi_h\|_{L^4(\Omega)} \le C\|\tilde{\varphi}-\varphi_h\|^{1-\frac{2}{4}}_{L^2(\Omega)}\|\tilde{\varphi}-\varphi_h\|^{\frac{2}{4}}_{H_0^1(\Omega)}.$$

We then have

$$S_{u,h}(\tilde{\varphi}-\varphi_h,\Pi_h\tilde{\varphi}-\varphi_h) \le C_\rho\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)}\Big(h\|v\|_{L^2(\Omega)} + \|\tilde{\varphi}-\varphi_h\|^{\frac{1}{2}}_{L^2(\Omega)}\|\tilde{\varphi}-\varphi_h\|^{\frac{1}{2}}_{H_0^1(\Omega)}\Big),$$

which, together with (68) and (69), yields

$$\|\tilde{\varphi}-\varphi_h\|_{H_0^1(\Omega)} \le C_\rho\Big(h\|v\|_{L^2(\Omega)} + \|\tilde{\varphi}-\varphi_h\|^{\frac{1}{2}}_{L^2(\Omega)}\|\tilde{\varphi}-\varphi_h\|^{\frac{1}{2}}_{H_0^1(\Omega)}\Big).$$

Applying Young's inequality then gives (67).

Step 3: Existence of constants $\bar{h} \in (0, h_3)$ and $\bar{\rho} := \bar{\rho}(\bar{u}) \le \min\{\rho, \hat{\rho}, \rho_{\bar{p}}\}$.

To show this, we first obtain from the definition of $\Sigma(\bar{y})$ in (14) and of Σ_r in (46) the existence of a $r_* > 0$ such that

$$\|\Sigma_r(\bar{y})\|_{L^1(\Omega)} \le \Sigma(\bar{y}) + 1 \quad \text{for all } r \in (0, r_*). \tag{70}$$

This together with (ii) in Proposition 4.3 yields

$$\|V(\bar{y}, r)\|_{L^2(\Omega)} \le C_{\bar{y}} r^{1/2} (\Sigma(\bar{y}) + 1)^{1/2} \quad \text{for all } r \in (0, r_*). \tag{71}$$

Moreover, thanks to Theorem 2.2 and the embeddings $L^2(\Omega) \hookrightarrow W^{-1,p}(\Omega)$ and $W_0^{1,p}(\Omega) \hookrightarrow C(\overline{\Omega}) \cap H_0^1(\Omega)$ for some $p > 2$, one has

$$\|y_u - \bar{y}\|_{L^\infty(\Omega)} + \|y_u - \bar{y}\|_{H_0^1(\Omega)} \le C_\rho \|u - \bar{u}\|_{L^2(\Omega)}$$

for all $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}$ and some constant C_ρ. Now Proposition 4.3 (i), (45), and the monotonic growth of $V(y, \cdot)$ imply that

$$\begin{aligned} \Big\| V\big(y_u, \|y_h - y_u\|_{L^\infty(\Omega)}\big) \Big\|_{L^2(\Omega)} &\le \Big\| V\Big(\bar{y}, \|y_h - y_u\|_{L^\infty(\Omega)} + \|y_u - \bar{y}\|_{L^\infty(\Omega)}\Big) \Big\|_{L^2(\Omega)} + C_{\bar{y},\rho} \|y_u - \bar{y}\|_{H_0^1(\Omega)} \\ &\le \Big\| V\Big(\bar{y}, C_\infty h + C_\rho \|u - \bar{u}\|_{L^2(\Omega)}\Big) \Big\|_{L^2(\Omega)} + C_{\bar{y},\rho} C_\rho \|u - \bar{u}\|_{L^2(\Omega)} \end{aligned} \tag{72}$$

for some constant $C_{\bar{y},\rho}$ and for all $h \in (0, h_3)$ and $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}$. Besides, from (48) and the monotonic growth of $V(\bar{y}, \cdot)$, there holds

$$\begin{aligned} \|Z_{y_u,\bar{y}}\|_{L^2(\Omega)} &\le L_\rho \|u - \bar{u}\|_{L^2(\Omega)} + \Big\| V\Big(\bar{y}, \|y_u - \bar{y}\|_{L^\infty(\Omega)}\Big) \Big\|_{L^2(\Omega)} \\ &\le L_\rho \|u - \bar{u}\|_{L^2(\Omega)} + \Big\| V\Big(\bar{y}, C_\rho \|u - \bar{u}\|_{L^2(\Omega)}\Big) \Big\|_{L^2(\Omega)} \end{aligned}$$

for all $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}$. We then have

$$\begin{aligned} \|Z_{y_u,\bar{y}}\|_{L^2(\Omega)} + \Big\| V\Big(y_u, \|y_h - y_u\|_{L^\infty(\Omega)}\Big) \Big\|_{L^2(\Omega)} &\le C_{1,\rho} \|u - \bar{u}\|_{L^2(\Omega)} \\ &\quad + 2 \Big\| V\Big(\bar{y}, C_\infty h + C_\rho \|u - \bar{u}\|_{L^2(\Omega)}\Big) \Big\|_{L^2(\Omega)} \end{aligned} \tag{73}$$

for all $h \in (0, h_3)$ and $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \cap \mathcal{U}_{ad}$ with $C_{1,\rho} := (C_{\bar{y},\rho} C_\rho + L_\rho)$.

Next, fix $\bar{h} \in (0, h_3)$ and $\bar{\rho} \le \min\{\rho, \hat{\rho}, \rho_{\bar{p}}\}$ such that

$$\bar{h}^\gamma + 2C_{\bar{y}} \big[C_\infty \bar{h} + C_\rho \bar{\rho}\big]^{\frac{1}{2}} (\Sigma(\bar{y}) + 1)^{\frac{1}{2}} + C_{1,\rho} \bar{\rho} \le \frac{1}{2 C_{1,h_3,\rho} C_{2,h_3,\rho}} \quad \text{and} \quad C_\infty \bar{h} + C_\rho \bar{\rho} < r_*,$$

where $C_{1,h_3,\rho}$ and $C_{2,h_3,\rho}$ are defined in (57) and (67), respectively. From this, (71), and (72), we conclude that

$$\bar{h}^\gamma + \|Z_{y_u,\bar{y}}\|_{L^2(\Omega)} + \Big\| V\Big(y_u, \|y_h - y_u\|_{L^\infty(\Omega)}\Big) \Big\|_{L^2(\Omega)} \le \frac{1}{2 C_{1,h_3,\rho} C_{2,h_3,\rho}}$$

for all $h \in (0, \bar{h}), u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\rho}) \cap \mathcal{U}_{ad}$.

The combination of this with (57) and (67) yields $\|\tilde{\varphi} - \varphi_h\|_{H_0^1(\Omega)} \le 2C_{2,h_3,\rho} h \|v\|_{L^2(\Omega)}$, and together with (57), we obtain $\|\tilde{\varphi} - \varphi_h\|_{L^2(\Omega)} \le C \varepsilon_h^u \|v\|_{L^2(\Omega)}$. Combining the last two estimates with Lemma 4.2 and the triangle inequality, we arrive at the desired conclusion. □

5. Discretization of the control problem

In this section, we discretize the control problem (P), show convergence of the discretizations, and derive error estimates of the discrete optimal solutions. In the following, we will consider three different discretizations of the control:

(i) *Variational discretization*: $\mathcal{U}_h = L^\infty(\Omega)$ (see, *e.g.*, [35]);
(ii) *Piecewise constant discretization*:

$$\mathcal{U}_h = \mathcal{U}_h^0 := \{u \in L^\infty(\Omega) \mid u_{|T} \in \mathbb{R} \quad \text{for all } T \in \mathcal{T}_h\};$$

(iii) *Continuous piecewise linear discretization*:

$$\mathcal{U}_h = \mathcal{U}_h^1 := \{u \in L^\infty(\Omega) \mid u_{|T} \in \mathcal{P}_1 \quad \text{for all } T \in \mathcal{T}_h\}.$$

Unless specified, any claim for $\mathcal{U}_h$ should be understood to hold for all three cases. For any $h > 0$, we now set $\mathcal{U}_{ad,h} := \mathcal{U}_{ad} \cap \mathcal{U}_h$. If $\mathcal{U}_h = \mathcal{U}_h^0$, then by $\mathcal{I}_h$ we denote the linear projection from $L^2(\Omega)$ onto $\mathcal{U}_h^0$. If $\mathcal{U}_h = \mathcal{U}_h^1$, then $\mathcal{I}_h : L^2(\Omega) \to \mathcal{U}_h^1$ denotes the Carstensen quasi-interpolation operator [8]. In both situations, we have $\mathcal{I}_h u \to u$ strongly in $L^2(\Omega)$ as $h \to 0$ for all $u \in L^2(\Omega)$ and $\mathcal{I}_h u \in \mathcal{U}_{ad,h}$ for all $u \in \mathcal{U}_{ad}$; see, *e.g.*, [27].

For any $h > 0$, we define the discretized optimal control problem

$$\min\{J(y_h, u_h) : u_h \in \mathcal{U}_{ad,h} \quad \text{and } y_h \in V_h \quad \text{satisfies (22) for } u := u_h\} \tag{P_h}$$

with

$$J(y, u) := \int_\Omega L(x, y(x))\,\mathrm{d}x + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)}.$$

Note that the discrete operator S_h defined in (25) does not appear in (P_h), since this operator is well-defined only locally by Theorem 3.2.

5.1. Convergence of discrete minimizers

We first have the convergence of minimizers of (P_h).

Theorem 5.1 (cf. [17], Thm. 4.1)**.** *Assume that a satisfies the growth condition*

$$|a(t)| \le C_0 + C_1|t|^m \quad \text{for all } t \in \mathbb{R} \tag{74}$$

for some positive constants C_0, C_1 and $m \ge 1$. Then there exists an $\tilde{h}_0 > 0$ such that (P_h) admits at least one global minimizer $(\bar{y}_h, \bar{u}_h)$ for all $0 < h < \tilde{h}_0$. Moreover, if $\{(\bar{y}_h, \bar{u}_h)\}_{0<h<\tilde{h}_0}$ is a sequence of solutions to (P_h), then there exists a subsequence that converges strongly in $H_0^1(\Omega) \times L^2(\Omega)$ to some $(\bar{y}, \bar{u})$ as $h \to 0$, where $\bar{u}$ is a global solution to (P).

Proof. The existence of discrete solutions to (P_h) is proven similarly to Claim 1 in the proof of Theorem 4.1 in [17]. Moreover, there exists a constant $\tilde{h}_0 > 0$ such that (P_h) admits at least one minimizer $(\bar{y}_h, \bar{u}_h)$ for all $0 < h < \tilde{h}_0$. The remainder of this proof is now divided into three steps as follows:

– *Claim 1: Weak convergence of $\{(\bar{y}_h, \bar{u}_h)\}$ to $(\bar{y}, \bar{u})$ satisfying (3).* Indeed, the boundedness of $\{\bar{u}_h\}$ in $L^\infty(\Omega)$ and thus in $L^2(\Omega)$ is due to the L^∞-boundedness of $\mathcal{U}_{ad}$. Since $(\bar{y}_h, \bar{u}_h)$ satisfies (22), there holds

$$\int_\Omega (b + a(\bar{y}_h))\nabla \bar{y}_h \cdot \nabla v_h\,\mathrm{d}x = \int_\Omega \bar{u}_h v_h\,\mathrm{d}x \qquad \text{for all } v_h \in V_h. \tag{75}$$

In particular, one has

$$\int_\Omega (b + a(\bar{y}_h))\nabla \bar{y}_h \cdot \nabla \bar{y}_h \,\mathrm{d}x = \int_\Omega \bar{y}_h \bar{u}_h \,\mathrm{d}x.$$

Combining this with Assumptions (A3) and (A2), we deduce from the Cauchy–Schwarz and Poincaré inequalities the boundedness of $\{\bar{y}_h\}$ in $H_0^1(\Omega)$. From this and the compact embedding $H_0^1(\Omega) \Subset L^p(\Omega)$ for any $p \geq 1$, we can take a subsequence, denoted in the same way, of $\{(\bar{y}_h, \bar{u}_h)\}_{0<h<\tilde{h}_0}$ that satisfies

$$(\bar{y}_h, \bar{u}_h) \rightharpoonup (\bar{y}, \bar{u}) \ \text{ in } H_0^1(\Omega) \times L^2(\Omega), \quad \bar{y}_h(x) \to \bar{y}(x) \ \text{ for a.e. } x \in \Omega, \ \text{ and } \ \bar{y}_h \to \bar{y} \ \text{ in } L^p(\Omega) \tag{76}$$

for some $(\bar{y}, \bar{u}) \in H_0^1(\Omega) \times \mathcal{U}_{ad}$ and for any $p \geq 1$, for instance, $p := 4m$ with constant m defined in (74). Thanks to the growth condition (74), we conclude from the last limit and the generalized Lebesgue Dominated Convergence Theorem that

$$a(\bar{y}_h) \to a(\bar{y}) \quad \text{strongly in } L^4(\Omega). \tag{77}$$

Now take any $v \in H_0^1(\Omega) \cap W^{2,4}(\Omega)$ and choose $v_h := \Pi_h v \in V_h$. Then $v_h \to v$ in $W^{1,4}(\Omega)$ as $h \to 0$; see, *e.g.*, [20]. Letting $h \to 0$ in the equation (75) and exploiting the first limit in (76) as well as the limit in (77) then yields

$$\int_\Omega (b + a(\bar{y}))\nabla \bar{y} \cdot \nabla v \,\mathrm{d}x = \int_\Omega \bar{u} v \,\mathrm{d}x \quad \text{for all } v \in H_0^1(\Omega) \cap H^2(\Omega),$$

which, together with the density of $H_0^1(\Omega) \cap W^{2,4}(\Omega)$ in $H_0^1(\Omega)$, implies that $\bar{y} = S(\bar{u})$.

– *Claim 2: Optimality of* $\bar{u}$. Let us show that $\bar{u}$ is a global solution of (P). For that purpose, we first observe from the inclusion $\mathcal{U}_{ad,h} \subset \mathcal{U}_{ad}$ and from the first limit in (76) that $\bar{u} \in \mathcal{U}_{ad}$. Take $u \in \mathcal{U}_{ad}$ arbitrarily and choose $u_h := u$ if $\mathcal{U}_h = L^\infty(\Omega)$ and $u_h := \mathcal{I}_h u$ if $\mathcal{U}_h = \mathcal{U}_h^i$ with $i = 0, 1$. One has $u_h \in \mathcal{U}_{ad,h}$ and $u_h \to u$ strongly in $L^2(\Omega)$ as $h \to 0$. In view of Theorem 3.1, for h small enough there exists at least one solution $y_h(u_h)$ of (22) such that $y_h(u_h) \to y_u := S(u)$ strongly in $H_0^1(\Omega) \cap C(\overline{\Omega})$. From this and the optimality of $(\bar{y}_h, \bar{u}_h)$, we have

$$\liminf_{h\to 0} J(\bar{y}_h, \bar{u}_h) \leq \liminf_{h\to 0} J(y_h(u_h), u_h) \leq \limsup_{h\to 0} J(y_h(u_h), u_h) = J(y_u, u) = j(u). \tag{78}$$

On the other hand, it follows from the limits in (76), Fatou's lemma, and the weak lower semicontinuity of the L^2-norm that

$$\liminf_{h\to 0} J(\bar{y}_h, \bar{u}_h) \geq J(\bar{y}, \bar{u}) = j(\bar{u}), \tag{79}$$

which together with (78) gives $j(\bar{u}) \leq j(u)$. Since $u \in \mathcal{U}_{ad}$ was arbitrary, $\bar{u}$ is a global optimal solution to (P).

– *Claim 3: Strong convergence in* $H_0^1(\Omega) \times L^2(\Omega)$. In fact, by plugging $u := \bar{u}$ into (78) and (79) and using the limits for $\bar{y}_h$ in (76), we can conclude that $\|\bar{u}_h\|^2_{L^2(\Omega)} \to \|\bar{u}\|^2_{L^2(\Omega)}$. Combining this with the limit for $\bar{u}_h$ in (76) yields

$$\bar{u}_h \to \bar{u} \quad \text{strongly in } L^2(\Omega). \tag{80}$$

It remains to prove the strong convergence of $\{\bar{y}_h\}$ in $H_0^1(\Omega)$. To this end, by the weak lower semicontinuity of the functional $H_0^1(\Omega) \ni y \mapsto \int_\Omega (b(x) + a(y(x)))\nabla y(x) \cdot \nabla y(x) \,\mathrm{d}x \in \mathbb{R} \cup \{\infty\}$; see, *e.g.*, Theorem 1.3 in [25], we deduce from the first limit in (76) that

$$\int_\Omega (b + a(\bar{y}))\nabla \bar{y} \cdot \nabla \bar{y} \,\mathrm{d}x \leq \liminf_{h\to 0} \int_\Omega (b + a(\bar{y}_h))\nabla \bar{y}_h \cdot \nabla \bar{y}_h \,\mathrm{d}x. \tag{81}$$

Moreover, (75) and (80) imply that

$$\liminf_{h\to 0} \int_\Omega (b + a(\bar{y}_h))\nabla \bar{y}_h \cdot \nabla \bar{y}_h \,\mathrm{d}x \leq \limsup_{h\to 0} \int_\Omega (b + a(\bar{y}_h))\nabla \bar{y}_h \cdot \nabla \bar{y}_h \,\mathrm{d}x$$

$$= \limsup_{h\to 0} \int_\Omega \bar{u}_h \bar{y}_h \,\mathrm{d}x = \int_\Omega \bar{y}\bar{u} \,\mathrm{d}x$$
$$= \int_\Omega (b + a(\bar{y}))\nabla\bar{y} \cdot \nabla\bar{y} \,\mathrm{d}x.$$

Combining this with (81), one has

$$\lim_{h\to 0} \int_\Omega (b + a(\bar{y}_h))\nabla\bar{y}_h \cdot \nabla\bar{y}_h \,\mathrm{d}x = \int_\Omega (b + a(\bar{y}))\nabla\bar{y} \cdot \nabla\bar{y} \,\mathrm{d}x. \tag{82}$$

Moreover, thanks to Assumption (A2) and by the nonnegativity of a, there hold $b(x) - \frac{b}{2} \geq \frac{b}{2} > 0$ and $\frac{b}{2} + a(\bar{y}_h(x)) \geq \frac{b}{2} > 0$ for all $x \in \Omega$. We then deduce from the limit in (80), the weak limit of $\{\bar{y}_h\}$ in $H_0^1(\Omega)$, and Theorem 1.3 in [25] that

$$\begin{aligned}
\int_\Omega \left(b(x) - \frac{b}{2}\right)|\nabla\bar{y}|^2 \,\mathrm{d}x &\leq \liminf_{h\to 0} \int_\Omega \left(b(x) - \frac{b}{2}\right)\nabla\bar{y}_h \cdot \nabla\bar{y}_h \,\mathrm{d}x \\
&\leq \limsup_{h\to 0} \int_\Omega \left(b(x) - \frac{b}{2}\right)\nabla\bar{y}_h \cdot \nabla\bar{y}_h \,\mathrm{d}x \\
&= \limsup_{h\to 0} \left[\int_\Omega \bar{u}_h\bar{y}_h \,\mathrm{d}x - \int_\Omega \left(\frac{b}{2} + a(\bar{y}_h)\right)\nabla\bar{y}_h \cdot \nabla\bar{y}_h \,\mathrm{d}x\right] \\
&= \int_\Omega \bar{u}\bar{y} \,\mathrm{d}x - \liminf_{h\to 0} \int_\Omega \left(\frac{b}{2} + a(\bar{y}_h)\right)\nabla\bar{y}_h \cdot \nabla\bar{y}_h \,\mathrm{d}x \\
&\leq \int_\Omega \bar{u}\bar{y} \,\mathrm{d}x - \int_\Omega \left(\frac{b}{2} + a(\bar{y})\right)\nabla\bar{y} \cdot \nabla\bar{y} \,\mathrm{d}x \\
&= \int_\Omega \left(b(x) - \frac{b}{2}\right)|\nabla\bar{y}|^2 \,\mathrm{d}x,
\end{aligned}$$

where we have employed the equations for $\bar{y}_h$ and $\bar{y}$ to derive the first and the last identities. We therefore have

$$\lim_{h\to 0} \int_\Omega \left(b(x) - \frac{b}{2}\right)\nabla\bar{y}_h \cdot \nabla\bar{y}_h \,\mathrm{d}x = \int_\Omega \left(b(x) - \frac{b}{2}\right)|\nabla\bar{y}|^2 \,\mathrm{d}x,$$

which, along with the weak limit, yields, the strong convergence of $\{\bar{y}_h\}$ in $H_0^1(\Omega)$. □

Next, we prove a kind of converse theorem. More precisely, we assume that $\bar{u} \in \mathcal{U}_{ad}$ is a *strict local minimum* of (P) with associated state $\bar{y}$, *i.e.*, there exists a constant $\bar{\varepsilon} > 0$ such that

$$j(\bar{u}) < j(u) \quad \text{for all } u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon}) \cap \mathcal{U}_{ad} \ \text{ with } u \neq \bar{u}.$$

We can obviously assume that $\bar{\varepsilon} < \rho$. Here ρ is defined in Theorem 3.2. We therefore can put the discrete operator S_h into (P_h). Then, for any $h \in (0, h_2)$, we consider the discretized optimal control problem defined *via*

$$\min_{u_h \in \mathcal{U}_{ad,h} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon})} j_h(u_h) \tag{$P_h^{\bar{\varepsilon}}$}$$

and the discretized cost functional given by

$$j_h : \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \to \mathbb{R}, \qquad j_h(u) := \int_\Omega L(x, (S_h(u))(x)) \,\mathrm{d}x + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)}.$$

Using Theorem 3.8 and Assumption (A4), we can show differentiability of j_h. The proof of the following result is straightforward and therefore omitted.

Theorem 5.2. *For any $h \in (0, h_2)$, the discrete cost functional $j_h : \overline{B}_{L^2(\Omega)}(\bar{u}, \rho) \to \mathbb{R}$ is of class C^1, and its derivative at $u \in B_{L^2(\Omega)}(\bar{u}, \rho)$ is given by*

$$j_h'(u)w = \int_\Omega (\varphi_h(u) + \nu u)w \,\mathrm{d}x \qquad \text{for all } w \in L^2(\Omega),$$

where $\varphi_h(u) \in V_h$ is the unique solution to (37) with $v := \frac{\partial L}{\partial y}(\cdot, S_h(u))$.

Since $\mathcal{I}_h \bar{u} \to \bar{u}$ strongly in $L^2(\Omega)$ as $h \to 0$, there exists a constant $\tilde{h} = \tilde{h}(\bar{\varepsilon}) > 0$ such that the admissible set of $(P_h^{\bar{\varepsilon}})$ is nonempty for all $0 < h < \tilde{h}$. We now provide a result on the existence of global minimizers and the associated optimality conditions of $(P_h^{\bar{\varepsilon}})$. Its proof is elementary and is thus omitted.

Theorem 5.3. *There exists a constant $h_* \in (0, h_2)$ such that for any $h \in (0, h_*)$, $(P_h^{\bar{\varepsilon}})$ admits at least one global minimizer $\bar{u}_h \in \mathcal{U}_{ad,h} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon})$. Moreover, there exists a function $\bar{\varphi}_h \in V_h$ that together with $\bar{u}_h$ and $\bar{y}_h := S_h(\bar{u}_h)$ satisfies*

$$\int_\Omega (b + a(\bar{y}_h))\nabla\bar{\varphi}_h \cdot \nabla w_h + \mathbb{1}_{\{\bar{y}_h \neq \bar{t}\}} a'(\bar{y}_h) w_h \nabla\bar{y}_h \cdot \nabla\bar{\varphi}_h \,\mathrm{d}x = \int_\Omega \frac{\partial L}{\partial y}(x, \bar{y}_h) w_h \,\mathrm{d}x, \tag{83a}$$

$$\int_\Omega (\bar{\varphi}_h + \nu\bar{u}_h)(u_h - \bar{u}_h) \,\mathrm{d}x \geq 0 \quad \text{for all } w_h \in V_h, u_h \in \mathcal{U}_{ad,h} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon}). \tag{83b}$$

We now state a convergence result in $L^2(\Omega)$, whose proof is similar to that of Theorems 4.2 and 4.3 in [26].

Theorem 5.4. *Let $\{\bar{u}_h\}$ be the sequence of discrete solutions to $(P_h^{\bar{\varepsilon}})$, defined in Theorem 5.3. Then $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} \to 0$ as $h \to 0^+$.*

Proof. Since $\{\bar{u}_h\}_{h_* > h > 0}$ is bounded in $L^2(\Omega)$, there exists a subsequence $\{h_n\}$ and an element $\tilde{u} \in \mathcal{U}_{ad} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon})$ such that $h_n \to 0^+$ and $\bar{u}_{h_n} \rightharpoonup \tilde{u}$ in $L^2(\Omega)$ as $n \to \infty$. Since $L^2(\Omega)$ is compactly embedded in $W^{-1,\tilde{p}}(\Omega)$, Theorem 2.2 implies that$S(\bar{u}_{h_n}) \to S(\tilde{u})$ in $W_0^{1,\tilde{p}}(\Omega)$. Moreover, one has $\bar{u}_{h_n} \in \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon}) \subset \overline{B}_{L^2(\Omega)}(\bar{u}, \rho)$ for all $n \geq 1$. From this and Theorems 3.1 and 3.2, there holds

$$\|S(\bar{u}_{h_n}) - S_{h_n}(\bar{u}_{h_n})\|_{W_0^{1,\tilde{p}}(\Omega)} \leq C h_n^{2/\tilde{p}} \to 0 \quad \text{as } n \to \infty.$$

This, together with a triangle inequality, yields $S_{h_n}(\bar{u}_{h_n}) \to S(\tilde{u})$ in $W_0^{1,\tilde{p}}(\Omega)$. There then holds that $S_{h_n}(\bar{u}_{h_n}) \to S(\tilde{u})$ in $C(\overline{\Omega})$, thanks to the continuous embedding $W_0^{1,\tilde{p}}(\Omega) \hookrightarrow C(\overline{\Omega})$ (due to $\tilde{p} \geq 4$). On the other hand, Theorems 3.1 and 3.2 yield $S_{h_n}(\bar{u}) \to S(\bar{u})$ in $C(\overline{\Omega})$. From the last two limits, Assumption (A4), the optimality of $\bar{u}_{h_n}$, and the weak lower semicontinuity of the norm in $L^2(\Omega)$, we obtain

$$j(\tilde{u}) \leq \liminf_{n\to\infty} j_{h_n}(\bar{u}_{h_n}) \leq \limsup_{n\to\infty} j_{h_n}(\bar{u}_{h_n}) \leq \lim_{n\to\infty} j_{h_n}(\bar{u}) = j(\bar{u}).$$

Since $\bar{u}$ is a strict local minimum of (P) and $\tilde{u} \in \mathcal{U}_{ad} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon})$, we have $\tilde{u} = \bar{u}$. A standard subsequence argument then shows the convergence of the whole sequence $\{\bar{u}_h\}_{h_* > h > 0}$ in $L^2(\Omega)$ to $\bar{u}$ as $h \to 0^+$. □

Remark 5.5. By Theorem 5.4, it holds that $\|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} \leq \frac{\bar{\varepsilon}}{2}$ for all $h \in (0, \tilde{h})$ and for some $\tilde{h} > 0$. Now for any $h \in (0, \tilde{h})$ and $u_h \in \mathcal{U}_{ad,h}$, we have that $w_h := t(u_h - \bar{u}_h) + \bar{u}_h \in \overline{B}_{L^2(\Omega)}(\bar{u}_h, \frac{\bar{\varepsilon}}{2})$ for $t > 0$ small enough and hence that $w_h \in \mathcal{U}_{ad,h} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon})$. The variational inequality (83b) then implies that

$$\int_\Omega (\bar{\varphi}_h + \nu\bar{u}_h)(u_h - \bar{u}_h) \,\mathrm{d}x \geq 0 \quad \text{for all } u_h \in \mathcal{U}_{ad,h}. \tag{84}$$

Remark 5.6. In view of Theorem 5.4, there exists a constant $h_{\bar{\varepsilon}} > 0$ such that any solution $\bar{u}_h$ of $(P_h^{\bar{\varepsilon}})$ belongs to the open ball $B_{L^2(\Omega)}(\bar{u}, \bar{\varepsilon})$. By Theorem 3.2, $(S_h(\bar{u}_h), \bar{u}_h)$ is thus a local minimizer of (P_h). We have therefore shown that any strict local solution of (P) can be approximated by local optimal controls of (P_h).

In order to show convergence in $L^\infty(\Omega)$, we first need the following lemma.

Lemma 5.7. *Let $q > 2$ be given and let $\bar{h}$ and $\bar{\rho}$ be defined in Theorem 4.5. If $\Sigma(\bar{y}) < \infty$, then for any $h \in (0, \bar{h})$ and $u, v \in \mathcal{U}_{ad}$ such that $v \in B_{L^2(\Omega)}(\bar{u}, \bar{\rho}) \cap \mathcal{U}_{ad}$, there hold*

$$\|y_u - y_h(v)\|_{L^2(\Omega)} + h\|y_u - y_h(v)\|_{H_0^1(\Omega)} \leq C\Big(h^2 + \|u - v\|_{L^2(\Omega)}\Big), \tag{85}$$

$$\|\varphi_u - \varphi_v\|_{H^2(\Omega)} \leq C\|u - v\|_{L^2(\Omega)}, \tag{86}$$

$$\|\varphi_u - \varphi_h(v)\|_{L^2(\Omega)} \leq C\Big(\varepsilon_h^v + \|u - v\|_{L^2(\Omega)}\Big), \tag{87}$$

$$\|\varphi_u - \varphi_h(v)\|_{H_0^1(\Omega)} \leq C\Big(h + \|u - v\|_{L^2(\Omega)}\Big), \tag{88}$$

$$\|\varphi_u - \varphi_h(v)\|_{L^\infty(\Omega)} \leq C\Big(h^{1-\frac{2}{q}} + \|u - v\|_{L^2(\Omega)}\Big), \tag{89}$$

for some constant C independent of u, v, and h. Here $y_u := S(u)$ and $y_h(v) := S_h(v)$, while φ_u is the unique solution to (6) for $v := \frac{\partial L}{\partial y}(\cdot, y_u)$ and $\varphi_h(v)$ is the unique solution to (37) for $y_h := S_h(v)$ and $v := \frac{\partial L}{\partial y}(\cdot, y_h(v))$.

Proof. We first observe from Theorem 2.2 that $S(\mathcal{U}_{ad})$ is bounded in $C(\overline{\Omega})$. Moreover, Theorems 2.2, 3.1, and 3.2 and a triangle inequality imply that for any $v \in B_{L^2(\Omega)}(\bar{u}, \bar{\rho}) \cap \mathcal{U}_{ad}$ and $h \in (0, \bar{h})$,

$$\|y_h(v)\|_{C(\overline{\Omega})} \leq \|y_h(v) - y_v\|_{C(\overline{\Omega})} + \|y_v - \bar{y}\|_{C(\overline{\Omega})} \leq C\Big(h + \|v - \bar{u}\|_{L^2(\Omega)}\Big) \leq C\big(\bar{h} + \bar{\rho}\big).$$

Consequently, there holds

$$\|y_u\|_{C(\overline{\Omega})}, \|y_h(v)\|_{C(\overline{\Omega})} \leq M \quad \text{for all } u \in \mathcal{U}_{ad}, v \in B_{L^2(\Omega)}(\bar{u}, \bar{\rho}) \cap \mathcal{U}_{ad}, h \in \big(0, \bar{h}\big), \tag{90}$$

and for some constant $M > 0$. Moreover, a triangle inequality in combination with (5) and (23) yields (85). For the other estimates, let $\varphi_{v,h}$ be the solution to (6) for $v := \frac{\partial L}{\partial y}(\cdot, y_h(v))$ and y_u replaced by $y_v := S(v)$. We need to show that

$$\|\varphi_{v,h} - \varphi_u\|_{H^2(\Omega)} \leq C\Big(\|u - v\|_{L^2(\Omega)} + h^2\Big) \tag{91}$$

for some constant $C > 0$ independent of u, v, and h. To this end, we subtract the equations for φ_u and $\varphi_{v,h}$ to obtain that $\varphi_{v,h} - \varphi_u \in H_0^1(\Omega)$ and

$$-\operatorname{div}[(b + a(y_v))\nabla(\varphi_{v,h} - \varphi_u)] + \mathbb{1}_{\{y_v \neq \bar{t}\}} a'(y_v)\nabla y_v \cdot \nabla(\varphi_{v,h} - \varphi_u) = g_{u,v,h} \tag{92}$$

with

$$g_{u,v,h} := -\operatorname{div}[(a(y_u) - a(y_v))\nabla\varphi_u] + Z_{y_u,y_v} \cdot \nabla\varphi_u + \frac{\partial L}{\partial y}(\cdot, y_h(v)) - \frac{\partial L}{\partial y}(\cdot, y_u).$$

Theorem 2.3 and Assumption (A4) imply that $\varphi_u \in H^2(\Omega)$. From this, the product formula, the chain rule [33], and the finiteness of the set $\{\bar{t}\}$, we deduce that

$$\operatorname{div}[(a(y_u) - a(y_v))\nabla\varphi_u] = Z_{y_u,y_v} \cdot \nabla\varphi_u + (a(y_u) - a(y_v))\Delta\varphi_u.$$

This shows that

$$g_{u,v,h} = (a(y_v) - a(y_u))\Delta\varphi_u + \frac{\partial L}{\partial y}(\cdot, y_h(v)) - \frac{\partial L}{\partial y}(\cdot, y_u) \in L^2(\Omega).$$

The standard stability estimate for the solution $\varphi_{v,h} - \varphi_u$ to (92) thus gives

$$\|\varphi_{v,h} - \varphi_u\|_{H^2(\Omega)} \leq C\|g_{u,v,h}\|_{L^2(\Omega)} \leq C\Big[\|y_u - y_v\|_{L^\infty(\Omega)} + \|y_h(v) - y_u\|_{L^2(\Omega)}\Big]$$

for some constant $C > 0$ not depending on u, v, and h, where we have employed (90), the fact that $\|\Delta\varphi_u\|_{L^2(\Omega)} \le C$ due to Theorem 2.3, and Assumptions (A3) and (A4) to derive the last estimate. From this, (85), and the fact that $\|y_u - y_v\|_{L^\infty(\Omega)} \le C\|u - v\|_{L^2(\Omega)}$ (due to (5)), we obtain (91). The estimate (86) is shown by a similar argument.

We now prove (87)–(89). According to the triangle inequality, (91), Theorem 4.5, and the boundedness in $L^2(\Omega)$ of $\{\frac{\partial L}{\partial y}(\cdot, y_h(v)) \mid v \in B_{L^2(\Omega)}(\bar u, \bar\rho) \cap \mathcal{U}_{ad}, h \le h_2\}$, we obtain (87) and (88). Finally, for (89), we first see from the continuous embedding $H_0^1(\Omega) \hookrightarrow L^q(\Omega)$, the interpolation error and inverse estimates [7] for $N = 2$ that

$$\begin{aligned}
\|\varphi_{v,h} - \varphi_h(v)\|_{L^\infty(\Omega)} &\le \|\varphi_{v,h} - \Pi_h\varphi_{v,h}\|_{L^\infty(\Omega)} + \|\Pi_h\varphi_{v,h} - \varphi_h(v)\|_{L^\infty(\Omega)}\\
&\le Ch^1\|\varphi_{v,h}\|_{H^2(\Omega)} + Ch^{-\frac{2}{q}}\|\Pi_h\varphi_{v,h} - \varphi_h(v)\|_{L^q(\Omega)}\\
&\le Ch^1\|\varphi_{v,h}\|_{H^2(\Omega)} + Ch^{-\frac{2}{q}}\|\Pi_h\varphi_{v,h} - \varphi_h(v)\|_{H_0^1(\Omega)}\\
&\le C\Big[h^1\|\varphi_{v,h}\|_{H^2(\Omega)} + h^{-\frac{2}{q}}\|\Pi_h\varphi_{v,h} - \varphi_{v,h}\|_{H_0^1(\Omega)} + h^{-\frac{2}{q}}\|\varphi_{v,h} - \varphi_h(v)\|_{H_0^1(\Omega)}\Big],
\end{aligned}$$

which together with Theorem 4.5 and the interpolation error estimate from Theorem 4.4.20 in [7] yields

$$\|\varphi_{v,h} - \varphi_h(v)\|_{L^\infty(\Omega)} \le C\left[h^{1-\frac{2}{q}}\|\varphi_{v,h}\|_{H^2(\Omega)} + h^{1-\frac{2}{q}}\left\|\frac{\partial L}{\partial y}(\cdot, y_h(v))\right\|_{L^2(\Omega)}\right].$$

Combining this with the uniform boundedness of $\varphi_{v,h}$ in $H^2(\Omega)$ and of $\frac{\partial L}{\partial y}(\cdot, y_h(v))$ in $L^2(\Omega)$ for all $v \in B_{L^2(\Omega)}(\bar u, \bar\rho) \cap \mathcal{U}_{ad}$ and $h \le h_2$, we conclude that $\|\varphi_{v,h} - \varphi_h(v)\|_{L^\infty(\Omega)} \le Ch^{1-\frac{2}{q}}$. From this, (91), and the embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$, the triangle inequality thus leads to (89). □

From Theorem 5.4, Remark 5.6, and the estimate (89) in Lemma 5.7, we obtain the desired convergence result in $L^\infty(\Omega)$. Its proof is similar to that of Theorem 5.3 in [13], Corollary 5.6 in [17], and Sections 3, 4 in [41] with some modifications, and left to Appendix C.

Theorem 5.8. *Let $\{(\bar y_h, \bar u_h)\}$ be the sequence of discrete solutions to (P_h) converging strongly to $(\bar y, \bar u)$ in $H_0^1(\Omega) \times L^2(\Omega)$ (see Thm. 5.4 and Rem. 5.6). If $\Sigma(\bar y) < \infty$, then $\|\bar u_h - \bar u\|_{L^\infty(\Omega)} \to 0$ as $h \to 0^+$.*

5.2. Error estimates for discrete minimizers

We finally turn to error estimates for discrete local minimizers $\bar u_h$ under the second-order sufficient optimality condition (19). We need the following technical lemma.

Lemma 5.9. *Let $\bar u \in \mathcal{U}_{ad}$ and $\bar y := S(\bar u)$. Assume that $\{\bar y = \bar t\}$ decomposes into finitely many connected components and that on each such connected component $\mathcal{C}$, either (17) or (18) is fulfilled. Let $\{v_n\} \subset L^2(\Omega)$, $v \in L^2(\Omega)$, $\varphi \in C^1(\overline\Omega) \cap W^{2,1}(\Omega)$, and $\{s_n\} \in c_0^+$ be arbitrary such that $\bar u + s_n v_n \in \mathcal{U}_{ad}$, $\|v_n\|_{L^2(\Omega)} = 1$ and $v_n \rightharpoonup v$ in $L^2(\Omega)$. Setting $u_n := \bar u + s_n v_n$, $y_n := S(u_n)$ for $n \ge 1$ and $w := S'(\bar u)v$, then the following assertions hold:*

(a) *If $w_n \to w$ in $C(\overline\Omega)$, then*

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{s_n}\int_\Omega Z_{y_n,\bar y}\cdot\nabla\varphi w_n\,\mathrm{d}x = &-[a_0'(\bar t) - a_1'(\bar t)]\int_{\{\bar y=\bar t\}} \mathbb{1}_{\{|\nabla\bar y|>0\}} w^2 \frac{\nabla\bar y\cdot\nabla\varphi}{|\nabla\bar y|}\,\mathrm{d}\mathcal{H}^1(x)\\
&+\int_\Omega \big[\mathbb{1}_{\{\bar y\ne\bar t\}} a''(\bar y)w^2\nabla\bar y + a'(\bar y; w)\nabla w\big]\cdot\nabla\varphi\,\mathrm{d}x.
\end{aligned}$$

(b) *If in addition* (8b) *is fulfilled for* $\bar\varphi := \varphi$, *then*

$$\liminf_{n\to\infty} \frac{j'(u_n)v_n - j'(\bar u)v_n}{s_n} = 2Q(\bar u, \bar y, \varphi; v) + \nu\Big(1 - \|v\|^2_{L^2(\Omega)}\Big).$$

Proof. We first observe from Theorem 2.2 and the embedding $L^2(\Omega) \hookrightarrow W^{-1,\tilde p}(\Omega)$ that

$$\frac{y_n - \bar y}{s_n} \to w = S'(\bar u)v \quad \text{in } W_0^{1,\tilde p}(\Omega) \hookrightarrow C(\overline\Omega). \tag{93}$$

Moreover, by Theorem 2.2 and the compact embedding $W^{2,p}(\Omega) \Subset C^1(\overline\Omega)$ for some $p > 2$, we deduce from the boundedness in $L^\infty(\Omega)$ of $\mathcal{U}_{ad}$ that

$$y_n \to \bar y \quad \text{strongly in} \quad C^1(\overline\Omega). \tag{94}$$

Ad (a): We first deduce from (5), for $n \in \mathbb{N}$ large enough, that

$$\tau_n := \|y_n - \bar y\|_{C(\overline\Omega)} \le C s_n < \delta, \tag{95}$$

where δ is the constant defined in (10). Moreover, there exists a constant $M > 0$ such that $\|y_n\|_{C(\overline\Omega)}, \|\bar y\|_{C(\overline\Omega)} \le M$ for all $n \ge 1$. According to Lemma 3.4, we have

$$Z_{y_n,\bar y} = Z^{(1)}_{y_n,\bar y} + Z^{(2)}_{y_n,\bar y} + Z^{(3)}_{y_n,\bar y} + Z^{(4)}_{y_n,\bar y} \tag{96}$$

with $Z^{(k)}_{y_n,\bar y}$, $k = 1,2,3,4$, defined in Lemma 3.4. For $Z^{(1)}_{y_n,\bar y}$, using (95) yields

$$\begin{aligned} Z^{(1)}_{y_n,\bar y} &= \mathbb{1}_{\{\bar y\in(-\infty,\bar t),\, y_n\in(-\infty,\bar t)\}}[a_0'(y_n)\nabla y_n - a_0'(\bar y)\nabla\bar y] + \mathbb{1}_{\{\bar y\in(\bar t,\infty),\, y_n\in(\bar t,\infty)\}}[a_1'(y_n)\nabla y_n - a_1'(\bar y)\nabla\bar y] \\ &= \big[\mathbb{1}_{\{\bar y\in(\infty,\bar t)\}} - \mathbb{1}_{\{\bar y\in(-\infty,\bar t),\, y_n\in[\bar t,\bar t+\delta)\}}\big][(a_0'(y_n) - a_0'(\bar y))\nabla y_n + a_0'(\bar y)\nabla(y_n - \bar y)] \\ &\quad + \big[\mathbb{1}_{\{\bar y\in(\bar t,\infty)\}} - \mathbb{1}_{\{\bar y\in(\bar t,\infty),\, y_n\in(\bar t-\delta,\bar t]\}}\big][(a_1'(y_n) - a_1'(\bar y))\nabla y_n + a_1'(\bar y)\nabla(y_n - \bar y)]. \end{aligned}$$

Since

$$\mathbb{1}_{\{\bar y\in(-\infty,\bar t),\, y_n\in[\bar t,\bar t+\delta)\}}, \mathbb{1}_{\{\bar y\in(\bar t,\infty),\, y_n\in(\bar t-\delta,\bar t]\}} \le \mathbb{1}_{\{0<|\bar y-\bar t|\le\tau_n\}} \to 0$$

a.e. in Ω, we have from (93), (94), and the Lebesgue dominated convergence theorem that

$$\begin{aligned} &\frac{1}{s_n}\int_\Omega Z^{(1)}_{y_n,\bar y}\cdot\nabla\varphi w_n\,\mathrm{d}x \to \int_\Omega \mathbb{1}_{\bar y\in(-\infty,\bar t)}[a_0''(\bar y)w\nabla\bar y + a_0'(\bar y)\nabla w]\cdot\nabla\varphi w\,\mathrm{d}x \\ &\quad + \int_\Omega \mathbb{1}_{\bar y\in(\bar t,\infty)}[a_1''(\bar y)w\nabla\bar y + a_1'(\bar y)\nabla w]\cdot\nabla\varphi w\,\mathrm{d}x = \int_\Omega \mathbb{1}_{\{\bar y\ne\bar t\}}\big[a''(\bar y)w^2\nabla\bar y + a'(\bar y)w\nabla w\big]\cdot\nabla\varphi\,\mathrm{d}x. \end{aligned} \tag{97}$$

For $Z^{(2)}_{y_n,\bar y}$, we see from (95) and the fact $\nabla\bar y = 0$ a.e. on $\{\bar y = \bar t\}$ that

$$Z^{(2)}_{y_n,\bar y} = \mathbb{1}_{\{\bar y=\bar t,\, y_n\in(\bar t-\delta,\bar t)\}}a_0'(y_n)\nabla(y_n - \bar y) + \mathbb{1}_{\{\bar y=\bar t,\, y_n\in(\bar t,\bar t+\delta)\}}a_1'(y_n)\nabla(y_n - \bar y).$$

Setting $\hat w_n := \frac{y_n-\bar y}{s_n}$ and exploiting (93) yields $w_n - \hat w_n \to w - w = 0$ in $C(\overline\Omega)$. From this, the limit $y_n \to \bar y$ in $C(\overline\Omega)$, and the continuity of a_0' and a_1', the dominated convergence theorem implies that

$$\lim_{n\to\infty}\frac{1}{s_n}\int_\Omega Z^{(2)}_{y_n,\bar y}\cdot\nabla\varphi w_n\,\mathrm{d}x = \lim_{n\to\infty}\int_\Omega\big[\mathbb{1}_{\{\bar y=\bar t,\, y_n\in(\bar t-\delta,\bar t)\}}a_0'(\bar t)\hat w_n + \mathbb{1}_{\{\bar y=\bar t,\, y_n\in(\bar t,\bar t+\delta)\}}a_1'(\bar t)\hat w_n\big]\nabla\hat w_n\cdot\nabla\varphi\,\mathrm{d}x.$$

As a result of (2) and the fact that $\hat w_n < 0$ on $\{\bar y = \bar t, y_n \in (\bar t - \delta, \bar t)\}$, there holds

$$\mathbb{1}_{\{\bar y=\bar t,\, y_n\in(\bar t-\delta,\bar t)\}}a_0'(\bar t)\hat w_n = \mathbb{1}_{\{\bar y=\bar t,\, y_n\in(\bar t-\delta,\bar t)\}}a'(\bar t;\hat w_n).$$

Similarly, one has

$$\mathbb{1}_{\{\bar{y}=\bar{t},y_n\in(\bar{t},\bar{t}+\delta)\}}a_1'(\bar{t})\hat{w}_n=\mathbb{1}_{\{\bar{y}=\bar{t},y_n\in(\bar{t},\bar{t}+\delta)\}}a'(\bar{t};\hat{w}_n).$$

We thus have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{s_n}\int_\Omega Z^{(2)}_{y_n,\bar{y}}\cdot\nabla\varphi w_n\,\mathrm{d}x&=\lim_{n\to\infty}\int_\Omega\mathbb{1}_{\{\bar{y}=\bar{t},y_n\in(\bar{t}-\delta,\bar{t})\cup(\bar{t},\bar{t}+\delta)\}}a'(\bar{t};\hat{w}_n)\nabla\hat{w}_n\cdot\nabla\varphi\,\mathrm{d}x\\&=\lim_{n\to\infty}\int_\Omega\mathbb{1}_{\{\bar{y}=\bar{t}\}}a'(\bar{t};\hat{w}_n)\nabla\hat{w}_n\cdot\nabla\varphi\,\mathrm{d}x,\end{aligned}$$

where we have used (95) and the fact that $\nabla y_n=\nabla\bar{y}=0$ and so $\nabla\hat{w}_n=0$ a.e. on $\{y_n=\bar{y}=\bar{t}\}$ to obtain the last identity. We thus conclude from the continuity of $a'(\bar{t};\cdot)$ due to Proposition 2.49 in [6], (93), and the dominated convergence theorem that

$$\lim_{n\to\infty}\frac{1}{s_n}\int_\Omega Z^{(2)}_{y_n,\bar{y}}\cdot\nabla\varphi w_n\,\mathrm{d}x=\int_\Omega\mathbb{1}_{\{\bar{y}=\bar{t}\}}a'(\bar{y};w)\nabla w\cdot\nabla\varphi\,\mathrm{d}x.\tag{98}$$

For $Z^{(4)}_{y_n,\bar{y}}$, we have from (29) that

$$\left|Z^{(4)}_{y_n,\bar{y}}\right|\le C_M[|y_n-\bar{y}||\nabla\bar{y}|+|\nabla(y_n-\bar{y})|]\left(\mathbb{1}_{\Omega^2_{y_n,\bar{y}}}+\mathbb{1}_{\Omega^3_{y_n,\bar{y}}}\right)\quad\text{a.e. in }\Omega.$$

This, together with the fact that $\mathbb{1}_{\Omega^2_{y_n,\bar{y}}}+\mathbb{1}_{\Omega^3_{y_n,\bar{y}}}\to 0$ a.e. in Ω as well as (93), yields

$$\lim_{n\to\infty}\frac{1}{s_n}\int_\Omega Z^{(4)}_{y_n,\bar{y}}\cdot\nabla\varphi w_n\,\mathrm{d}x=0.\tag{99}$$

It remains to estimate $Z^{(3)}_{y_n,\bar{y}}$. To this end, we first deduce from (52) and the coarea formula for Lipschitz mappings (see, *e.g.*, [30], Thm. 2, p. 117) or [1], Sect. 2.7) that

$$\begin{aligned}s_n^{-1}\left|\int_\Omega\mathbb{1}_{\Omega^2_{y_n,\bar{y}}}\nabla\bar{y}\cdot\nabla\varphi(w_n-\hat{w}_n)\,\mathrm{d}x\right|&\le s_n^{-1}\|w_n-\hat{w}_n\|_{L^\infty(\Omega)}\|\nabla\varphi\|_{L^\infty(\Omega)}\int_\Omega\mathbb{1}_{\{0<\bar{y}-\bar{t}\le\tau_n\}}|\nabla\bar{y}|\,\mathrm{d}x\\&=s_n^{-1}\|w_n-\hat{w}_n\|_{L^\infty(\Omega)}\|\nabla\varphi\|_{L^\infty(\Omega)}\int_{\bar{t}}^{\bar{t}+\tau_n}\int_{\{\bar{y}=t\}}\mathrm{d}\mathcal{H}^1(x)\,\mathrm{d}t\\&\le Cs_n^{-1}\tau_n\|w_n-\hat{w}_n\|_{L^\infty(\Omega)}\|\nabla\varphi\|_{L^\infty(\Omega)}\to 0,\end{aligned}$$

where we have used (17) and (18) to derive the last estimate and exploited (95) and the fact that $w_n-\hat{w}_n\to 0$ in $C(\overline{\Omega})$ to obtain the last limit. Similarly, $s_n^{-1}|\int_\Omega\mathbb{1}_{\Omega^3_{y_n,\bar{y}}}\nabla\bar{y}\cdot\nabla\varphi(w_n-\hat{w}_n)\,\mathrm{d}x|\to 0$. From these limits and the definition of $Z^{(3)}_{y_n,\bar{y}}$, we deduce that

$$s_n^{-1}\int_\Omega\left|Z^{(3)}_{y_n,\bar{y}}\cdot\nabla\varphi(w_n-\hat{w}_n)\right|\mathrm{d}x\to 0$$

and thus

$$\lim_{n\to\infty}\frac{1}{s_n}\int_\Omega Z^{(3)}_{y_n,\bar{y}}\cdot\nabla\varphi w_n\,\mathrm{d}x=\lim_{n\to\infty}\frac{1}{s_n}\int_\Omega Z^{(3)}_{y_n,\bar{y}}\cdot\nabla\varphi\hat{w}_n\,\mathrm{d}x=\lim_{n\to\infty}\frac{1}{s_n^2}\int_\Omega Z^{(3)}_{y_n,\bar{y}}\cdot\nabla\varphi(y_n-\bar{y})\,\mathrm{d}x\tag{100}$$

provided that one of these three limits exists. For $\zeta_i(\bar{u},\bar{y};s_n,v_n)$, $i=0,1$, defined in (11), one has

$$\begin{aligned}P_n&:=Z^{(3)}_{y_n,\bar{y}}(y_n-\bar{y})+2\sum_{i=0}^1\zeta_i(\bar{u},\bar{y};s_n,v_n)\nabla\bar{y}\\&=[a_0'(\bar{t})-a_1'(\bar{t})]\left[\mathbb{1}_{\Omega^2_{y_n,\bar{y}}}-\mathbb{1}_{\Omega^3_{y_n,\bar{y}}}\right]\nabla\bar{y}(y_n-\bar{y})+2[a_0'(\bar{t})-a_1'(\bar{t})](\bar{t}-y_n)\left[\mathbb{1}_{\Omega^2_{y_n,\bar{y}}}-\mathbb{1}_{\Omega^3_{y_n,\bar{y}}}\right]\nabla\bar{y}\\&=[a_0'(\bar{t})-a_1'(\bar{t})](2\bar{t}-\bar{y}-y_n)\left[\mathbb{1}_{\Omega^2_{y_n,\bar{y}}}-\mathbb{1}_{\Omega^3_{y_n,\bar{y}}}\right]\nabla\bar{y}.\end{aligned}$$

By (21) and (94), we deduce that

$$\lim_{n\to\infty} \frac{1}{s_n^2} \int_\Omega P_n \cdot \nabla\varphi \,\mathrm{d}x = 0.$$

Combining this with (100) and (16), we can conclude that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{s_n} \int_\Omega Z^{(3)}_{y_n,\bar y} \cdot \nabla\varphi w_n \,\mathrm{d}x &= \limsup_{n\to\infty} \left[\frac{1}{s_n^2} \int_\Omega P_n \cdot \nabla\varphi \,\mathrm{d}x - \frac{2}{s_n^2} \int_\Omega \sum_{i=0}^{1} \zeta_i(\bar u, \bar y; s_n, v_n) \nabla\bar y \cdot \nabla\varphi \,\mathrm{d}x \right] \\ &= -2\tilde Q(\bar u, \bar y, \varphi; \{s_n\}, v). \end{aligned} \tag{101}$$

Together with (96)–(99), we obtain assertion (a) from (20).

Ad (b): Defining the functional $G : L^\infty(\Omega) \to \mathbb{R}$ *via* $G(y) := \int_\Omega L(x, y(x)) \,\mathrm{d}x$ and employing Assumption (A4), we deduce that G is of class C^2 and that its derivatives are given by

$$G'(y)y_1 := \int_\Omega \frac{\partial L}{\partial y}(x, y(x)) y_1(x) \,\mathrm{d}x \quad \text{and} \quad G''(y)y_1y_2 := \int_\Omega \frac{\partial^2 L}{\partial y^2}(x, y(x)) y_1(x) y_2(x) \,\mathrm{d}x$$

for all $y, y_1, y_2 \in L^\infty(\Omega)$. We see from the chain rule that for any $u, v \in L^2(\Omega)$,

$$j'(u)v = G'(S(u))S'(u)v + \nu \int_\Omega uv \,\mathrm{d}x.$$

This, together with a Taylor expansion and the fact that $\|v_n\|_{L^2(\Omega)} = 1$, yields

$$\begin{aligned}\frac{1}{s_n}[j'(u_n)v_n - j'(\bar u)v_n] &= \frac{1}{s_n}[G'(y_n)S'(u_n)v_n - G'(\bar y)S'(\bar u)v_n] + \nu \int_\Omega v_n^2 \,\mathrm{d}x \\ &= \frac{1}{s_n}[G'(y_n) - G'(\bar y)]S'(\bar u)v_n + \frac{1}{s_n} G'(y_n)[S'(u_n)v_n - S'(\bar u)v_n] + \nu \\ &= \frac{1}{s_n} \int_0^1 G''(\bar y + s(y_n - \bar y))(y_n - \bar y) S'(\bar u)v_n \,\mathrm{d}s + \frac{1}{s_n} G'(\bar y)[S'(u_n)v_n - S'(\bar u)v_n] \\ &\quad + \frac{1}{s_n}[G'(y_n) - G'(\bar y)][S'(u_n)v_n - S'(\bar u)v_n] + \nu. \end{aligned} \tag{102}$$

Obviously, the third term on the right-hand side of (102) tends to 0 since $v_n \to v$ in $W^{-1,\tilde p}(\Omega)$, S' is continuous, and G is of class C^2. Moreover, it follows from Assumption (A4), (93), and the dominated convergence theorem that the first term on the right-hand side of (102) tends to $G''(\bar y)(S'(\bar u)v)^2$. It remains to estimate the limes inferior of the second term on the right-hand side of (102). Subtracting the equations for $z_n^{(1)} := S'(u_n)v_n$ and $z_n^{(2)} := S'(\bar u)v_n$, we find that $z_n := z_n^{(1)} - z_n^{(2)} \in H_0^1(\Omega)$ satisfies

$$-\operatorname{div}\big[(b + a(\bar y))\nabla z_n + \mathbb{1}_{\{\bar y \neq \bar t\}} a'(\bar y)\nabla\bar y z_n\big] = \operatorname{div}\Big[(a(y_n) - a(\bar y))\nabla z_n^{(1)} + Z_{y_n,\bar y} z_n^{(1)}\Big] =: g_n. \tag{103}$$

We then have $z_n = S'(\bar u)g_n$, which together with (8b) yields

$$G'(\bar y)\Big[z_n^{(1)} - z_n^{(2)}\Big] = \langle G'(\bar y), z_n\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = \langle g_n, \varphi\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = -(B_n + C_n) \tag{104}$$

for $B_n := \int_\Omega (a(y_n) - a(\bar y))\nabla z_n^{(1)} \cdot \nabla\varphi \,\mathrm{d}x$ and $C_n := \int_\Omega Z_{y_n,\bar y} \cdot \nabla\varphi z_n^{(1)} \,\mathrm{d}x$. As a result of Theorem 2.2 and the fact that $v_n \to v$ in $W^{-1,\tilde p}(\Omega)$, there holds $z_n^{(1)} \to S'(\bar u)v$ in $W_0^{1,\tilde p}(\Omega)$. Besides, from (93) and Lemma 3.5 in [23], we have

$$\frac{a(y_n(x)) - a(\bar y(x))}{s_n} \to a'(\bar y(x); (S'(\bar u)v)(x)) \quad \text{for all } x \in \overline{\Omega}.$$

The dominated convergence theorem thus implies that

$$\frac{1}{s_n}B_n \to \int_\Omega a'(\bar y; S'(\bar u)v)\nabla(S'(\bar u)v)\cdot\nabla\varphi\,\mathrm{d}x.$$

This, along with (104) and assertion (a), ensures that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{s_n}G'(\bar y)[S'(u_n)v_n - S'(\bar u)v_n] &= [a_0'(\bar t) - a_1'(\bar t)]\int_{\{\bar y=\bar t\}}\mathbb{1}_{\{|\nabla\bar y|>0\}}(S'(\bar u)v)^2\frac{\nabla\bar y\cdot\nabla\varphi}{|\nabla\bar y|}\,\mathrm{d}\mathcal{H}^1(x)\\ &\quad - 2\int_\Omega a'(\bar y; S'(\bar u)v)\nabla(S'(\bar u)v)\cdot\nabla\varphi\,\mathrm{d}x - \int_\Omega \mathbb{1}_{\{\bar y\neq\bar t\}}a''(\bar y)(S'(\bar u)v)^2\nabla\bar y\cdot\nabla\varphi\,\mathrm{d}x.\end{aligned}$$

Using these limits, (102), and the identity for Q in (19) (see also [24], Thm. 3.22), we arrive at (b). □

The following theorem is one of main results of the paper, which extends Theorem 2.14 in [14] (see also, [17], Lem. 5.2) to the case where the cost functional j is of class C^1 but not necessarily C^2.

Theorem 5.10. *Let $\{(\bar y_h, \bar u_h)\}$ be the sequence of discrete solutions to (P_h) converging strongly to $(\bar y, \bar u)$ in $H_0^1(\Omega)\times L^2(\Omega)$. Assume that $\{\bar y = \bar t\}$ decomposes into finitely many connected components and that on each such connected component $\mathcal{C}$, either (17) or (18) is fulfilled. Assume further that the second-order sufficient condition (19) is fulfilled. Then there exist constants $C>0$ and $\hat h\in(0,\min\{\bar h, h_*\})$ such that*

$$\|\bar u_h - \bar u\|^2_{L^2(\Omega)} \le C\Big[\big(\varepsilon_h^{\bar u_h}\big)^2 + \|\bar u - u_h\|^2_{L^2(\Omega)} + j'(\bar u)(u_h - \bar u)\Big] \tag{105}$$

for all $h\in(0,\hat h)$ and $u_h\in\mathcal{U}_{ad,h}\cap\overline{B}_{L^2(\Omega)}(\bar u,\hat\varepsilon)$ with $\hat\varepsilon := \min\{\bar\varepsilon,\bar\rho\}$. Here ε_h^u is defined as in (56).

Proof. We first observe from Proposition 2.9 that $\Sigma(\bar y)<\infty$. For simplicity of notation, we set $\varepsilon_h := \varepsilon_h^{\bar u_h}$. We first show that

$$\begin{aligned}[j'(\bar u_h) - j'(\bar u)](\bar u_h - \bar u) &\le j'(\bar u)(u_h - \bar u)\\ &\quad + C\Big[\varepsilon_h\|\bar u_h - \bar u\|_{L^2(\Omega)} + \varepsilon_h\|u_h - \bar u\|_{L^2(\Omega)} + \|u_h - \bar u\|_{L^2(\Omega)}\|\bar u_h - \bar u\|_{L^2(\Omega)}\Big]\end{aligned} \tag{106}$$

for some constant $C>0$, for all $u_h\in\mathcal{U}_{ad,h}\cap\overline{B}_{L^2(\Omega)}(\bar u,\hat\varepsilon)$ and $h\in(0,\min\{\bar h,h_*\})$. To this end, let us take any $h\in(0,\min\{\bar h,h_*\})$, $u\in\mathcal{U}_{ad}\cap\overline{B}_{L^2(\Omega)}(\bar u,\hat\varepsilon)$, and $u_h\in\mathcal{U}_{ad,h}\cap\overline{B}_{L^2(\Omega)}(\bar u,\hat\varepsilon)$. We deduce from (7), Theorem 5.2, Lemma 5.7, and the Cauchy–Schwarz inequality that

$$|[j_h'(u) - j'(u)](u_h - \bar u)| = \left|\int_\Omega(\varphi_h(u) - \varphi_u)(u_h - \bar u)\,\mathrm{d}x\right| \le C\varepsilon_h^u\|u_h - \bar u\|_{L^2(\Omega)}. \tag{107}$$

Moreover, we deduce from $j_h'(\bar u_h)(\bar u_h - u_h)\le 0$ and $j'(\bar u)(\bar u - \bar u_h)\le 0$ that

$$\begin{aligned}&[j'(\bar u_h) - j'(\bar u)](\bar u_h - \bar u)\\ &\quad= [j_h'(\bar u_h) - j'(\bar u_h)](\bar u - \bar u_h) + [j_h'(\bar u_h) - j'(\bar u)](u_h - \bar u) + j_h'(\bar u_h)(\bar u_h - u_h) + j'(\bar u)(u_h - \bar u_h)\\ &\quad\le [j_h'(\bar u_h) - j'(\bar u_h)](\bar u - \bar u_h) + [j_h'(\bar u_h) - j'(\bar u)](u_h - \bar u) + j'(\bar u)[(u_h - \bar u) + (\bar u - \bar u_h)]\\ &\quad\le [j_h'(\bar u_h) - j'(\bar u_h)](\bar u - \bar u_h) + [(j_h'(\bar u_h) - j'(\bar u_h)) + (j'(\bar u_h) - j'(\bar u))](u_h - \bar u) + j'(\bar u)(u_h - \bar u).\end{aligned} \tag{108}$$

Applying (107) yields that

$$\begin{cases}|[j_h'(\bar u_h) - j'(\bar u_h)](\bar u - \bar u_h)| \le C\varepsilon_h\|\bar u - \bar u_h\|_{L^2(\Omega)},\\ |[j_h'(\bar u_h) - j'(\bar u_h)](u_h - \bar u)| \le C\varepsilon_h\|\bar u - u_h\|_{L^2(\Omega)}.\end{cases} \tag{109}$$

Using (7), (86), $H^2(\Omega) \hookrightarrow L^2(\Omega)$, and the Cauchy–Schwarz inequality yields

$$|[j'(\bar{u}_h) - j'(\bar{u})](u_h - \bar{u})| = \left| \int_\Omega [(\varphi_{\bar{u}_h} - \bar{\varphi}) + \nu(\bar{u}_h - \bar{u})](u_h - \bar{u}) \, \mathrm{d}x \right| \leq C \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} \|u_h - \bar{u}\|_{L^2(\Omega)}. \quad (110)$$

From this, (108), and (109), we derive (106).

We now prove the conclusion of the theorem by contradiction. To that purpose, we suppose that there exist $h_n \to 0^+$ and $u_{h_n} \in \mathcal{U}_{ad,h_n} \cap \overline{B}_{L^2(\Omega)}(\bar{u}, \hat{\varepsilon})$ such that

$$\|\bar{u}_{h_n} - \bar{u}\|^2_{L^2(\Omega)} > n\Big[(\varepsilon_{h_n})^2 + \|\bar{u} - u_{h_n}\|^2_{L^2(\Omega)} + j'(\bar{u})(u_{h_n} - \bar{u})\Big] \quad \text{for all } n \geq 1$$

or, equivalently, with $s_n := \|\bar{u}_{h_n} - \bar{u}\|_{L^2(\Omega)}$ that

$$\frac{1}{n} > \frac{(\varepsilon_{h_n})^2}{s_n^2} + \frac{\|\bar{u} - u_{h_n}\|^2_{L^2(\Omega)}}{s_n^2} + \frac{j'(\bar{u})(u_{h_n} - \bar{u})}{s_n^2} \quad \text{for all } n \geq 1. \quad (111)$$

By setting $v_n := \frac{\bar{u}_{h_n} - \bar{u}}{s_n}$, and by extracting a subsequence if necessary, we have

$$s_n \to 0^+, \quad \|v_n\|_{L^2(\Omega)} = 1, \quad \bar{u}_{h_n} = \bar{u} + s_n v_n, \quad v_n \rightharpoonup v \text{ in } L^2(\Omega) \quad \text{for some } v \in L^2(\Omega).$$

We first show that v is an element of the critical cone $\mathcal{C}(\mathcal{U}_{ad}; \bar{u})$ defined in (9). To this end, we first deduce that $v \geq 0$ a.e. on $\{\bar{u} = \alpha\}$ and $v \leq 0$ a.e. on $\{\bar{u} = \beta\}$. Moreover, since $j'(\bar{u})v_n \geq 0$, there holds $j'(\bar{u})v \geq 0$. On the other hand, from (109) and (110) for $u_h := \bar{u}_{h_n}$, we obtain that

$$\limsup_{n\to\infty} \big\{ [j'(\bar{u}_{h_n}) - j'_{h_n}(\bar{u}_{h_n})]v_n + [j'(\bar{u}) - j'(\bar{u}_{h_n})]v_n \big\} \leq \lim_{n\to\infty} C(\varepsilon_{h_n} + s_n) = 0,$$

which yields

$$\begin{aligned} j'(\bar{u})v = \lim_{n\to\infty} j'(\bar{u})v_n &= \lim_{n\to\infty} [j'_{h_n}(\bar{u}_{h_n})v_n + (j'(\bar{u}_{h_n}) - j'_{h_n}(\bar{u}_{h_n}))v_n + (j'(\bar{u}) - j'(\bar{u}_{h_n}))v_n] \\ &\leq \lim_{n\to\infty} j'_{h_n}(\bar{u}_{h_n})v_n = \lim_{n\to\infty} \frac{1}{s_n}[j'_{h_n}(\bar{u}_{h_n})(u_{h_n} - \bar{u}) + j'_{h_n}(\bar{u}_{h_n})(\bar{u}_{h_n} - u_{h_n})]. \end{aligned}$$

From this and the fact that $j'_{h_n}(\bar{u}_{h_n})(\bar{u}_{h_n} - u_{h_n}) \leq 0$, we obtain

$$j'(\bar{u})v \leq \lim_{n\to\infty} \frac{1}{s_n} j'_{h_n}(\bar{u}_{h_n})(u_{h_n} - \bar{u}) \leq \lim_{n\to\infty} \|\varphi_{h_n}(\bar{u}_{h_n}) + \nu \bar{u}_{h_n}\|_{L^2(\Omega)} \frac{\|u_{h_n} - \bar{u}\|_{L^2(\Omega)}}{s_n} \to 0,$$

where we have used Theorem 5.2 and the Cauchy–Schwarz inequality to derive the last estimate and the boundedness of $\{\|\varphi_{h_n}(\bar{u}_{h_n}) + \nu \bar{u}_{h_n}\|_{L^2(\Omega)}\}$ (due to Lem. 5.7) as well as (111) to pass to the limit. There therefore holds that $j'(\bar{u})v = 0$. This and Lemma 4.11 in [3] lead to $v(x) = 0$ whenever $\bar{\varphi}(x) + \nu\bar{u}(x) \neq 0$. We thus have $v \in \mathcal{C}(\mathcal{U}_{ad}; \bar{u})$.

We now derive a contradiction and thus complete the proof. To this end, we divide (106) (with $h := h_n$) by s_n^2 to obtain

$$\frac{1}{s_n}[j'(\bar{u}_{h_n}) - j'(\bar{u})]v_n \leq \frac{j'(\bar{u})(u_{h_n} - \bar{u})}{s_n^2} + C\left(\frac{\varepsilon_{h_n}}{s_n} + \frac{\varepsilon_{h_n}}{s_n} \frac{\|u_{h_n} - \bar{u}\|_{L^2(\Omega)}}{s_n} + \frac{\|u_{h_n} - \bar{u}\|_{L^2(\Omega)}}{s_n} \right).$$

Taking the limes inferior as $n \to \infty$, employing (111), and using Lemma 5.9 (ii), we conclude that

$$2Q(\bar{u}, \bar{y}, \bar{\varphi}; v) + \nu\Big(1 - \|v\|^2_{L^2(\Omega)}\Big) \leq 0. \quad (112)$$

Combining this with (19) and the fact that $\|v\|_{L^2(\Omega)} \leq \liminf_{n\to\infty} \|v_n\|_{L^2(\Omega)} = 1$, we have $v = 0$. Inserting this into (112) leads to $0 < \nu \leq 0$, which is the desired contradiction. □

Theorem 5.11 (Variational discretization). *Assume that $\mathcal{U}_h = L^\infty(\Omega)$. Let $\{(\bar y_h, \bar u_h)\}$ be the sequence of discrete solutions to (P_h) converging strongly to $(\bar y, \bar u)$ in $H_0^1(\Omega) \times L^2(\Omega)$. Under all assumptions of Theorem 5.10, there exists a constant $C > 0$ such that for any $h \in (0, \hat h)$,*

$$\|\bar u_h - \bar u\|_{L^2(\Omega)} \le C\Big(h^{1+\gamma} + \sigma_0^{\frac12}\|\Sigma_{\kappa_h}(\bar y)\|_{L^1(\Omega)}^{\frac12} h^{\frac32}\Big) \le Ch^{1+\gamma} \tag{113}$$

with $\kappa_h := C_\infty h + \|S(\bar u_h) - \bar y\|_{L^\infty(\Omega)}$, and σ_0, $\gamma \in (0, \frac12)$, and Σ_{κ_h} defined in (15), (54), and (46), respectively.

Proof. We first observe from Proposition 2.9 that $\Sigma(\bar y) < \infty$. By (5), there holds $\kappa_h \le C_\infty h + C\|\bar u_h - \bar u\|_{L^2(\Omega)} \to 0$ as $h \to 0^+$. This, in combination with (70), yields

$$\|\Sigma_{\kappa_h}(\bar y)\|_{L^1(\Omega)} \le \Sigma(\bar y) + 1 \quad \text{for all } h \text{ small enough.} \tag{114}$$

Choosing now $u_h := \bar u$ in (105) yields

$$\|\bar u_h - \bar u\|_{L^2(\Omega)} \le C\varepsilon_h^{\bar u_h} \quad \text{for all } h \in (0, \hat h). \tag{115}$$

Setting $r_h := \|S_h(\bar u_h) - S(\bar u_h)\|_{L^\infty(\Omega)}$ and using (45) yield $r_h \le C_\infty h$. Exploiting (56), the Cauchy–Schwarz inequality, Proposition 4.3, the estimate (48), and the monotonic growth of $V(\bar y, \cdot)$, there holds

$$\begin{aligned}
\big(\varepsilon_h^{\bar u_h}\big)^2 &\le 3h^{2+2\gamma} + 3h^2\|V(S(\bar u_h), r_h)\|_{L^2(\Omega)}^2 + 3h^2\big\|Z_{S(\bar u_h),\bar y}\big\|_{L^2(\Omega)}^2 \\
&\le 3h^{2+2\gamma} + Ch^2\left[\Big\|V\Big(\bar y, r_h + \|S(\bar u_h) - \bar y\|_{L^\infty(\Omega)}\Big)\Big\|_{L^2(\Omega)}^2 + \|S(\bar u_h) - \bar y\|_{H_0^1(\Omega)}^2 + \|\bar u_h - \bar u\|_{L^2(\Omega)}^2\right] \\
&\le 3h^{2+2\gamma} + Ch^2\left[\Big\|V\Big(\bar y, C_\infty h + \|S(\bar u_h) - \bar y\|_{L^\infty(\Omega)}\Big)\Big\|_{L^2(\Omega)}^2 + \|S(\bar u_h) - \bar y\|_{H_0^1(\Omega)}^2 + \|\bar u_h - \bar u\|_{L^2(\Omega)}^2\right] \\
&\le C\Big[h^{2+2\gamma} + h^2\sigma_0\|\Sigma_{\kappa_h}(\bar y)\|_{L^1(\Omega)}\Big(C_\infty h + \|S(\bar u_h) - \bar y\|_{L^\infty(\Omega)}\Big) + h^2\|S(\bar u_h) - \bar y\|_{H_0^1(\Omega)}^2 + h^2\|\bar u_h - \bar u\|_{L^2(\Omega)}^2\Big] \\
&\le C\Big[h^{2+2\gamma} + h^2\sigma_0\|\Sigma_{\kappa_h}(\bar y)\|_{L^1(\Omega)}\Big(h + \|\bar u_h - \bar u\|_{L^2(\Omega)}\Big) + h^2\|\bar u_h - \bar u\|_{L^2(\Omega)}^2\Big],
\end{aligned}$$

where we have used Theorem 2.2 to derive the last inequality. From this, (114), and (115), a simple computation gives (113). ☐

Remark 5.12. If the constant p_* in Theorem 2.2 is large enough, then, for arbitrary small $\varepsilon > 0$, we can take $\gamma := \frac12 - \varepsilon$ in (54) by choosing p_0 close to 2 enough. Therefore, the order of convergence in Theorem 5.11 becomes $O(h^{\frac32 - \varepsilon})$. This order is less than the one for the smooth situation investigated in [14], there the authors showed that the order of convergence associated with the variational discretization is $O(h^2)$. This fact can be attributed to the nondifferentiability of the function a in the state equation as we will see later in the numerical example section.

Similarly, we obtain from Theorem 5.10 error estimates for piecewise constant and continuous piecewise linear controls.

Theorem 5.13 (Piecewise constant discretization and continuous piecewise linear discretization). *Assume that $\mathcal{U}_h = \mathcal{U}_h^i$, $i = 0, 1$. Let $\{(\bar y_h, \bar u_h)\}$ be the sequence of discrete solutions to (P_h) converging strongly to $(\bar y, \bar u)$ in $H_0^1(\Omega) \times L^2(\Omega)$. Under all assumptions of Theorem 5.10, there exist constants $C > 0$ and $\hat h_* \in (0, \hat h)$ such that*

$$\|\bar u_h - \bar u\|_{L^2(\Omega)} \le Ch \quad \text{for all } h \in (0, \hat h_*). \tag{116}$$

Proof. According to Theorem 2.5, $\bar\varphi$ and $\bar u$ are Lipschitz continuous on $\overline\Omega$. Hence constants $C, C_1 > 0$ and $\hat h_* \in (0, \hat h)$ exist such that for any $h \in (0, \hat h_*)$, there exists a $u_h \in \mathcal{U}_{ad,h}$ satisfying $\|\bar u - u_h\|_{L^\infty(\Omega)} \le C_1 h$ and $j'(\bar u)(\bar u - u_h) = 0$ for the case $\mathcal{U}_h = \mathcal{U}_h^0$; see, *e.g.*, Lemma 4.17 in [16], as well as $|j'(\bar u)(\bar u - u_h)| \le Ch^2$ for the case $\mathcal{U}_h = \mathcal{U}_h^1$; see the proof of Theorem 5.4 in [17]. Combining this with (105) and the fact that $\varepsilon_h^{\bar u_h} \le C_2 h$ for all $h \in (0, \hat h)$ yields (116). ☐

5.3. Numerical example

We conclude this section with a preliminary numerical example for the variational discretization of the optimal control problem. Specifically, we consider the problem

$$\begin{cases} \min\limits_{u\in L^\infty(\Omega)} j(u) := \dfrac{1}{2}\displaystyle\int_\Omega (y_u(x) - y_d(x))^2\,\mathrm{d}x + \dfrac{\nu}{2}\int_\Omega u(x)^2\,\mathrm{d}x \\ \quad \text{s.t.} \quad -\operatorname{div}[(b(x) + \max\{y_u - m, 0\})\nabla y_u] = u \quad \text{in } \Omega, \quad y_u = 0 \quad \text{on } \partial\Omega, \\ \qquad 0 \le u(x) \le 4\pi^2 \quad \text{a.e. } x \in \Omega, \end{cases} \tag{117}$$

where $\Omega := (0,1)^2 \subset \mathbb{R}^2$, $m \in (0,1]$, $\nu := 1$, $b(x) := b(x_1, x_2) = 2 - \max\{\sin \pi x_1 \sin \pi x_2 - m, 0\}$, and

$$y_d(x) := (1 + 16\nu\pi^4)\sin \pi x_1 \sin \pi x_2 + 4\nu\pi^4 \mathbb{1}_{\{\sin \pi x_1 \sin \pi x_2 > m\}}(\cos^2 \pi x_1 \sin^2 \pi x_2 + \sin^2 \pi x_1 \cos^2 \pi x_2).$$

This problem fits the general setting with $\alpha = 0$, $\beta = 4\pi^2$, $a(t) = \max\{t - m, 0\}$; *i.e.*, $a_0(t) := 0$, $a_1(t) := t - m$, and $\bar t := m$. Setting $\bar u := 4\pi^2 \bar y$ with $\bar y := \sin \pi x_1 \sin \pi x_2$ and $\bar\varphi := -\nu\bar u$, it is straightforward to verify that $(\bar u, \bar y, \bar\varphi)$ satisfies the first-order optimality condition (8) associated with (117). We shall now show that there exists an $\varepsilon > 0$ such that if $0 \le 1 - m < \varepsilon$ then all assumptions of Theorem 2.7, and thus of Theorem 5.11, are fulfilled. First, for $\bar t = m = 1$ we have

$$\{\bar y = \bar t\} = \{\bar y = 1\} = \{(0.5, 0.5)\} \quad \text{and} \quad \{\bar y < \bar t\} = \Omega \backslash \{(0.5, 0.5)\}.$$

Moreover, $\nabla \bar y$ obviously vanishes on $\{\bar y = 1\}$, and Lemma A.1 in the appendix shows that the structural assumption in (18) holds. For $\bar t = m \in (0,1)$, a simple computation shows that

$$|\nabla \bar y(x)| > 0 \quad \text{for all} \quad x = (x_1, x_2) \in \{\bar y = \bar t\},$$

which validates (17). It remains to show the existence of a number $\varepsilon > 0$ such that the second-order sufficient optimality condition (19) is fulfilled, provided that $0 \le 1 - m < \varepsilon$. To this end, by virtue of (2), there holds for a.e. $(x_1, x_2) \in \Omega$ that $a'(\bar y(x_1, x_2); s) = \mathbb{1}_{\{\bar y(x_1, x_2) > \bar t\}} s$ for all $s \in \mathbb{R}$. From this, the fact that $\mathbb{1}_{\{\bar y \ne \bar t\}} a'' \equiv 0$, and the explicit expression of Q in (19), we have

$$\begin{aligned} Q(\bar u, \bar y, \bar\varphi; v) = {} & \frac{1}{2}\|S'(\bar u)v\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|v\|^2_{L^2(\Omega)} - \int_\Omega \mathbb{1}_{\{\bar y > \bar t\}} S'(\bar u)v \nabla S'(\bar u)v \cdot \nabla\bar\varphi\,\mathrm{d}x \\ & + \frac{1}{2}[a_0'(\bar t) - a_1'(\bar t)] \int_{\{\bar y = \bar t\}} \mathbb{1}_{\{|\nabla \bar y| > 0\}} (S'(\bar u)v)^2 \frac{\nabla \bar y \cdot \nabla \bar\varphi}{|\nabla \bar y|}\,\mathrm{d}\mathcal{H}^1(x) \quad \text{for all } v \in L^2(\Omega). \end{aligned}$$

Since $\bar\varphi = -\nu\bar u = -4\nu\pi^2 \bar y$, there holds

$$\frac{1}{2}[a_0'(\bar t) - a_1'(\bar t)] \int_{\{\bar y = \bar t\}} \mathbb{1}_{\{|\nabla \bar y| > 0\}} (S'(\bar u)v)^2 \frac{\nabla \bar y \cdot \nabla \bar\varphi}{|\nabla \bar y|}\,\mathrm{d}\mathcal{H}^1(x) = 2\nu\pi^2 \int_{\{\bar y = \bar t\}} (S'(\bar u)v)^2 |\nabla \bar y|\,\mathrm{d}\mathcal{H}^1(x) \ge 0$$

for all $v \in L^2(\Omega)$. Consequently, we have

$$Q(\bar u, \bar y, \bar\varphi; v) \ge \frac{1}{2}\|S'(\bar u)v\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|v\|^2_{L^2(\Omega)} + 4\nu\pi^2 \int_\Omega \mathbb{1}_{\{\bar y > \bar t\}} S'(\bar u)v \nabla S'(\bar u)v \cdot \nabla \bar y\,\mathrm{d}x \tag{118}$$

for all $v \in L^2(\Omega)$. We now estimate the last term in the right-hand side of (118). For that purpose, we observe that

$$\begin{aligned} \mathbb{1}_{\{\bar y > \bar t\}} |\nabla \bar y|^2 &= \pi^2 \mathbb{1}_{\{\sin \pi x_1 \sin \pi x_2 > m\}} (\cos^2 \pi x_1 \sin^2 \pi x_2 + \sin^2 \pi x_1 \cos^2 \pi x_2) \\ &= \pi^2 \mathbb{1}_{\{\sin \pi x_1 \sin \pi x_2 > m\}} (\sin^2 \pi x_1 + \sin^2 \pi x_2 - 2\sin^2 \pi x_1 \sin^2 \pi x_2) \end{aligned}$$

$$\leq \pi^2(2-2m^2)$$

and thus

$$\mathbb{1}_{\{\bar y>\bar t\}}|\nabla \bar y| \leq \pi\sqrt{2(1-m^2)} \leq 2\pi\sqrt{1-m}$$

in Ω, where we have exploited that $1 < 1+m \leq 2$. From this and the Cauchy–Schwarz inequality, there holds

$$\begin{aligned}\left|\int_\Omega \mathbb{1}_{\{\bar y>\bar t\}} S'(\bar u)v\nabla S'(\bar u)v\cdot\nabla\bar y\,\mathrm{d}x\right| &\leq \left\|\mathbb{1}_{\{\bar y>\bar t\}}\nabla\bar y\right\|_{L^\infty(\Omega)}\|S'(\bar u)v\|_{L^2(\Omega)}\|\nabla S'(\bar u)v\|_{L^2(\Omega)}\\ &\leq 2\pi C^2\sqrt{1-m}\|v\|^2_{L^2(\Omega)}\end{aligned}$$

for all $v \in L^2(\Omega)$ and for some constant C independent of m and v. Here we have used the fact that $\|S'(\bar u)v\|_{H^1_0(\Omega)} \leq C\|v\|_{L^2(\Omega)}$ due to Theorem 2.2. Combining this with (118) yields

$$Q(\bar u,\bar y,\bar\varphi;v) \geq \frac12\|S'(\bar u)v\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|v\|^2_{L^2(\Omega)} - 8C^2\nu\pi^3\sqrt{1-m}\|v\|^2_{L^2(\Omega)} \quad \geq \frac12\|S'(\bar u)v\|^2_{L^2(\Omega)} + \frac{\nu}{4}\|v\|^2_{L^2(\Omega)} > 0$$

for all $v \in L^2(\Omega)$, $v \neq 0$, provided that $0 \leq 1-m < \varepsilon$ with positive constant ε satisfying

$$\frac14 - 8C^2\pi^3\sqrt{\varepsilon} = 0.$$

We have therefore verified that all hypotheses of Theorems 2.7 and 5.11 are fulfilled. Finally, Lemma A.2 in the appendix shows that $\Sigma(\bar y) = 0$ if $\bar t = m = 1$ and $\Sigma(\bar y) > 0$ if $\bar t = m \in (0,1)$.

We now consider the discrete approximation $\bar u_h$ of $\bar u$. Thanks to Remarks 5.5 and 5.6, we have $\bar u_h = \mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar\varphi_h)$. From this, (22), and (83a), $\bar y_h$ and $\bar\varphi_h$ satisfy

$$\begin{cases}\displaystyle\int_\Omega (b(x)+\max\{\bar y_h - m, 0\})\nabla\bar y_h\cdot\nabla w_h\,\mathrm{d}x = \int_\Omega \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar\varphi_h\right)w_h\,\mathrm{d}x,\\ \displaystyle\int_\Omega (b(x)+\max\{\bar y_h - m, 0\})\nabla\bar\varphi_h\cdot\nabla v_h + \mathbb{1}_{\{\bar y_h>m\}}\nabla\bar y_h\cdot\nabla\bar\varphi_h v_h\,\mathrm{d}x = \int_\Omega(\bar y_h - y_d)v_h\,\mathrm{d}x\end{cases} \tag{119}$$

for all $w_h, v_h \in V_h$. Since the nonlinearities of (119) are semi-smooth, it is reasonable to solve this system by a semi-smooth Newton (SSN) method; see, *e.g.*, [38, 45] as well as [23]. Setting $y_0 := 128x_1x_2(1-x_1)(1-x_2)$, we notice that the sets $\{y_0 > \bar t\}$ and $\{y_0 < \bar t\}$ have positive measures for all $m \in (1-\varepsilon, 1]$. The starting point for the discrete SSN method solved (119) is then taken as $(y_{h,0},\varphi_{h,0}) := (\mathrm{Proj}_{V_h}(y_0), 0)$ for different mesh sizes h, where Proj_{V_h} stands for the L^2 projection mapping onto V_h. The integrals over elements are approximated with a quadrature scheme. This introduces a variational crime which however does not reduce the expected approximation order for piecewise linear functions. In all our tests, the SSN method converged in four or five or six iterations.

We report the resulting discretization errors $\|\bar u_h - \bar u\|_X$ for $X = L^2(\Omega)$ and for $h \in \{\frac{\sqrt2}{2^n} \mid 4 \leq n \leq 10\}$ as well as the experimental order of convergence

$$\mathrm{EOC}_{L^p}(n) := \frac{\log\left(\left\|\bar u - \bar u_{h_{n+1}}\right\|_{L^p(\Omega)}\right) - \log\left(\|\bar u - \bar u_{h_n}\|_{L^p(\Omega)}\right)}{\log(h_{n+1}) - \log(h_n)}$$

in Table 1 for both cases $m = 1$ and $m = 0.95$. For the situation $m = 1$, the results indicate an EOC of 2, which indicates that we are observing a superconvergence property; compare [40]. The EOC for this case is consistent with the guaranteed rate of $O(h^2)$ shown for the smooth problem in [14]. An suitable explanation for this could be that in the case $m = 1$, we have the following identity

$$a(\bar y(x_1,x_2)) = \max\{\bar y(x_1,x_2) - 1, 0\} = 0 \quad \text{for all } (x_1,x_2) \in \Omega,$$

and thus the coefficient $a(y)$ of the state equation in (117) is in fact smooth at the optimal state. When $m = 0.95$ (and the coefficient is nonsmooth), the values of EOC are not stable, however their minimum value is approximately equal to 1.5, which fits the theoretical study shown in Remark 5.12.

TABLE 1. discretization errors and experimental orders of convergence (EOC) in L^2 for the optimal control $\bar{u}$ in dependence of h_n

	$m = 1$		$m = 0.95$	
h_n	$\|\bar{u} - \bar{u}_{h_n}\|_{L^2(\Omega)}$	$\mathrm{EOC}_{L^2}(n)$	$\|\bar{u} - \bar{u}_{h_n}\|_{L^2(\Omega)}$	$\mathrm{EOC}_{L^2}(n)$
0.08838834764831845	0.18868399994780682	1.9842206675459904	0.16611404078219316	2.30149884431476
0.04419417382415922	0.047689759843375215	1.9960412352638899	0.03369660566403936	1.499478604870065
0.02209708691207961	0.011955200141487647	1.9990094098708553	0.01191785556126703	1.703223336241136
0.011048543456039806	0.002990852924165744	1.9997522957092797	0.0036599639201972353	2.2944527855422274
0.005524271728019903	0.0007478416210831529	1.9999380703342444	0.000746066761816203	1.5712691960881326
0.0027621358640099515	0.0001869684309751765	1.9999845171567776	0.0002510605828845355	2.277237268364241
0.0013810679320049757	4.674260937760375e−05		5.179189573581347e−05	

6. Conclusions

We have studied the numerical approximation of an optimal control problem governed by a quasilinear elliptic equation with nonsmooth coefficient in the divergence part. The convergence of a sequence of minimizers of some discrete control problems to a global minimizer of the original problem is shown. *A priori* error estimates for three types of discretizations (variational, piecewise constant, and continuous piecewise linear discretizations) are derived under an explicit second-order sufficient condition for the continuous optimal control problem and a structural assumption on the optimal state. The estimate for variational discretization corrobates the proven rate, although the observed rate is higher, which motivates follow-up work on rate optimality or superconvergence properties for optimal control of nonsmooth quasilinear equations.

Acknowledgments

Part of this work was completed during a visit of the second author to the Vietnam Institute for Advanced Study in Mathematics (VIASM). The second author would like to thank the VIASM for their financial support and hospitality.

Funding

This work was supported by the DFG under the grants CL 487/2-1 and RO 2462/6-1, both within the priority programme SPP 1962 "Nonsmooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization".

Data availability statement

The research data associated with this article are included in the article.

References

[1] G. Alberti, S. Bianchini and G. Crippa, Structure of level sets and Sard-type properties of Lipschitz maps. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **XII** (2013) 863–902.

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

[3] T. Bayen, J.F. Bonnans and F.J. Silva, Characterization of local quadratic growth for strong minima in the optimal control of semi-linear elliptic equations. *Trans. Amer. Math. Soc.* **366** (2014) 2063–2087.

[4] A. Behzadan and M. Holst, Multiplication in Sobolev spaces, revisited. *Ark. Mat.* **59** (2021) 275–306.

[5] A. Bejan, Convection Heat Transfer, 4 edition. J. Wiley & Sons (2013).

[6] J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems. Springer-Verlag, Berlin, Heidelberg (2000).

[7] S.C. Brenner and L.R. Scott, The Mathematical Theory of Finite Element Methods. *Texts in Applied Mathematics*. Springer-Verlag, New York (2008).

[8] C. Carstensen, Quasi-interpolation and a posteriori error analysis in finite element methods. *ESAIM: Math. Modell. Numer. Anal.* **33** (1999) 1187–1202.

[9] E. Casas and V. Dhamo, Error estimates for the numerical approximation of a quasilinear Neumann problem under minimal regularity of the data. *Numer. Math.* **117** (2011) 115–145.

[10] E. Casas and V. Dhamo, Error estimates for the numerical approximation of Neumann control problems governed by a class of quasilinear elliptic equations. *Comput. Optim. Appl.* **52** (2012) 719–756.

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

[12] 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** (2009) 688–718.

[13] E. Casas and F. Tröltzsch, Numerical analysis of some optimal control problems governed by a class of quasilinear elliptic equations. *ESAIM:COCV* **17** (2011) 771–800.

[14] E. Casas and F. Tröltzsch, A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. *Comput. Optim. Appl.* **53** (2012) 173–206.

[15] E. Casas, M. Mateos and F. Tröltzsch, Error estimates for the numerical approximation of boundary semilinear elliptic control problems. *Comput. Optim. Appl.* **31** (2005) 193–219.

[16] E. Casas, M. Mateos and J.-P. Raymond, Error estimates for the numerical approximation of a distributed control problem for the steady-state Navier–Stokes equations. *SIAM J. Control Optim.* **46** (2007) 952–982.

[17] E. Casas, M. Mateos and A. Rösch, Numerical approximation of control problems of non-monotone and non-coercive semilinear elliptic equations. *Numer. Math.* **149** (2021) 305–340.

[18] M. Chipot, Elliptic Equations: An Introductory Course. Birkhäuser Verlag, Basel (2009).

[19] C. Christof and C. Meyer, A note on a priori L^p-error estimates for the obstacle problem. *Numer. Math.* **139** (2018) 27–45.

[20] P.G. Ciarlet, The Finite Element Method for Elliptic Problems. *Classics in Applied Mathematics*. SIAM (2002).

[21] S. Clain, Elliptic operators of divergence type with Hölder coefficients in fractional Sobolev spaces. *Rend. Mat. Appl.* **17** (1997) 207–236.

[22] C. Clason, V.H. Nhu and A. Rösch, No-gap second-order optimality conditions for optimal control of a non-smooth quasilinear elliptic equation. *ESAIM: COCV* **27** (2021) 62.

[23] C. Clason, V.H. Nhu and A. Rösch, Optimal control of a non-smooth quasilinear elliptic equation. *Math. Control Related Fields* **11** (2021) 521–554.

[24] C. Clason, V.H. Nhu and A. Rösch, Numerical analysis of a nonsmooth quasilinear elliptic control problem: I. Explicit second-order optimality conditions. *ESAIM: Math. Modell. Numer. Anal. (ESAIM: M2AN)* **58** (2024) 993–1029.

[25] B. Dacorogna, Direct Methods in the Calculus of Variations. *Applied Mathematical Sciences*, 2 edition. Springer (2008).

[26] J.C. De Los Reyes and V. Dhamo, Error estimates for optimal control problems of a class of quasilinear equations arising in variable viscosity fluid flow. *Numer. Math.* **132** (2016) 691–720.

[27] J.C. De Los Reyes, C. Meyer and B. Vexler, Finite element error analysis for state-constrained optimal control of the Stokes equations. *Control Cybern.* **37** (2008) 251–284.

[28] K. Deckelnick and M. Hinze, Convergence of a finite element approximation to a state-constrained elliptic control problem. *SIAM J. Numer. Anal.* **45** (2007) 1937–1953.

[29] T. Dupont and R. Scott, Polynomial approximation of functions in Sobolev spaces. *Math. Comput.* **34** (1980) 441–463.

[30] L.C. Evans and R.F. Gariepy, Measure Theory and Fine Properties of Function, 4th edition. CRC Press, New York (1992).

[31] R.S. Falk, Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.* **44** (1973) 28–47.

[32] T. Geveci, On the approximation of the solution of an optimal control problem governed by an elliptic equation. *RAIRO. Anal. Numér.* **13** (1979) 313–328.

[33] D. Gilbarg and N.S. Trudinger, Elliptic Partial Differential Equations of Second Order. Springer-Verlag, Berlin, Heidelberg (2001).

[34] P. Grisvard, Elliptic Problems in Nonsmooth Domains. Pitman Advanced Pub. Program (1985).

[35] M. Hinze, A variational discretization concept in control constrained optimization: the linear-quadratic case. *Comput. Optim. Appl.* **30** (2005) 45–61.

[36] M. Hinze and F. Tröltzsch, Discrete concepts versus error analysis in PDE-constrained optimization. *GAMM-Mitteilungen* **33** (2010) 148–162.

[37] M. Holtmannspötter and A. Rösch, A priori error estimates for the finite element approximation of a nonsmooth optimal control problem governed by a coupled semilinear PDE-ODE system. *SIAM J. Control Optim.* **59** (2021) 3329–3358.
[38] K. Ito and K. Kunisch, Lagrange Multiplier Approach to Variational Problems and Applications. *Advances in Design and Control*. SIAM (2008).
[39] A. Kufner, O. John and S. Fučík, Function Spaces. Noordhoff International Publishing, Leyden; Academia, Prague (1977). Monographs and Textbooks on Mechanics of Solids and Fluids; Mechanics: Analysis.
[40] C. Meyer and A. Rösch, Superconvergence properties of optimal control problems. *SIAM J. Control Optim.* **43** (2004) 970–985.
[41] C. Meyer and A. Rösch, L^∞-estimates for approximated optimal control problems. *SIAM J. Control Optim.* **44** (2005) 1636–1649.
[42] C. Meyer and O. Thoma, A priori finite element error analysis for optimal control of the obstacle problem. *SIAM J. Numer. Anal.* **51** (2013) 605–628.
[43] I. Neitzel, J. Pfefferer and A. Rösch, Finite element discretization of state-constrained elliptic optimal control problems with semilinear state equation. *SIAM J. Control Optim.* **53** (2015) 874–904.
[44] G. Savaré, Regularity results for elliptic equations in Lipschitz domains. *J. Funct. Anal.* **152** (1998) 176–201.
[45] M. Ulbrich, Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. *MOS-SIAM Series on Optimization*. SIAM (2011).
[46] Y.B. Zel'dovich and Y.P. Raizer, Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena. Academic Press (1966).

Please help to maintain this journal in open access!

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting subscribers@edpsciences.org.

More information, including a list of supporters and financial transparency reports, is available at https://edpsciences.org/en/subscribe-to-open-s2o.

Appendix A. Verification of a structural assumption and computation of the jump functional

Lemma A.1. *Let* $\Omega := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_1, x_2 < 1\}$ *and let* $y(x_1, x_2) := \sin(\pi x_1)\sin(\pi x_2)$. *Then there exists a constant* $c_s > 0$ *such that*

$$\operatorname{meas}_{\mathbb{R}^2}(\{|y - 1| < r\}) \leq c_s r \tag{A.1}$$

for all $r > 0$ *small enough.*

Proof. Take any sufficiently small $r > 0$ satisfying $(1 - r, 1) \subset (0, 1)$. A simple computation yields

$$\{|y-1| < r\} = \left\{ \tfrac{1}{\pi}\arcsin\tfrac{1-r}{\sin \pi x_2} < x_1 < 1 - \tfrac{1}{\pi}\arcsin\tfrac{1-r}{\sin \pi x_2}, \tfrac{1}{\pi}\arcsin(1-r) < x_2 < 1 - \tfrac{1}{\pi}\arcsin(1-r) \right\}.$$

We thus obtain

$$\operatorname{meas}_{\mathbb{R}^2}(\{|y - \bar{t}| < r\}) = \int_{\frac{1}{\pi}\arcsin(1-r)}^{1-\frac{1}{\pi}\arcsin(1-r)} \left(1 - \tfrac{2}{\pi}\arcsin\tfrac{1-r}{\sin \pi x_2}\right) \mathrm{d}x_2 \leq \left[1 - 2\tfrac{1}{\pi}\arcsin(1-r)\right]^2, \tag{A.2}$$

where we have used the fact that

$$0 \leq 1 - \tfrac{2}{\pi}\arcsin\tfrac{1-r}{\sin \pi x_2} \leq 1 - 2\tfrac{1}{\pi}\arcsin(1-r) \quad \text{for all } 0 < x_2 < 1, r \in (0, 1).$$

L'Hospital's rule then shows that

$$\lim_{r\to 0^+} \frac{1-2\frac{1}{\pi}\arcsin(1-r)}{\sqrt{r}} = \frac{4}{\pi}\lim_{r\to 0^+}\sqrt{\frac{r}{1-(1-r)^2}} = \frac{2\sqrt{2}}{\pi},$$

which, along with (A.2), yields (A.1). □

Lemma A.2. *Let* $\Omega := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_1, x_2 < 1\}$ *and let* $y(x_1, x_2) := \sin(\pi x_1)\sin(\pi x_2)$. *Then*

$$\Sigma(y) = \begin{cases} 8\sigma_0\left(1 - \frac{2}{\pi}\arcsin\bar{t}\right) & \text{if} \quad \bar{t} \in (0,1), \\ 0 & \text{if} \quad \bar{t} = 1, \\ 4\sigma_0 & \text{if} \quad \bar{t} = 0. \end{cases}$$

Proof. We distinguish the following cases:

(i) For any $\bar{t} \in (0,1)$ and any sufficiently small $r > 0$ satisfying $(\bar{t} - r, \bar{t} + r) \subset (0,1)$, we have

$$\begin{aligned}\{|y - \bar{t}| < r, \partial_{x_1} y > 0\} &= \{\bar{t} - r < \sin(\pi x_1)\sin(\pi x_2) < \bar{t} + r, \cos(\pi x_1) > 0\} \\ &= \left\{\frac{\bar{t}-r}{\sin(\pi x_2)} < \sin(\pi x_1) < \frac{\bar{t}+r}{\sin(\pi x_2)}, \frac{\bar{t}+r}{\sin(\pi x_2)} \le 1, 0 < x_1 < \frac{1}{2}\right\} \\ &\cup \left\{\frac{\bar{t}-r}{\sin(\pi x_2)} < \sin(\pi x_1), \frac{\bar{t}+r}{\sin(\pi x_2)} > 1, 0 < x_1 < \frac{1}{2}\right\}.\end{aligned}$$

A simple computation therefore shows that

$$\begin{aligned}\{|y - \bar{t}| < r, \partial_{x_1} y > 0\} &= \left\{\tfrac{1}{\pi}\arcsin\tfrac{\bar{t}-r}{\sin \pi x_2} < x_1 < \tfrac{1}{\pi}\arcsin\tfrac{\bar{t}+r}{\sin \pi x_2}, \tfrac{1}{\pi}\arcsin(\bar{t}+r) \le x_2 \le 1 - \tfrac{1}{\pi}\arcsin(\bar{t}+r)\right\} \\ &\cup \left\{\tfrac{1}{\pi}\arcsin\tfrac{\bar{t}-r}{\sin \pi x_2} < x_1 < \tfrac{1}{2}, \tfrac{1}{\pi}\arcsin(\bar{t}-r) < x_2 < \tfrac{1}{\pi}\arcsin(\bar{t}+r)\right\} \\ &\cup \left\{\tfrac{1}{\pi}\arcsin\tfrac{\bar{t}-r}{\sin \pi x_2} < x_1 < \tfrac{1}{2}, 1 - \tfrac{1}{\pi}\arcsin(\bar{t}+r) < x_2 < 1 - \tfrac{1}{\pi}\arcsin(\bar{t}-r)\right\}.\end{aligned}$$

We thus obtain

$$\begin{aligned}\int_\Omega \mathbb{1}_{\{|y-\bar{t}|<r, \partial_{x_1} y > 0\}} |\partial_{x_1} y| \,\mathrm{d}x &= -\frac{2}{\pi}[\cos(\arcsin(\bar{t}+r)) - \cos(\arcsin(\bar{t}-r))] \\ &\quad - \frac{2}{\pi}(\bar{t}-r)[\arcsin(\bar{t}+r) - \arcsin(\bar{t}-r)] + 2r\left(1 - \frac{2}{\pi}\arcsin(\bar{t}+r)\right).\end{aligned}$$

Applying L'Hospital's rule yields

$$\lim_{r\to 0^+} \frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-\bar{t}|<r, \partial_{x_1} y > 0\}} |\partial_{x_1} y| \,\mathrm{d}x = 2\left(1 - \frac{2}{\pi}\arcsin\bar{t}\right).$$

Similarly, there hold

$$\begin{aligned}\lim_{r\to 0^+} \frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-\bar{t}|<r, \partial_{x_1} y < 0\}} |\partial_{x_1} y| \,\mathrm{d}x &= 2\left(1 - \frac{2}{\pi}\arcsin\bar{t}\right), \\ \lim_{r\to 0^+} \frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-\bar{t}|<r, \partial_{x_2} y > 0\}} |\partial_{x_2} y| \,\mathrm{d}x &= 2\left(1 - \frac{2}{\pi}\arcsin\bar{t}\right), \\ \lim_{r\to 0^+} \frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-\bar{t}|<r, \partial_{x_2} y < 0\}} |\partial_{x_2} y| \,\mathrm{d}x &= 2\left(1 - \frac{2}{\pi}\arcsin\bar{t}\right).\end{aligned}$$

By adding these four limits, we obtain

$$\lim_{r\to 0^+}\frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-\bar t|<r\}}[|\partial_{x_1}y|+|\partial_{x_2}y|]\,\mathrm{d}x = 8\Big(1-\frac{2}{\pi}\arcsin\bar t\Big).$$

(ii) For $\bar t = 0$ and for any $r\in(0,1)$ sufficiently small, we see from a straightforward calculation that

$$\begin{aligned}\{|y-0|<r,\partial_{x_1}y>0\} = &\Big\{0<x_1<\tfrac{1}{\pi}\arcsin\tfrac{r}{\sin\pi x_2},\tfrac{1}{\pi}\arcsin r\le x_2\le 1-\tfrac{1}{\pi}\arcsin r\Big\}\\ &\cup\Big\{0<x_1<\tfrac12,0<x_2<\tfrac{1}{\pi}\arcsin r\Big\}\cup\Big\{0<x_1<\tfrac12,1-\tfrac{1}{\pi}\arcsin r<x_2<1\Big\}.\end{aligned}$$

Consequently, it holds that

$$\int_\Omega \mathbb{1}_{\{|y-0|<r,\partial_{x_1}y>0\}}|\partial_{x_1}y|\,\mathrm{d}x = -\frac{2}{\pi}[\cos(\arcsin r)-1]+r\Big(1-\frac{2}{\pi}\arcsin r\Big).$$

L'Hospital's rule then shows that

$$\lim_{r\to 0^+}\frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-0|<r,\partial_{x_1}y>0\}}|\partial_{x_1}y|\,\mathrm{d}x = 1.$$

Similarly, we can conclude that

$$\lim_{r\to 0^+}\frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-0|<r\}}[|\partial_{x_1}y|+|\partial_{x_2}y|]\,\mathrm{d}x = 4.$$

(iii) For $\bar t = 1$ and for any $r>0$ such that $r\in(0,1)$, we have

$$\{|y-1|<r,\partial_{x_1}y>0\} = \Big\{\tfrac{1}{\pi}\arcsin\tfrac{1-r}{\sin\pi x_2}<x_1<\tfrac12,\tfrac{1}{\pi}\arcsin(1-r)\le x_2\le 1-\tfrac{1}{\pi}\arcsin(1-r)\Big\}.$$

There therefore holds that

$$\int_\Omega \mathbb{1}_{\{|y-1|<r,\partial_{x_1}y>0\}}|\partial_{x_1}y|\,\mathrm{d}x = \frac{2}{\pi}[\cos(\arcsin(1-r))+(1-r)\arcsin(1-r)]+r-1.$$

Again, L'Hospital's rule shows that

$$\lim_{r\to 0^+}\frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-1|<r,\partial_{x_1}y>0\}}|\partial_{x_1}y|\,\mathrm{d}x = 0.$$

Similarly, we can deduce that

$$\lim_{r\to 0^+}\frac{1}{r}\int_\Omega \mathbb{1}_{\{|y-1|<r\}}[|\partial_{x_1}y|+|\partial_{x_2}y|]\,\mathrm{d}x = 0.$$

□

Appendix B. $W^{1+\gamma,2}$-regularity of solutions to the adjusted linearized state equation

Let p_0, p_1, and γ be given as in (54). In order to show the $W^{1+\gamma,2}$-regularity of solutions to the adjusted linearized state equation (55), we need the following result for the function

$$\mathbb{f}_{\bar u} := \mathbb{1}_{\{\bar y\neq\bar t\}}a'(\bar y)\nabla\bar y, \tag{B.1}$$

where $\bar y := S(\bar u)$ is the optimal state corresponding to the control $\bar u$.

Proposition B.1. *Assume that $\bar{u} \in L^\infty(\Omega)$ satisfies $\Sigma(\bar{y}) < \infty$. Then there holds*

$$\mathbb{f}_{\bar{u}} \in (W^{\gamma,p_0}(\Omega))^2 = (W_0^{\gamma,p_0}(\Omega))^2. \tag{B.2}$$

Proof. Since $\gamma < \frac{1}{p_0}$, there holds that

$$W^{\gamma,p_0}(\Omega) = W_0^{\gamma,p_0}(\Omega);$$

see, *e.g.*, Corollary 1.4.4.5 in [34]. Thanks to Theorem 2.2, one has $\bar{y} \in W^{2,p_1}(\Omega) \hookrightarrow C^1(\overline{\Omega})$ and thus $\bar{y} \in C^1(\overline{\Omega})$. Therefore, we have $\mathbb{f}_{\bar{u}} \in L^\infty(\Omega)^2 \hookrightarrow L^{p_0}(\Omega)^2$. From the definition of the Sobolev spaces of fractional order; see *e.g.*, Definition 6.8.2 in [39], it then suffices to show that

$$\iint_{\Omega\times\Omega} \frac{|\mathbb{f}_{\bar{u}}(x) - \mathbb{f}_{\bar{u}}(\hat{x})|^{p_0}}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} < \infty. \tag{B.3}$$

Since $\bar{y} \in C^1(\overline{\Omega})$, there holds for some constant $L > 0$ that

$$|\bar{y}(x) - \bar{y}(\hat{x})| \le L|x-\hat{x}| \quad \text{for all } x, \hat{x} \in \overline{\Omega}. \tag{B.4}$$

Applying (B.4) then yields

$$\iint_{\Omega\times\Omega} \frac{|\bar{y}(x) - \bar{y}(\hat{x})|^{p_0}}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} \le \iint_{\Omega\times\Omega} L^{p_0}|x-\hat{x}|^{-2+p_0(1-\gamma)}\,\mathrm{d}x\,\mathrm{d}\hat{x} < \infty, \tag{B.5}$$

due to the fact that $2 - p_0(1-\gamma) < 2 = N$; see *e.g.*, page 331 in [39]. Moreover, from the Sobolev embedding $W^{1,p_1}(\Omega) \hookrightarrow C^{0,1-\frac{2}{p_1}}(\overline{\Omega})$, there holds $\nabla\bar{y} \in C^{0,1-\frac{2}{p_1}}(\overline{\Omega})^2$ and we obtain

$$|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})| \le L'|x-\hat{x}|^{1-\frac{2}{p_1}} \quad \text{for all } x, \hat{x} \in \overline{\Omega}$$

and for some constant $L' > 0$. From this, we have

$$\iint_{\Omega\times\Omega} \frac{|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})|^{p_0}}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} \le C\iint_{\Omega\times\Omega} |x-\hat{x}|^{-2+p_0(1-\frac{2}{p_1}-\gamma)}\,\mathrm{d}x\,\mathrm{d}\hat{x} < \infty, \tag{B.6}$$

in the view of (54). We now set

$$Z(x,\hat{x}) := \mathbb{f}_{\bar{u}}(x) - \mathbb{f}_{\bar{u}}(\hat{x}) = \mathbb{1}_{\{\bar{y}\neq\bar{t}\}}(x)a'(\bar{y}(x))\nabla\bar{y}(x) - \mathbb{1}_{\{\bar{y}\neq\bar{t}\}}(\hat{x})a'(\bar{y}(\hat{x}))\nabla\bar{y}(\hat{x}) \tag{B.7}$$

for all $x, \hat{x} \in \Omega$. Define the functions

$$\chi_{=}(x,\hat{x}) = \begin{cases} 1 & \text{if} \quad \bar{y}(\hat{x}) = \bar{t}, \\ 0 & \text{otherwise}, \end{cases} \quad \chi_{<,<}(x,\hat{x}) = \begin{cases} 1 & \text{if} \quad \bar{y}(\hat{x}), \bar{y}(x) < \bar{t}, \\ 0 & \text{otherwise}, \end{cases} \quad \chi_{>,>}(x,\hat{x}) = \begin{cases} 1 & \text{if} \quad \bar{y}(\hat{x}), \bar{y}(x) > t, \\ 0 & \text{otherwise} \end{cases}$$

and

$$\chi_{>,\le}(x,\hat{x}) = \begin{cases} 1 & \text{if} \quad \bar{y}(\hat{x}) > \bar{t}, \bar{y}(x) \le \bar{t}, \\ 0 & \text{otherwise}, \end{cases} \quad \chi_{<,\ge}(x,\hat{x}) = \begin{cases} 1 & \text{if} \quad \bar{y}(\hat{x}) < \bar{t}, \bar{y}(x) \ge \bar{t}, \\ 0 & \text{otherwise}. \end{cases}$$

We then have the identity

$$\chi_{=}(x,\hat{x}) + \chi_{<,<}(x,\hat{x}) + \chi_{>,>}(x,\hat{x}) + \chi_{>,\le}(x,\hat{x}) + \chi_{<,\ge}(x,\hat{x}) = 1 \tag{B.8}$$

for all $x, \hat{x} \in \Omega$. Since $\nabla\bar{y} = 0$ a.e. on $\{\bar{y} = \bar{t}\}$,

$$|Z(x,\hat{x})|^{p_0}\chi_{=}(x,\hat{x}) = \left|\mathbb{1}_{\{\bar{y}\neq\bar{t}\}}(x)a'(\bar{y}(x))[\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})]\right|^{p_0}\chi_{=}(x,\hat{x}) \le C|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})|^{p_0}.$$

From this and (B.6), we have

$$\iint_{\Omega\times\Omega} \frac{|Z(x,\hat{x})|^{p_0}\chi_{=}(x,\hat{x})}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} < \infty. \tag{B.9}$$

Moreover, by a simple computation, there holds for a.e. $\hat{x}, x \in \{\bar{y} < \bar{t}\}$ that

$$\begin{aligned} Z(x,\hat{x}) &= a_0'(\bar{y}(x))\nabla\bar{y}(x) - a_0'(\bar{y}(\hat{x}))\nabla\bar{y}(\hat{x}) \\ &= a_0'(\bar{y}(x))[\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})] + [a_0'(\bar{y}(x)) - a_0'(\bar{y}(\hat{x}))]\nabla\bar{y}(\hat{x}). \end{aligned}$$

We then deduce from the fact that $\bar{y} \in C^1(\overline{\Omega})$, the Lipschitz continuity of a_0' on bounded sets, and (B.5), (B.6) that

$$\begin{aligned} \iint_{\Omega\times\Omega} \frac{|Z(x,\hat{x})|^{p_0}\chi_{<,<}(x,\hat{x})}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} &\le C\iint_{\Omega\times\Omega} \frac{|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})|^{p_0}}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} \\ &\quad + C\iint_{\Omega\times\Omega} \frac{|\bar{y}(x) - \bar{y}(\hat{x})|^{p_0}}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} < \infty. \end{aligned} \tag{B.10}$$

Similarly, there holds

$$\iint_{\Omega\times\Omega} \frac{|Z(x,\hat{x})|^{p_0}\chi_{>,>}(x,\hat{x})}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} < \infty. \tag{B.11}$$

On the other hand, for a.e. $\hat{x} \in \{\bar{t} < \bar{y}\}$ and $x \in \{\bar{y} \le \bar{t}\}$, we conclude from the fact $\nabla\bar{y} = 0$ a.e. on $\{\bar{y} = \bar{t}\}$ that

$$\begin{aligned} |Z(x,\hat{x})| &= |a_0'(\bar{y}(x))\nabla\bar{y}(x) - a_1'(\bar{y}(\hat{x}))\nabla\bar{y}(\hat{x})| \\ &= |a_0'(\bar{y}(x))[\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})] + [a_0'(\bar{y}(x)) - a_0'(\bar{t})]\nabla\bar{y}(\hat{x}) + [a_0'(\bar{t}) - a_1'(\bar{t})]\nabla\bar{y}(\hat{x}) + [a_1'(\bar{t}) - a_1'(\bar{y}(\hat{x}))]\nabla\bar{y}(\hat{x})| \\ &\le C[|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})| + |\bar{y}(x) - \bar{t}| + |\bar{y}(\hat{x}) - \bar{t}| + |a_0'(\bar{t}) - a_1'(\bar{t})||\nabla\bar{y}(\hat{x})|] \\ &\le C_2[|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})| + |\bar{y}(x) - \bar{y}(\hat{x})| + \sigma_0|\nabla\bar{y}(\hat{x})|], \end{aligned}$$

where we have used the Lipschitz continuity on bounded sets of a_0' and a_1' and the definition of σ_0 in (15). This, along with the fact that $\bar{y} \in C^1(\overline{\Omega})$, yields

$$|Z(x,\hat{x})|^{p_0}\chi_{>,\le}(x,\hat{x}) \le C[|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})|^{p_0} + |\bar{y}(x) - \bar{y}(\hat{x})|^{p_0} + \sigma_0|\nabla\bar{y}(\hat{x})|\chi_{>,\le}(x,\hat{x})]$$

for all $x, \hat{x} \in \Omega$. Analogously, there holds

$$|Z(x,\hat{x})|^{p_0}\chi_{<,\ge}(x,\hat{x}) \le C[|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})|^{p_0} + |\bar{y}(x) - \bar{y}(\hat{x})|^{p_0} + \sigma_0|\nabla\bar{y}(\hat{x})|\chi_{<,\ge}(x,\hat{x})]$$

for all $x, \hat{x} \in \Omega$. We therefore have

$$|Z(x,\hat{x})|^{p_0}[\chi_{>,\le}(x,\hat{x}) + \chi_{<,\ge}(x,\hat{x})] \le C[|\nabla\bar{y}(x) - \nabla\bar{y}(\hat{x})|^{p_0} + |\bar{y}(x) - \bar{y}(\hat{x})|^{p_0} + \sigma_0|\nabla\bar{y}(\hat{x})|e(x,\hat{x})] \tag{B.12}$$

for all $x, \hat{x} \in \Omega$. Here

$$\begin{aligned} e(x,\hat{x}) := \chi_{>,\le}(x,\hat{x}) + \chi_{<,\ge}(x,\hat{x}) &\le \begin{cases} 1 & \text{if} \quad 0 < |\bar{y}(\hat{x}) - \bar{t}| \le |\bar{y}(\hat{x}) - \bar{y}(x)|, \\ 0 & \text{otherwise} \end{cases} \\ &\le \begin{cases} 1 & \text{if} \quad 0 < |\bar{y}(\hat{x}) - \bar{t}| \le L|x - \hat{x}|, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

due to the definition of $\chi_{>,\le}(x,\hat{x})$, $\chi_{<,\ge}(x,\hat{x})$, and (B.4). We now conclude from the definition of $\Sigma(\bar{y})$ in (14) that there exists a constant $r_0 > 0$ such that

$$\frac{\sigma_0}{r}\int_\Omega \mathbb{1}_{\{0<|\bar{y}(\cdot)-\bar{t}|<r\}}(\hat{x})|\nabla\bar{y}(\hat{x})|d\hat{x} \le \Sigma(\bar{y}) + 1 \quad \text{for all } 0 < r < r_0.$$

We then have

$$\begin{aligned}
&\iint_{\Omega\times\Omega} \frac{\sigma_0 e(x,\hat{x})|\nabla\bar{y}(\hat{x})|}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x}\\
&\qquad= \iint_{\Omega\times\Omega\cap\{r_0/L\le|x-\hat{x}|\}} \frac{\sigma_0 e(x,\hat{x})|\nabla\bar{y}(\hat{x})|}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x} + \iint_{\Omega\times\Omega\cap\{|x-\hat{x}|<r_0/L\}} \frac{\sigma_0 e(x,\hat{x})|\nabla\bar{y}(\hat{x})|}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x}\\
&\qquad\le C\iint_{\Omega\times\Omega\cap\{r_0/L\le|x-\hat{x}|\}} \frac{1}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x}\\
&\qquad\quad+\sigma_0\iint_{\Omega\times\Omega\cap\{|x-\hat{x}|<r_0/L\}} \frac{1}{|x-\hat{x}|^{2+p_0\gamma}}\mathbb{1}_{\{0<|\bar{y}(\cdot)-\bar{t}|<L|\cdot-x|\}}(\hat{x})|\nabla\bar{y}(\hat{x})|\,\mathrm{d}\hat{x}\,\mathrm{d}x\\
&\qquad\le C(r_0,L)+\sigma_0\int_{\{\xi\in\mathbb{R}^2:|\xi|<\frac{r_0}{L}\}} \frac{1}{|\xi|^{2+p_0\gamma}}d\xi\int_\Omega \mathbb{1}_{\{0<|\bar{y}(\cdot)-\bar{t}|<L|\xi|\}}(\hat{x})|\nabla\bar{y}(\hat{x})|\,\mathrm{d}\hat{x}\\
&\qquad= C(r_0,L)+L\int_{\{\xi\in\mathbb{R}^2:|\xi|<\frac{r_0}{L}\}} \frac{d\xi}{|\xi|^{1+p_0\gamma}}\frac{\sigma_0}{L|\xi|}\int_\Omega \mathbb{1}_{\{0<|\bar{y}(\cdot)-\bar{t}|<L|\xi|\}}(\hat{x})|\nabla\bar{y}(\hat{x})|\,\mathrm{d}\hat{x}\\
&\qquad\le C(r_0,L)+L\frac{2\pi}{1-p_0\gamma}\Big(\frac{r_0}{L}\Big)^{1-p_0\gamma}(\Sigma(\bar{y})+1)\\
&\qquad<\infty,
\end{aligned}$$

due to (54). From this, (B.5), (B.6), and (B.12), we have

$$\iint_{\Omega\times\Omega} \frac{|Z(x,\hat{x})|^{p_0}[\chi_{<,\ge}(x,\hat{x})+\chi_{>,\le}(x,\hat{x})]}{|x-\hat{x}|^{2+p_0\gamma}}\,\mathrm{d}x\,\mathrm{d}\hat{x}<\infty.$$

Combining this with (B.8), (B.9), (B.10), and (B.11), yields (B.3). □

Lemma B.2. *Assume that $\bar{u}\in L^\infty(\Omega)$ satisfies $\Sigma(\bar{y})<\infty$. Then for any $z\in H^1(\Omega)$, there holds*

$$\operatorname{div}(\mathbb{f}_{\bar{u}}z)\in W^{-1+\gamma,2}(\Omega).$$

Moreover,

$$\|\operatorname{div}(\mathbb{f}_{\bar{u}}z)\|_{W^{-1+\gamma,2}(\Omega)}\le C\|\mathbb{f}_{\bar{u}}\|_{W^{\gamma,p_0}(\Omega)}\|z\|_{H^1(\Omega)} \tag{B.13}$$

for some constant C independent of z.

Proof. By Proposition B.1, one has $\mathbb{f}_{\bar{u}}\in(W^{\gamma,p_0}(\Omega))^2$. We now apply the multiplication theorem for Sobolev spaces; see, *e.g.*, Theorem 1.4.4.2 in [34] and Theorem 7.4 in [4] for $s_1=\gamma$, $s_2=1$, $s=\gamma$, $p_1=p_0>2$, $p_2=2$, $p=2$, and $n=2$ to obtain

$$\mathbb{f}_{\bar{u}}z\in(W^{\gamma,2}(\Omega))^2.$$

From this and the continuity of the divergence operator from $(W^{\gamma,2}(\Omega))^2$ to $W^{-1+\gamma,2}(\Omega)$ for $\gamma\ne\frac12$; see, *e.g.*, Theorem 1.4.4.6 in [34], we have $\operatorname{div}(\mathbb{f}_{\bar{u}}z)\in W^{-1+\gamma,2}(\Omega)$. Finally, we also have (B.13) from Theorem 7.4 in [4] and Theorem 1.4.4.6 in [34]. □

Proposition B.3. *The following assertions hold:*

(i) *Let $p>2$ be fixed but arbitrary. Then, for any $u\in L^2(\Omega)$ and $v\in W^{-1,p}(\Omega)$, (55) admits a unique solution $\tilde{z}_{u,v}\in W_0^{1,p}(\Omega)$. Furthermore, there exists a constant $\rho_p>0$ such that*

$$\|\tilde{z}_{u,v}\|_{W_0^{1,p}(\Omega)}\le C\|v\|_{W^{-1,p}(\Omega)}\quad\text{for all } u\in\overline{B}_{L^2(\Omega)}(\bar{u},\rho_p) \tag{B.14}$$

and for some constant C independent of u and v.

(ii) *Assume that* $\bar{u} \in L^\infty(\Omega)$ *satisfies* $\Sigma(\bar{y}) < \infty$. *Then, for any* $u \in \mathcal{U}_{ad}$ *and any* $v \in W^{-1+\gamma,2}(\Omega)$, *the equation* (55) *admits a unique solution* $\tilde{z}_{u,v}$ *in* $W^{1+\gamma,2}(\Omega)$. *Moreover, there exists a constant* $\hat{\rho}$ *such that*

$$\|\tilde{z}_{u,v}\|_{W^{1+\gamma,2}(\Omega)} \leq C\Big(1 + \|\mathbb{f}_{\bar{u}}\|_{W^{\gamma,p_0}(\Omega)}\Big)\|v\|_{W^{-1+\gamma,2}(\Omega)} \tag{B.15}$$

for all $u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \hat{\rho}) \cap \mathcal{U}_{ad}$ *and for some constant* C *independent of* u *and* v.

Proof. Ad (i): For any $u \in L^2(\Omega) \hookrightarrow W^{-1,p}(\Omega)$, by Theorem 2.2, we have $y_u = S(u) \in W_0^{1,p}(\Omega)$. From this, and Remark 2.9 in [12], we deduce the existence and uniqueness of $\tilde{z}_{u,v}$ in $H_0^1(\Omega)$ to (55).

We now rewrite (55) as follows

$$\begin{cases} -\operatorname{div}[(b + a(y_u))\nabla \tilde{z}_{u,v} + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u)\tilde{z}_{u,v}\nabla y_u] = v - \operatorname{div}(Z_{y_u,\bar{y}}\tilde{z}_{u,v}) & \text{in } \Omega, \\ \tilde{z}_{u,v} = 0 & \text{on } \partial\Omega, \end{cases}$$

where $Z_{y_u,\bar{y}}$ is defined as in (27). In other words, one has

$$\tilde{z}_{u,v} = S'(u)(v - \operatorname{div}(Z_{y_u,\bar{y}}\tilde{z}_{u,v})), \tag{B.16}$$

due to Theorem 2.2. By Lemma 3.4, we have that $Z_{y_u,\bar{y}} \in L^p(\Omega)^2$. Since $\tilde{z}_{u,v} \in H_0^1(\Omega) \hookrightarrow L^q(\Omega)$ with $q > \frac{2p}{p-2}$, we then deduce that $Z_{y_u,\bar{y}}\tilde{z}_{u,v} \in L^r(\Omega)^2$ with $r = \frac{pq}{p+q} \in (2,p)$. There therefore holds

$$v - \operatorname{div}(Z_{y_u,\bar{y}}\tilde{z}_{u,v}) \in W^{-1,r}(\Omega).$$

Applying Theorem 2.2 yields $\tilde{z}_{u,v} \in W_0^{1,r}(\Omega) \hookrightarrow C(\overline{\Omega})$. We then have $Z_{y_u,\bar{y}}\tilde{z}_{u,v} \in L^p(\Omega)^2$ and thus

$$v - \operatorname{div}(Z_{y_u,\bar{y}}\tilde{z}_{u,v}) \in W^{-1,p}(\Omega).$$

This, together with Theorem 2.2, gives $\tilde{z}_{u,v} \in W_0^{1,p}(\Omega)$.

We now prove (B.14). To this end, by Theorem 2.2 and the compact embedding $L^2(\Omega) \Subset W^{-1,p}(\Omega)$, we first have

$$\sup\{\|S'(u)\|_{\mathbb{L}(W^{-1,p}(\Omega),W_0^{1,p}(\Omega))} \mid u \in \overline{B}_{L^2(\Omega)}(\bar{u},1)\} \leq C_1,$$

which gives for all $u \in \overline{B}_{L^2(\Omega)}(\bar{u},1)$ that

$$\begin{aligned} \|\tilde{z}_{u,v}\|_{W_0^{1,p}(\Omega)} &\leq C_1\|v - \operatorname{div}(Z_{y_u,\bar{y}}\tilde{z}_{u,v})\|_{W^{-1,p}(\Omega)} \leq C_1\Big[\|v\|_{W^{-1,p}(\Omega)} + \|Z_{y_u,\bar{y}}\tilde{z}_{u,v}\|_{L^p(\Omega)}\Big] \\ &\leq C_1\Big[\|v\|_{W^{-1,p}(\Omega)} + \|\tilde{z}_{u,v}\|_{C(\overline{\Omega})}\|Z_{y_u,\bar{y}}\|_{L^p(\Omega)}\Big] \\ &\leq C_1\Big[\|v\|_{W^{-1,p}(\Omega)} + \|\tilde{z}_{u,v}\|_{W_0^{1,p}(\Omega)}\|Z_{y_u,\bar{y}}\|_{L^p(\Omega)}\Big], \end{aligned} \tag{B.17}$$

where we have just used the embedding $W_0^{1,p}(\Omega) \hookrightarrow C(\overline{\Omega})$. From Lemma 3.4 and the fact that $\|y_u - \bar{y}\|_{W_0^{1,p}(\Omega)} \to 0$ as $\|u - \bar{u}\|_{W^{-1,p}(\Omega)} \leq C\|u - \bar{u}\|_{L^2(\Omega)} \to 0$ (see Thm. 2.2), we have

$$\|Z_{y_u,\bar{y}}\|_{L^p(\Omega)} \to 0 \quad \text{as} \quad \|u - \bar{u}\|_{L^2(\Omega)} \to 0.$$

Then, there exists a constant $\rho_p \in (0,1]$ such that

$$\|Z_{y_u,\bar{y}}\|_{L^p(\Omega)} \leq \frac{1}{2C_1} \quad \text{for all } u \in \overline{B}_{L^2(\Omega)}(\bar{u},\rho_p).$$

Combining this with (B.17) yields

$$\|\tilde{z}_{u,v}\|_{W_0^{1,p}(\Omega)} \leq 2C_1\|v\|_{W^{-1,p}(\Omega)} \quad \text{for all } u \in \overline{B}_{L^2(\Omega)}(\bar{u},\rho_p).$$

Ad (ii): We first show the $W^{1+\gamma,2}$-regularity of $\tilde{z}_{u,v}$. To this end, we now rewrite (55) as follows

$$\begin{cases} -\operatorname{div}[(b+a(y_u))\nabla \tilde{z}_{u,v}] = v + \operatorname{div}[\mathbb{f}_{\bar{u}}\tilde{z}_{u,v}] & \text{in } \Omega, \\ \tilde{z}_{u,v} = 0 & \text{on } \partial\Omega \end{cases}$$

or, equivalently,

$$\begin{cases} -\Delta \tilde{z}_{u,v} = \dfrac{1}{b+a(y_u)}(\nabla b + \mathbb{1}_{\{y_u \neq \bar{t}\}} a'(y_u)\nabla y_u)\cdot \nabla \tilde{z}_{u,v} + \dfrac{1}{b+a(y_u)}[v + \operatorname{div}(\mathbb{f}_{\bar{u}}\tilde{z}_{u,v})] & \text{in } \Omega, \\ \tilde{z}_{u,v} = 0 & \text{on } \partial\Omega. \end{cases} \tag{B.18}$$

From Proposition B.1 and Lemma B.2, there holds $\operatorname{div}[\mathbb{f}_{\bar{u}}\tilde{z}_{u,v}] \in W^{-1+\gamma,2}(\Omega)$ and thus

$$v + \operatorname{div}[\mathbb{f}_{\bar{u}}\tilde{z}_{u,v}] \in W^{-1+\gamma,2}(\Omega). \tag{B.19}$$

Since $\mathcal{U}_{ad}$ is bounded in $L^\infty(\Omega)$ and thus in $L^{p_0}(\Omega)$, we deduce from Theorem 2.2 that $y_u, \bar{y} \in W^{2,p_0}(\Omega) \hookrightarrow C^1(\overline{\Omega})$ and then that

$$\|y_u\|_{W^{2,p}(\Omega)} + \|y_u\|_{C^1(\overline{\Omega})} \leq M \quad \text{for all } u \in \mathcal{U}_{ad} \tag{B.20}$$

for some constant $M > 0$. Here p_0 is the constant in (54). Thanks to Assumptions (A2) and (A3), one has $b + a(y_u), (b+a(y_u))^{-1} \in W^{1,\infty}(\Omega)$ and

$$\|b + a(y_u)\|_{W^{1,\infty}(\Omega)} + \left\|(b+a(y_u))^{-1}\right\|_{W^{1,\infty}(\Omega)} \leq C_M \quad \text{for all } u \in \mathcal{U}_{ad}. \tag{B.21}$$

Combining this with (B.19) yields

$$\frac{1}{b+a(y_u)}[v + \operatorname{div}[\mathbb{f}_{\bar{u}}\tilde{z}_{u,v}]] \in W^{-1+\gamma,2}(\Omega),$$

where we have employed the fact that the multiplication bilinear mapping $W^{1,\infty}(\Omega) \times W^{-1+\gamma,2}(\Omega) \to W^{-1+\gamma,2}(\Omega)$ is continuous; see, *e.g.*, [4]. The second term in the right hand side of (B.18) then belongs to $W^{-1+\gamma,2}(\Omega)$. Moreover, the first one also belongs to $W^{-1+\gamma,2}(\Omega)$, since it is in $L^2(\Omega) \hookrightarrow W^{-1+\gamma,2}(\Omega)$. We therefore apply the $W^{1+\gamma,2}$-regularity of solutions to (B.18) to have that $\tilde{z}_{u,v} \in W^{1+\gamma,2}(\Omega)$; see, *e.g.*, Theorem 3 in [44] and Theorem 4.1 in [21].

It remains to show the existence of a constant $\hat{\rho} > 0$ satisfying (B.15). To this end, we first see from (54) that $1 > \frac{2}{p_0} \geq 1 - \gamma$ and $1 < \frac{p_0}{p_0 - 1} < 2$. From this and the Sobolev embedding theorem; see *e.g.*, Theorem 3.8 in [4], we deduce that $W_0^{1,\frac{p_0}{p_0-1}}(\Omega) \hookrightarrow W_0^{\frac{2}{p_0},2}(\Omega) \hookrightarrow W_0^{1-\gamma,2}(\Omega)$ and thus

$$W^{-1+\gamma,2}(\Omega) \hookrightarrow W^{-1,p_0}(\Omega). \tag{B.22}$$

We now apply assertion (i) for $p := p_0$ to derive that $\tilde{z}_{u,v} \in W_0^{1,p_0}(\Omega)$ and that

$$\|\tilde{z}_{u,v}\|_{W_0^{1,p_0}(\Omega)} \leq C\|v\|_{W^{-1,p_0}(\Omega)} \leq C\|v\|_{W^{-1+\gamma,2}(\Omega)} \quad \text{for all } u \in \overline{B}_{L^2(\Omega)}(\bar{u}, \rho_{p_0}) \cap \mathcal{U}_{ad}. \tag{B.23}$$

Moreover, by applying Theorem 3 in [44] (see also [21], Thm. 4.1) to (B.18), and using estimates (B.13) and (B.21), as well as the embeddings $W_0^{1,p_0}(\Omega) \hookrightarrow H_0^1(\Omega)$ and $W_0^{1,p_0}(\Omega) \hookrightarrow W^{-1+\gamma,2}(\Omega)$, there is a constant $C = C(M)$ such that

$$\begin{aligned} \|\tilde{z}_{u,v}\|_{W^{1+\gamma,2}(\Omega)} &\leq C(M)\left[\|\tilde{z}_{u,v}\|_{W^{-1+\gamma,2}(\Omega)} + \|v\|_{W^{-1+\gamma,2}(\Omega)} + \left(1 + \|\mathbb{f}_{\bar{u}}\|_{W^{\gamma,p_0}(\Omega)}\right)\|\tilde{z}_{u,v}\|_{H_0^1(\Omega)}\right] \\ &\leq C(M)\left[\|v\|_{W^{-1+\gamma,2}(\Omega)} + \left(1 + \|\mathbb{f}_{\bar{u}}\|_{W^{\gamma,p_0}(\Omega)}\right)\|\tilde{z}_{u,v}\|_{W_0^{1,p_0}(\Omega)}\right] \quad \text{for all } u \in \mathcal{U}_{ad}. \end{aligned} \tag{B.24}$$

Setting now $\hat{\rho} := \rho_{p_0}$ and combining the last inequality with (B.23) yields (B.15). □

Appendix C. Proof of Theorem 5.8

First, (8c) leads to

$$\bar{u}(x) = \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}(x)\right) \quad \text{for all } x \in \overline{\Omega}. \tag{C.1}$$

Assume now that $\{(\bar{y}_h, \bar{u}_h)\}$ is the sequence of discrete solutions to (P_h) converging strongly to $(\bar{y}, \bar{u})$ in $H_0^1(\Omega) \times L^2(\Omega)$ as obtained from Theorem 5.4 and Remark 5.6. In order to prove

$$\|\bar{u}_h - \bar{u}\|_{L^\infty(\Omega)} \to 0, \tag{C.2}$$

we now split the proof into three cases.

Case 1: $\mathcal{U}_h = L^\infty(\Omega)$. As a result of (84), $\bar{u}_h$ can be expressed as

$$\bar{u}_h(x) = \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_h(x)\right) \quad \text{for all } x \in \overline{\Omega}.$$

This, in combination with (C.1), yields

$$|\bar{u}(x) - \bar{u}_h(x)| = \left|\mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}(x)\right) - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_h(x)\right)\right| \le \frac{1}{\nu}|\bar{\varphi}(x) - \bar{\varphi}_h(x)|$$

for all $x \in \overline{\Omega}$. Taking the L^∞-norm and using (89), we arrive at

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} \le C\left(h^{1-\frac{2}{q}} + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}\right).$$

This, along with Theorem 5.4, yields (C.2).

Case 2: $\mathcal{U}_h = \mathcal{U}_h^0$. Since $\bar{u}$ is Lipschitz continuous according to Theorem 2.5, there holds

$$|\bar{u}(x_1) - \bar{u}(x_2)| \le L_{\bar{u}}|x_1 - x_2| \quad \text{for all } x_1, x_2 \in \overline{\Omega} \tag{C.3}$$

and for some constant $L_{\bar{u}} > 0$. For any $T \in \mathcal{T}_h$, we have from (84) that

$$\bar{u}_{h|T} = \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu\,\mathrm{meas}_{\mathbb{R}^2}(T)}\int_T \bar{\varphi}_h \,\mathrm{d}x\right).$$

By the mean value theorem, there exists an element $x_T \in \overline{T}$ such that $\int_T \bar{\varphi}_h \,\mathrm{d}x = \mathrm{meas}_{\mathbb{R}^2}(T)\bar{\varphi}_h(x_T)$, and thus there holds $\bar{u}_{h|T} = \mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar{\varphi}_h(x_T))$. We then have from (C.1), for any $\tilde{x} \in T$, that

$$\begin{aligned}
\left|\bar{u}(\tilde{x}) - \bar{u}_{h|T}\right| &\le \left|\bar{u}(x_T) - \bar{u}_{h|T}\right| + |\bar{u}(\tilde{x}) - \bar{u}(x_T)| \\
&= \left|\mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}(x_T)\right) - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_h(x_T)\right)\right| + |\bar{u}(\tilde{x}) - \bar{u}(x_T)| \\
&\le \frac{1}{\nu}|\bar{\varphi}(x_T) - \bar{\varphi}_h(x_T)| + |\bar{u}(\tilde{x}) - \bar{u}(x_T)| \\
&\le \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{C(\overline{\Omega})} + |\bar{u}(\tilde{x}) - \bar{u}(x_T)|.
\end{aligned}$$

Moreover, (C.3) leads to

$$|\bar{u}(\tilde{x}) - \bar{u}(x_T)| \le L_{\bar{u}}|\tilde{x} - x_T| \le L_{\bar{u}}h.$$

In summary, we derive for any $\tilde{x} \in T$ that

$$\left|\bar{u}(\tilde{x}) - \bar{u}_{h|T}\right| \le L_{\bar{u}}h + \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)}.$$

Consequently, $\|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} \le L_{\bar{u}}h + \frac{1}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)}$. From this, (89), and Theorem 5.4, we have (C.2).

Case 3: $\mathcal{U}_h = \mathcal{U}_h^1$**.** In this situation, $\bar{u}_h$ cannot be computed from $\bar{\varphi}_h$ by a pointwise projection formula. In order to have the limit (C.2), we shall follow the argument from the results in Sections 3 and 4 in [41]. For that purpose, for each $h > 0$, let E_i be an arbitrary vertex of the triangulation $\mathcal{T}_h$ and let $e_i \in \mathcal{U}_h^1$ define a system of nodal basis functions, *i.e.*, $e_i(E_j) = \delta_{ij}$, where δ_{ij} is the Kronecker symbol. Setting $\bar{u}_i := \bar{u}_h(E_i)$ and $\bar{\varphi}_i := \bar{\varphi}_h(E_i)$, we now show that

$$M := \max_i \left| \bar{u}_i - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_i\right)\right| \le C\left[h^{1-2/q} + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}\right] \tag{C.4}$$

for all h small enough and for some constant $C > 0$ independent of h. In fact, it suffices to consider the case that $M > 0$. By Lemma 3.8 in [41], there exist indices i_0 and k_0 such that E_{k_0} is a neighboring vertex of E_{i_0} and that

$$M = \left|\bar{u}_{i_0} - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_{i_0}\right)\right| \quad \text{and} \quad \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_{k_0}\right) \ne -\frac{1}{\nu}\bar{\varphi}_{k_0}. \tag{C.5}$$

Moreover, there holds that

$$-\left[\bar{u}_{k_0} - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_{k_0}\right)\right] \le M \le \bar{u}_{i_0} + \frac{1}{\nu}\bar{\varphi}_{i_0} \le -\left(\bar{u}_{k_0} + \frac{1}{\nu}\bar{\varphi}_{k_0}\right) \tag{C.6}$$

in case (A) $\bar{u}_{i_0} - \mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar{\varphi}_{i_0}) > 0$ and that

$$\bar{u}_{k_0} - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_{k_0}\right) \le M \le -\left(\bar{u}_{i_0} + \frac{1}{\nu}\bar{\varphi}_{i_0}\right) \le \bar{u}_{k_0} + \frac{1}{\nu}\bar{\varphi}_{k_0}$$

in case (B) $\bar{u}_{i_0} - \mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar{\varphi}_{i_0}) < 0$; see Lemma 3.7 in [41]. We now discuss here only case (A) as case (B) can be treated in the same way. By the second relation in (C.5) and (C.6), we have

$$\beta = \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_{k_0}\right) < -\frac{1}{\nu}\bar{\varphi}_{k_0};$$

see the estimate (3.21) in [41]. From this, the Lipschitz continuity of $\bar{\varphi}$ (due to Thm. 2.5), and (89), in view of Theorem 5.4 there holds

$$\begin{aligned}
-\frac{1}{\nu}\bar{\varphi}_{i_0} &= -\frac{1}{\nu}\bar{\varphi}_{k_0} + \frac{1}{\nu}[\bar{\varphi}_{k_0} - \bar{\varphi}(E_{k_0})] + \frac{1}{\nu}[\bar{\varphi}(E_{k_0}) - \bar{\varphi}(E_{i_0})] + \frac{1}{\nu}\left[\bar{\varphi}(E_{i_0}) - \frac{1}{\nu}\bar{\varphi}_{i_0}\right] \\
&\ge -\frac{1}{\nu}\bar{\varphi}_{k_0} - \frac{L_{\bar{\varphi}}}{\nu}h - \frac{2}{\nu}\|\bar{\varphi} - \bar{\varphi}_h\|_{L^\infty(\Omega)} \\
&\ge -\frac{1}{\nu}\bar{\varphi}_{k_0} - \frac{L_{\bar{\varphi}}}{\nu}h - \frac{C}{\nu}\left[h^{1-2/q} + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}\right] \\
&> \beta - \frac{L_{\bar{\varphi}}}{\nu}h - \frac{C}{\nu}\left[h^{1-2/q} + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}\right] \\
&> \alpha
\end{aligned} \tag{C.7}$$

for all h small enough and for some constants $L_{\bar{\varphi}}, C > 0$. Moreover, since $\bar{u}_{i_0} - \mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar{\varphi}_{i_0}) > 0$, we have $\mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar{\varphi}_{i_0}) < \bar{u}_{i_0} \le \beta$ and thus

$$-\frac{1}{\nu}\bar{\varphi}_{i_0} < \beta.$$

Consequently, $\mathrm{Proj}_{[\alpha,\beta]}(-\frac{1}{\nu}\bar{\varphi}_{i_0}) = -\frac{1}{\nu}\bar{\varphi}_{i_0} \in (\alpha,\beta)$. We then deduce from the identity in (C.5), (C.7), and the fact $\bar{u}_{i_0} \le \beta$ that

$$M = \bar{u}_{i_0} - \mathrm{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\nu}\bar{\varphi}_{i_0}\right) \le \beta + \frac{1}{\nu}\bar{\varphi}_{i_0} \le C\left[h^{1-2/q} + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}\right],$$

which shows (C.4).

Exploiting now (C.4), the Lipschitz continuity of $\bar{u}$ (due to Thm. 2.5), and the argument used in Section 4 in [41], we arrive at

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Omega)} \leq C\left[h^{1-2/q} + \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}\right]$$

for all h small enough and for some constant $C > 0$ independent of h. Theorem 5.4 then implies (C.2). □